Future by design

A framework for introducing radical change in urban rail systems

by

Marcelo Blumenfeld Mendonca

A thesis submitted to the University of Birmingham
for the degree of Doctor of Philosophy

September 2018

School of Electronic, Electrical and Systems Engineering
College of Engineering and Physical Sciences
University of Birmingham

ABSTRACT

Urban mobility is increasingly becoming accepted as a basic human need, when socio-economic opportunities depend on the ability to reach places within acceptable times. On the other hand, the emergence of megalopoleis as dominant features of the global landscape has been increasing commuting effort to unprecedented levels. These ever-larger urban areas allowed by the dominance of the automobile and their associated travel distances highlight important shortcomings in the operation of mass transport systems. Public transport users in megalopoleis spend up to two times longer than drivers for covering similar distances, exacerbating important social and economic inequalities and reinforcing the preference for private modes.

However, even though there is an assumption that the problem can be easily overcome by increasing the speed of transport systems, advocates of this approach overlook important utility trade-offs that arise from the conflict between greater vehicle speeds and the additional time required to access the services. The first original aspect of the thesis is the deeper understanding of the inherent limitations of paradoxes in urban rail systems. For instance, metro systems are inherently constrained by a paradox between access and in-vehicle speeds, which prevents them from offering sufficient door-to-door speeds to cover long distances within acceptable travel times. It becomes clear that these systemic limitations can only be solved by radical innovation, especially in cases where the systems environment is rapidly changing.

The first part of the research comprises a literature review on the foundations of engineering to understand how to achieve radical change in socio-technical systems. This in turn leads to the second original aspect of the thesis: a novel heuristic framework that combines the backcasting method with a system engineering approach to develop innovative solutions that are equally robust and resilient in face of the uncertainty of future scenarios. With that, normative scenario building becomes a quantitative process in which benefits, performance, and risks can be analysed and optimised according to different parameters.

The second part of the thesis, and its third main original aspect, illustrates the framework in a specific case study of metro systems in megalopoleis. Models are used to identify the functional paradoxes that are used to develop a proposed concept that comprises three main operational foundations. Firstly, an operational strategy where autonomous vehicles stop in different patterns along the line to reduce access times without an impact on in-vehicle times. Secondly, stations are located off the main line to guarantee that all passengers can board their preferred services within minimum headways. Finally, the operational concept adopts autonomous vehicles that travel in platoons and are controlled by vehicle-to-vehicle communication algorithms similarly to automated highways. Results show that this type of solution can potentially improve door-to-door journey times in metro systems if practical barriers can be overcome. In theory, it can reduce the distance between stations to a minimum and thus reduce access time by 50%, while simultaneously increasing in-vehicle speeds by 45% and reduce door-to-door journey times by up to 31% compared to conventional operations. Moreover, capacity can also be increased between 20% and 40% compared to current systems.

Therefore, this thesis proposes a series of heuristic steps rooted in normative scenarios to develop operational concepts which are not only innovative but also robust, in a quantitative and verifiable manner. Systems can be functionally modelled, allowing specific technical requirements and specifications to be met in the future. With that, the limitations of current capabilities are reversed from their original position of functional constraint, to a position of normative functional guidelines for development. By focusing on what tools to develop for an ideal system rather than a system that adapts to current tools, this research is a starting point to a new perspective on developing future urban systems.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors Prof. Clive Roberts and Prof. Felix Schmid. You decided to take a bet on a Brazilian transport planner with a design background, and my transformation into an academic systems engineer would not have been possible without your orientation, support, challenges, many corrections, and teachings.

Secondly, to my wife Chloe. For you are always there, in moments of joy and moments of despair. Thank you for the constant encouragement and the kind words when all seemed lost. Thank you for making every day worth living, and the inspiration to do so for the common good. Also, for being an incredible moral compass. I have and always will have utmost admiration for you.

To my family, who I would not have reached today without. To my mother Marli, for the enormous effort and commitment in providing the best education I could have, and for being a role model with impeccable work ethics which I aspire to. To my brother Michel, for the partnership and teaching me that doing something for others is the best way to help oneself.

To my late grandmother Bella, who patiently cultivated my insatiable thirst for knowledge and passion for books which proved very useful in pursuing this degree.

To my colleagues in the Birmingham Centre for Railway Research and Education, for making workdays enjoyable and challenging, and for the great Christmas parties in summer!

Finally, a posthumous thank you to Carl Sagan, Alan Watts, and Jacques Fresco, who despite never having met in person, have inspired me to pursue a life of righteousness, science, and teaching for a better world.

Contents

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Determinism	A paradigm dictated by causal relationships in which a future event is determined only by prior events
epistemology	branch of philosophy concerned with the theory of knowledge and its methods
GHG	Greenhouse gas
heuristics	a problem-solving activity that is based on the exploration of possible solutions rather than following a pre-defined set of steps for problem-solving
ISO	International Organisation for Standardisation
MBSE	Model Based Systems Engineering
MoP	Measures of Performance
normative	an approach that designates actions or states as desirable or undesirable based on a certain judgement
SysML	Systems Modelling Language
teleology	contrary to determinism, explains actions according to their purpose rather than their cause
Tpkm	Tera passenger-km
TRIZ	Theory of Inventive Problem Solving
TRL	Technical Readiness Levels
TRM	Technology Roadmapping

1. INTRODUCTION

1.1. Motivations, aims and objectives

This thesis was conducted in face of the impacts of transport behaviour on the social, economic, and environmental sustainability of large cities; and how the unsustainable patterns of urbanisation has made it very difficult for public transport systems to compete with the dominant private and semi-private modes. These issues have highlighted the need for radical innovation in infrastructure systems such as urban rail.

Therefore, the main aim of the research was to develop a framework for the implementation of radical innovation in complex systems, where inherent trade-offs and constantly changing environments keep incremental evolution from achieving its goals. With that, the thesis focuses on the inherent systemic limitations of existing metro systems in offering acceptable door-to-door journey times and capacity for long distance trips.

The objectives of this research are two-fold: firstly, to conduct an epistemological review and develop a framework in which radically innovative systems can be designed in a robust and reliable manner. Secondly, to apply the framework in the case study of metro systems in megalopoleis to illustrate and analyse its ability to devise radically innovative concepts that may turn into potential engineering solutions.

1.2. Context

1.2.1. Megalopoleis

While the title mentions railway-based urban transport systems, this thesis focuses specifically on large urban areas as they seem to exacerbate transport externalities in terms of travel times, energy consumption and pollution (Kennedy, et al., 2015; West & Bettencourt, 2010; Van Wee, et al., 2006). Such reality offers a fruitful context for the application of the developed framework in a context that is known to require radical change. From there, an inductive reasoning on the wider applications of the processes suggested can be discussed based on the outcomes of this research.

One of the reasons for the challenging trends we observe unfold can be accounted to the distinctive patterns of growth. Cities are often depicted as complex systems, and as such they follow similar patterns of non-linear growth as natural systems, arising from bottom-up emergent properties of self-organisation (Mumford, 1961; Meadows, 2008; Tero, et al., 2010; Yates, 1987). In contrast, man-made systems are usually engineered from the top-down, following incremental evolutionary processes (Blanchard & Fabrycky, 2006). One important outcome which is crucial to this research is that the exponential growth of externalities can lead to collapse when such systems become sufficiently large. Moreover, cities are constantly evolving, and while anthropogenic infrastructure systems grow in a different way than organisms (Tero, et al., 2010), this disparity, at a certain point, becomes unsustainable and will require radical action to break from undesirable trends.

The adoption of the term megalopoleis in this research is intentional because it encapsulates the distinctive features of these large urban areas. The term was coined by Geddes (1915) and made prominent by Gottmann (1961) in his study of American cities. As opposed to the centralised structure of the classical metropolis, a megalopolis is commonly defined as a polycentric urban region comprising cities and towns that are physically separated, but functionally connected (Lang & Knox, 2009; Hall & Pain, 2009). On the other hand, it distinguishes itself from a megacity because its definition is based on its function rather than its population size. Megacities as definitions have been the focus of continuous scrutiny as there has not been any unanimous numerical metric for a city to be considered a megacity

(Forstall, et al., 2009). For example, values ranging between 5 and 10 million inhabitants have been found in literature (IUGS, 2005; Moavenzadeh & Markow, 2007; UN-DESA, 2014). Moreover, the numerical thresholds in use are arbitrary and do not acknowledge the more complex spatial structure of these regions (Forstall, et al., 2009; Sorensen & Okata, 2011).

1.2.2. Radical change

Although innovation is normally seen as part of the optimisation of systems, the type of innovation highlighted here is one of structural changes in face of complexity, long-term uncertainty, and trends leading to undesirable future scenarios. Saviotti (1986) points out that when the environment is constant, incremental adaptations in systems may suffice, yet under changing conditions, adaptation can only be achieved by a radical redesign leading to a completely new internal structure.

The term radical within radical change derives from the Latin word radix and refers to the root of a concept, which highlights structural changes in systems rather than the incremental development of their parts. Related to systems thinking, it acknowledges the emergent properties that arise from complex interactions between the parts, and thus consider also the interfaces rather than components themselves.

In dealing with a changing world, it is necessary to change our epistemological approach to engineering urban systems. In face of the pressing challenges of sustainability in large urban areas, new methodologies are required, integrating engineering and approaches from social sciences to create a new framework for innovation and resilience (Walsh, et al., 2015).

1.2.3. Heuristics

Decision making processes in face of the uncertainty of long-term futures usually involves heuristic approaches to problem solving (Kuhn, 1962; Tversky & Kahneman, 1974). Heuristics in this thesis relates to a problem-solving activity which is based on the exploration for solutions rather than following a step-by-step approach of the context of justification (Lakatos, 1976; Dreborg, 1996). Chapter 3 defines heuristics in further detail.

This research is also presented as metaheuristic, because there is no algorithm for innovation. Logically, an algorithm only exists for solutions within a known spectrum of problems, and this research stands at the other end of the spectrum looking for new problems to be solved. As an open-ended exploratory quest, the research began with what was perceived as technical solution to systemic problems in public transport systems. However, when confronted with the wider methodological, philosophical, and technical issues of the problem, it was understood that the research question was in fact much deeper than simply an ‘eureka moment’ which does not reflect the logic of discovery (Carmichael, 1930). That in turn led to a change in the research enquiry, focusing on an epistemological framework that can successfully address radical change in man-made systems rather than a technical solution for a specific problem.

1.3. Scope

This thesis defines its scope in several levels. Firstly, among all urban systems, this thesis will focus its theme on urban transport systems because of their intrinsic relationship with the spatial organisation of a city and consequently the outcomes in urban indicators (Hall, 2014; Marchetti, 1994; Schaeffer & Sclar, 1980). Within that spectrum, the focus will be towards passenger services rather than freight, even though both are essential for urban sustainability. More specifically, the work will be directed towards urban guided systems such as metros, for their more common application to large urban areas in terms of speed and capacity.

The methodology section will adopt a wider view on normative future scenarios but acknowledging that the area of futures studies is even wider. Similarly, the relationship with the Vee systems engineering process does not discredit other processes. The choice of relating the framework with the Vee was due to the synergy of the latter with a backcasting approach, and the appropriateness to working in higher levels of abstraction.

Finally, the scope of the case study was limited to addressing the aspects of door-to-door journey times and capacity of urban rail systems. It is acknowledged that the research does not address very important aspects of railway operations such as costs, reliability, and safety in the same level of detail. The case study is thus an illustration of the main topic of the thesis (the framework), where the resulting operational concept is but one option between various radically innovative solutions for the matter. The narrow technical scope of the case study helps emphasise its role as a first iteration in a series of engineering processes for decision making, as discussed in Chapter 6.

1.4. Structure

This thesis is organised in three parts. Chapters 2 and 3 contains the main literature review of the research topics. Starting from the main question on how to overcome systemic shortcomings of public transport cities that arise with increasing distances, I will discuss the current epistemological and methodological approaches found in the literature. Firstly, the thesis will briefly introduce the different philosophical stances of heuristic and algorithmic approaches and ponder on their impact on innovation in engineering projects. Secondly, I will highlight foresight methods commonly used in other fields and how they can be used to promote radical change in urban systems. Finally, the thesis will compare the traditional reductionist approach used in the physical sciences to the more recent holistic perspectives from systems theory that have recently gained prominence, especially in the context of complexity such as the case in this thesis.

Chapter 4 is a direct result from the previous chapter and presents and explains the framework created in order to synergistically merge the three main pillars of the approach: namely, heuristics, normative forecasting, and systems engineering, in order to promote radical change in a systematic and traceable environment.

Part two focuses on the application of the method developed into a case study of metro systems in large urban areas, and the impacts that the size of cities have on door-to-door journey times. Chapter 5 runs the framework in such context as a means to achieve a radically innovative operational concept that can overcome inherent shortcomings. As an outcome of the process, an operational concept is proposed, followed by the definition of functional requirements, and an optimal selection of technologies.

Part three, composed of Chapters 6 and 7, brings an evaluation of the framework in terms of its operational robustness and validity. Also, the author discusses the potential applications of teleological approaches in engineering, and future research needed to further refine both the framework and the operational concept presented in the case study.

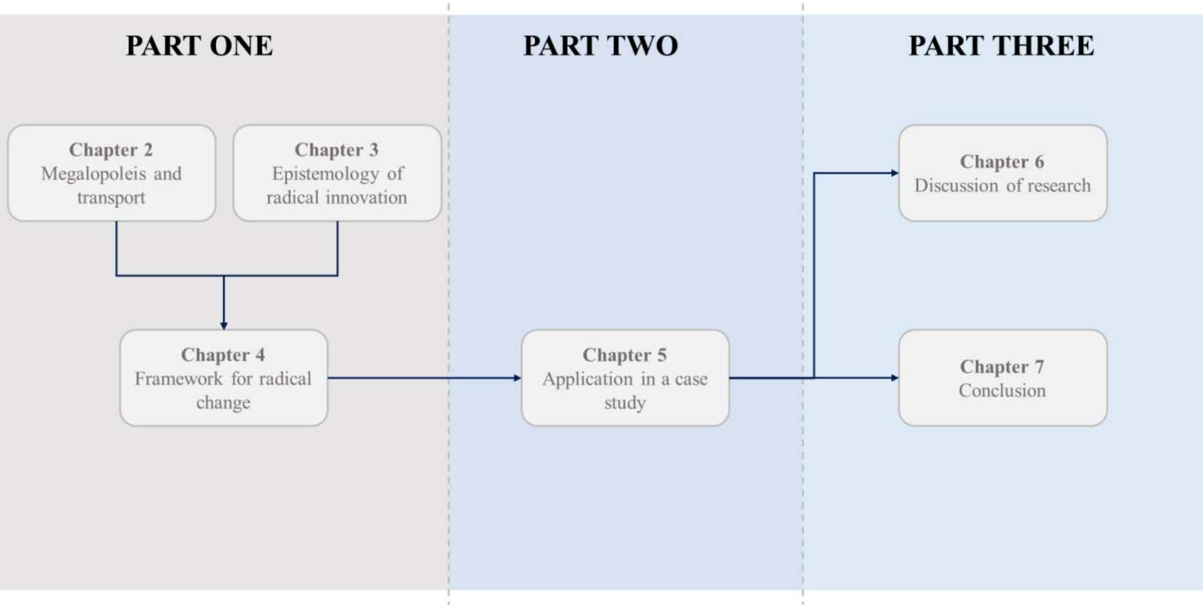


Figure 1. Illustration of thesis structure

1.5. Publications

The following list contains the six articles written and successfully published by the author during his PhD research. The first four papers are directly related to the topic of this thesis, each comprising a different aspect of it. The remaining two articles were written (and co-written in one case) using the knowledge acquired during the thesis and applied to specific projects.

- Blumenfeld, M., C. Roberts and F. Schmid. 2016. A systems approach to developing a new type of metro for megalopolis. Proceedings of the Institution of Civil Engineers – Transport. 169 (4). pp. 225-238.

- Blumenfeld, M., F. Schmid and C. Roberts. 2016.A systems approach to developing future urban systems: creating a metro that fits the megalopolis. Proceedings of the IET International Conference in Railway Engineering, 12-13 May 2016, Brussels, Belgium.

- Blumenfeld, M., C. Roberts and F. Schmid. 2016. A systems approach to developing a metro fit for 2050. 11th World Congress in Railway Research, 29 may- 2 June 2016, Milan, Italy

- Blumenfeld, M., C. Roberts and F. Schmid. 2016. An operational strategy to increase average door-to-door speeds of metro systems in megalopoleis, in Proceedings of the International Conference on Smart Infrastructure and Construction, 27-29 June 2016, Cambridge, UK.

- R. Evans, M. Blumenfeld, C. Roberts and F. Schmid. 2016. Systems Thinking for Delivering Long Term Visions for Rail. 11th World Congress in Railway Research, 29 may- 2 June 2016, Milan, Italy.

- Blumenfeld, M., E. Stewart and E. Lathios. 2016. Using Model Based Systems Engineering to increase the robustness of monitoring systems in railways. 7th IET Conference on Railway Condition Monitoring, 27-28 September 2016, Birmingham, UK.

2. INTRODUCTION AND CONTEXT

2.1. Background

Although large cities have been part of the human context for millennia, the appearance and subsequent dominance of megalopoleis in the urban landscape is unprecedented. Not only the majority of the world population now lives in cities, but also urban areas are rapidly growing in number and size, and this transition has been raising concerns over its impacts on social, economic and environmental sustainability at a global scale (UN-DESA, 2014). While in 1950 there were only 7 cities with more than 5 million inhabitants, there are 85 of these as of 2017 (UN-DESA, 2014; Cox, 2017). Moreover, trends indicate that by 2030, there will be more than a hundred cities above the 5 million inhabitant threshold, housing more than 1.2 billion people, mostly in developing countries (UN-DESA, 2014). By 2050, projections forecast 8 cities with more than 30 million inhabitants, and by the end of the century, we might witness cities with populations exceeding 80 million (Hoornweg & Pope, 2016).

The relevance of urbanisation trends is not due to the size of modern cities, but to their increasing economic and social importance to the international landscape. Large cities now drive the world economy, holding on average 75% of the economic activity of their respective country. By 2025, six hundred cities alone will account for 60% of the global GDP (McKinsey Global Institute, 2011).

Research shows that, at an aggregate level, cities can benefit from population growth (West, 2017; Bettencourt, et al., 2010). Cities are catalysts of social and economic interactions and can lead to economies of scale in infrastructure needs (West & Bettencourt, 2010). Their inherent complexity suggests that they follow similar patterns of growth as natural systems (Tero, et al., 2010). It is not surprising that cities are therefore often deemed analogous to biological organisms or whole ecosystems (Mumford, 1961; Dendrinos & Mullally, 1985;

Samaniego & Moses, 2008; Meadows, 2008). While often depicted as complex systems, it appears that their dynamics and relationships are anything but random. Urban form and density and urban form and scale can be explained by simple functions of economies of scale and increasing returns, and travel time and cost budgets (Batty, 2005; West, 2017; Marchetti, 1994).

Studies have found increasing returns in socioeconomic parameters and economies of scale in material infrastructure as cities experience population growth. For instance, the requirements for total road surface, length of electrical cables, water pipes or number of petrol stations usually increase at 15% less than the expected linear growth, showing significant economies of scale (West & Bettencourt, 2010). At the same time, per capita social indicators such as wages, creative jobs, and number of patents produced increase by a factor of 1.15 compared to linear population growth (Bettencourt, et al., 2010).

On the other hand, the superlinear scaling of socio-economic parameters also feeds into a reinforcing feedback that attracts further population growth and economic activity, making them thirsty for resources. Although an increased population count can lead to economies of scale in infrastructure, cities and especially megalopoleis still account for disproportional social and environmental impacts in terms of travel times, waste production and energy consumption (Gwilliam, 2002; Moss & Qing, 2012; Kennedy, et al., 2015). They occupy less than 3 per cent of the land surface, yet use 75 per cent of the available resources, and account for about 67 per cent of all greenhouse gas emissions (UN-HABITAT, 2012). Should these patterns remain unchanged, energy consumption would increase up by 61%, resulting in a 48% increase in GHG emissions (Ligtvoet & Hilderink, 2014; World Energy Council, 2013; EIA, 2016).

As previously highlighted, urbanisation can theoretically benefit societies with economies of scale and increasing socio-economic returns (West & Bettencourt, 2010). However, in practice most western cities grew in an uncoordinated way in terms of area, a process known as the phenomenon of urban sprawl (Batty, et al., 2003). Much of the spatial transformation of cities can be accounted to technological developments in modes of transport that have changed the nature and the boundaries of mobility (Schaeffer & Sclar, 1980; Green, 1988). It was only when systems were able to move passengers and goods more efficiently that villages could expand into cities and cities could transform into the current urban giants we see today. As increases in travel speeds are generally transferred to the coverage of greater distances rather than in the reduction of total travel times, the potential area of cities grows quadratically with speed increments (Zahavi, 1974; Marchetti, 1994).

Subsequently, just as the railways were responsible for transforming cities into metropoleis, the automobile was responsible for the birth of the megalopoleis. Since private and semi-private transport can offer higher door-to-door speeds, the affordability of the automobile in the twentieth century has allowed urban areas to spread almost boundlessly as they did not require public transport infrastructure to do so. Laube et al (1999) show very strong correlations that indicate that the greater the road provision, the higher the traffic speeds, and consequently, the more kilometres per person travelled. As a result, urban areas have quickly expanded beyond 2,000 km^2, requiring ever greater mobility to have sufficient access to economic opportunities (Cox, 2017).

Moreover, transport intensity has accompanied such growth and is now an urgent challenge to the sustainability of cities under current trends. Urban passenger demand has grown eighteen-fold since 1950 to 25 Tera passenger-km (Tpkm) in 2015 and is estimated to double to 50 Tpkm by 2050 (ITF, 2017; Moriarty & Honnery, 2008). Unsurprisingly, the transport sector is responsible for 36% of the total energy consumed globally, and road transport alone is responsible for approximately 60% of that amount (International Energy Agency, 2017). Left unchanged and added to the increasing motorisations levels in the developing world, the environmental burden from the transport sector could increase substantially, comprising an increase in up to 82% in energy consumption and 79% in CO_2 emissions by 2050 (World Energy Council, 2011). While there is general support for low mobility scenarios where non-motorised modes are prioritised, large urban areas are unlikely to be able to adhere completely, as their size requires at least some degree of motorised transport to guarantee travel speeds for an acceptable level of access.

This situation has proven difficult to revert because of the deep structural changes that private motorisation has imposed on cities, especially those in developing countries lacking public transport infrastructure. The greater door-to-door speeds enabled by private modes allowed urban areas to stretch beyond the reach of public transport, thus promoting a more energy intensive paradigm of low density that reinforces private mobility over other transport modes (McCahill & Garrick, 2012). In such reality of low density and longer distances, public transport systems cannot compete with the automobile in terms of door-to-door speeds because of systemic shortcomings that will be discussed in further detail in this thesis. As found by Laube et al (1999) and Gyimesi et al (2011), speed is a crucial element for mode choice because it directly impacts the amount of time available to other activities within the limited hours of a day.

As discussed on the next chapters, many technical and technological solutions emerged in times when the current systems could not cope with the extraordinary new dynamics of urbanisation and ever-increasing distances. The reason why most have not been able to revolutionise the field can arguably be illustrated by two main characteristics. Firstly, all solutions seem to focus solely on one aspect of the problem and not the whole system that comprises the journey, therefore failing to deal with all variables of the system simultaneously. More specifically, these solutions also treat the transport system solely as a technical system where the infrastructure components are all that matters, rather than a socio-technical one, where user behaviour and non-motorised components are also taken into account. With that, important variables such as access distances and time penalties are usually left out of the equation. Secondly, these examples adopt a traditional algorithmic approach to the problem (one of optimising existing solutions) in a context of uncertainty where no optimal solution is yet known. They approach the technicality of the solution from a causal perspective rather than using a teleological stance. In other words, concepts seek to find the best possible solution using the currently available tools rather than envisioning what tools are needed to create the right solution.

It becomes clear that incremental technological enhancements to current systems are not sufficient to overcome the inherent limitations that arise from internal conflicts under changing conditions. Mostly this is not due to insufficient technical capabilities, but because the conventional methods used in science cannot accommodate the necessary innovation under extraordinary conditions. As highlighted by Kuhn (1962), conventional methods tend to insulate themselves from problems which are novel and extraordinary. In that sense, the reality of megalopoleis as urban areas which are now outstretching the ability of transport systems to connect cities within acceptable travel times is extraordinary to transport research. Therefore, new approaches and longer time horizons must be adopted in order to break away from undesirable trends and adapt to the future needs of a highly urbanised world (Moriarty & Honnery, 2008; Banister & Hickman, 2013).

2.2. Transport and the emergence of the megalopoleis

One of the most compelling features of urban mobility around the globe and across history is that cities have generally grown in size proportionally to the average speed of travel at the time. In other words, cities have generally been around ‘one-hour wide’, meaning that their radii have been, in km, around half of the average speeds (in km/h) achievable with the transport modes available at the time (Marchetti, 1994). The apparent reason for such stability is that urban dwellers tend to devote a fixed amount of time per day to general mobility when they are able to. This has been widely documented that, throughout history and across countries, people tend to spend an average of 70 minutes on all trips performed in a day, usually referred to as *travel time budget* (Zahavi, 1974; Zahavi & Talvitie, 1980; Schäfer, 2000; Bieber, et al., 1994; Marchetti, 1994; Gyimesi, et al., 2011).

Such extensive observations are yet not free from debate (Goodwin, 1981; Gunn, 1981). Mokhtarian and Chen (2004) highlight variations in travel time according to income, gender, mode of transport, and urban form. Roth and Zahavi (1981) also found that values in developing countries seemed to vary greatly depending on income and car ownership. In addition, van Wee et al. (2006) found that average daily travel time has increased in the Netherlands. However, on the aggregate level, those amounts are in accordance with the 70-minute theory. More importantly, these debates shed light on a wider question which is central to this research: are longer travel times a voluntary decision when cities grow? If those not owning a car spend more time travelling, is it by choice or by chance? These questions refer to a deeper inquiry on whether cities should work in order to maintain the stability in travel times within people's budgets. It seems that, given the option, people will choose to remain within their time budgets.

If we assume humans to be territorial animals whose objective, among others, is to expand their area of resource possibilities, then one would travel as far as possible in order to explore the greater amount of resources one is able to (Winterhalder, 1981; Ausubel & Marchetti, 2001). However, spatial consumption is bounded by finite capabilities, either in terms of energy, time, or money. For instance, Kelly (1983) and Jones & Madsen (1989) show that the length of trips of hunter gatherers were already limited by calorific returns before cities even existed. It is a logical issue in economics: for hunter gatherers and every other animal, the energy expenditure of looking for food must not exceed the calorific return of the food collected, otherwise it will result in overall energetic loss. For urban dwellers, on the other hand, the disutility of travel cannot exceed a certain budget of the utility of the activity that follows, or it incurs overall losses and thus cannot be sustained in the long-term among other daily costs (Laube, et al., 1999). As researchers found, monetary budgets follow this premise, and people tend to dedicate a certain share of total expendable income that will ensure mobility yet within their financial limitations (Zahavi & Talvitie, 1980; Marchetti, 1994; Laube, et al., 1999). In the same way that calories and money are finite goods, so is time. It can be inferred then that the stability of travel time budgets derives from the general perception of travel as a disutility which takes time from productive activities within a finite number of hours of a day (Szalai, 1972; Laube, et al., 1999).

Moreover, when given the opportunity to reduce their travel time by faster modes of transport, people tend to choose more travel rather than travel time savings, which reinforces the notion of an invariant time budget threshold (Chen & Mokhtarian, 1999). Zahavi (1974) adds that, when afforded, faster travel speeds tend to be converted to longer distances rather than more trips. Two main notions can then be expanded from this: first, that the average distance travelled across a whole population will be a product of the time and costs budgets and the absolute speed of the transport network (Laube, et al., 1999); second, that people will look for the highest speed they can afford in economic terms, in order to expand their area of potential socio-economic exploration (Marchetti, 1994).

One of the reasons behind this desire to travel farther in the urban context can be accounted to urban density and the perceived quality of urban living. As cities have in the previous centuries been seen as crowded and slum-like areas, those who could afford to travel faster would move out to the countryside to enjoy more space and greenery (Kellett, 1969). Travel time budgets then act as spatial limitations to the distances that people can live from places while still participating in the urban context. Consequently, once technological developments permitted faster travel with the industrial revolution, urban areas expanded accordingly, from the pedestrian villages to the railways driven metropoleis (Schaeffer & Sclar, 1980). The railways, in that sense, permitted the conception of the suburbs (Howard, 1902). Until the 1840s, travel was performed by foot or on horseback, which means that even big cities had a diameter no larger than 5 km (Banister, 2011; Marchetti, 1994).

London is perhaps a fruitful example of the geographical expansion of cities that followed the increases in transport speeds. Before the opening of the metropolitan railways in the second half of the 19th century, London was constrained to walking distances and rarely stretched beyond a mile from the Thames; conversely, commutes longer than 5 km were not uncommon once the railways were in place (Taylor & Green, 2001; Green, 1988). In a specific company on Saville Row, Green (1988) found that, in 1857, 95% of the workers travelled less than 5 km and the longest commute was 7 km. After the establishment of the London Underground, by the end of the century more than half of the employees travelled more than 5 km and the longest commute was over 10 km (Figure 2).

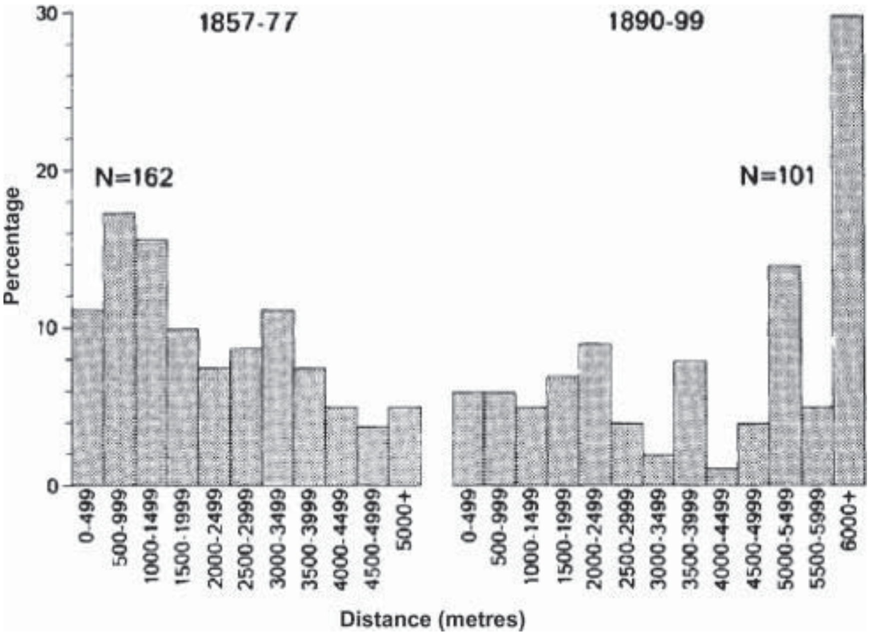


Figure 2. Henry Poole's workers commute distance (Green, 1988)

Subsequently, new technologies and more efficient operations enabled the radical transformation of the metropolis into the megalopolis. In the twentieth century, rapid population and economic growth, added by the increased affordability of faster travel especially by private and semi-private motorisation, expanded the boundaries of most urban areas even further, giving way to extreme suburbanisation with a new paradigm of low-density sprawl (Hall, 2009). While a person on foot has a potential area of exploration of 20 km^2 within their travel time budget, a person in a car can exceed 1,000 km^2 within the same time (Ausubel, 2014). Even more, the urban transformation created by route flexibility and convenience of private and semi-private motorisation has also led to a new type of urban dynamics. The main distinctive characteristic of automobiles is not necessarily higher speeds in absolute terms, but higher average door-to-door speeds because the car simplifies the journey to virtually motorised-only components. Most megalopoleis reflect the nature of automobile use, where cities and towns become part of a polycentric region functionally connected by the higher speeds the transport infrastructure can supply within people's travel time budgets. These new dynamics were responsible for the expansion and reshaping of the 'common' city. As Newman and Kenworthy (1991) found, in 1980 most of jobs in the world's major cities had moved outside their central business districts, and less than half of the population were living in the inner urban area.

As a consequence of decentralised geographies, urban areas could expand beyond the 'one-hour wide' paradigm because trips broke from the usual radial pattern. For instance, the average speed on the road network does not generally exceed 45 km/h (Newman & Kenworthy, 1991). If they followed Marchetti's (1994) constant, the average area of these cities should then be around 1,600 km², yet the average area of the world's largest 75 cities is now at approximately 2,241 km², which results on an average radius of 26.7 km were they perfectly round (Cox, 2017). In fact, researchers have found strong correlation between levels of car use, urban area and urban density (Laube, et al., 1999). There is considerable difference between the urban fabric of Los Angeles and Dhaka, which reflects the different levels of access to private motorisation (Cox, 2017; World Health Organisation, 2015). Even in cities where high levels of railway usage and strict planning boundaries were retained, such as London, there are significant differences in mode share between the central and suburban areas (Transport for London, 2017). As a result, these metropolitan regions have also spread out much further than the speeds of their average transport systems, also due to natural limitations of population density (Newman & Kenworthy, 1991; Cox, 2017).

Figures Figure 3 to Figure 8 show the growth of six of the largest cities in the world over time, to highlight the impact of the increased travel speeds on urban sprawl (images adapted from ESRI (2015)). Even though the widespread availability of private motorisation occurred at different times and by different forces, the results in terms of urban growth are shared among all. Population density and socio-economic demographic trends tend to correlate with the levels of automobile use and public transport availability.

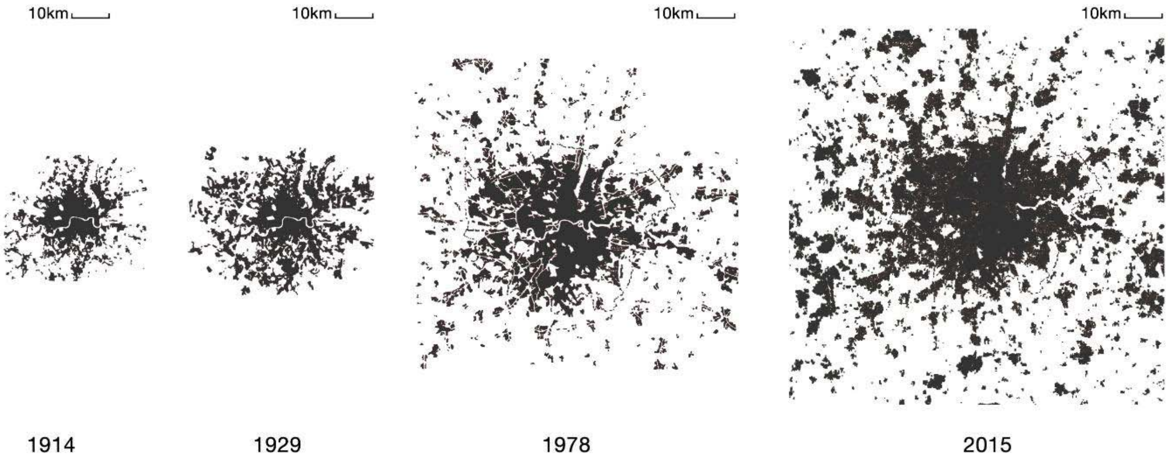


Figure 3. Urban area of London over time

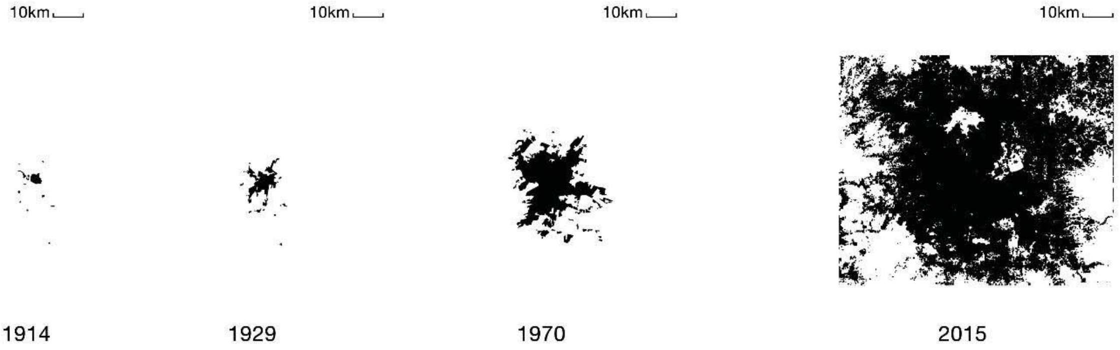


Figure 4. Urban area of Mexico City over time

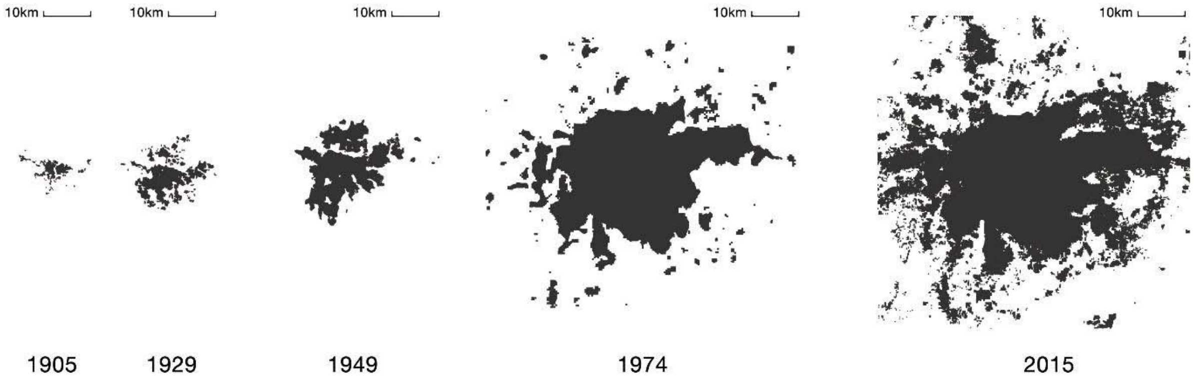


Figure 5. Urban area of São Paulo over time

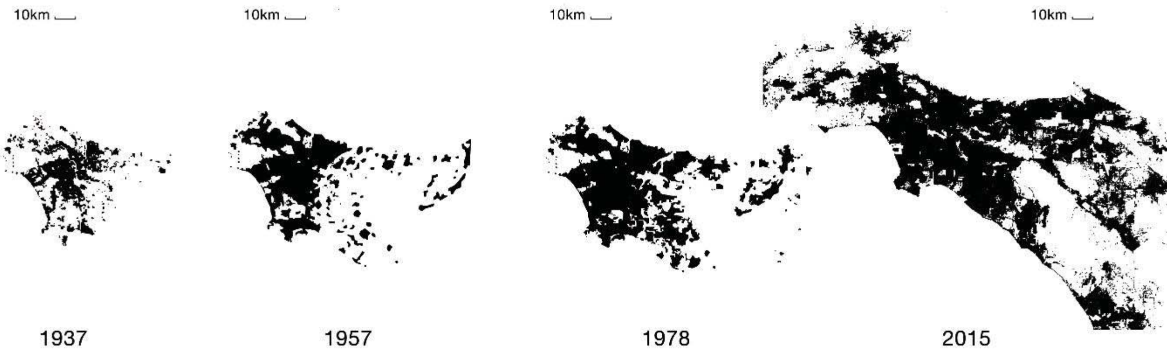


Figure 6. Urban area of Los Angeles over time

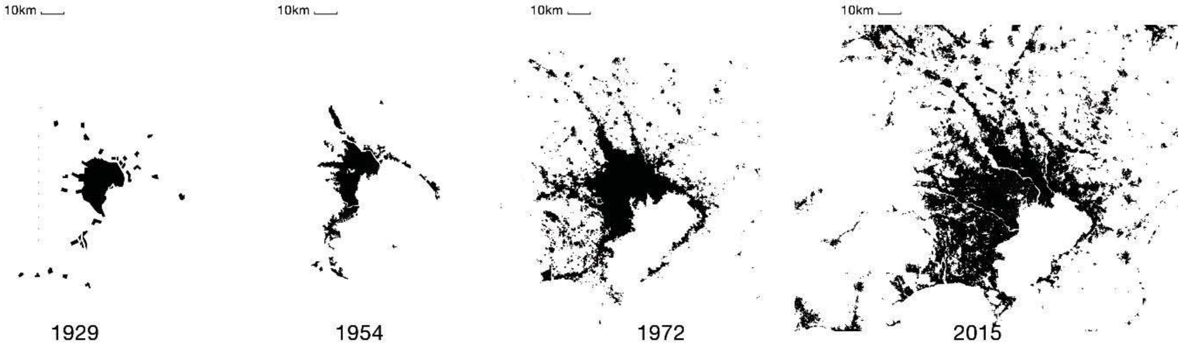


Figure 7. Urban area of Tokyo over time

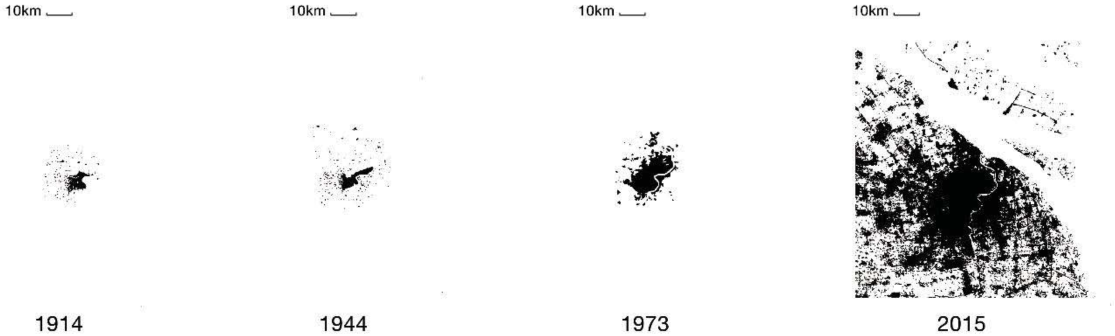


Figure 8. Urban area of Shanghai over time

The particular relevance of the megalopoleis as an urban phenomenon lies on the social, environmental, and economic burdens that have accompanied the disparity between distance and speed currently seen in many of the world's largest cities. These areas, such as London, São Paulo, and Tokyo, are now imposing longer travel times for all dwellers than their compatriot smaller counterparts, as shown in Figure 9. In developing countries, where the provision of transport infrastructure is lacking, the effects of urban mobility are even more accentuated (Gwilliam, 2003). It is not surprising that now several megalopoleis witness the phenomenon of supercommuting, in which dwellers spend three or more hours of travelling to and from work per day (Rigby, 2011; Moss & Qing, 2012).

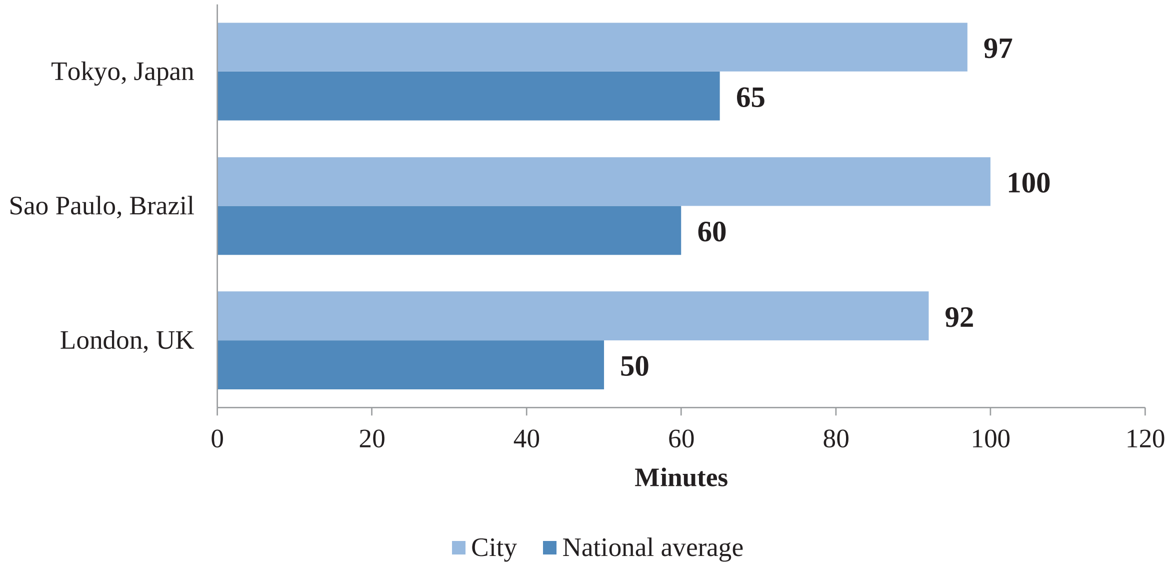


Figure 9. Average commuting times in London, São Paulo, and Tokyo compared to their national averages (Department for Transport, 2015; Metro/SP, 2013; Kobayashi, et al., 2011)

Adopting average travel times is a simplification of the highly complex nature of movement within cities, but these values help illustrate the generally longer times found in megalopoleis that break from the stability found in other scales of urban agglomeration. This perspective builds on the work of Zahavi (1974), Marchetti (1994), Laube et al. (1999), and others where the average journey time is used to understand the interaction between travel speed and urban area, even though trip lengths vary by purpose and type. A broader view is thus adopted by this research as a starting point for capability definition later on.

Longer travel times affect more than just the generalised costs of travel. They have significant impact on people’s health and well-being. Research has shown that longer commutes directly affect people’s sense of happiness and levels of anxiety, especially those lasting between 61 and 90 minutes (Office for National Statistics, 2014). In addition, longer commutes have been found to increase blood pressure and calorie consumption, as well as to reduce levels of fitness activities (White & Rotton, 1998; Royal Society for Public Health, 2016). More specifically, those commuting by car seem to suffer from higher levels of stress and negative mood (Wener & Evans, 2011).

Moreover, where urban growth followed the increase in car ownership, private motorised modes became a necessity rather than a choice. Consequently, the distribution of the impacts has not been shared equally among the population of these cities. Firstly, the socio-economic consequences from limitless urban expansion have been considerably more severe on those without access to on-demand transport, who suffer a greater impact both in terms of travel times and the area they can reach within their budgets (Schaeffer & Sclar, 1980; Banister, 2011). As found by Gyimesi et al. (2011), drivers in large urban areas still tend to be able to maintain their travel times within the normal budgets. Yet, this is not the case for public transport users. Zahavi (1974) had already recognised such imbalance in which households not owning a car that travel by public transport reach their travel time budget much before they reach even half of their cost budget. He also adds that cities in developing countries are farther from equilibrium when compared to their developed counterparts.

In São Paulo, public transport users spend 134 minutes commuting compared to 62 minutes spent by those driving (Metro/SP, 2013). The situation is not much different in Mexico City, where public transport users spend between 118 and 162 minutes commuting, in contrast to 81 of drivers (Instituto Nacional de Estadística, Geografía e Informática, 2007). However, robust transport infrastructure and developed economies do not seem to create any better results. Londoners who travel by rail spend 55% longer than those who drive and New Yorkers on public transport spend almost twice the time of drivers (Department for Transport, 2015; McKenzie & Rapino, 2011).

These results seem to confirm the notions that: (1) urban dwellers tend to dedicate about 70 minutes per day when they can, travelling further rather than saving travel times (Zahavi & Talvitie, 1980; Marchetti, 1994); and (2) that dwellers in public transport end up exceeding their travel time budgets not by choice but because they cannot afford higher speeds (Roth & Zahavi, 1981). Drivers who can afford faster travel are more likely to remain within or close to the 70 minutes estimated by research (Zahavi, 1974; Laube, et al., 1999). This shows that, whenever possible, people will try to remain within their travel time budgets. In addition, the data also suggests that public transport users spend longer on daily travel because they have to, and not because they want to.

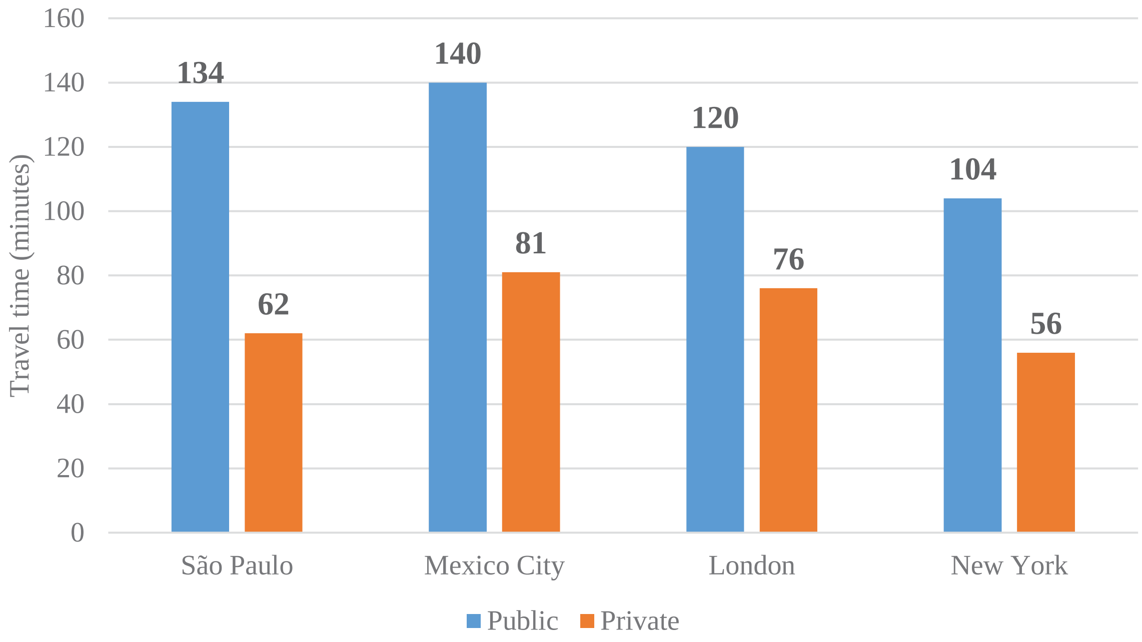


Figure 10. Average travel times of public transport users and drivers in São Paulo, Mexico City, London, and New York (Metro/SP, 2013; Instituto Nacional de Estadística, Geografía e Informática, 2007; Department for Transport, 2015; McKenzie & Rapino, 2011)

Secondly, the discrepant levels of access within an urban area will inevitably create a strong correlation between income and levels of accessibility. When the urban fabric begins to impose longer travel times, location becomes a premium and a defining characteristic whether one can remain within their travel time budgets. Research has shown that prices rise around locations that will reflect shorter travel times. In cities with a robust transport infrastructure, this is usually the case along rail corridors (Loo, et al., 2010). On the other hand, in developing countries where all travellers have to endure higher levels of road congestion due to the lack of options, there tends to be an inversion of the urban geography in which those on higher incomes live in the inner areas and those on low incomes are pushed to the peripheries where access is scarce (Gwilliam, 2003). One example of this particular instance is São Paulo, where traffic speeds average between 15 km/h and 20 km/h during peak times (Companhia de Engenharia de Tráfego de São Paulo - CET/SP, 2017). Unsurprisingly, those on higher incomes populate the inner urban area, stretching up to 12 km from the city centre in order to secure their travel time budgets (Instituto Brasileiro de Geografia e Estatística, 2011).

In essence, the assumption that people will travel as far as they can afford only works when the transport network can provide adequately high speeds. Whenever there is discrepancy in speed between modes, the urban fabric will readjust where people will seek ways with which they can remain within their budgets. When conjunctural, Strano et al. (2015) have found that mode readjustments happen even at the microscopic level, with daily choices between the car and the Underground in London according to the congestion experienced the day before. But the luxury of choice remains available only to those who can afford all modes of transport. Nonetheless, when the disruption is structural and often as the previous example of São Paulo, then people will relocate accordingly, thus reinforcing a feedback loop of inequality and lack of accessibility.

Thirdly, the environmental burden of the dominance of private motorisation in most of large urban areas has created a paradigm of high levels of mobility and consequently of high energy consumption. When one considers the size of megalopoleis, long-distance motorised trips seem inevitable, yet the extensive use of private modes exacerbates the issue. It is widely documented that across the world energy consumption and greenhouse gas emissions per passenger-km are significantly lower for public transport users (Kenworthy, 2003). Unsurprisingly, the transport sector is currently responsible for 36% of the total energy consumed globally, and road transport alone is responsible for approximately 60% of that amount (International Energy Agency, 2017). Left unchanged and added to the increasing motorisations levels in the developing world, the environmental burden from the transport sector can increase substantially, comprising an increase in up to 82% in energy consumption and 79% in CO_2 emissions by 2050 (World Energy Council, 2011).

Finally, what acts as a solution in low density settings becomes an issue when populations grow. There are inherent capacity limitations to road networks which create severe impacts in terms of time and costs. In context of high population in megalopoleis, corridors quickly saturate, leading to congestion and subsequently lower average speeds which stretch travel times further than the allocated budgets (Trigg, 2015). Whereas cities expanded on the basis of traffic-free average speeds, the induction of congestion severely affects the average speeds of road-based modes, especially bus users. Researchers have found that congestion costs can add up to up to 7.5% of the city's GDP (Timilsina & Dulal, 2010; Cintra, 2014).

It logically follows that a city that offers efficient public transport can limit traffic congestion, reduces pollution, conserves energy and promotes social equity (Cervero, 1998). Socially, it increases access for non-drivers, and this freedom of access has been central to the liberal notions of equal access and of economic opportunity (Sommer, 2012). Since it does not involve the ownership and maintenance of a private vehicle, public transport promotes a more democratic access to opportunities and is unlikely to price users out based on fares. Moreover, public transport systems such as metros can offer a much higher capacity than road networks, a feature which is key to ensure acceptable travel times in megalopoleis.

However, the same liberal interpretation of democracy also encompasses the freedom of choice. The choice of transport is influenced by several factors, including irrational feelings that tend to favour the car as providing a better service (Handy, et al., 2005; Steg, 2005; Beirão & Cabral, 2007). More importantly, there is a strong correlation between the ratio of end-to-end public and private transport speeds and the utilisation of public transport, as shown in Figures Figure 11 and Figure 12. In other words, people tend to rationally choose whichever mode is fastest, considering the overall door-to-door speeds. Consequently, significant differences in door-to-door speeds between private and public transport are known to perpetuate a vicious cycle that reinforces the sprawling urban structure and increasingly hinders the efficiency of public modes of transport (Laube, et al., 1999).

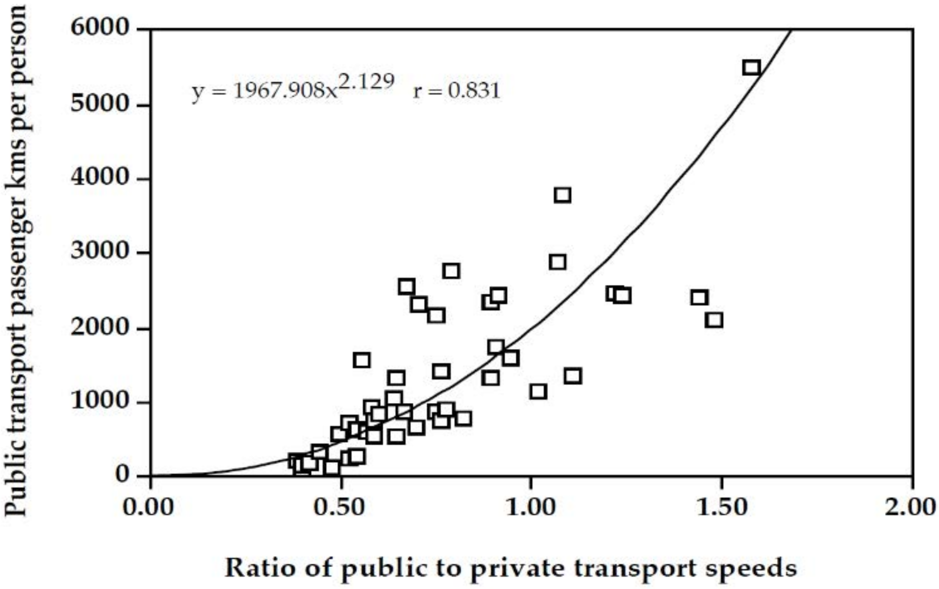


Figure 11. Relationship between the relative speed of public and private transport and the use of public transport in cities (Laube, et al., 1999)

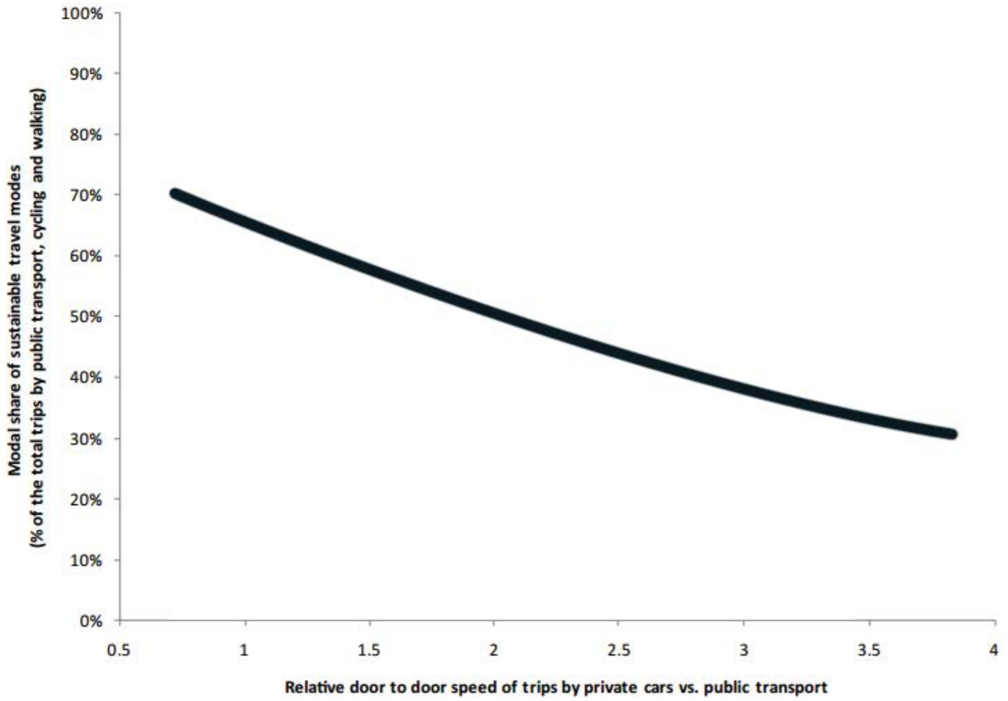


Figure 12. Modal share of sustainable travel modes according to the ratio of the door-to-door speed of private cars and public transport *(*adapted from (Pourbaix, et al., 2015))

Therefore, it is not surprising that, throughout the decades, many have proposed various solutions for increasing the door-to-door speed of public transport. In various distinct circumstances, engineers have proposed specific technical solutions to shorten trip components so as to adapt to the ever-increasing distances. Tough and O'Flaherty (1971) compiled many ideas developed in the late 19th and early 20th centuries that resemble passenger conveyor belts, such as the famous Adkins and Lewis 'Never Stop Railways' showcased at the Wembley exhibition in 1925. More recently, attempts focused on reducing the number of intermediate stops between origin and destination in order to increase average travel speeds, prominently exemplified by Personal Rapid Transit systems (Anderson, 2000). Some others focused on increasing the maximum line speed in order to connect ever greater distances into megaregions within shorter travel times (Blum, et al., 1997; Musk, 2013; Pagliara, et al., 2015; Rutkin, 2016; Zheng & Kahn, 2013). However, these visions of boundless futures tend to overlook important paradoxes that arise from the complex interactions between trip components and not necessarily reduce door-to-door travel times as a whole (Givoni & Banister, 2012). As covered in more detail later in this thesis, although these systems successfully overcome some of the specific limitations of trip components, they would incur extra penalties elsewhere in terms of time or capacity.

While an increase in the average-door-to-door speeds of public transport systems is crucial to attract users from private modes and, consequently contribute to a more sustainable urban environment, a paradigm shift as such is not a simple task. Just as the aforementioned attempts have failed to fully address the issue, other robust operational strategies which will be discussed in further detail on Section 4.2 have also fallen short of a system-wide solution. It seems, therefore, that more than technical and technological solutions, the emergence of the megalopoleis requires a paradigm shift in our epistemological approach to problem solving.

2.3. Summary

Over the centuries, urbanisation seems to have been an inevitable process due to scaling laws of enhanced socio-economic activity and resource efficiency. Considering the stability in travel times, cities have generally been about one-hour wide, meaning that they grew quadratically to the door-to-door speeds offered by modes of transport. It logically follows that, the faster people can travel, the larger urban areas can expand. The widespread affordability of private and semi-private modes can be accounted to the birth of most megalopoleis, simply for the fact that they simplify trip components and offer greater overall travel speeds. However, the relatively recent process that culminated in the dominance of megalopoleis challenge the sustainability of cities in the future. Such dominance has led to significant concerns in terms of economic stability, social equity, and environmental health. It created a paradigm of low accessibility to public transport users that impact their economic opportunities and reinforce the preference for private modes in an already saturated road network. Most experts in the literature agree that the key lies in improving public transport systems to offer the same door-to-door travel speeds, as people rationally choose the fastest mode they can afford to. Nonetheless, many advocate simplistic solutions that overlook the complex interactions between trip components. It becomes clear that a new approach is needed, one that enables radical innovation to the new reality of large cities.

3. LITERATURE REVIEW

3.1. Introduction to the epistemology of engineering

The emergence of megalopoleis and their heightened impacts on urban mobility seem to require a different approach to problem solving to conventional processes in engineering projects. Despite the fact that these cities incrementally grew in size, it does not necessarily mean that they are just a problem of proportion and that incremental upgrades to infrastructure will suffice. As it will be explained in detail in Chapter 4, these urban areas have created anomalous dynamics, which reflect in brand new problems for transport systems. It is perhaps necessary to take a step back and analyse the epistemological foundations of engineering in order to understand and formalise the paths to radical innovation.

As Koen (1985) highlights, the field of engineering is intrinsically related to innovation as it aims to create change with the use of resources in the form of new solutions. Yet, its methodological definitions are not as clear as those for traditional scientific methods. Engineering, consequently, often derives many of its methods from the physical sciences, purposefully or not, even though the aims of the fields are somewhat distinct. In that sense, one could argue that confusion arises when engineering methods are mistaken with those from the physical sciences (Koen, 1985). One of the aims of this chapter is to highlight the important distinctions between the traditional methods in the philosophy of science and the aims in the practice of engineering of future change under uncertain conditions. Firstly, there are the different contexts of justification and discovery regarding ideas and hypotheses (Koen, 1985; Schickore, 2014); secondly, the important distinction between deterministic and teleological approaches to problem solving (Dreborg, 1996; Banister & Hickman, 2013); and finally, methods that are based either on reductionist or holistic views (von Bertalanffy, 1967). These dichotomies will be explored in the following literature review sections in order to convey necessary epistemological understandings for achieving specific results in engineering processes.

3.2. Heuristics and the logic of discovery

For centuries, the literature of philosophy of science has been debating over two main approaches for creating and testing new hypotheses, namely the logic of discovery and the logic of justification. The main contrast between the two stances is that the former concerns the generation of a new hypothesis, and the latter tests and verifies it (Schickore, 2014). In other words, the context of justification relates to the validation of hypothesis using known processes whereas the context of discovery uses abductive reasoning and exploratory methods to create new ideas. In midst of the debate, Popper (1959) was one of the main critics of the logic of discovery and advocated for the limitation of science to problems of justification. For him, only problems which can be analysed using well-defined scientific methods can be deemed as scientific. For them, the 'logic' of discovery was not scientific because it involved irrational and intuitive processes that cannot be examined logically (Reichenbach, 1938; Braithwaite, 1953).

In that perspective, scientific investigation could only be guaranteed by standards and procedures that enable the falsifiability of hypotheses (Popper, 1959). This is an important foundation for scientific advancements because it adds to the scope and precision of scientific procedures. This epistemic view relates to an algorithmic approach, which tests hypotheses by following specific step-by-step instructions (Ormrod, 2007). It reflects the philosophical stances of logical positivism and is pervasive in the physical sciences thanks to its characteristics of replication, falsifiability, and empiricism (Kuhn, 1962). The main aspect of algorithms is its ability to test hypothesis using knowledge that has already been proven so that an optimal solution can be found and verified.

However, although the logic of proof and the context of justification bring important benefits to endeavours in the physical sciences, it may not always be the case in the engineering realm. That is because while the physical sciences are mainly concerned with the explanation of observable facts in an external context, the practice of engineering is inherently attached to the creation or discovery of new solutions to societal problems (Koen, 1985). Engineering is naturally about problem solving. In that sense, one can argue that the logic of justification insulates the scientific paradigm from innovation and protects it from the possibility of new discoveries.

> "We have already seen, however, that one of the things a scientific community acquires with a paradigm is a criterion for choosing problems that, while the paradigm is taken for granted, can be assumed to have solutions. To a great extent these are the only problems that the community will admit as scientific or encourage its members to undertake" (Kuhn, 1962, p. 37)

Kuhn (1962) likens the context of justification to puzzle solving rather than problem solving, exactly because the logic of proof behind it limits the types of problems that can be solved to only those that already fit the current scientific paradigms. Traditional research, thus, is a cumulative process and its success derives from regularly selecting problems that can be solved with conceptual and instrumental techniques which already exist. While this contributes to rigour in the testing of hypotheses, it is only concerned with the extension and articulation of existing paradigms, and thus it does not aim at novelty nor prepares the scientific paradigm to deal with anomalies or extraordinary phenomena (Schickore, 2014).

This is perhaps well illustrated in the work of Altshuller (1984) in the attempt to create an algorithm for technical invention in engineering processes. The resulting methodology, namely TRIZ (translated from Russian as 'theory of the resolution of invention-related task'), proposes 40 algorithms for problem solving (Altshuller, 1999). As highly applied processes, these algorithms seem to be generic enough to be applied to a variety of situations, but too broad to provide universally replicable step-by-step instructions for problems of all kind.

In contrast, engineering often deals exactly with the novelty of unprecedented circumstances which require a new solution (Koen, 1985). In these cases, not all problems have a defined set of steps for an optimal solution, or some problems require a degree of innovation that the logic of justification cannot lead to. They can be so due to the inherent complexity of a problem, the extraordinary nature of a problem, or the absence of certainty regarding the existence of an optimal solution (Kuhn, 1962; Tversky & Kahneman, 1974; Polya, 1945; Michalewcz & Fogel, 2004; Ormrod, 2007; Schickore, 2014). From that point, the context of justification and the context of discovery follow distinct and sometimes almost opposite paths in their epistemic foundations.

Schickore (2014) highlights that the particular purpose of the process of discovery is explaining anomalies or surprising phenomena. Contrary to the logical positivists, Hanson (1958) argued that there is in fact a ‘logic’ of discovery in the form of a process of abductive inferences. Despite the unusual name, abductive reasoning is not uncommon for engineers. It works backwards from the novel problem to an explanatory hypothesis, considering that the problem has been solved and therefore is not new anymore. In that sense, scientific discovery becomes a form or problem-solving by constructing heuristics in order to efficiently search for solutions (Schickore, 2014). There is a logical reason for that, since there cannot be a pre-defined set of steps that leads to an unprecedented destination. Therefore, a new idea is hardly ever or never just an increment to what is already known, but a revolution in itself (Kuhn, 1962).

The context of discovery involves a set of heuristic principles, in that they become significantly helpful in the realisation of novelty for paradigm shifts (Kuhn, 1962; Schickore, 2014). The definition of heuristics can take many different forms in the literature. Heuristics can be seen broadly as an exploratory endeavour in contrast to the specific set of steps or rules of algorithms (Carmichael, 1930). They have an important part in the field of mathematics and computing, where problems can require prohibitive processing power or time. For these cases, heuristics provide a set of rules of thumb to guide problem solving, which avoid undesirable states in the problem space (Lakatos, 1976; Schickore, 2014). In computer programming, heuristic algorithms perform exploratory searches for solutions which may or may not be optimal in a context where an optimal solution can be unfeasible (Lin & Kernighan, 1973). In mathematics, Polya (1945) and Michalewicz and Fogel (2004) propose heuristics as various alternative methods to solving novel mathematical problems for which an algorithm process is not known. They list many procedures, such as analogy, and also the recurrent theme of working backwards from a hypothetical solution. The reason for doing so is a natural mental process explained by psychologists. Tversky and Kahneman (1974) define heuristics as judgements we make in situations of uncertainty that surpass our immediate cognitive ability. Amongst the many examples, Kahneman (2011) shows that when confronted with problems that one is not sure there is a solution, the natural reaction is to reduce the problem to a less complex analogous problem, and then infer the solution back to the first.

Nonetheless, the context of discovery can have a more significant impact when considered as an epistemological foundation for problem solving. In a broader sense, the context of discovery and its respective heuristics widen the potential solution space for exploration. As they have the ability to detach from current processes and scientific paradigms, they can prevent cases of '*solutioneering*' in engineering problems, that is, jumping to solutions without defining the problem (James, 1984). From an epistemological point of view, these events happen when one applies the logic of justification in the development of a new idea or hypothesis. Armed only with the tools already in use, it becomes impossible to tackle problems of a different paradigm.

The result is then limited to an optimised version of previous solutions for old problems rather than a solution for a new problem.

It is difficult to devise universal methods for engineering because heuristics and the logic of discovery are an identity of it (Koen, 1985). Thus, the relevance of this approach to engineering is that it allows engineers to project novelty in a logical manner. Engineering projects are born from the context of discovery because they are inherently related to solutions to novel problems in society. However, the subsequent processes of investigation and validation of projects also requires the traditional methods of the context of justification, because they are supported by the natural laws of the physical world.

It seems that engineering has two sides, and simultaneously accommodates both the context of justification and the context of discovery in its processes. The context of discovery relates to situations during the project lifecycle in which decisions of change have to be made. On the other hand, these decisions need to be verified and validated using the context of justification. These include problems of optimisation, for which algorithms can be devised to run and test the solution. But at that point a solution has been already devised so that there cannot be levels of uncertainty, otherwise it is impossible to find the best solution when not all parameters are known.

However, for circumstances of radical and future change like the one approached in this research, the extraordinary nature of the problems is unlikely to be solved with algorithmic processes rooted in the context of justification. While heuristics cannot be assessed by the merits of falsifiability, they carry an internal logic that can be judged by the merits of fallibility (Lakatos, 1978). It is impossible to deduce causality in projects which are going to be finished in the future (Koen, 1985). For the same reason, the context of discovery serves as an epistemological foundation to radical change, especially when combined with heuristic methods that permit open-ended exploration of solutions in a normative way.

3.3. Foresight methods and radical change

> "If there is such a thing as growing human knowledge, then we cannot anticipate today what we shall know only tomorrow. […] No scientific predictor - whether a human scientist or a calculating machine - can possibly predict, by scientific methods, its own future results."
>
> (Popper, 1961, pp. xii-xiii)

The challenge in causing change is that goals and environments also change during the engineering process, as a natural result of the long time between the start and the completion of a project. Using the example of megalopoleis, the continuous process of urban expansion and changes in social and spatial dynamics add extra challenges to the efficacy of engineering projects. In that sense, the exact final state of a project tends to be uncertain at its beginning (Koen, 1985). Technological developments, disruptive events, and societal changes in direction can all transform the purpose or the relevance of man-made systems. Walsh et al. (2015) highlight that both developed and developing countries are facing significant challenges in their critical infrastructure systems due to changing environments in the form of population growth, financial constraints, and accelerated environmental degradation.

Radical changes in man-made systems are needed when current capabilities no longer fulfil the purposes of the system, or when external trends lead to undesirable future outcomes. Christensen (2000) points out that the limitations of incremental innovations are based on the law of diminishing returns. It means that, at a certain point of maturity, greater engineering efforts are required to achieve increasingly small marginal enhancements in system performance because the technologies involved are reaching their maximum capability.

Contrary to the bathtub model which focuses on the reliability of a specific system, the S-shaped curve in Figure 13 illustrates the life-cycle of technological capabilities. For that reason, Saviotti (1986) points out that when the environment is constant, incremental adaptations in systems may suffice, yet under changing conditions, adaptation can only be achieved by a radical redesign leading to a completely new internal structure. Although emerging disruptive technologies might initially provide a lower performance, they differ from current technologies in their potential to achieve future demands. Incremental evolution then reaches a point where performance can only be surpassed by another radical redesign.

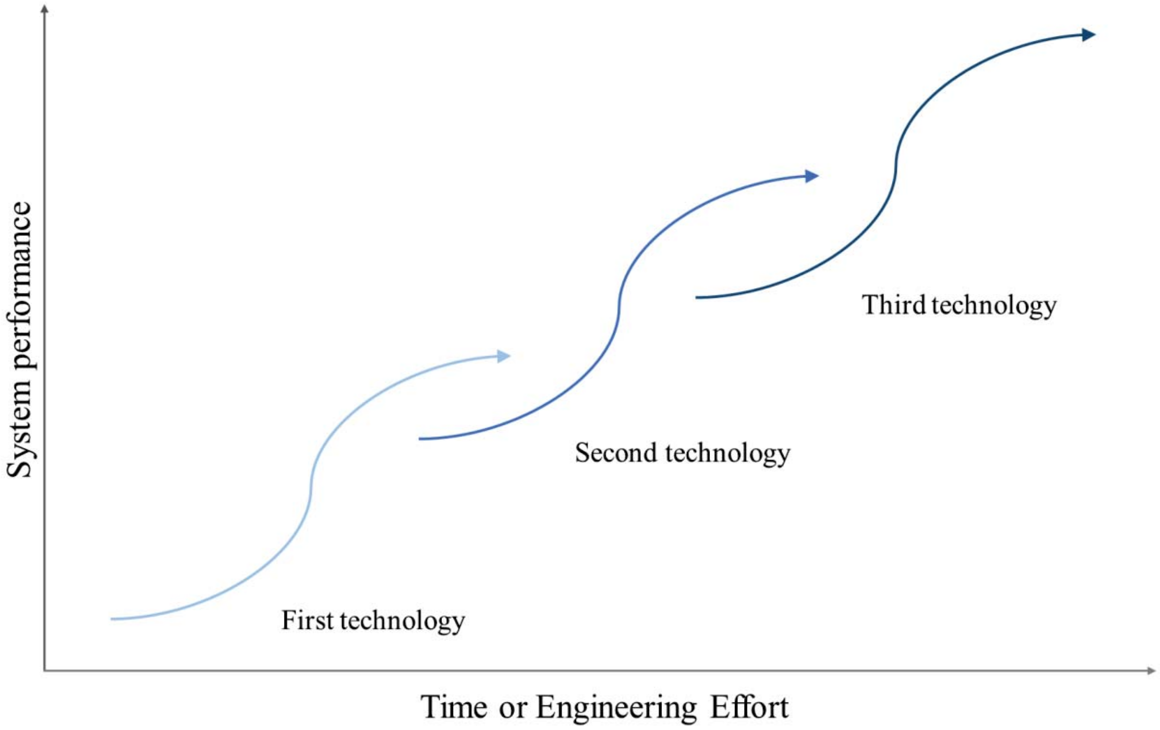


Figure 13. Patterns of radical innovation for engineering effort and performance (Christensen, 2000)

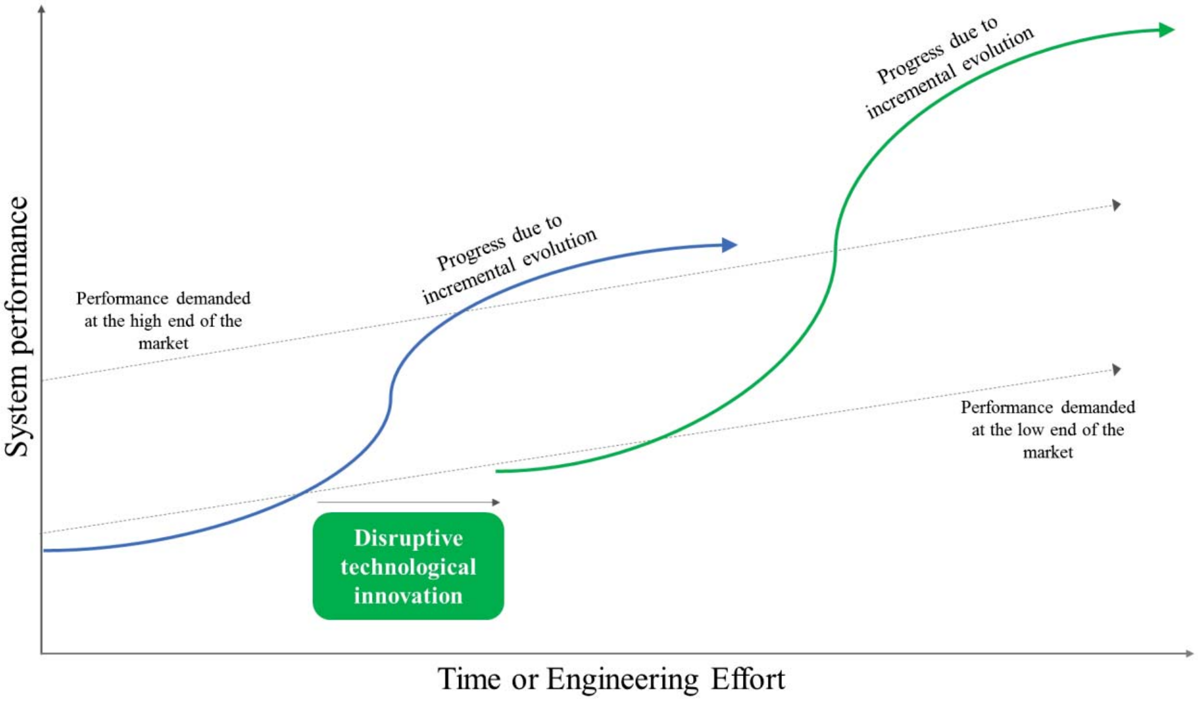


Figure 14. The effect of disruptive technologies on system performance over time (Christensen, 2000)

In infrastructure systems, radical change has to be associated with the future due to the large scale of these engineering projects (Koen, 1985). In the literature, however, the term radical change seems to be more prevalent in the areas of management and business (Perez, 1983; Orlikowski, 1993; Norman & Verganti, 2014). The less prevalent discussion about radical change in engineering could be associated with the fact that the methods of the natural sciences (those in the context of justification) have been the main drivers of engineering pursuits (Simon, 1981; Banister & Hickman, 2013; Dreborg, 1996; Masini, 2006).

As a consequence, such radical changes are highly unlikely in the current approach to infrastructure systems because the perspective is fragmented without an overall vision of what they should be. This in turn weakens their resilience against the changing paradigms in social organisation and the physical environment (Council for Science and Technology, 2009). Moreover, there are strong concerns regarding the usefulness of forecasting in the study of highly complex, long-term problems where futures in which braking with trends are required

(Banister & Hickman, 2007). By differentiating the natural sciences from the ‘sciences of the artificial’, Simon (1981) points out an important distinction in the approaches of determinism and teleology which are the basis for the different methods of foresight.

Foresight as an interest is a natural characteristic of mankind. In fact, the ability to mentally simulate potential scenarios and act upon them has been praised as the ultimate evolutionary advantage that makes us uniquely human (Suddendorf & Corballis, 2007). Humans, of all animals, are among the few if not the only capable of anticipating actions based on the mental modelling of scenarios, and that had an enormous impact on the shaping of society (Suddendorf, 2013). As a discipline and a method of investigation, however, foresight traces back to the World War II and nuclear warfare (Masini, 2006; Anderson, 1973). At that point, concern began to grow not only about what futures were probable, but also what futures were possible, and even more, what futures were desirable. With the technological advancements that followed the industrial and economic development of the time, added to the increasing concerns with the impacts of human action, futurism became a prominent area of interest, highlighted in the works of von Kármán (1946), Kahn and Wiener (1967), Bertrand de Jouvenel (1967), and Meadows et al (1972). Nonetheless, these studies which peaked in the 1960s have since slowly lost momentum and been relegated among other disciplines (Marien, 2002).

One formalisation of foresight takes the name of futures studies and can follow three different approaches for investigation: deterministic, exploratory, and normative (Masini, 2006; Banister & Hickman, 2013). The different approaches are not exclusive and in fact will be shown in the chapter to work better if used in synergy. This section will explore the differences between them and discuss their aptitude to enable radical change.

3.3.1. Deterministic approaches

Deterministic approaches are dominant in the natural sciences and seem to take prominence in the field of engineering due to their strong tradition. The epistemic stance of determinism is one of pure objectivity and closely related to the natural sciences, where the observer has no ability or intention to change or influence the observed phenomena. In many of the sciences such as physics, chemistry, and biology, determinism is the tool which ensures that context of justification is ensured. Based on Popper's (1959) proposal of falsifiability and replication as the basis of scientific knowledge, only predictive hypothesis can be tested. Thus, as the logic of justification looks for direct causality, the deterministic approach consequently adopts a perspective in which the observer cannot interfere in the observed phenomena so that the hypothesis can be unbiasedly confirmed or refuted. To do so reliably, deterministic processes need to operate in closed systems and controlled environments.

Therefore, deterministic approaches are inherently based on forecasting methods to predict the future. A forecast is a probabilistic statement, on a relatively high confidence level, about the future (Jantsch, 1967). The traditional way in which forecasts have been used is to identify the trends between the past and present, and then 'run the trend' or extrapolate them into the future (Banister & Hickman, 2013). Bertrand de Jouvenel (1967) critically refers to this process as trend extrapolation, which derives from the assumption that processes move in the same direction and at the same pace as during a past period. Forecasts are consequently projective processes and must be irrespective of change because they do not compute any scenario but that of extrapolation. Anderson (1973) adds that forecasting can be fruitful for conditions of extreme mechanical isolation, where causality can be confirmed, and hypothesis scientifically tested. On the other hand, many have highlighted that forecasting methods cannot cope with the complexity of open systems and do not carry the ability to direct action away from undesirable trends when necessary (Robinson, 1988; Höjer & Mattson, 2000; Masini, 2006;

Banister & Hickman, 2005). As a result, forecasting is unlikely to generate creative and radical solutions to current challenges (Banister & Hickman, 2013).

Ironically perhaps, forecasting is widely and prominently used in societal affairs even though it is concerned with how things are and not with how things ought to be (Tetlock & Gardner, 2016; Simon, 1981). Much of the anticipation of future states is conducted in the form of forecasts. In Britain, these activities have been conducted mainly in the form of horizon scanning, in which futurists seek to understand the impact of technologies on societies (Niiniluoto, 2001). Horizon scanning can also be used in the evaluation of different scenarios, as discussed in the next sub-section. The results of such projections are then used to steer governmental decisions. However, although the scanning process does lead to action in the face of future possibilities, it is still a deterministic perspective, because it adopts the future as an inevitable extrapolation of the present. Besides the main difference in the respect of change, forecasting inevitably diverges in terms of objectivity and complexity to planning.

The future is stochastic and not a simple projection of the past (Niiniluoto, 2001). For that reason, it is almost impossible to have a science of the future in the shape of the logic of justification (Jouvenel, 1967). Even famous forecasters make wrong predictions, which shows that forecasts are not an exact science especially when dealing with the complexity of societal dynamics (Tetlock & Gardner, 2016). Kahneman (2011) lists many instances, from political experts to stock traders to medical diagnostics, where the forecasts performed equally or even worse than random guessing. In his words, ‘errors of prediction are inevitable because the future is unpredictable’ (Kahneman, 2011). Popper (1961) himself acknowledged that a deterministic stance about the future of society is contradictory for discoveries and new ideas. Godet (1986) highlights the logical incongruence about the nature of forecasts in societal matters where they become either wrong or prove themselves a self-fulfilling prophecy: if a projection highlights the rising of a future problem, then action will be solved and the forecast therefore wrong; in contrast, if a projection points at a positive possibility, then all action is taken considering the assumption to be true and therefore makes the forecast a reality. In that sense, an accurate forecast cannot exist in societal affairs because of the intrinsic interaction between the observer and the observed phenomena. In addition, there seems to be strong correlation between the accuracy of forecasts and the time span of the prediction, meaning that long-term forecasts are wrong more often than not. This has perhaps nothing to do with the techniques but with the complexity of the events they try to predict.

3.3.2. The Delphi method

The Delphi method has been widely mentioned in futures studies and seems to have been suggested for the three different approaches. Nonetheless, it has been placed here because the more common outcomes of the method are either projective or exploratory. Anderson (1973) refers to the method as a means to enhance the reliability of technological forecasts. Melander (2018) provides a comprehensive analysis of Delphi methods in exploratory studies using scenarios. Furthermore, Robinson et al. (2011) and Quist et al. (2011) have suggested a similar participatory method for normative scenarios.

The Delphi method gathers experts' opinions in a systematic way in order to achieve higher reliability in future scenarios. First described by Linstone and Turoff (1975) , the method works as a sequence of *n* iterations where experts anonymously answer questions and give comments. Subsequently, all answers are evaluated and fed again into the individual opinion process. The repeated evaluations are seen as a fruitful heuristic tool for situations of uncertainty such as future scenarios. The first applications of Delphi were in technological forecasts, while now they have been widely used in various other areas. The reason for the wide acclaim of the method is that it expands the pool of possible scenarios for evaluation and thus enhances the robustness of the projections.

By promoting an iterative exchange of ideas between experts, the heuristic potential for innovative ideas increases significantly. Nonetheless, the Delphi method as a tool itself is arguably insufficient for reliable forecasts or robust scenario evaluation for long-term projections. Firstly, because even experts have not shown better results in forecasting long-term events. In several instances, Tversky and Kahneman (1974), and Tetlock and Gardner (2016) have found that experts scored equally or lower than random guessing in predicting long-term events in complex contexts. Secondly, because projections of futures and elaboration of alternative scenarios do not promote change unless followed by actions. As a result, exploratory approaches have flourished with or without the use of the Delphi method.

3.3.3. Exploratory approaches

The main defining characteristic of exploratory approaches when compared to their deterministic counterparts is the acknowledgement of the likelihood of change in the future. They have been the most commonly used form of scenario-based methods, also known as the French approach or *La Prospective* (Godet, 1979; Masini, 2006; Niiniluoto, 2001; Banister & Hickman, 2013). Exploratory approaches were developed as a response to the shortcomings of determinism in face of the complexity and uncertainty of the uncontrolled environments of the real world. In contrast to situations of mechanical isolation necessary for forecasts, the exploratory approach embraces the fact that the future in societal matters is complex and likely to change (Godet, 1986; Niiniluoto, 2001).

From an epistemological perspective, prospective methods depart from the traditional forecast-based approaches because their thinking is based on the ideas of non-determination and actions towards the future (Godet, 1986). The reason for doing so is to challenge current practices and trigger new thinking (Banister & Hickman, 2013). In La Prospective, the main focus is on the envisioning of alternative futures besides the most probable one. These futures are described by different scenarios based on changes that can possibly or plausibly occur in the meantime. Futures scenarios are therefore heuristic tools for foresight to envision possible futures in order to improve decision-making (Berkhout & Hertin, 2002). The great value of a scenario is being able to take complex elements and weave them into a story which is coherent, systematic, comprehensive, and plausible (Coates, 2000). Masini (2006) points out that while visions of the future are rooted in emerging changes that need to be identified, it is equally important to transform the outcomes of these visions into projects of action in order to break from undesired possible trends that are identified in the process.

There seems to be much less debate on the definitions and uses of scenarios than on the methods for using them. Kahn and Wiener (1967), and Durance and Godet (2010) agree on scenarios as hypothetical future events created to clarify the impacts of their respective decision points (Figure 15). Berkhout and Fisher (2002) add that scenarios are relevant because novelty and surprise are also inescapable features of the future, and more importantly because humans are able to change actions in sight of undesirable trends identified in the array of possible futures. In that sense, scenarios need to lead to practical choices, policies, and alternative actions in order to deal with the outcomes of such prospections (Coates, 2000).

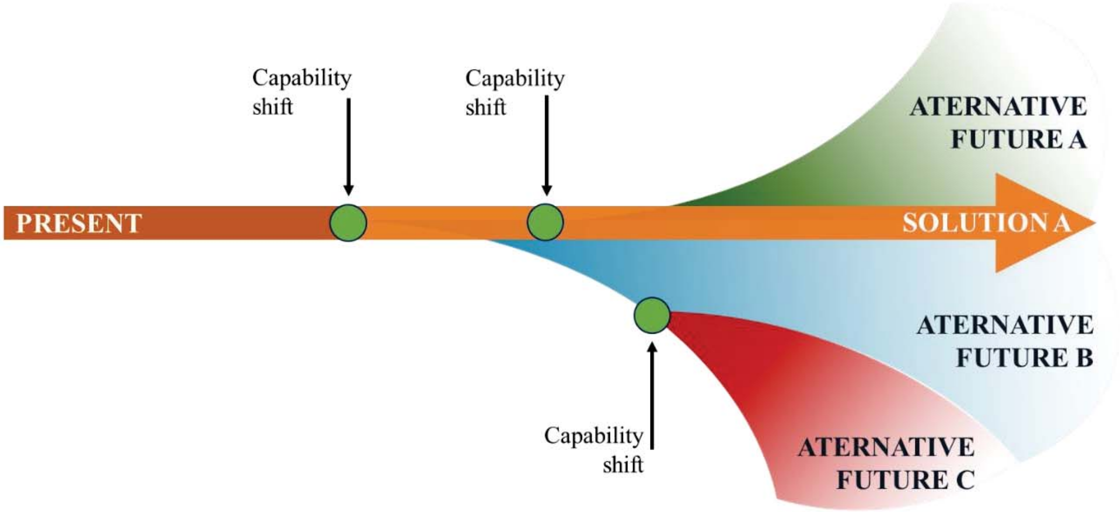


Figure 15. Scenarios as future outcomes of their respective decision points

On the other hand, the number of scenarios necessary for a robust evaluation of possible futures and their outcomes seems to generate more debate in the literature. Banister and Hickman (2013) suggest that exploratory methods are generally comprised of a two-dimensional matrix within which four scenarios are developed. Figures Figure 16 and Figure 17 illustrate the use of four different future scenarios to evaluate changes in selected variables. On the other hand, Durance and Godet (2010) and those of the original *La Prospective* argue that four scenarios are too restrictive and not entirely appropriate for the complexity of social issues. It seems, therefore, that the discussion is not necessarily based on the nature of scenarios but on their application. Banister and Hickman (2013), Gazibara et al. (2010), Gallopin et al. (1997), and Berkhout and Fisher (2002) investigate scenarios from socio-technical perspectives of action, either through policy or projects.

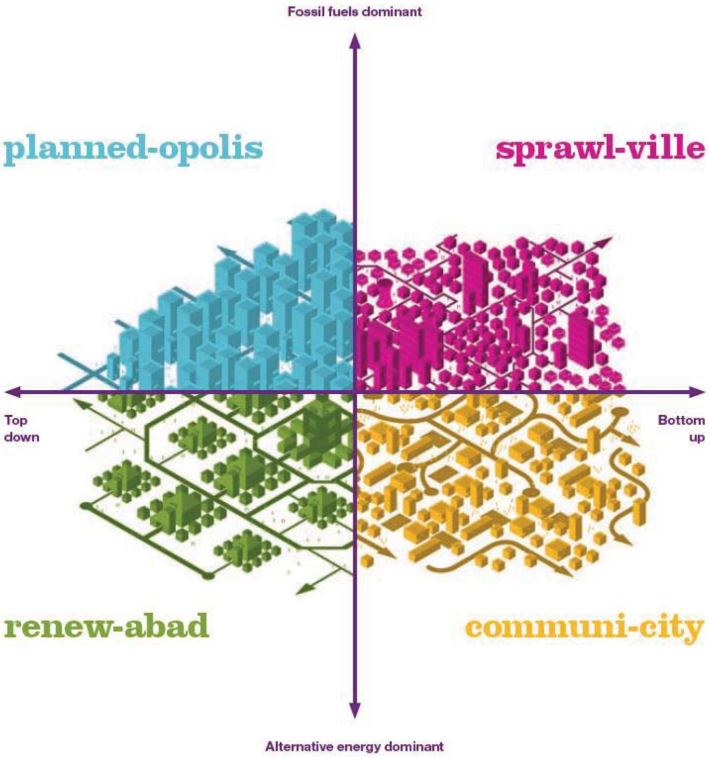
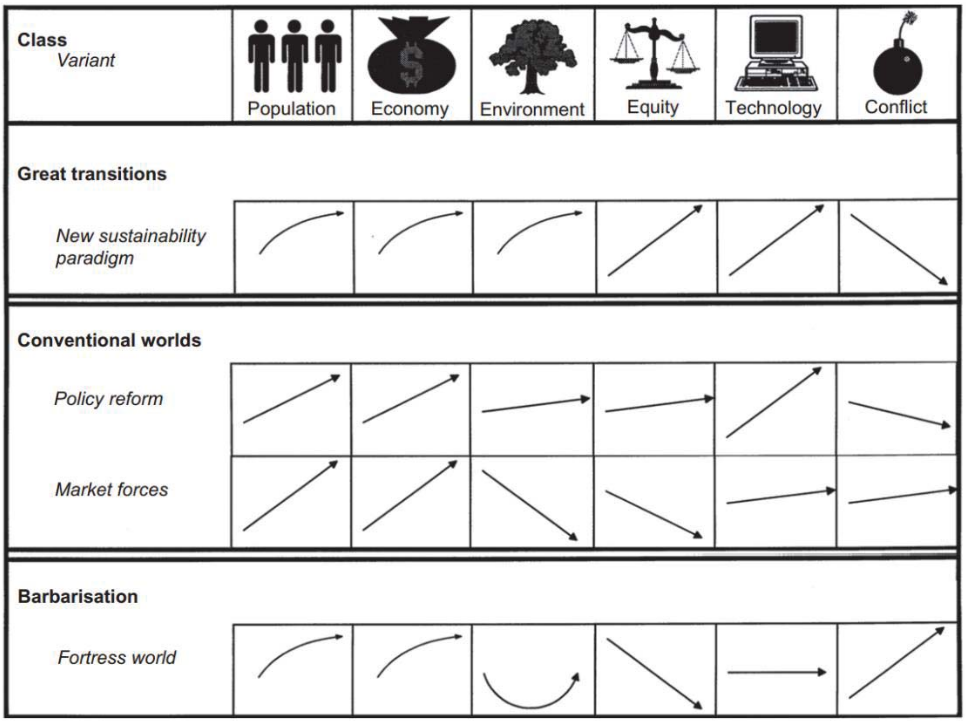


Figure 16. Four scenarios for urban mobility in 2040 (Gazibara, et al., 2010)



Class *Variant*	Population	Economy	Environment	Equity	Technology	Conflict
Great transitions						
New sustainability paradigm	⤴	⤴	⤴	↗	↗	↘
Conventional worlds						
Policy reform	↗	↗	→	→	↗	↘
Market forces	↗	↗	↘	↘	→	→
Barbarisation						
Fortress world	⤴	⤴	⤻	↘	→	↗

Figure 17. Evaluation of the impacts of different scenarios on selected variables (Gallopin, et al., 1997)

Regardless of the number of scenarios used, exploratory approaches are more aligned with the complexities of society and the inevitable changes that follow. Knowing that the unique trait of foresight in humans enables us to change the course of action based on mental models of the future, then it logically follows that the future of human affairs can never be deterministic. As previously said, forecasts of complex issues can never be right, because once forecasts indicate any undesirable trend, actions are taken in order to change the path of events. In that sense, scenarios take a step further from this realisation and investigate the outcomes of such diversions. As a result, planners can evaluate potential futures and respectively prepare for action.

However, although scenarios offer a more realistic and flexible approach to foresight, there are important considerations on the robustness of exploratory approaches regarding their application to engineering projects. Firstly, Durance and Godet (2010) point out that although scenarios are devoid of values, they claim that personal judgement is the most reliable element available, and thus scenario development is inherently attached to the work of the experts involved in their creation. Even if the process is based on the outcomes of a Delphi method, it still inevitably involves expert judgement. On that, it is perhaps important to highlight the inversely proportional accuracy of expert forecasts in comparison to the length of the projection (Tetlock & Gardner, 2016). Secondly, although there is a framework for their formulation, the resulting number of possible scenarios can increase the complication of the evaluation process, which then might require those scenarios to remain broad and shallow for practical reasons. While this is manageable from a policy perspective, it does not provide a specific blueprint for action even in the normative approaches that will be discussed in the next section (Dreborg, 1996; Banister & Hickman, 2005) . Finally, and perhaps consequential of the previous, is that even though exploration prepares for several different futures, it does not necessarily indicate a well-defined path to the most desirable alternative. The more devoid of value a scenario is, the closer it becomes to a deterministic perspective in terms of action, and thus the further it stands from a normative stance which can compute and generate radical change. Therefore, by exploring all alternatives, decisions remain between the plausible, the possible, and the probable.

3.3.4. Normative approaches

At the same time when the exploration of alternative futures rose to prominence, another debate regarding scenarios branched out and created a third approach stretching the contrast with determinism to a maximum. One of the reasons for the sudden revolution within the stance to the future in the 1960s can be attributed to the socio-political situation of the time, in which science and technology were making significant leaps of change. Martino (1969) argued that technology was at that point advancing so fast that it was growing by its own dynamism rather than in response to human needs or desires. In that sense, technological prowess could promise virtually limitless futures which were radically different than a simple continuation of the present, but engineers and scientists simply ignored the impacts of their work on society. As a result, authors began to promote a more structured recognition of the importance of technology in society and a search for methods to manage resources for research and development (Roberts, 1969; Gabor, 1969; Martino, 1969).

The logic behind such claims is that if technology can take human capabilities to virtually any direction, then perhaps the process of foresight should be inverted in that technological development would work in favour of societal goals rather than despite them. Gabor (1969) emphasised that "*technology had become so powerful in shaping society that we cannot plan technology without planning normatively and inventively for the whole of human life*". Roberts (1969) added that the methods used for creating alternative futures were inherently more robust than those used for predicting them, for the sense that exploratory approaches still carry uncertainty while normative approaches are directed towards a specific goal.

Consequently, the concept of normative forecasting became an approach to futures studies in which the focus is on what the future should look like rather than what it could look like. Between the possible and the probable emerges the desirable, which creates a new level of thinking in forms of projects and not probabilities (Masini, 2006). Therefore, the main distinction between deterministic and normative approaches lies on the differences between causality and teleology. Whereas the first explains what has happened and what is likely to happen given initial conditions, the second deals with what should happen, and the actions needed to achieve that desired future scenario.

The reasoning behind normative approaches based on the epistemological notion of teleology is perhaps summarised by a famous quote by Gabor (1963), which says that *"the future cannot be predicted, but futures can be invented"*. From the acknowledgement that current actions will inevitably change the outcome of the future, a normative approach transforms the future into a project rather than predictions as in traditional forecasts. It is impossible to accurately predict the future, especially long-term ones under changing conditions. And, even if it was possible, such information would not be an end in itself because the crucial question is what to do with the range of choices available (Robinson, 1988). Nonetheless, even though normative approaches seemed to more logically fit the almost infinite capabilities unleashed by technological advancements, there was a hiatus between their first appearance in the literature and their renaissance twenty years later. Linstone (1969) describes the resistance to the new approach as a consequence of the natural human inertia to change and fear of the unknown.

The debate between deterministic and normative approaches was restored by Robinson (1988), who introduced the term 'backcasting' for the method opposed to forecasts traditionally used.

Since then, the debate for backcasting concerns the shortcomings of determinism in policy issues, especially when trends lead towards undesirable outcomes. While exploratory methods are known as the French approach, normative methods are also known as the Swedish approach as this is where most development and use of the approach has taken place (Banister & Hickman, 2013). The debate gained prominence in the following years and decades as the concerns about sustainability and trends in energy and transport were found to lead to social, economic, and environmental damage (Dreborg, 1996; Höjer & Mattson, 2000; Hickman & Banister, 2007; Neuvonen, et al., 2014). As Robinson explains:

> "*The major distinguishing characteristic of backcasting analysis is a concern, not with what futures are likely to happen, but with how desirable futures can be attained. It is thus explicitly normative, involving working backwards from a particular desirable future end-point to the present in order to determine the physical feasibility of that future and what policy measures would be required to reach that point*". (Robinson, 1990, pp. 822-823)

In that sense, backcasting is a type of scenario approach, yet the distinction with the exploratory approach lies on the matter of deliberate change. This is not to say that these approaches are exclusive or that there are differences in their merit, but to emphasise the relevance of a normative perspective in certain circumstances. Dreborg (1996) points out that 'backcasting is particularly applicable to long-term complex issues, involving many aspects of society as well as technological innovations and change'. Table 1 describes the epistemological distinctions between normative and deterministic approaches to foresight, relating to the concepts discussed earlier in the chapter.

Table 1. Forecasting and backcasting – five levels (Dreborg, 1996).

	Forecasting	**Backcasting**
Philosophical views	Causality; Determinism; Context of justification;	Teleology; Partial indeterminacy; Context of discovery;
Perspective	Dominant trends; Likely futures;	Societal problems in need of solution; Desirable futures; Scope for human choice;
Approach	Extrapolate trends into the future;	Define interesting futures; Analyse consequences, and conditions for these futures to materialise;
Methods	Various econometric models;	Partial and conditional extrapolations highlighting interesting polarities and technological limits;
Techniques	Various mathematical algorithms;	Heuristic methods;

Backcasting has been widely discussed as a tool for future policy as a means to avoid undesired trends through radical change (Höjer & Mattson, 2000; Banister & Hickman, 2013). However, there is some debate surrounding its definition. Robinson (1990) proposes it as a method composed of well-defined consecutive steps divided into sub-steps (Figure 18).

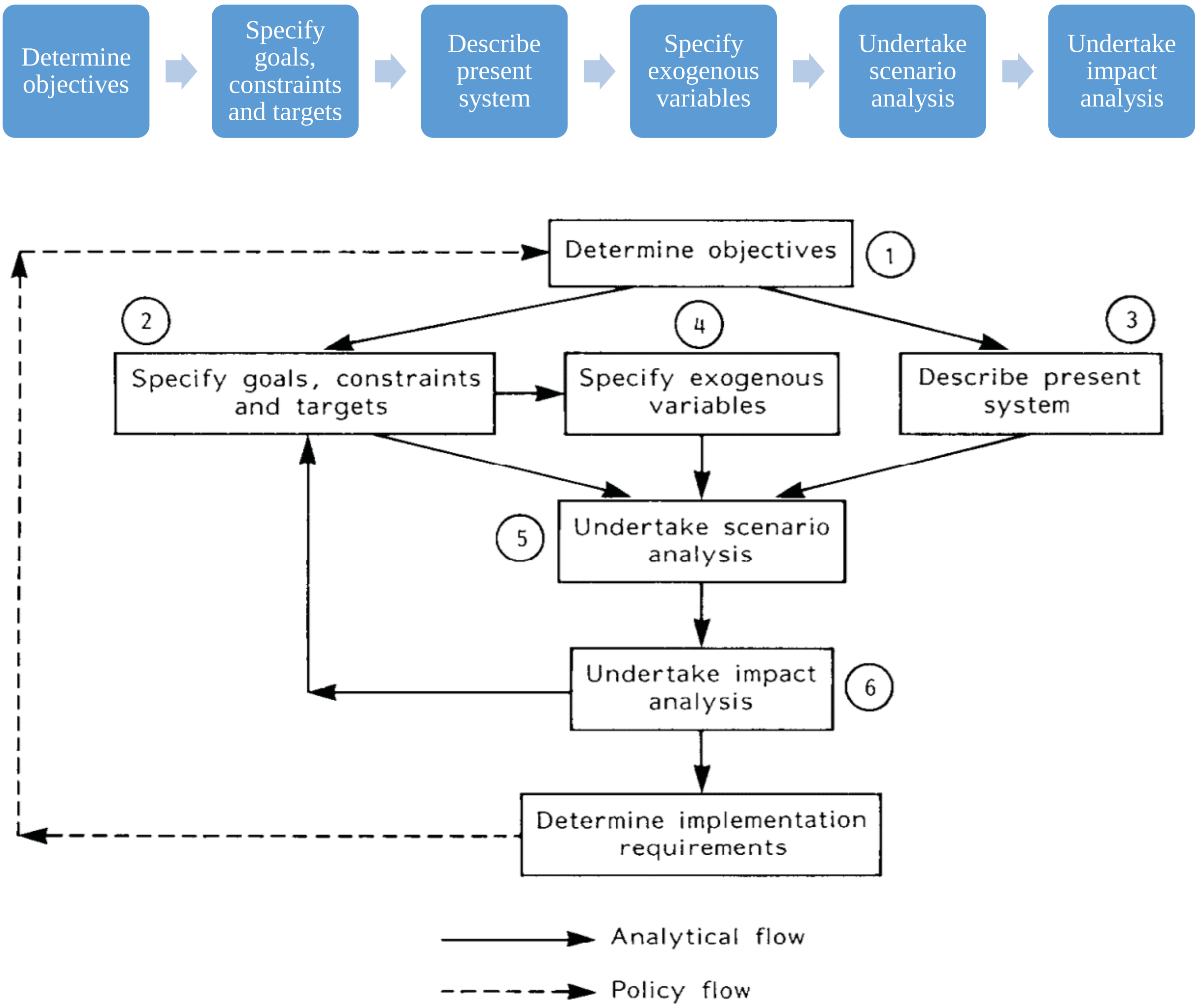


Figure 18. Outline of the backcasting method (Robinson, 1990)

Perhaps one of the reasons for the reluctance from scientists and engineers to adopt backcasting approaches lies exactly on its heuristic nature. Even though forecasts cannot predict more accurately than normative propositions, they still encounter much more support because they adopt the *status quo* and can be moulded within a replicable process, and thus agreeing with the logical positivism of the context of justification. However, Robinson (1990) points out important incongruences to forecasting methods in response to their perceived higher reliability. For instance, even the most neutral 'business as usual' forecasts rely on the assumption that circumstances ought to remain the same, which is in fact equally a matter of chance. In controlled environments where human action is irrelevant, that might be an accurate premise with higher probabilities, yet in societal matters changing conditions are a natural part of the timeline. Despite such contradictions, deterministic approaches are still dominant in the realm of socio-technical systems (Robinson, 1988; Dreborg, 1996; Banister & Hickman, 2013).

Another reason for the preference for forecasting methods is that few backcasting studies are used as a blueprint of the desirable future or a cut and dried action plan (Banister & Hickman, 2005). That is a critical issue when considering engineering projects because engineering and technology are the main drivers of change in the physical realm of society. Since a normative approach treats future scenarios as objectives to be attained through change, one would envision backcasting as an intrinsic part of the purpose of engineering described by Koen (1985). Nonetheless, Dreborg (1996) acknowledges that the approach has remained mostly in the field of policy and planning due to the lack of a technical framework.

The use of normative approaches in a technical context appeared in the literature under the name of Technology Roadmapping (TRM). The term was introduced by Galvin (1998) in an article highlighting its successful use at Motorola and in the semi-conductor industry. Since then, it has been used in business and market research for directing research and development in technology areas (Phaal, et al., 2004; Rinne, 2004; Amer & Daim, 2010; McDowall, 2012). Galvin (1998) explains that roadmaps work by setting future targets to be attained by industry in the form of technical progress or standards. McDowall (2012) adds that TRMs are crucial to innovation and ensures successful outcomes because they comprise expectations (what is thought likely to happen), desires (what is hoped will happen) and promises (what will be made to happen). As in backcasting, the method adopts a normative point of view for the development of industrial capabilities, based on the gap between a view of the future and the current status of a technology.

However, the benefits of standards and specific targets in a normative approach are only straightforward in simple technical systems. As the literature shows, TRM has been used mainly for products and components based on market predictions, and usually with only a few stakeholders from the same industry. In contrast, urban systems such as metros are increasingly complex and complicated systems of systems, with multiple stakeholders and objectives that are not always aligned with market forces (Dreborg, 1996). They require not only the normative approach and the technological independence of the back-casting method, but also a systems engineering process to maintain the robustness of the solution and its reliability, given the uncertainty of future scenarios.

Therefore, while normative approaches can be fruitful in situations of undesirable trends and changing conditions, there still seems to be a gap in the literature that can transform desired scenarios into promised futures through an engineering project. On the one hand, the use of backcasting in order to identify the necessary steps that will diverge undesirable trends into desired states can be very broad and technically insufficient. On the other hand, technology roadmaps ensure specific and verifiable technical steps in order to achieve a certain capability, but the methods rely on a reductionist view of a certain component that involves a small number of stakeholders who usually share similar perspectives. The case of urban transport systems seems to encompass both sides, in that their socio-technical structure carries the complexity of multiple stakeholders, yet their functions rely on their technical and technological capabilities. It can logically be inferred that radical change from the inherent externalities that arise from changing conditions requires the combination of backcasting with approaches that adopt a systems perspective (Dreborg, 1996).

3.4. Systems engineering and the holistic view of things

So far, this chapter has debated two epistemological dichotomies of the philosophy of science and their relevance to engineering projects. Firstly, regarding the division of problems between the logic of justification and the logic of discovery, and subsequently between algorithmic and heuristic methods. Secondly, concerning the different ways to look at the future and their impact on radical change, comprising several approaches that stand within the spectrum between determinism and teleology. Finally, this section will analyse the third dichotomy of the author's epistemic triangles, one between the traditional reductionism of the natural sciences in contrast to the holistic view of the world embracing systems as a whole. It can be seen that these aspects form an interrelated epistemological structure of the understanding of the world and its natural and artificial systems, as shown in Figure 19.

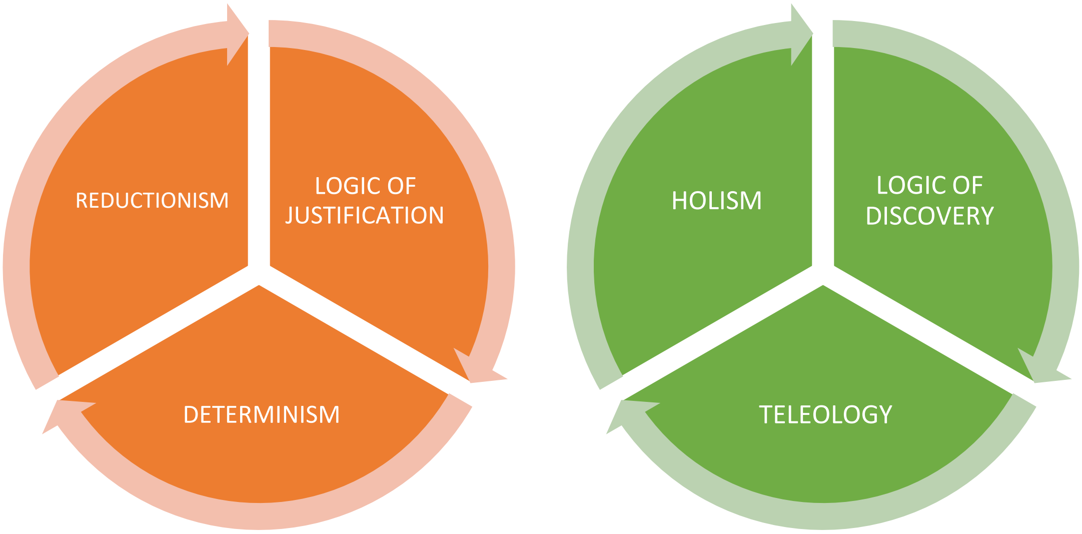


Figure 19. Different epistemological triads in the philosophy of science

Similarly to foresight approaches, the discussion between reductionism and holism as a discipline dates back to the period after World War II. The root of the debate emerged when traditional scientific methods were questioned about their efficacy in dealing with the multiple variables encountered in the real world, not only regarding natural systems but also the increasingly complicated and complex artificial systems in society. In his seminal paper, Weaver (1948) argued that science until the 1900s could only deal with problems of simplicity, which means separating parts from the whole in order to make general mathematical statements about their predictable behaviour. This mechanistic approach was how the physical sciences could ensure the replicability and determinism of experiments rooted in the logic of discovery.

Blanchard and Fabrycky (2006) explain that the traditional worldview is based on two main ideas. The first is reductionism, which provides an analytical understanding of the world where everything can be reduced into smaller parts. By splitting things into smaller units, the perceived level of simplicity increases, and causal relationships can be found. The second idea, which derives from the first, is that of the mechanistic approach, in which the whole is seen as merely the assembly of the smaller parts. Such an approach has been prominent in the physical sciences because teleological notions (such as purpose, needs, etc) are irrelevant to the observation and explanation of natural phenomena.

While the approach of isolation and simplification is relevant to the physical sciences, not all problems in the real world can be solved using such analytical process because systems and their emergent properties are more than the sum of their parts (von Bertalanffy, 1967). The mechanistic view of the world overlooks the important distinction between complication and complexity: complication is a large group of simple relationships in isolation where all entities and their relationships are known at all times, while complexity is an irreducible, unpredictable dynamically changing array of interrelationships. (Snowden, 2002). Based on this inability of scientific methods to deal with the reality of complexity in the real world, Weaver (1948) then classified problems in two other domains besides simplicity, namely organised complexity and disorganised complexity. The first logical perception of complexity is that, since it deals with wholes rather than a conjunction of parts, causality cannot be deduced, and thus it moves away from the logic of justification that can be found in isolated relationships. Disorganised complexity involves problems of uncontrollable dimensions, in which the behaviour can only be estimated using probabilistic methods. They are situations of randomness, which surpass our ability to capture and understand all interactions taking place (Crutchfield & Wiesner, 2010). In that, the traditional methods of scientific investigation could deal well with problems of simplicity and disorganised complexity, as they were either reducible to causal links or chaotic so that nothing else than probabilities could be inferred (von Bertalanffy, 1967).

In contrast, organised complexity comprises problems which involve dealing simultaneously with a sizable number of factors which are interrelated into an organic whole. They are situations where the number of variables is too large to separate into single causalities, but not too large that its behaviour can be understood and modelled. From a philosophical perspective, the distinction of organised and disorganised complexity is not ontological but epistemological, in that randomness derives from technical and technological limitations of human ability (Spinoza, 1996). However, for practical reasons, this discussion will be left aside in this thesis.

The notion of organised complexity is seen as the foundation of systems science, which studies not solely the properties of components but mainly the dynamic interactions between them (Crutchfield & Wiesner, 2010). The definition of a system relates closely to that of organised complexity. The Oxford Dictionary defines it as ‘a set of things working together as parts of a mechanism or an interconnecting network; a complex whole’. Similarly, to the International Council of Systems Engineering, a system is ‘a combination of interacting elements organized to achieve one or more stated purposes’ (INCOSE, 2011). From that, two distinctive characteristics of systems can be highlighted: one of organisation and one of teleology. The first is inherently rooted in the holistic perceptions of the whole as more than just the sum of its parts acknowledged in the Gestaltian psychology (Wertheimer, 1938). Systems science understands that systems are impossible to be understood via a collection of isolated causal relationships because their emergent properties derive exactly from the interaction between components. The second characteristic concerns its purpose. Isolated parts in the traditional methods are neutral. When isolated from their context, they become devoid of any purpose, which is in fact the necessary conditions for the causality of logical positivism to be deduced. On the other hand, systems are inherently purposeful (von Bertalanffy, 1967). While their parts might be neutral by themselves, their interactions with the rest of the system create the emergent functions of purpose.

Subsequently, in a similar way that systems science provided a more comprehensive understanding of the world, systems engineering provides a new understanding of artificial systems which focuses on the growing interactions and interfaces between parts. The Federal Aviation Administration of the United States summarises systems engineering as

> *"A discipline that concentrates on the design and application of the whole (system) as distinct from the parts. It involves looking at a problem in its entirety, taking into account all the facets and all the variables and relating the social to the technical aspect."* (Federal Aviation Administration, 2014)

In other words, systems engineering is the process of creating and analysing man-made systems of organised complexity. This does not mean that systems engineering is exclusive and irrespective of the other engineering fields, but quite the opposite as they complement each other. As Weiss (1969) notes, "*only by shuttling back and forth between the worm's eye view of detail and the bird's eye of the scenery can the scientist gain and retain a sense of perspective and proportion*". Systems engineering is thus not an exhaustive activity in itself, but most fruitful when applied synergistically with the traditional reductionist methods.

Moreover, this definition also stresses the role of social aspects in systems engineering. Drawing from the Cynefin model by Snowden (2002), it can be argued that systems engineering is a systematic process to iteratively analyse a system from complexity to complication and later to simplicity in order to manage and design its parts. It then joins analysis and synthesis in its approach towards the whole life-cycle of the system. Snowden (2002), however, seems to underrepresent the role of human and social aspects in large systems, and therefore underestimate their complexity on top of their complication. Many, if not all, of the man-made systems are in fact socio-technical systems that do not function autonomously but are also the outcome of the activities of human actors (Geels, 2004). These systems, of which transport is one example, have been named socio-technical in order to stress the reciprocal interrelationship between humans and machines that encompasses the production, distribution and use of technology (Ropohl, 1999; Geels, 2004).

Considering the complication and complexity of these systems, the value of systems engineering lies on the added layers of validation and verification throughout the project which results in direct traceability between each component and each requirement and each goal. Even though systems engineering processes rely on physical and functional decomposition of the whole system in manageable pieces, traceability approaches and verification methods maintain the holistic view necessary to deal with the problem. Thanks to common language of requirements during the development phase, systems engineers can achieve more reliable sets of requirements that increases the efficiency and reduces the risks and errors in both the development and introduction into service of a project (Blumenfeld, et al., 2016a). According to INCOSE (2011), cost and schedule overrun of projects reduce with increasing systems engineering (SE) effort and appear to minimise beyond 10% SE effort. Similar findings have been reported by McNulty (2011) in the specific context of the railways (Figure 20). The study conducted by the Department for Transport and chaired by Sir Roy McNulty (2011) highlights that a critical success factor for engineering projects depends on getting early concepts and designs right, because by the time the project has spent 15% of its budget, it has committed over 80% of its total costs.

Figure 20 illustrates the influence of each stage of major projects on overall whole-life costs. Early stages of design and development require a small portion of the costs incurred, yet have a significant impact on the costs committed for the programme. Conversely, once the system design has been agreed and delivery has started, incurred costs increase significantly. This means that changes in the later stages of the programme will incur significant additional cost and time burdens to the project. In addition, it also highlights the importance of systems engineering in the cost-effectiveness of whole-life costs.

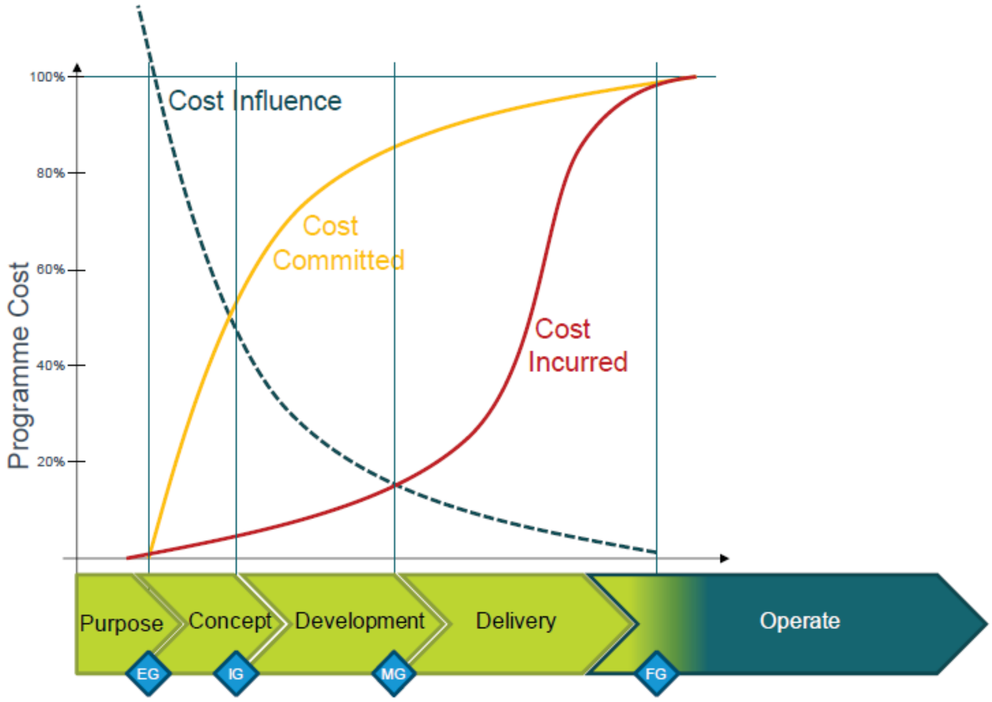


Figure 20. Cost influence, commitment, and spend against programme phases (ATKINS, 2011)
EG – Entry Gate – Authority to Analyse Options, IG – Initial Gate – Authority to Develop an Option, MG – Main Gate – Authority to Invest, and FG – Final Gate – Authority to Transfer Accountability

Therefore, the field of systems engineering is concerned with ‘building the right system right’, so that all requirements are identified and later fulfilled, and the overarching purpose is elicited and then met. In the literature, common characteristics of systems engineering include: (1) a top-down approach that understands systems as wholes and not parts put together (Blanchard

& Fabrycky, 2006); (2) a life-cycle solution that addresses all phases from design to disposal and satisfies customer expectations and meets public acceptability (Schmidt, 1993); and (3) an interdisciplinary approach throughout the system design and development process (von Bertalanffy, 1967; Blanchard & Fabrycky, 2006).

Various processes have been suggested for modelling a system's life-cycle. A general process has now been standardised by the International Organisation for Standardisation (ISO), under the reference number ISO/IEC/IEEE 15289:2017. Yet, several more specific models have been proposed so far. One of the first models was Royce's (1970) waterfall model for software development (Figure 21). It summarises the order of the development process, acknowledging some feedback between steps. However, the model does not compute for efficient verification and validation, as these are only performed between consecutive steps and not between the synthesis and the analysis phases of the project.

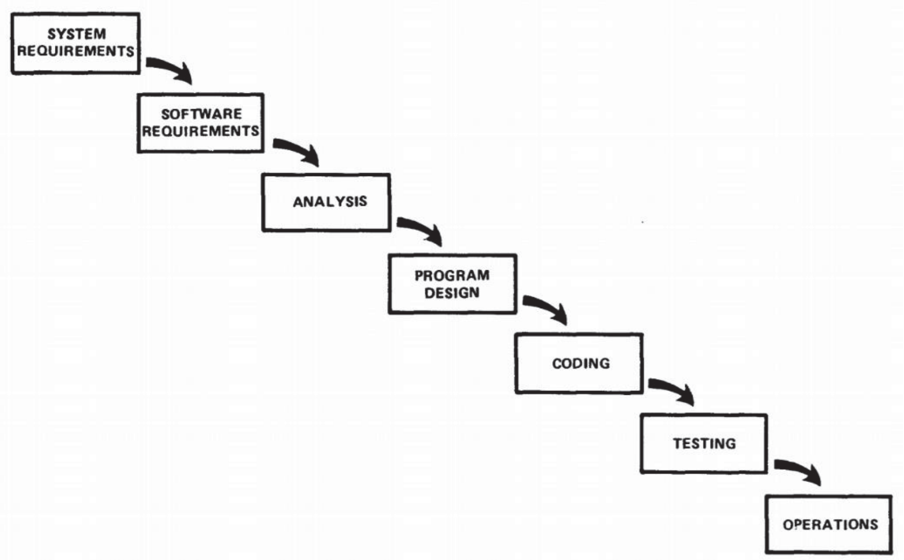


Figure 21. The waterfall life-cycle model (Royce, 1970)

For that reason, Boehm (1988) proposed a spiral model for software development which focuses on risk management (Figure 22). It illustrates a cumulative process, both on the incurred costs (represented by the axis) and project stages (represented by the angle). Each spiral addresses a similar sequence of steps, yet each adapted to the respective stage of the project. An important characteristic to be observed is that the model was created to illustrate software development, which is inherently different than a project of a physical system. For instance, the repeated development of prototypes in each cycle would incur significant costs in a large technical system such as a railway line. The verification of requirements and design, therefore, requires a different approach that is not cumulative such as in the development of virtual solutions where costs are mainly in human hours.

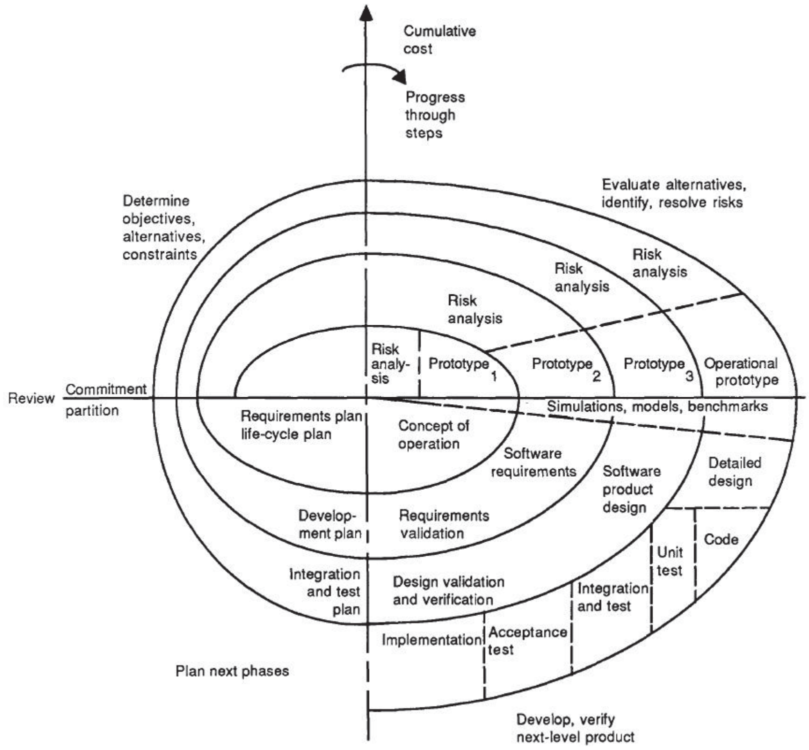


Figure 22. Spiral model of software process (Boehm, 1988)

Moreover, the two models do not clearly indicate the role of systems engineering (Fosberg & Mooz, 1991). With that in mind, systems engineers tend to adopt another life-cycle model, namely the Vee model. First introduced by Fosberg and Mooz (1991), it is now widely used within the International Council of Systems Engineering (INCOSE, 2011). The Vee model, as shown in Figure 23, builds from the waterfall model, but with the difference that the synthesis part is shown going upwards rather than continuing the slide downwards. Nonetheless, this is more than a merely aesthetic change. The V-shape represents graphically the relationship between the analytical and the synthetical stages of the project. In addition, the vertical axis illustrates the level of abstraction involved in each of the processes, which goes from the general to the detailed in the design stages, and takes the inverse direction on integration, verification, and validation stages of the project.

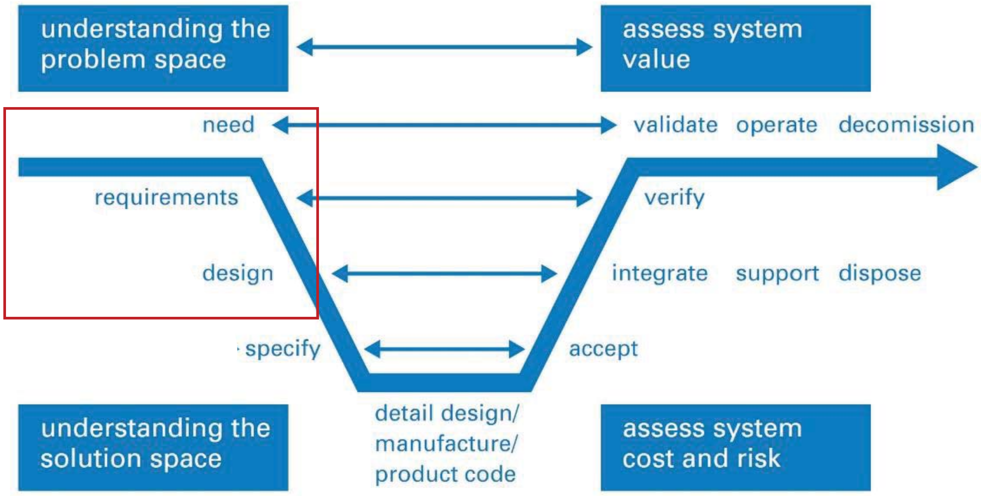


Figure 23. The Vee life-cycle model (INCOSE UK, 2009). Red square indicates the scope of the research within the Vee diagram

The Vee life-cycle model organises the development process from design to implementation. The first half of the cycle is analytical, starting from a holistic need which incurs a defined purpose. The purpose highlights the overall capability desired of the system, and each capability must have a single clear statement of purpose and required outcomes (INCOSE UK, 2014). At this point, as the diagram suggests, the process focuses on clearly understanding the problem.

The desired capability and required outcomes, according to each stakeholder, will then create a set of requirements for the system and navigate towards the solution realm. The Oxford Dictionary defines a requirement as “a thing that is compulsory; a necessary condition”. The definition within the Systems Engineering field is similar in essence but adds technical terms. INCOSE (2011) suggests requirement as ‘a statement that identifies a system, product or process characteristic or constraint […] and is deemed necessary for stakeholder acceptability”.

Evans (2015) focuses on expectations from different parties involved: “A description of what is needed from a system or process, including the expectations of relevant stakeholders”.

It must be noted that requirements must be written in a clear, consistent, and unambiguous manner so that misinterpretations can be avoided (Evans, 2015). Good requirements not only ensure that the design and specification processes are conducted in accordance with the general guidelines of Systems Engineering, but also that the traceability between all component designs can be verified to fit the needs of the system. Without them, projects fail, are late, go over budget, or end up never being used (Alexander & Stevens, 2002). For instance, a kitchen scale would have requirements concerning weight capacity, precision, reliability, etc. In that, a few of them could be described as:

1. The scale shall measure items up to 5,000g
2. The scale shall measure items with a precision of 0.1g
3. The scale shall measure items at an accuracy of 99% and reliability of 99.97%.

With such requirements in place, the specific design and specification of components is improved by ensuring an extra level of verification and later a layer of validation against the purpose of the system. In addition, they help identifying capability gaps in components that need addressing, usually conducted in the form of measures of performances (Blanchard & Fabrycky, 2006). The process of component selection or design then becomes traceable where choices for bespoke or off-the-shelf items rely on their compliance with the necessary requirements for the fulfilment of the main purposes of the system.

However, the Vee model and systems engineering processes in general are also not devoid of criticisms. Firstly, Douglass (2015) highlights that the process is highly static and that each step requires a semantic approval, which is highly unreliable. Most requirements and specifications are document-based and in written form, in a way that increases the subjectivity and ambiguity in the process. In addition, document-centric processes make traceability along the Vee very difficult to manage when the number of specifications and requirements usually reach the thousands. For such reason, model-based systems engineering (MBSE) has been gaining momentum among systems engineers as a more accurate and even more reliable practice (Douglass, 2015). MBSE uses models rather than documents to trace high-level systems requirements down to the specific design. It logically follows that, since the models are related to each other, changing one element in one diagram automatically updates the corresponding elements in other diagrams. Consequently, this interdependence provides the extra level of verification that is not present in document-centred processes (Acheson, et al., 2013).

Secondly, even though systems engineering is a teleological activity in which systems are developed to fulfil a certain purpose, there is a degree of determinism when it comes to addressing technological capabilities. The elicitation of requirements in systems engineering depends on the current feasibility of design capability, in the sense that specifications are traditionally set to be met rather than developed. The selection of technical specifications is generally limited to what is available at the design stage of the project. Therefore, although systems engineering processes are comprehensive in the system design and validation, they allow little space for radical innovation. The perceived limitation in the function and the range of technical and technological capabilities end up steering the project towards an incremental activity and away from the possibility of radical redesigns. Kaplan (1964) coined a famous phrase that sums this up: “if the only tool you have is a hammer, then every problem becomes a nail”. In that, even a systems engineering project does not become entirely a solution looking for a problem, but a list of possible solutions that are looking to be fit onto a problem.

3.5. Summary

Chapter 3 reviewed the literature to discuss three main paradoxes in the epistemology of engineering. These dichotomies are now exacerbated by the very distinct dynamics in the current environment of socio-technical systems. The first paradox concerns the logic of justification of scientific methods and the nature of problems dealt by engineering projects. Engineering aims at novel solutions for social needs unprecedented circumstances, whereas methods rooted in the logic of justification can only accommodate problems of known scope in which replication and falsifiability can be maintained. Therefore, there is a logical paradox between the overarching goals of engineering and the methods that it traditionally applies. Step-to-step processes cannot be applied to problems of high complexity where a solution is unknown, such as those experienced by socio-technical systems. Consequently, it is argued that heuristic processes rooted in the logic of discovery are more suitable to the new paradigms faced by engineers in the 21st century.

The second paradox relates to the perspective taken in large engineering projects for socio-technical systems. Those projects share two main characteristics: (1) they are entirely man-made, and (2) they take several years from development to deployment. In that sense, there is a logical conflict with the traditional methods in science that take a deterministic approach. Determinism is intrinsic to research in the physical sciences because it protects the subject from the observer's interference. However, engineering projects require interference because they are not natural phenomena of chance, but rather projects of choice. Consequently, adopting a deterministic approach limits the extent of innovation as radical change cannot be simply forecast. Furthermore, engineering projects are inherently forward-looking as they take long to be finished, and since they are man-made, normative scenarios seem more appropriate as they begin from a desired state and work backwards to find the actions needed to achieve it.

The third and final paradox between the epistemology of engineering and the methods from the physical sciences involves the approach to problem solving. Traditionally, scientific methods have long applied reductionist approaches, meaning that they divide the whole into manageable parts that can be investigated in their causal relationships. However, socio-technical systems are complex, and their complexity emerges exactly from the interaction between parts. Historically, the choice for reductionism and the look for simplicity has also been a result of a lack of technical and technological capability to analyse problems of organised complexity. With the recent and rapid advances in technological prowess, the fields of systems science and systems engineering have taken prominence, for they offer a holistic perspective to problems which is essential for radical innovation.

4. METHODOLOGY

4.1. Introduction

This section builds on the discussion in the previous chapter regarding the necessary epistemological foundations for radical change in engineering projects. From the review conducted, one can see how the rapid process of urbanisation is challenging the efficiency of current transport systems at its core. As a consequence, increasingly rapid radical innovation is required to adapt to ever faster changing conditions. From the literature review, three main points can be highlighted:

1. Innovation cannot follow a pre-defined method. Radical change is rooted in the logic of discovery which involves a heuristic approach to problems. In that, radical innovation is an ad hoc process of discovery rather than an algorithm for optimisation.

2. Deterministic methods cannot accommodate radical change in the future because they depend on things remaining the same. Thus, a teleological perspective using normative methods is needed to direct changes from undesirable trends and towards a desired state.

3. A mechanistic approach based on isolation cannot fully accommodate the increasing complexity of artificial systems and the inherent conflicts that arise from the numerous interactions. Systems engineering processes are then crucial in that they enable the modelling and verification of such issues before the implementation phase, thus reducing risks in face of future uncertainty.

Therefore, there cannot be a scientific method for radical innovation in complex systems. The resulting methodology is in fact a framework to ensure that critical guidelines are acknowledged, yet without the rigidity of traditional scientific methods in order to allow a degree of flexibility necessary for innovative assumptions. It derives from the epistemological triad of discovery, teleology, and holism, in that they depend on each other to achieve synergistic results. The framework combines the teleology and material freedom of a backcasting approach with the holism, traceability, and verifiability of systems engineering processes.

The framework sequence works backwards and forwards simultaneously, assessing the current capabilities of the system in face of the selected objectives and desired state in order to develop an operational concept through the systems engineering process (Blumenfeld, et al., 2016b). From a systems engineering perspective, it focuses on the first stages of the left-hand side of the Vee diagram. More specifically, the framework covers the steps from purpose/need identification, to requirement elicitation, to design specification. As a result, a six-step framework is suggested as the foundation of radical change in complex systems such as urban transport modes. It is intentionally kept at a high level of abstraction in order to allow for specific methods to be applied within the process according to the nature of the problem, as explained later in this chapter. Figure 24 illustrates the suggested process in its logical sequence, while Figure 25 illustrates the framework in terms of innovation and time.

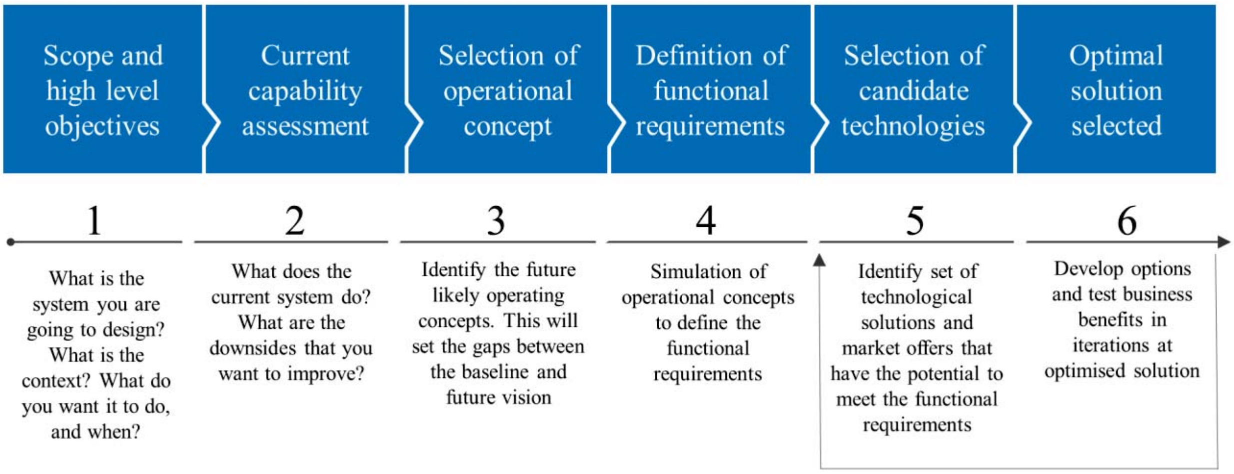


Figure 24. The proposed framework in a logical sequence

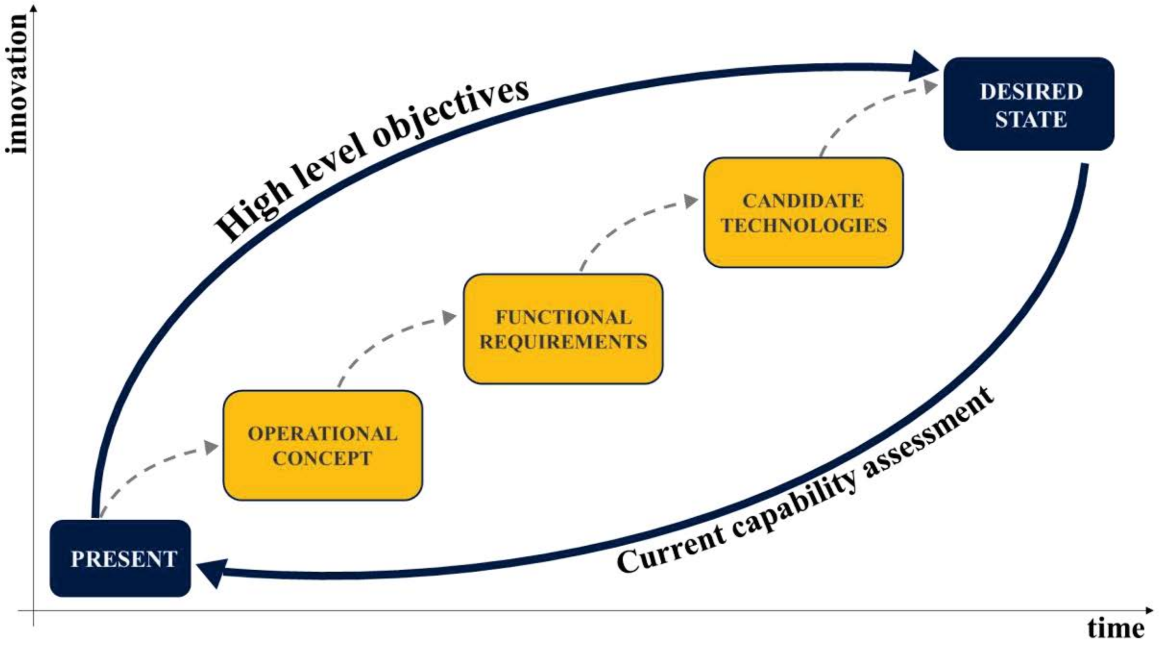


Figure 25. The proposed framework in terms of innovation and time

4.2. Scope and high-level objectives

The first stage of the proposed process is similar to the future visions in a backcasting approach and the purposes used in capability systems engineering processes. As also stated both by Robinson (1990) and INCOSE (2014), the initial vision begins with a clear definition of purpose and scope. In the realm of socio-technical systems, this is achieved by statements of desired outcomes of the operation of the given system. From the notion of socio-technical system, high-level objectives are defined by stakeholder needs. High-level objectives must be written in a clear, measurable, and unambiguous form, so that the operational concept can be validated properly.

Drawing from the backcasting approach, purposes as statements of desired futures shall not be bound by current capabilities. They should express, in functional terms, the desired state of the system. It logically follows that a timespan should also be expressed so as to measure the gap in capability between the present and the desired future. On the other hand, the backcasting approach described by Robinson (1990) works with a set of scenarios. While this can be used for policies, socio-technical systems can only take a single desired scenario in the form of the ideal system as the goal.

For that reason, it is important to define a clear delimitation of the scope of the system in order to limit the number of variables. This will ensure the problem remains within the realm of organised complexity where the number of variables is manageable. In addition, this stage greatly benefits from the understanding of the boundaries of the system in its environment, where the interactions with external entities and the internal interfaces between subsystems are identified. To do so, context diagrams as suggested by Kossiakoff et al. (2011) are a fruitful tool to identify the external entities and their respective interactions with the systems. Figure 26 exemplifies the use of context diagrams in a case study of a train and its interactions with the entities around it.

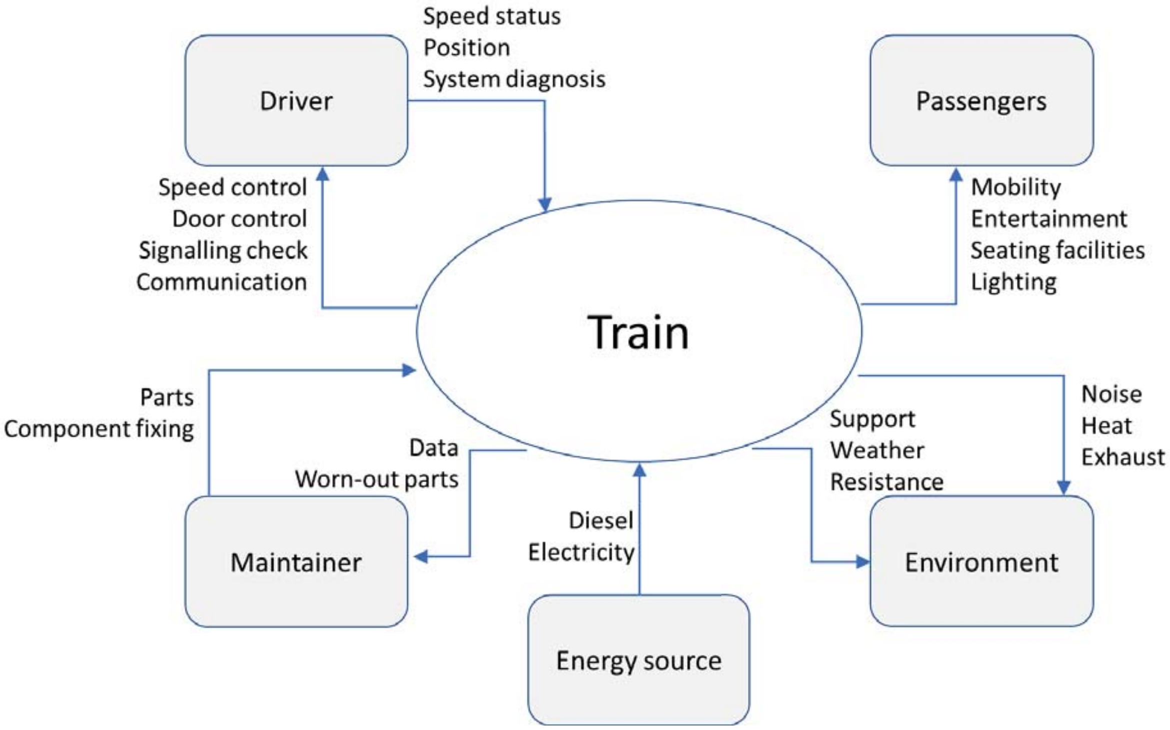


Figure 26. Context diagram for a train

4.3. Current capability assessment

The second stage provides an examination of the current ability of the system in relation to the desired outcomes from the previous stage. For instance, if an engineering project envisions running trains at 500 km/h, then it is necessary to assess the current capability of trains in terms of speed so that the technical and technological gaps between the present and the desired state can be identified. In addition, the traditional prognostic methods of forecasting and horizon scanning can identify performance limitations of current systems. As Christensen (2000) shows, the incremental evolution of technologies follows a sigmoid function where they reach a saturation point of performance increase regardless of extra engineering effort. The S-shape of incremental evolution derives from internal conflicts between sub-systems that limit indefinite linear enhancements in performance according to engineering efforts.

These conflicts, when found, will highlight the critical functional aspects that require radical redesign in order to overcome systemic limitations, which creates a new path for incremental evolution of another applied technology. An example of the inherent limitations of incremental evolution in systems is found on jet engines in commercial aeroplanes. While the size of the fan helps reducing the amount of energy consumption, bigger engines increase the drag and inevitably require more energy to push through wind resistance. In that sense, incremental improvements, regardless of advancements, are inherently limited by such paradox. Therefore, the only way out of these limitations, especially if the system were in constant change, is by a radical redesign of the system.

Yet, the redesign depends on the correct identification of issues with the current system. The most common methods for capability evaluation in both systems engineering and backcasting rely on models and simulations. It is important to remain attentive to a wide quantitative analysis of current capabilities so that the gap between present and desired future can be correctly measured. These can relate to the interfaces within the system, or technological limitations of subsystems and components. For instance, the maximum speed of trains is bounded by the rolling stock interfaces with the track and power infrastructure.

In that sense, models, prototypes, and simulations are important tools to identify the functions constrained by internal paradoxes in their components and sub-systems. In particular, systems dynamics modelling tools such as causal loop models (see Morecroft (1982)) will help identifying the functions where higher engineering effort will not reflect on proportional performance gains.

4.4. Development of operational concept

Once the limitations of current systems have been identified, it becomes possible to conceptualise an ideal system in which these internal paradoxes have been overcome. This involves the aforementioned inductive reasoning where the system is envisioned as having solved the problem. At this stage, it is important to focus on the functional aspects of the system otherwise physical constraints of current components would bring inevitable limitations to the operational concept. Operational concepts can be defined as the scenarios of socio-technical systems. Their main objective is in fact the accurate description of the structure and dynamics of systems in order to verify and validate their components and functions. In the context of radical innovation for future systems, they also assist on the identification of necessary technical and technological capabilities which are not currently available or in use. It logically follows that evaluating operational concepts against current capabilities can only fulfil incremental evolutionary curves and does not allow for radical change.

This is where the normative approach of backcasting significantly benefits the engineering framework. Since the desired state is a future entity, it inherently allows for technological independence to develop a solution that is higher on the innovation axis than it would be when forecasting incremental developments (Blumenfeld, et al., 2016b). At this point, current technical and technological limitations can be intentionally overlooked in order to allow for a complete solution regardless of current feasibility.

The focus of the normative approach is the inverse of traditional processes. It is not on how far current technological capabilities can take the system performance, but what technical capabilities are required in order to fulfil a certain desired performance. Being a model itself, an operational concept carries the ability to perform such task because its virtual nature permits future technological assumptions. Thus, the approach enables the development of an operational concept that fully meets the objectives defined in the first step that would be infeasible had the current capabilities been taken into consideration.

On the other hand, it is important to ensure that this stage will not lead to ‘black-box thinking’, where the operational concept relies on technical and technological capabilities which are either infeasible or not clearly specified. The term black-box thinking is used here to denote an idealised assumption of an unspecified technology that can solve problems without an explanation on how it does it. It is a tempting thinking in foresight. Since the potential of technologies is virtually limitless when considering all possibilities in the future, one can easily end up assuming that there will be a solution for every problem and a disentanglement to every internal conflict. However, such a position adds nothing more than a picture of the end goal without the respective path to achieve it. To add the necessary steps, systems engineering techniques and tools can assist in maintaining the robustness and reducing project risks in face of future uncertainty. In addition, the following stages borrow techniques from systems engineering processes in order to evaluate the operational concept and iteratively assess the robustness of the system.

The most important outcome underlying the modelling of the operational concept is a clear, detailed, and quantified description of the regular operation of the system. These statements, in form of models or simulations, will define the functional requirements which will later lead to technology selection based on risk and benefits of the capabilities needed. Models can take various forms. Müller (2009) broadly defines it as a representation of an entity in relation to its structure or its purpose, while Hybertson (2009) provides a more specific definition for a systems engineering context:

> "[A model is an] explicit approximation, representation, or idealisation of selected aspects of the structure, behaviour, operation, properties, or other characteristics that can be associated with one or more systems" (Hybertson, 2009, p. 63)

Of the various ways to model the operational concept of a system, the traditional descriptive written approach in backcasting lacks the necessary technical robustness that lead to specific requirements (Dreborg, 1996). System modelling can be achieved by a number of different tools, such as SysML, stock-and-flow, causal loops, and others (Holt & Perry, 2014; Meadows, 2008; Morecroft, 1982). These tools are essential for the verification and validation of operational concepts in a reliable manner which assist on informed decisions.

As in the framework suggested by Robinson (1990), more than one operational concept may be modelled. This variety is rooted on the understanding of equifinality in complex systems described in the previous chapter, where various different arrangements can achieve a similar outcome. From a backcasting perspective, all operational concepts carry the same value once they achieve the same goal. Nonetheless, the subsequent stages rooted in systems engineering and analysis add the technical robustness for calculated decisions based on risks, benefits and impacts.

4.5. Definition of functional requirements

Following from the operational concept, the purpose of this stage is to add the robustness of traceability, verification, and validation of systems engineering processes into the ambitious visions of normative scenarios in foresight. Once the operational concept has shown to fulfil the high-level objectives, requirements can be captured from the models. Requirements should follow the same patterns of those in systems engineering processes: they should be clear, concise, unambiguous, measurable, and traceable positive statements.

An important aspect of this stage is that the focus should be kept on the functional aspects and not on physical specifications, otherwise technological limitations to the operational concept will inevitably constrain the design process. Logically, physical specifications are only known for components that are already commercially available. Therefore, if requirements describe the physical aspects of a component rather than the function to be performed, the normative aspect of radical innovation cannot be attained, and the process becomes one of incremental evolution.

4.6. Selection of candidate technologies

The definition of functional requirements from the operational concept will result in technological specifications of physical components. At this point, there are three potential situations for the selection of technologies, depending on the functional requirements previously defined for components and subsystems, and their respective performance available (Figure 27). These outputs are not mutually exclusive in that a requirement can provide two different technological needs for two different specifications.

A. A component or subsystem can be already available and fulfil the necessary specification, which offers low risks where no radical innovation is needed. This situation is for requirements for which currently available (off-the-shelf) components can deliver the performance needed.

B. A component or subsystem can be technically feasible, but not at the operational level desired. In this case, functional requirements will state the necessary performance for the desired operational concept to be realised. Priority shall be given to technologies that offer the required performance with the lowest risk, however there are added layers of comparison to be taken into account in the next stage.

C. At last, a component or subsystem might demand a completely new technology that is not yet available when the incremental evolution of current technologies is unlikely to match the performance required. In this case, the technology selection acquires a completely normative approach to technological development, by determining the capability to be achieved. Such cases should not be discarded immediately. A normative guidance to technological development shall be considered amongst other factors in order to assess its risk-benefit ratio.

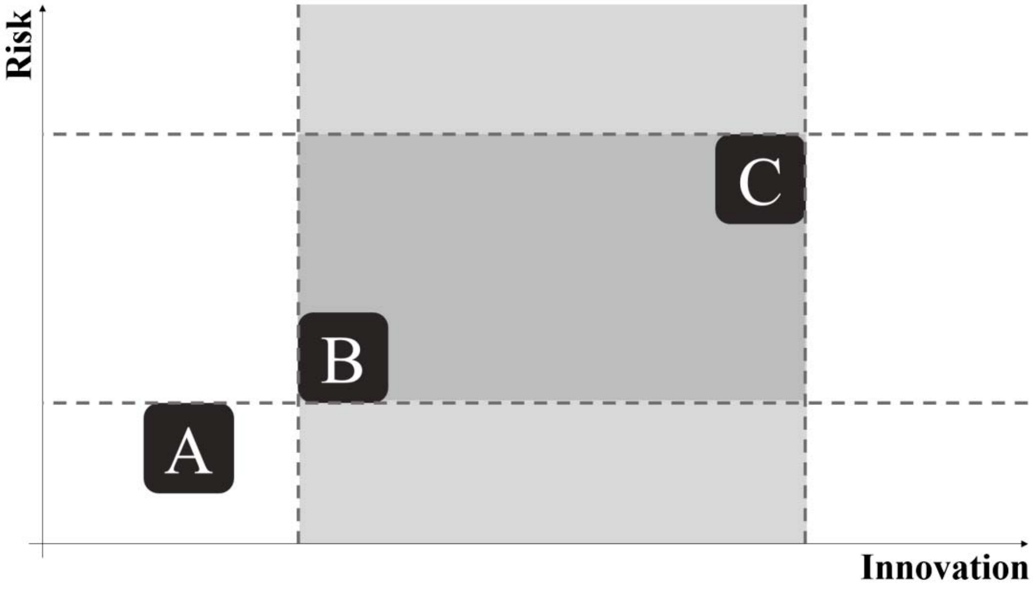


Figure 27. Innovation versus risk in technological requirements

Drawing a table of Measures of Performance (MoP) will help identifying the level of availability of each component so that risks and a performance timespan can be inferred. Blanchard and Fabrycky (2006) explain measures of performance as the comparison between the current capability and the functional requirements of the system.

This framework covers the range between B and C as highlighted, in that off-the-shelf technologies do not require a normative process of backcasting in order to define the required future capabilities. This framework induces innovation but not at a higher risk than necessary to fulfil the functional requirements of the system. The area highlighted in Figure 27 is conceptual, meaning that it does not carry any value in its proportions apart from illustration purposes. The main point of the graph is to illustrate the combination of innovation and risk: beyond the current capabilities in order to leapfrog current trends and to meet future objectives and limited by the minimum necessary technological development in order to limit risks and uncertainty. Its application in future-oriented engineering processes guides the choice of technologies in order to improve cost influence curves without performance loss. As previous discussions show that technologies are bound to advance over time, projects looking to anticipate them should look beyond B, but not ahead of C in order to contain unnecessary risk and cost increases.

4.7. Selection of optimised solution

Once the candidate solutions have been identified with their respective degree of uncertainty, a bigger picture can be devised for the whole system. Firstly, a sensitivity analysis of each functional requirement can define the influence of each technology on the whole of the outcomes of the operational concept. Sensitivity analysis studies how variations and uncertainty in different inputs influence the variability and uncertainty of the outputs of a model (Saltelli, et al., 2008; Morio, 2011).

On the other hand, it is important to define the uncertainty and technological horizon of given performance requirements. A critical technology which performance has not yet been proven to be possible offers greater risk to project success, regardless of its protagonist role in the operational concept. Conversely, technologies which offer near-optimal outputs with greater certainty may be more beneficial depending on the level of performance achieved and the smaller gap in capability required.

To evaluate these aspects, the systems designer must keep simultaneously an open mind towards the functional requirements to allow for novel technologies to be considered, while maintain a pragmatic stance towards the robustness of the final system architecture. A fruitful tool for such is the Pugh Decision Matrix (Pugh, 1991; Pugh, 1996), which can be populated with parameters such as impact, dependability, and technical readiness levels (TRL).

The Pugh Decision Matrix was formally included in the process after its successful empirical application during the S-CODE project (S-CODE, 2017). Inferring from evidence, the large number of possible solutions based on a multitude of potential technologies requires a methodical and quantifiable process of selection. That way, decisions for C-type technologies which are still conceptual can be scrutinised and their inclusion, whenever needed, justified. The case study in the next chapter will briefly illustrate circumstances of such process at a high level, as the main purpose of this research is on method development and not on the finalised solution of the system within the case study per se.

5. APPLICATION IN A CASE STUDY – METRO SYSTEMS IN MEGALOPOLEIS

5.1. Introduction

The previous chapter described the framework for introducing radical change in complex systems such as transport. Thorough as the description is, a process rooted in the logic of discovery can only be described in abstract terms of heuristics. In a similar way to Polya (1945), it is argued here that the best way to explain a heuristic method is by analogy. Therefore, this chapter will introduce and illustrate the framework within a case study of metro systems in megalopoleis. This will further develop the discussion started in Chapter 2 in search of a radical solution for the challenges of public transport systems in megalopoleis.

5.2. Scope and high-level objectives

5.2.1. Introduction

Metro systems are an important asset in large urban areas. They facilitate movement across the city to ensure that longer distances can be covered within reasonable time. They must operate timely, safely, and in an economic manner to cater for the stakeholders involved. Due to the high costs for construction, operation, and maintenance, metro systems are usually subsidised as they are systems of critical value to areas of great mobility demand such as megalopoleis. Figure 28 exemplifies the context diagram of metro systems, where passengers and society should be seen as the primary group.

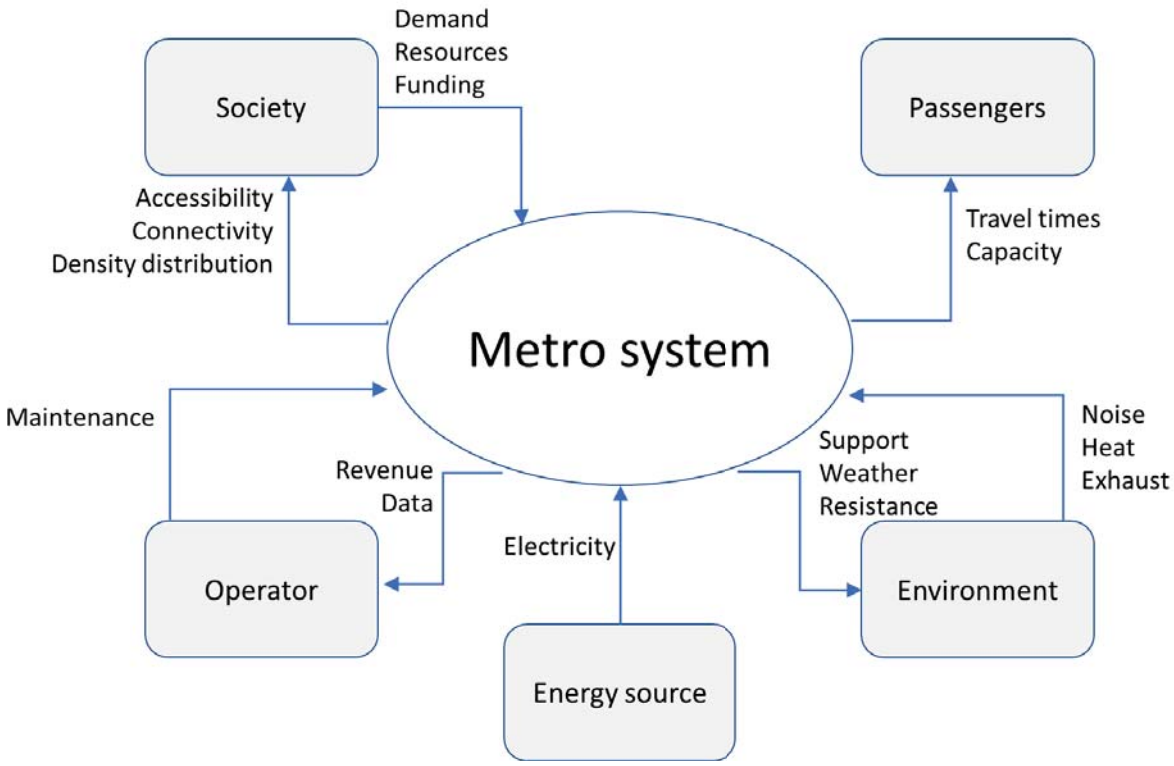


Figure 28. Context diagram of a metro system

Of the aspects of metro system operations that are critical for the wider benefit of society, two were chosen as per the scope of this research: travel times and capacity. Their choice does not undermine other technical aspects such as reliability and safety as less important, bus understands them as subsequent iterations of the framework for the resulting operational concept. As previously discussed, passengers opt for the fastest mode they can afford to, so a slow public transport system will not fulfil its primary goal. In addition, a system that cannot cope with passenger demand imposes severe penalties in travel time and comfort, added to important safety concerns. Therefore, these two aspects are discussed as the main objectives of the system.

5.2.2. Door-to-door travel times

The discussion in Chapter 2 provided the foundation of the purpose of an urban rail system. If we assume that cities as wholes need to be entirely accessible to its dwellers, then transport systems need to ensure that they remain approximately 'one-hour wide' or else they become fragmented. In addition, considering that there is an inevitably urgent need to reduce the intensity of private transport for our future sustainability, and since users will choose the fastest mode they can afford, public transport systems should offer door-to-door trips within the 30 minutes range. In doing so, mode shift remains supported by the liberal notions of freedom of choice. In other words, public transport systems should provide trips that cover the same distance within the same time as drivers, including all components of access, waiting, movement, etc.

There is debate whether drivers and public transport users travel similar distances and thus require similar travel speeds. Zahavi (1974) observed the opposite, in that drivers commute longer distances because they can afford greater speeds. At that time however, the traditional perspective of spatial dynamics was that those on higher incomes would live in the suburbs and those on lower incomes would inhabit the central area surrounding the business district. However, as shown by the maps in Appendix A, much has changed since with the dominance of large cities and megalopoleis and respective travel patterns. In these urban areas, central locations now carry a premium in that they are better connected to jobs and opportunities.

Inevitably, those on lower incomes, who are exactly those more reliant on public transport, have been gradually forced to the fringes of the city. In some cases where transport infrastructure is insufficient, they in fact travel longer distances than those who can afford to drive. The reason for such is because although megalopoleis are essentially polycentric, jobs still tend to be centrally concentrated (as shown in Appendix A). It then feeds a feedback mechanism that reinforces disparities in access and constrains the geographic potential of cities.

Another debate that stirs controversy in the literature regards the utility or disutility of travel time. In opposition to Zahavi and Talvitie (1980), Marchetti (1994), Laube et al. (1999), and others, Mokhtarian (Mokhtarian, 2003), Lyons and Urry (2005) , and Lyons et al. (2007) have suggested the idea of useful travel time. In their views, passengers can use their journeys to perform useful activities such as sleeping, studying, reading, or working, which in turn would make their travel time a utility rather than ‘wasted’. In fact, Lyons and Urry (2005) acknowledge the significant behaviour change caused by advances in internet communications. Steer Davies Gleave (2002) adds that rail travel specifically can be seen as productive, as the mode offers the space and necessary conditions for such.

Nonetheless, these views seem to overlook the various types of urban rail modes and underestimate their important differences. While intercity rail, known as mainline in the UK, offer seats to most passenger and even tables in some cases (Figure 29), urban rail focuses heavily on capacity and therefore offers little or very little personal space for the productive tasks mentioned before (Figures Figure 30 and Figure 31). Therefore, even though the utility of travel times can be inferred for certain types of rail journeys, it is unlikely that they apply to the context of megalopoleis where it is a disutility to be mitigated.

Figure 29. Interior of London Midland train class 150 (Rail Technology Magazine, 2013)

Figure 30. Interior of Yamanote Line train in Tokyo (Experience Tokyo, 2015)

Figure 31. Yamanote Line train during rush hour in Tokyo (Kikuchi, 2017)

From the literature, there is extensive literature indicating that average journey times in cities usually aggregate at around half hour (Zahavi & Talvitie, 1980; McKenzie & Rapino, 2011; Laube, et al., 1999; Metro/SP, 2013). In addition, complementing the data from Figure 10, aggregate research findings show that drivers in twenty of the largest cities in the world commute an average distance of 19.7 km in approximately 33 minutes (Gyimesi, et al., 2011).

In that sense, the purpose of radical capability changes in urban rail systems in regards to travel times is that it adheres to the historical invariant travel time budgets regardless of size. Therefore, two main purpose statements can be inferred: (1) the system shall cover at least the same distance of drivers in megalopoleis (19.7 km) in approximately 30 minutes; and (2) the system must be scalable in that different distances can also be covered within a similar travel time or less.

5.2.3. Capacity

A second important consideration in urban rail systems is capacity. As Figure 32 illustrates, heavy metro systems are commonly found in megalopoleis because they are those that provide the greatest combination of commercial speeds and capacity (Figure 32). The large passenger demand and longer trips to be covered tend to justify the higher costs of construction and operation. In cities such as São Paulo, Tokyo, and London, lines carry more than one million passengers a day, achieving normally 60,000 and sometimes almost 100,000 passengers/hour/direction (Parkinson & Fisher, 1996; Metro/SP, 2018; Transport for London, 2018).

Capacity is an essential component of operations in urban rail systems because it directly impacts travel times, passenger comfort, and passenger safety. Firstly, crowded vehicles require longer dwell times to accommodate the more difficult passenger movement between carriage and platform. In some observed cases, boarding and alighting time per passenger can double in crowding situations (Tirachini, et al., 2013). The longer boarding times will result in longer dwell times, which considerably impact in-vehicle times as lines usually comprise many stops.

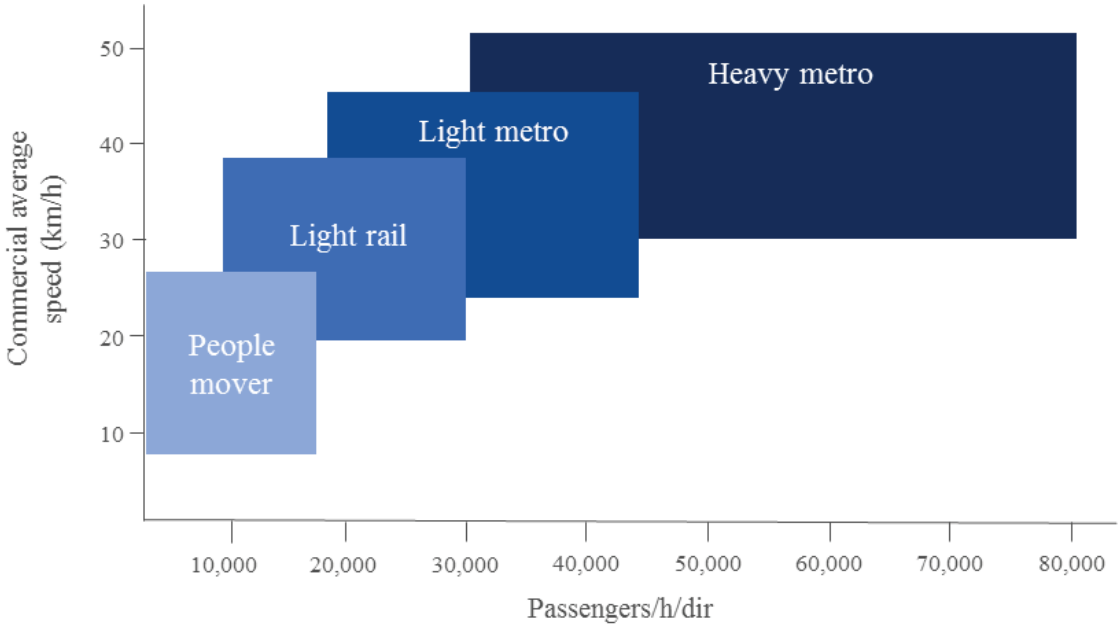


Figure 32. Public transport modes according to their commercial average speed and capacity (Connor, et al., 2015)

Secondly, the reasoning behind capacity also relates to passenger comfort. Crowding levels have considerable impacts on the disutility of in-vehicle times. Tirachini et al. (2016) report that passengers experience a significant increase in discomfort when crowding levels reach 6 passengers per square metre (6 passengers/m^2). Nonetheless, recent accounts in the media show that many systems currently face significantly greater demand for capacity than the system can supply, as shown on Table 2 (Lim, 2008; Transport for London, 2011; Pai, 2009; Barbosa, 2016). In fact, some systems even take such values as the operational threshold due to operating already over capacity, as is the case in São Paulo (Barbosa, 2016).

Table 2. Crowding levels at peak hours in megalopoleis

System	**Peak hour crowding level**
São Paulo	8.1 passengers/m²
Tokyo	7.0 passengers /m²
Singapore	4.9 passengers /m²
London	6.3 passengers /m²
Mumbai	11.0 passengers /m²

From that, the high-level objective of the system concerning capacity is that it shall be able to carry no less than 100,000 passengers per hour per direction. That provides a step up from most systems currently operating in megalopoleis. Moreover, capacity should be increased as much as possible in order to accommodate the growing populations especially in developing countries. Crush loads are an important factor on mode choice, in that its disutility should be reduced whenever possible.

5.2.4. Transport equity

While traditionally not much of an issue from an engineering perspective, the matter of equity in transport and access is seen as an important concern in the planning realm. The lack of access to adequate transport and consequently adequate travel speeds reinforces social disparities and imposes severe disadvantages on lower income households to seek equal economic opportunities (Lucas, et al., 2016). In that sense, the traditional predict and provide approach can only further increase such disparities, and urban railways as socio-technical systems need to be addressed in that sense. Martens (2006) explains that building infrastructure based on the projected demands and values of time only reinforces the provision of access to those who already benefit from the system.

It is a simple logical reasoning. If we know that population and jobs densify around existing stations, and that land increases in value and economic activity accelerates in these areas, then we can assume that the demand and the aggregate values of time will also be higher for those using that stop. It logically follows that, if operations and line planning are solely based on these projections, they will inevitably reinforce inequity as access will be improved to those already in better regards. This issue is exemplified by the operational strategies discussed on Section 5.3.

Therefore, it is important to ensure that the development of socio-technical systems also look at the social dynamics they incite. The provision of equitable access to all passengers should be seen as a goal whenever possible. Also, because there is evidence that demand in urban rail is usually created when not followed (Cervero & Sullivan, 2010).

5.3. Current capability assessment

An important realisation of journey travel times in public transport systems is that they comprise several distinct components which sometimes conflict with one another. In the planning literature, travel times are usually evaluated in a broader sense of generalised travel costs as these include the personal notion of value of time in monetary form (Wardman, 2004). Door-to-door generalised travel costs consist basically of five parts: (i) access time (T_a); (ii) entry and exit times (T_e); (iii) waiting time (T_w); (iv) in-vehicle time (T_v); and (v) interchange time (T_i). For our calculations, we assume the access time T_a to be the average time spent covering the distance to and from the means of transport at each end of the journey, and T_e to be the average time between ticket barriers and platforms, and vice-versa. To them are added specific weightings of value of time (γ, δ, φ, ω) in relation to in-vehicle time. Door-to-door travel time (T_t) can thus be expressed as:

$$(1)\ T_t = \gamma 2T_a + \delta T_w + T_v + \varphi 2T_e + \omega T_i$$

There is strong debate on the weightings and the value of travel time savings. These values, illustrated by the weightings (γ, δ, φ, ω), represent the willingness to pay to diminish travel time by one unit when compared to in-vehicle time (T_v) (Jara-Díaz & Guevara, 2003). Although Wardman (2004) provides a comprehensive list of values of travel time savings in public transport systems, there are three issues that prevent them from being directly used in this research. Firstly, there is significant deviation in the values which challenge the idea of universal values of travel time savings. Secondly, values of travel time savings as willingness to pay tend to be attached to personal choices and other elasticities, and thus can be inappropriate for social evaluation (Mackie, et al., 2001). Thirdly, and more specific to this research, it is logically unreliable to infer values of travel time savings for a system that does not exist yet, in that the different operations and added capabilities may significantly skew the willingness to pay for the reduction of time spent in trip components. Therefore, this research adopts a simplified version of Equation (1, as follows in Equation (2. In that, weightings can be later applied in an *ad hoc* manner according to the specific context in which the system is found.

$$(2)\ T_t = 2T_a + T_w + T_v + 2T_e + T_i$$

A driver, for example, would have a relatively short access time (T_a) from door to car and from car to door, no waiting or entry/exit times (T_w or T_e) as the vehicle is readily available, and the trip mainly consists of an in-vehicle component (T_v). It can also include interchange time (T_i) if the trip will also comprise another motorised component. Conversely, the trip on an urban rail system is more complicated: access time (T_a) from door to station and from station to door; entry and exit times (T_e) from the station entrance to the platform and vice-versa; waiting time (T_w) for the time at the platform; in-vehicle time (T_v), which comprises the time spent inside the vehicle; and interchange time (T_i), which accounts for the time to change between lines and/or between modes when necessary.

In all motorised modes, door-to-door travel times are inherently limited by trade-offs, so moving more quickly does not necessarily convert into shorter travel times. In the case of private modes, such as cars or motorcycles, maximum speeds and consequently travel times are limited by traffic density and speed restrictions, amongst other elements (Hall, 1996). That way, a megalopolis where private modes are the main form of travel not only are harmful to the environment but also cannot accommodate the large demand in networks because the available area for roads is limited. In public modes such as buses or trains, door-to-door travel time depends on both the time to access the mode and the in-vehicle travel time, among other elements as it will be discussed in detail below

These trade-offs have profound implications in the perception of urban boundaries. Regardless of the maximum speed achieved at a certain point, it is the average door-to-door speed, and consequently the overall journey time, that defines the city limits for a particular user of a transport mode within their travel time budget. For these and other economic reasons, aeroplanes have never been able to create commutable regions of 500 km radius because of the time spent on the non-flying components of the trip. Similarly, Cox (2017) points out that, although the high-speed rail route between Tokyo and Osaka is practically entirely urbanised, commuting efforts in terms of door-to-door time still prevents the corridor from forming a single urban region.

In the specific context of urban rail systems, trip components interact with each other often in conflicting ways. From the literature, we know that T_v is the sum of all times spent travelling between stations for *n* stops, according to maximum line speed (V, in m/s), acceleration (α, in m/s²), braking (β, in m/s²), and dwell time (T_d). Waiting time (T_w) is in average half the headway between trains. Jerk time (T_j) accounts for the time needed to comfortably transition from acceleration and braking. Finally, T_v also depends considerably on the distance between stations (D) (Vuchic, 2005):

$$(3)\ T_v = \sum_i^{n-1} \sum_j^n \frac{D_{ij}}{V} + \frac{V}{2}\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) + nT_j + T_d$$

The number of stations *n* is calculated simply by dividing journey distance (D_j) by the average distance between stations (D):

$$(4)\ n = \frac{D_j}{D}$$

From Equation 3, it is then possible to model the current capabilities of metro systems and identify the trade-offs that arise from conflicting components. For the model, journey distance is assumed to be 19.7 km based on Gyimesi et al. (2011), and operational parameters based on data from current metro lines as listed on Table 3 (Siemens, 2012; Karekla & Tyler, 2012; Transport for London, 2013; Powell & Palacín, 2015).

Acceleration and braking are based on the Munich metro train designed by Siemens (Siemens, 2012). Jerk is calculated as the maximum value where passenger stability is not compromised (Powell & Palacín, 2015). Some systems may operate at lower standards depending on the age of the components and physical limitations on the lines. It is seen as unlikely that acceleration, jerk, and braking can be increased significantly in the future due to passenger stability and safety.

As a baseline scenario, it is also assumed that users walk to stations and have an average of one interchange between lines. No specific or reliable set of data was found for the number of interchanges per passenger, so a simplified estimate of one was adopted on the basis that while some users will not require an interchange, others may require more than one. It is assumed that users will choose the route with the fewer number of interchanges. This assumption is supported by a study conducted by TfL using WiFi tracking (Transport for London, 2017b).

Table 3. Operating parameters of metro systems

Parameter	**Value**
Maximum line speed (V)	90 km/h (25 m/s)
Acceleration rate (α)	1.3 m/s²
Braking rate (β)	1.2 m/s²
Jerk rate (j)	1 m/s³
Interchange time between lines (T_i)	270 s
Waiting time (T_w)	60 s
Entry/exit time (T_e)	165 s
Dwell time (T_d)	30 s

Moreover, access distance (d) is assumed to be half of the distance between stations (D), as shown in Figure 33, to represent an average situation for users. It is recognised that there is usually an overlap between catchment areas, but this is generally counterbalanced by those travelling further than *d* to access the station. In addition, access distance is not considered part of the journey distance (meaning it would subtract from it) because not all passengers will live at such specific location. More importantly, research shows that station catchments are usually more complex that simple geometric entities. However, the simplistic model is used in this research for calculation purposes as the focus is not on the estimation of demand but on average journey times (Young & Blainey, 2017). Distance between stations (D) is then assumed to be 1,000 metres based on the median values from São Paulo Metro and London Underground (Transport for London, 2008; Mobilize Brasil, 2017);

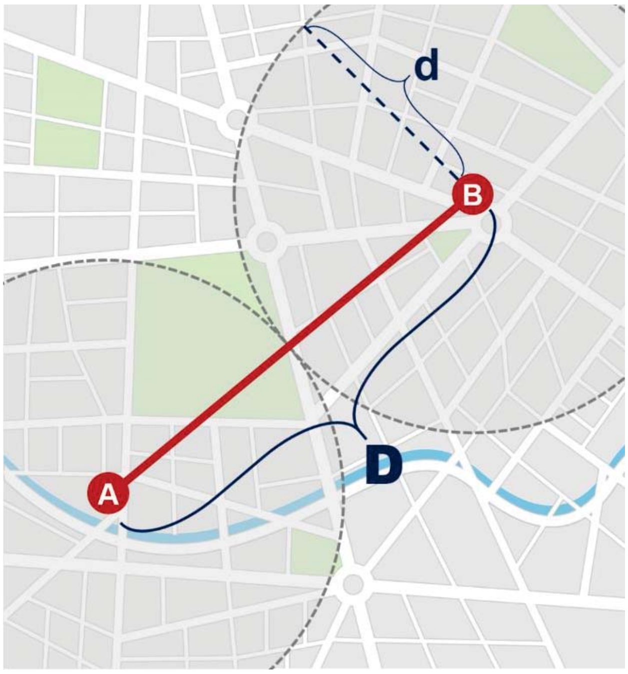


Figure 33. The relationship between interstation distance (D) and access distance (d)

The results of the modelling process using the aforementioned parameters indicate that a 19.7 km trip on a metro system would take approximately 49 minutes (Figure 34). While a generalised average, the value obtained is in accordance with those illustrated on Figure 10 in Chapter 2. Logically, there are many variables that make up for such similarity in results between model and reality. For instance, users may travel shorter distances but perform more than one interchange; users may have a longer access time to walk to a certain line in order to avoid interchanges; entry and exit times are dependent on station design; etc.

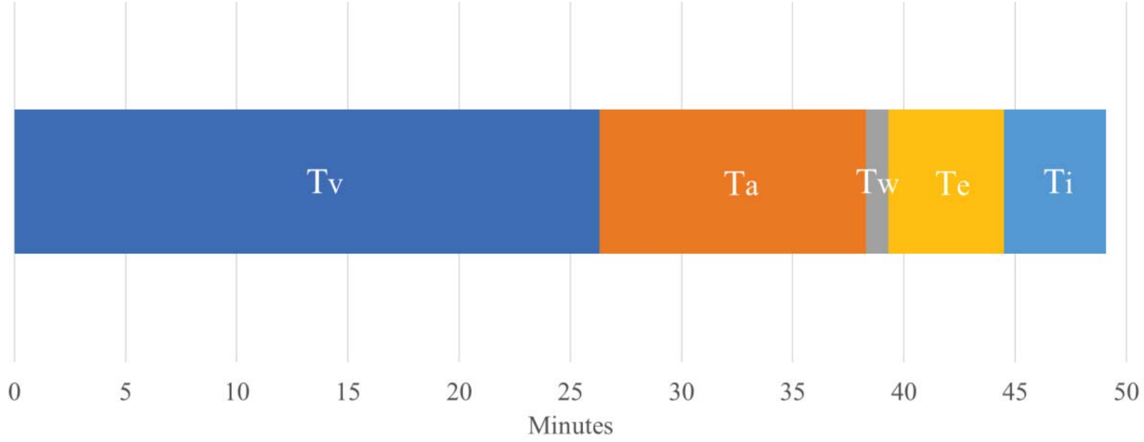


Figure 34. Breakdown of a 19.7 km model journey on a metro system

It is important to note that the motorised component accounts for about 55% of the door-to-door time, highlighting the influence of all other components on the total. Furthermore, the results also shed light on a crucial aspect of metro systems operations, referred here to as the *coverage paradox*. The journey on a metro system involves two types of components, namely those of non-motorised access to and within stations where distances are the critical factors, and in-vehicle time where speed is the critical factor.

Therefore, from a non-motorised perspective, stations should be placed close to each other in order to reduce access time (T_a). However, shorter interstation distances (D) inevitably affect in-vehicle times because of the extra stops the train has to make. Conversely, locating stations far apart could increase in-vehicle speeds, but the greater interstation distances (D) will require passengers to walk longer to access stations.

Considering that access speeds are limited by human abilities at about 80 m/min (Pachi & Ji, 2005), it seems logical that priority should be given to reducing the distance between stations (D). Givoni and Rietveld (2014) reached a similar conclusion in their study. However, the choice is not universally straightforward, especially considering the growing journey distances in the context of expanding megalopoleis. In longer journey distances, as shown in Figure 35, the share of door-to-door time spend in-vehicle (T_v) grows significantly. That is because of the time penalties imposed by the need to stop at more stations in between origin and destination. Vuchic (2005) calculates these penalties (T_l) using Equation (5):

$$(5)\ T_l = \frac{V}{2} * \left(\frac{1}{\alpha} + \frac{1}{\beta}\right) + T_j + T_d$$

Where Tj is calculated by (α+β)/j and equivalent to 2.5 seconds with the given parameters. Since there are limitations in acceleration and braking for passenger comfort, the increase in the number of stops can quickly add up to considerable time penalties. With the operating parameters of Table 3, each stop adds 52.5 seconds to in-vehicle travel time (T_V). Thus, in a 19.7 km journey, approximately 16 minutes are used for stopping at intermediate stations. As a result, stops add severe penalties to travel time. For that same 19.7 km journey distance, in-vehicle time accounts for approximately 55% of door-to-door journey time, while at 30 km, it would raise to approximately 66%.

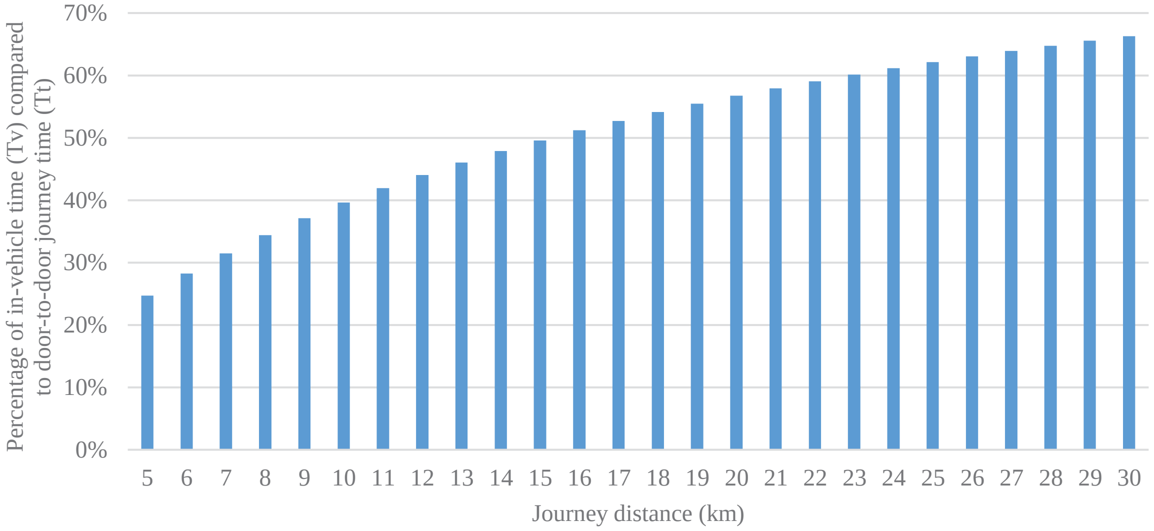


Figure 35. Share of door-to-door time (Tt) spent in-vehicle according to journey distance

Therefore, there is an inherent paradox in the operation of metro systems, and also other public transport modes, regarding the trade-offs between access times and in-vehicle speeds. Many scholars have thus investigated the subject in order to minimise the disutilities and reduce the generalised travel costs of public transport systems that pose barriers to their use (Givoni & Rietveld, 2014; Blainey, et al., 2012; Vuchic, 2005; Furth & Day, 1985). Nonetheless, discussed below, these come accompanied by inherent trade-offs in either travel times or wider impacts on urban structure and social, economic, and environmental sustainability.

5.3.1. Reducing the number of stops

Without changing any operational parameters, the only way to increase in-vehicle speeds is to reduce the number of stops between origin and destination. One of such ways to achieve that is by reducing the number of stops on the line. Givoni and Rietveld (2014) explore this scenario for commuter rail trips into Amsterdam, based on findings that users were not necessarily always using their closest station. However, this in turn increases the access distance (d) that users must travel to use the service, and since access speeds are significantly lower than in-vehicle speeds, the increase in travel times are expected to be substantial.

Figure 36 illustrates how these trade-offs prevent interstation distances from reducing door-to-door travel times. In this case, the minimum travel time achievable by urban rail is approximately 47 minutes, 42% longer than the time spend by drivers found by Gyimesi et al. (2011) and only marginally better than the baseline capability.

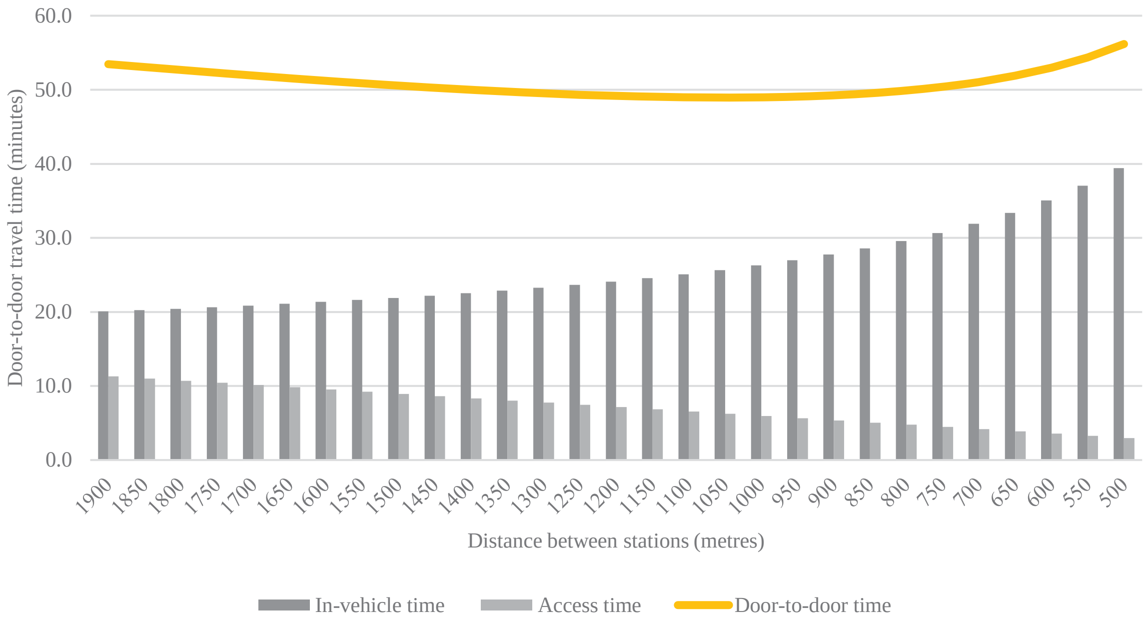


Figure 36. The trade-off between access time and interstation distances on a 19.7 km journey

Consequently, reducing the number of stops does not necessarily result in shorter travel times because of the conflict between access times and line speeds, and important considerations arise from it. Firstly, each 100m increase in interstation distance (D) reduces interstation time (T_s) by 4 seconds but imposes an extra 40 seconds to access time (T_a). Secondly, the value of access time also increases when distances become longer, mostly because these trips will likely include interchanges and certain aspects of unreliability (Wardman, 2004). Finally, greater interstation distances may require access by motorised modes, imposing extra interchange times and also barriers to those who do not or cannot drive (Blainey, et al., 2012).

5.3.2. Increasing maximum line speed

From an access and social equity perspective, it is better that stations are as close to each other as possible so that they can be accessed by non-motorised modes. It is well known that prices around stations are higher, so limited stations along the line can impose higher penalties to those who are more likely to depend on public transport (Loo, et al., 2010; Heres, et al., 2013; Mohammad, et al., 2013). Moreover, research shows that adding stations to a line can induce densification in population and jobs, greater public transport use, and economic development around the extra stations (Dittmar & Ohland, 2004; Handy, 2005; Cervero & Sullivan, 2010). Taking the general elements of urban dynamics into account, urban rail transport systems should focus on maintaining short interstation distances to promote a more balanced and sustainable polycentric distribution. This would enable all users to access stations by walking or cycling, thus reducing social, economic and environmental issues and increasing the health of individuals. In addition, shorter distances between stations promote denser regional development, which improves accessibility with lower mobility needs (Deweerdt, 2016). Therefore, one can model the minimum interstation distances by adding the distances required for acceleration, braking, and jerk. For calculation purposes at this stage, track equipment such as switches and crossings, as well as line-side equipment that might require speed limits, were overlooked.

Yet, the results below show that increasing maximum line speeds under minimum interstation distance is also unlikely to solve the trade-offs and provide door-to-door travel times within the normal travel time budgets, considering the distances to be covered and the number of stops on the line. Since acceleration and braking are limited for passenger comfort, increasing line speeds will inevitably require longer access time, due to longer distances to accelerate and brake. With that, the same problems arise as above, and the minimum door-to-door travel time achievable is approximately 45 minutes when maximum line speed is 135 km/h. In that case, access time would be approximately 8 minutes, but the number of stops prevents the system from achieving a more efficient door-to-door travel time. Moreover, on top of these issues, higher operating speeds also raise concerns over the increased energy consumption of such system.

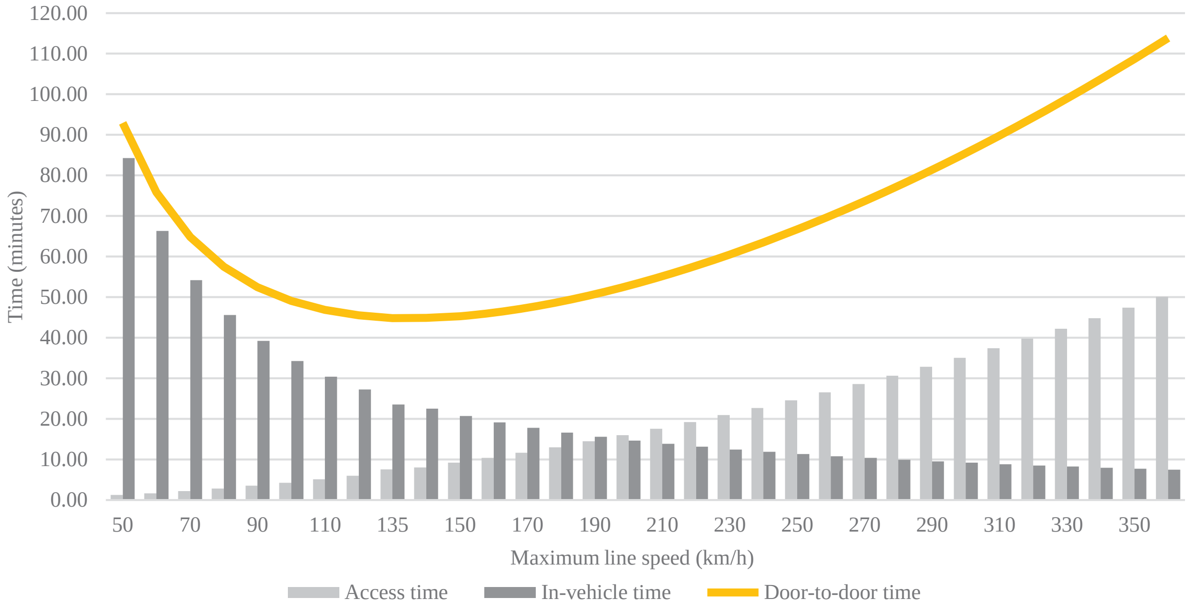


Figure 37. Door-to-door travel times based on maximum line speeds for a 19.7 km journey

5.3.3. Accelerated operation strategies

In face of these systemic shortcomings of mass transport systems, several studies have proposed strategies and technologies to increase average speeds without increasing the distance between stations. In terms of accelerated operation strategies, Furth and Day (1985), Fu et al. (2003), and Vuchic (2005) suggest three methods to increase speeds at constant interstation distances: (1) local/express; (2) zonal; and (3) skip-stop. These methods all share the same principle of attending only some of the stations on the line and thus limiting the number of intermediate stops between origin and destination and consequently increasing in-vehicle travel speeds.

Local/express services, perhaps the most common, comprise an array of services that follow patterns from stopping at every station (local service) to stopping only at those with higher demand (express service), and any combination in-between (Vuchic, 2005). Lines can have different types of express services in order to cater for different demands, such as the Nankai line in Osaka that offers six different stopping patterns along the same line (Figure 38. Nankai line map illustrating all the different services).

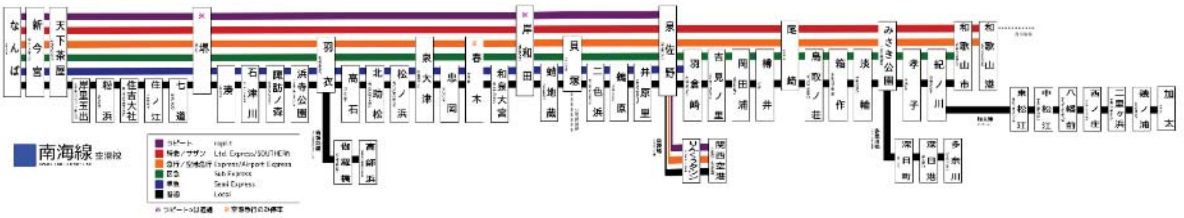


Figure 38. Nankai line map illustrating all the different services

Zonal operations, as the name suggests, divide the urban area in distinct zones and trains connect each zone to the centre. Journey time, consequently, is reduced as the train skips the stations in the zones between the centre and the destination zone. Finally, skip-stop operations assign stations of a line into fixed stopping patterns categories: A, B, and AB; or sometimes A, B, C, and Transfer as illustrated in Figures Figure 39 and Figure 40. Meanwhile, it logically follows that trains run in their assigned patterns: trains A stop at A and AB stations, trains B stop at B and AB stations, and trains C stop at C and Transfer (ABC) stations.

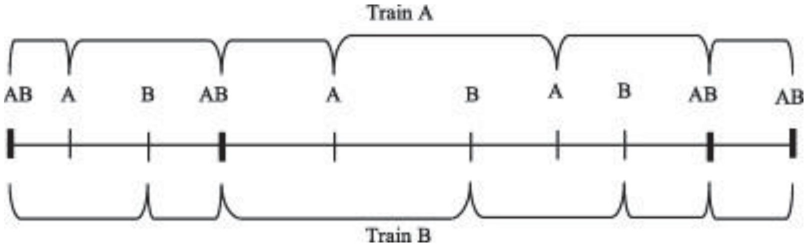


Figure 39. Illustration of a skip-stop strategy with A and B patterns (Freyss, et al., 2013)

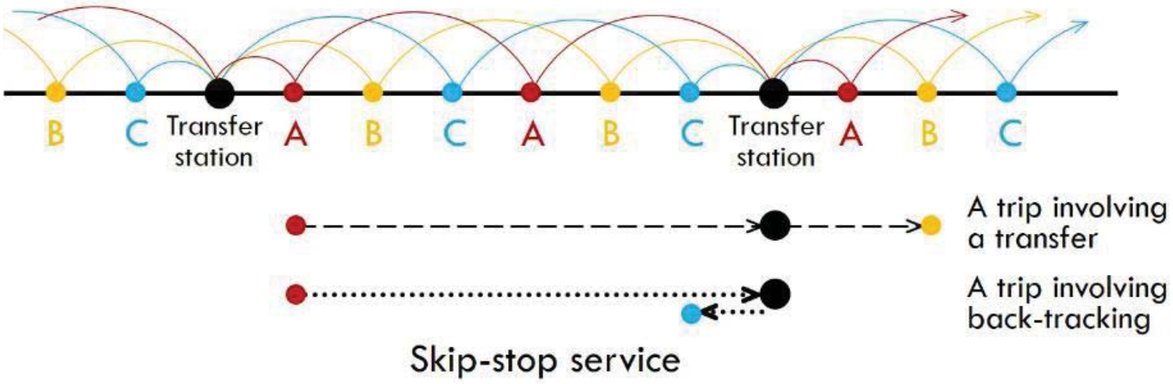


Figure 40. Illustration of a skip-stop strategy with A, B, and C patterns (Gu, 2015)

Although all strategies can potentially reduce the fleet size and operational costs, they have significant influence on the generalised cost of users. Firstly, there is a logical impact on those wanting to travel from stations A to B and B to A (or A to C, B to C, C to A in a three-type scheme). These passengers will have to transfer to another service in order to reach their destination, and sometimes even travel backwards to do so. For instance, Lee (2014) found that, after the adoption of skip-stop operations in line 4 of the Seoul Metro, while in-vehicle times became 20% to 26% shorter, waiting, transfer and additional access time increased by 24% to 38%.

Secondly, they will create significant disparities in travel times between users in different locations. Transfer stations will offer greater connectivity which in turn leads to a market premium in their surroundings, feeding a positive feedback between prices and demand, and potentially leading to crowding. Givoni and Rietveld (2012) found that passengers do not necessarily use their closest station but travel to those which offer better access. In addition, the appraisal of demand based on value of time can reinforce these disparities providing better access to those who already enjoy higher access as the disparity in connectivity can price out those who may need it most.

Thirdly, since overall frequency tends to be reduced, the service tends to operate below its capacity, which is an important factor for metro systems in large cities. Finally, such patterns predict savings on linear journeys where interchanges are limited, such as the case of centralised urban areas. Nonetheless, the polycentric nature of megalopoleis and the resulting complexity in travel patterns mean that accelerated strategies reduce the points of interchange for users and thus add important penalties to travel time and accessibility.

5.4. Development of operational concept

In face of the constraints faced by current capabilities, the operational concept should focus on overcoming the coverage paradox in order to achieve the main objectives of door-to-door journey times and theoretical capacity. In that sense, the system should ideally maintain stations close to each other in order to reduce access time, but simultaneously avoid the time penalties of the increasing number of extra stops. Moreover, as seen in the cases of local/express and skip-stop strategies, the system should also ensure that all users can reach all services at any station in order to distribute the benefits within the network, rather than offering time savings to a share of the passengers at the expense of time penalties to others. This section explains the conceptual solution investigated in this thesis, which is composed of three foundational aspects.

5.4.1. Operational strategy

Firstly, operations are divided in different services. Each service stops at stations along the line observing a certain pattern (P_x), not unlike a local/express operation. However, the main and crucial difference lies in the pattern algorithm to ensure that all stations are attended by all services. For that reason, the line need to be circular or operated as if it were an infinite loop. A line can take various forms to operate as an infinite loop manner. Figure 41 illustrates a traditional circular line as found operating in Moscow, Tokyo, Copenhagen, and other cities. Figure 42 represents a straight line where vehicles turn around at the ends to serve the other direction. And Figure 43 illustrates the ‘carrousel’ operation seen in São Paulo Metro, where trains switch tracks between the second to last and the last station of the line. That way, the trains are already on the return track for operating on the other direction. However, it must be highlighted that the ‘carrousel’ operation requires a greater level of automation in order to reduce possible conflicts between trains when switching within short headways.

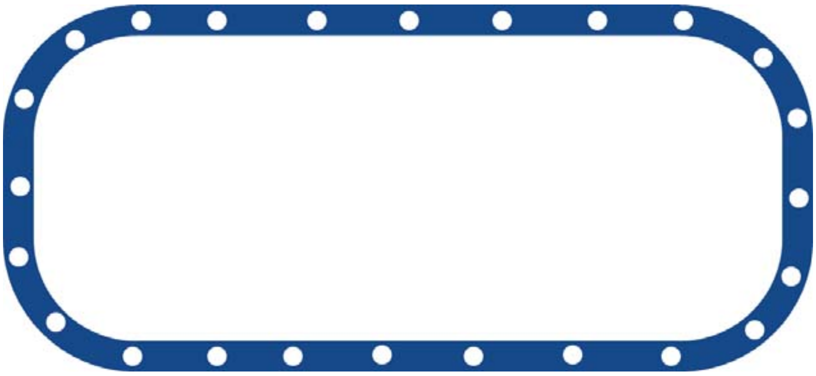


Figure 41. Illustration of a circular line

Figure 42. Illustration of an infinite line with loops

Figure 43. Illustration of 'carousel' style operations on a straight line

Regardless of the layout of the line, an important consideration from a functional perspective is that it must allow services to maintain the stopping pattern counter even when changing directions. In essence, a service does not address stations specifically, but simply follows its pattern indefinitely regardless of the direction of the line. For example, when a service that stops every three stations (P_3) departs from the second to last station in one direction of the line, it will skip the last station of that direction, the first station of the other direction, and the next stop will be the second station in the other direction. This is logically only needed in linear layouts, as circular lines automatically operate in a way that permits continuous forward movement.

Moreover, the number of stations and the stopping patterns need to be defined in a way that permits services to attend all stations. For that to happen, the number of stations must not by divisible by any of the patterns, and as a result each vehicle will eventually stop at every station, taking a number of 'laps' equal to its pattern in order to call at all stations. The mathematical rule of the stopping patterns can be defined by Equation (6.

$$(6)\ P_x = \{x \mid x < N_s/2, x \nmid N_s, x \in N\}$$

For instance, a service that stops every 3 stations (P_3) will take 3 laps to serve all stations on the line, regardless of the total number of stations. For the case illustrated in this thesis, five patterns were adopted: stopping at every station, or every two, three, five, or seven stations ($P_{1,2,3,5,7}$). Figure 44 illustrates the stopping patterns.

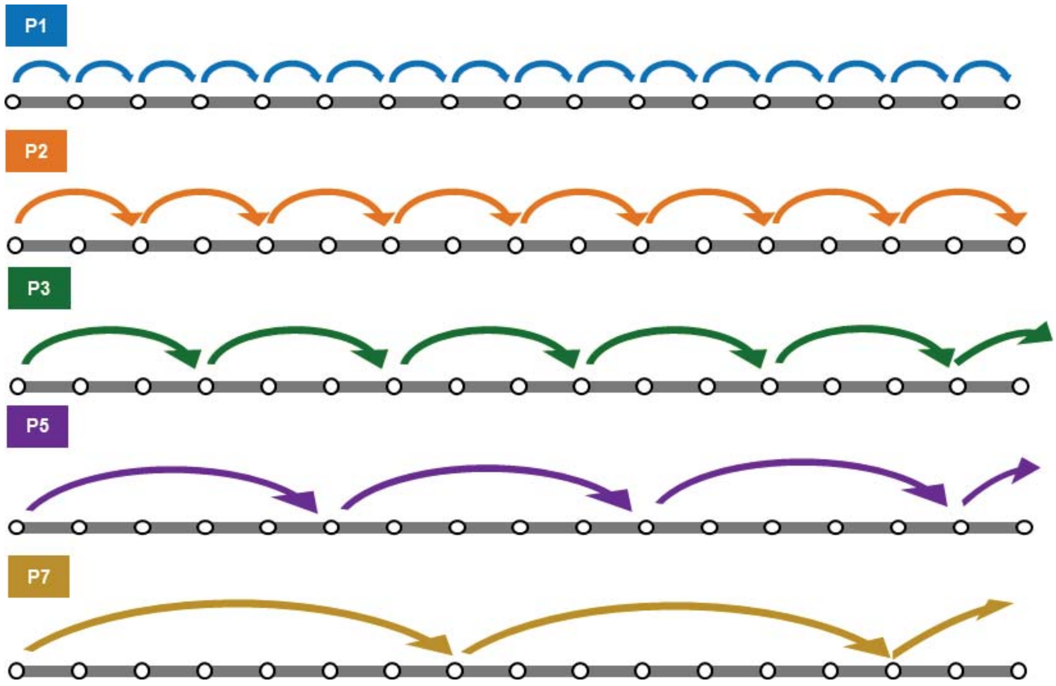


Figure 44. Example of stopping patterns

The proposed concept solves the issues of local/express and skip-stop operations as all stations are equally served by all patterns providing equally optimal travel times to all users, since any origin-destination journey on the same line requires at most one change. Time savings can be calculated by the reduction in the time penalty (T_l) that each stop adds to in-vehicle time (T_v). Using equation (5), the total time taken to travel from one station to the next is:

$$(7)\ T_s = \frac{3.6 * D_{min}}{V} + T_l$$

Where D_{min} is the minimum running distance between stations, V is the maximum speed of the line, and T_l is the time penalty from stopping at the station. Consequently, services that do not stop at every station will have fewer penalties:

$$(8)\ T_{s_x} = xT_s - (x-1)T_l$$

Given a number of stations determined by the journey length and the distance between stations (D), one can calculate total in-vehicle time (T_v). The distinction between D and D_{min} will be explained in the following section. In-vehicle time is a function of the time between stations (T_s), the number of stations (N_s), and the pattern (P_x) involved:

$$(9)\ T_{v_x} = \frac{T_{s_x} * \frac{(N_S - 1)}{P_x}}{60}$$

The extra time for an additional change between services is only needed for when the number of stations travelled is a prime number, representing a reduced portion of travellers and a smaller time penalty. This research adopts this time as platform interchange time (T_p) and assumes it to be equal to the headway between trains.

5.4.2. Line layout

Secondly, such operational strategy requires stations to be located off the main line in order to operate efficiently. In a similar way to other accelerated operational strategies, a four-track arrangement is necessary so that the slower services do not stand in the way of their faster counterparts and end up eliminating the advantages in travel speeds of the different stopping patterns. When stations are located off the main line, interstation distances can be reduced to a minimum without adding time penalties usually connected to extra stops. Vehicles that are not serving a particular station will continue on the main line while those that are bound to stop will move to the passing loop where the station is and start braking (Figure 45). When reaching a platform at any station, passengers are able change to the vehicle or vehicles that will serve their destination station. The operation of these different patterns in a two-track system would require more complicated timetables that may result in longer headways and reduced capacity, which explains the preference for a four-track system.

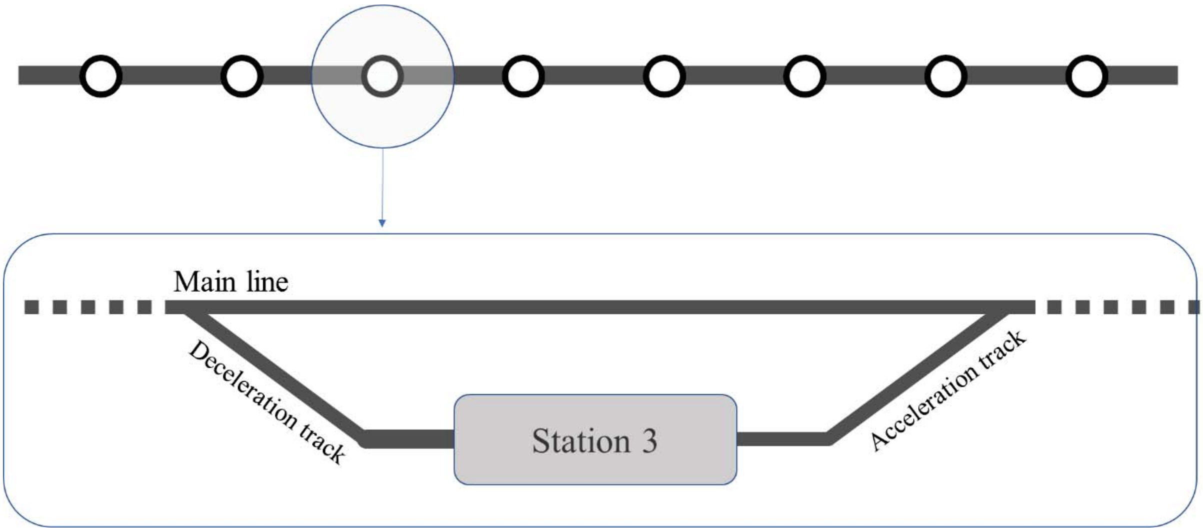


Figure 45. Illustration of off-line stations

In that arrangement, there is an added benefit that the distance between stations (D_s) becomes even smaller than it would be if the station were not in a loop. The physical limitation to the minimum running distance between stations (D_{min}) is calculated by the sum of acceleration, jerk, and braking distances, and clearance distance for the switch. Consequently, placing the station off the main line also contributes to closer access for passengers without interfering with the maximum line speed because the added track length needed for the loop results in a shorter interstation distance (Figure 46).

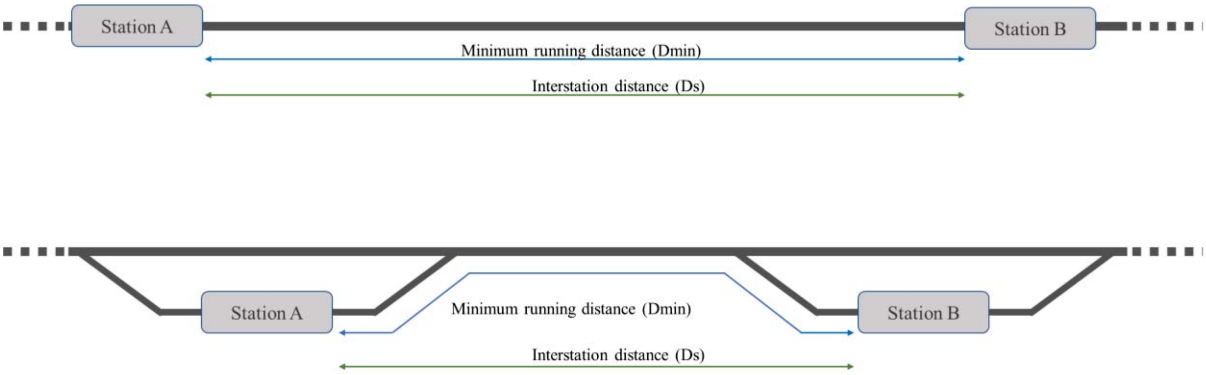


Figure 46. Illustration of minimum running distance and interstation distance in different line layouts

To calculate the minimum running distance (D_{min}), it is necessary to calculate the following parameters:

Acceleration distance (D_α), the distance needed to accelerate at a certain rate (α) from 0 km/h to maximum speed (V), calculated by:

$$(10)\ D_\alpha = \frac{V^2}{2\alpha}$$

While, braking distance (D_β), the distance needed to decelerate from maximum speed (V) to 0 km/h at a certain rate (β) is given by:

$$(11)\ D_\beta = \frac{V^2}{2\beta}$$

Jerk distance (D_j): distance required to transition from acceleration to braking in a comfortable way to passengers, based on the time to transition (t_j), acceleration (α), braking (β), and jerk (j) rates, and can be calculated as:

$$(12)\ t_j = \frac{|\alpha|+|\beta|}{j}$$

$$(13)\ D_j = V * t_j + \frac{\alpha * {t_j}^2}{2} + \frac{j * {t_j}^3}{6}$$

Switch clearance (L_c) corresponds to the length of horizontal track needed for the turnout track to reach a clearance margin (L_{cm}) from the main track, based on the minimum curve radius (R_{min}) (Figure 47). The latter is dependent on the maximum speed (V), cant (h_a), cant deficiency (h_b), track gauge (G), and gravity (g) approximated at 9.8 m/s².

$$(14)\ R_{min} = \frac{G*V^2}{g(h_a+h_b)}$$

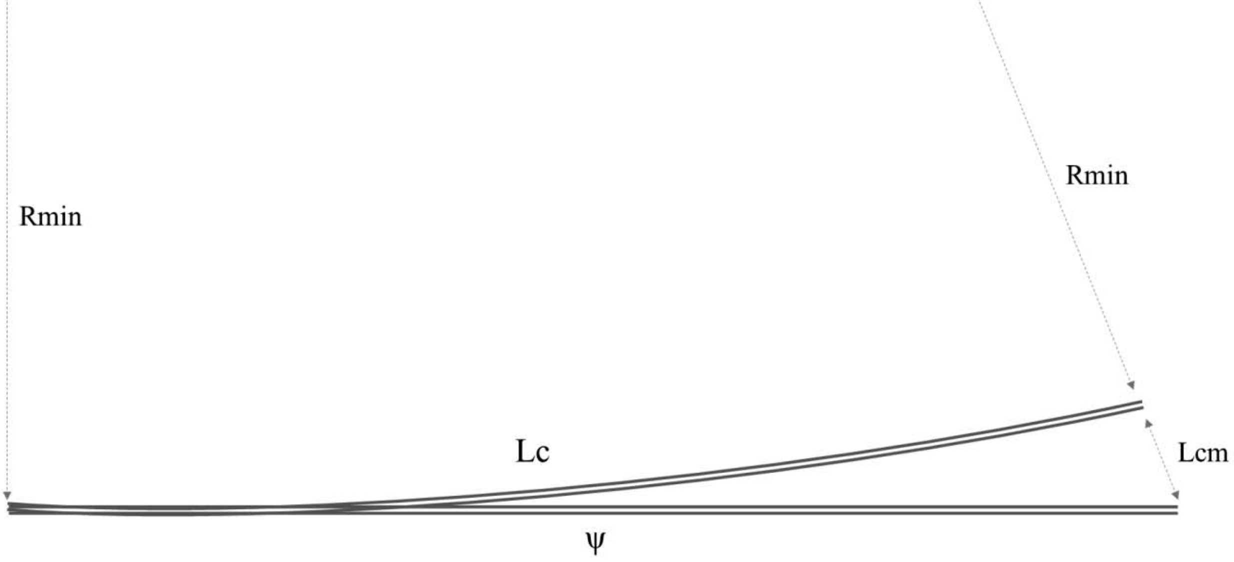


Figure 47. Distance needed to safely clear the turnout

Calculation for the switch clearance length (L_c) depends on the clearance margin, set for ensuring a safe horizontal distance between consecutive vehicles.

$$(15)\ L_c = \sqrt{(\psi - \left(\frac{\psi*L_{cm}}{R_{min}+L_{cm}}\right))^2 + (\frac{R_{min}*L_{cm}}{R_{min}+L_{cm}})^2}$$

With that, it is also possible to infer the distance between the platform (D_{lp}) and the main line by assuming that the angle at the end of L_c remains unchanged and a straight line starts from its end point. This distance is in fact responsible for the theoretical reduction of the interstation distance in relation to the minimum running distance. The greater D_{lp} is, the farther the platform is from the main line and consequently the smaller is the ratio between interstation distance (D) and minimum running distance (D_{min}).

$$(16)\ D_{lp} = \left(\frac{R_{min}*L_{cm}}{R_{min}+L_{cm}}\right) * \left(\frac{\psi+D_\beta}{\psi}\right)$$

Finally, the minimum running distance (D_{min}) is calculated by:

$$(17)\ D_{min} = D_{\alpha} + D_{\beta} + D_{j} + 2L_{c} + L_{p}$$

Where L_p is the platform length, dependent on the dimension of vehicles which will be discussed in the following section.

These equations show a trade-off between the clearance margin (L_{cm}) and the distance between line and platform (D_{lp}). A greater line-platform distance would reduce interstation distance in comparison to minimum running distance, yet at a cost of a bigger clearance margin that would affect the headway and capacity as speed reductions prior to the turnout would be required.

The clearance margin (L_{cm}) is dependent on the vehicle width (W) and track gauge (G), in order to allow vehicles to pass each other safely. A value of 40 cm was estimated to cater for the kinematic envelope of moving trains, at a greater amount than the guidance from the RSSB (2004) for safety purposes. This value, based on guidance for between 20 cm and 25 cm, has been used for calculation purposes only, and should be covered by specific research once rolling stock design is completed. The equation for L_{cm} is thus:

$$(18)\ L_{cm} = W + 0.4$$

Moreover, vehicle width (W) was set at 2.6 m based on the London Underground fleet (London Underground, 2007). Based on these parameters and those of Table 3, the distance between the line and platform must be 10.8 m for L_{cm} to be at its minimum value of 3 m. The calculations also assume that are no other curves between the switch and platform arrival in order to reduce lateral acceleration. Under these premises, the difference between the minimum running distance (D_{min}) and the distance between stations (D) is negligible because of the small angles.

While this track arrangement can optimise in-vehicle travel, platforms then become a bottleneck. In a normal arrangement, if all services shared the same platform, it would incur significant penalties in waiting time as users would have to wait for several services before boarding their intended train. The headway service interval between two trains of the same pattern increases and could exceed $(H_{min}+T_d)*(P_x+1)$ at certain points along the line, where H_{min} is the headway between trains, T_d is the dwell time at the stations, and P_x the number of patterns.

To avoid conflict and delays, services of different patterns need different platforms to operate. Considering the distribution of passengers in the different patterns, the number of platforms needed is equal to the number of patterns plus an emergency exit in case of faults or failures in the system. This in turn calls for investigation as land requirements and technical feasibility become an issue. If the distribution is done in parallel (Figure 48), the land required for the system becomes and important concern, and also the platform change distance between services would impose a significantly higher time penalty (including two sets of stairs or escalators and a walk of at least the distance between platforms). As it is known that walking and transfer times are seen more negatively than in-vehicle time, then such imposition would affect the attractiveness of the system. The combination of two services on adjacent platforms are, on the other hand, a benefit, since arranging the most likely transfers to be on the same platform would significantly reduce platform change time (T_p).

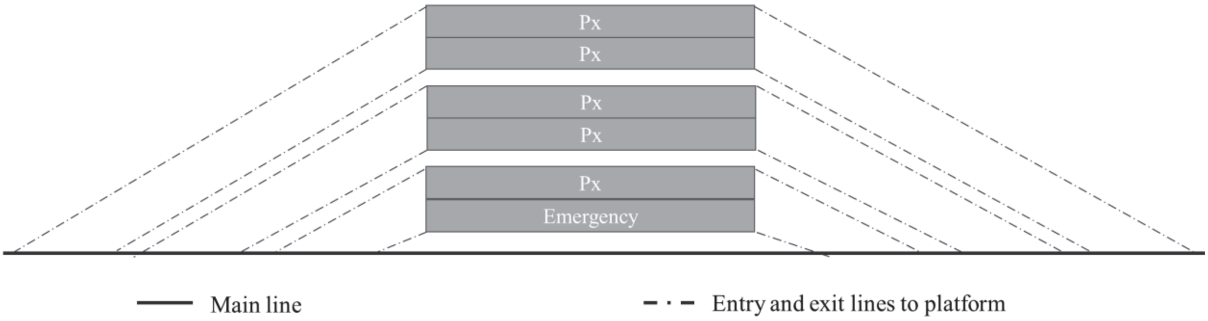


Figure 48. Possible platform arrangement in parallel

A more efficient layout to reduce land use and permit retrofitting of legacy stations is to distribute platforms linearly (Figure 49). That way, only one platform is required rather than at least (P_x+1)/2 platforms, and the other side of the platform can be used for another line, which in turn also reduces interchange times (T_i).

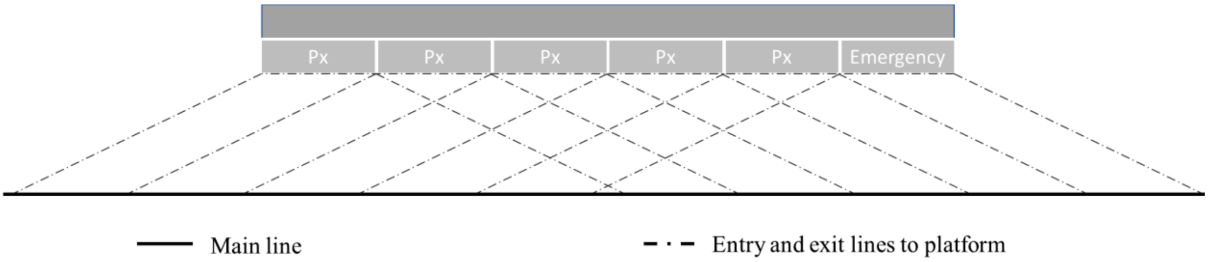


Figure 49. Linear platform arrangement

Moreover, platform change time (T_p) is reduced to only walking time between sections, with a maximum value in minutes calculated by Equation (19, where L_t is the length of each train, P_x is the number of different patterns operated, and walking speed is assumed to be 80 m/minute based on Pachi and Ji (2005).

$$(19)\ T_p = L_t P_x / 80$$

However, such arrangement, while more efficient from the perspective of passenger journey times, raises issues over the reliability and safety of the system. The presence of various crossings at short distances from each other increases the complication in operations, which can in turn escalate costs and maintenance requirements. Moreover, current traffic management technologies may not be able to cope with such tight control.

In addition, while this offers a more efficient land use, if each service is composed of regular metro trains that are approximately 120 metres long, the necessary platform length would become unfeasible, reaching almost the total distance between stations.

5.4.3. Rolling stock

The solution proposed, therefore, is the platooning of autonomous vehicles that are virtually coupled rather than physically connected carriages that form a conventional train. In this concept, one pattern is assigned to each vehicle (the equivalent of a car in a metro train), or a set of vehicles where demand is high, and vehicles of different patterns are virtually coupled to form a platoon. Each platoon is ideally composed of vehicles of all patterns so that the headway between services of the same pattern remains at its minimum (Figure 50).

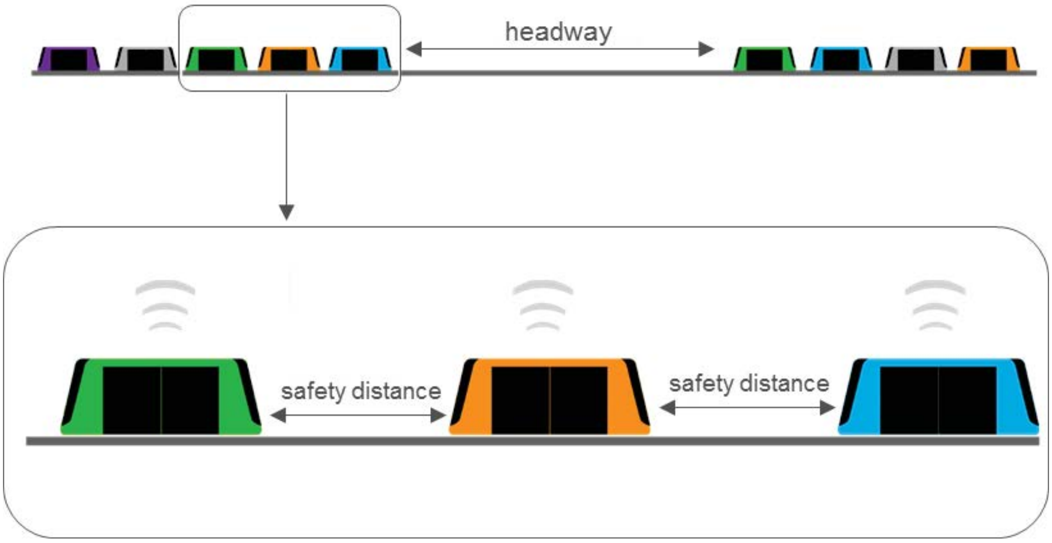


Figure 50. Illustration of platooning vehicles

This operational concept contains two distinct headways. Firstly, within the platoon each vehicle maintains itself from the vehicle ahead at a safety distance (d_{min}). To do so, vehicle-to-vehicle communication algorithms similar to those developed for automated highways are employed (Tank & Linnarts, 1997; Gehring & Fritz, 1997; Robinson, et al., 2010; Fernandes & Nunes, 2015).

Since all platoons contain vehicles of all patterns, this strategy guarantees that every station is serviced by all patterns at the minimum headway between trains. Vehicles that skip a certain stop will continue on the main line while those stopping will move to the loop and start braking (Figure 51). A set of algorithms by Fernandes and Nunes (2015) is used as a foundation to the operational concept as it also assumes stations placed off the main line. In this set, the movement and organisation of platoons is governed by four algorithms, which depend on the current position of the vehicle and its relationship with the platoon.

When a platoon approaches a station, vehicles will adjust their safety distance so that those attending the station can do so safely (b). The movement of vehicles will result in the identification of a leader as the vehicle in the front. The leader will assume its specific algorithm, and the vehicles will take as guidance and reassemble following Algorithm 2 (c). Once attended, the vehicles at the station will move towards the main line, where they may encounter another platoon to join (d,e). The algorithm controls their movement so that the virtual coupling can be conducted safely and at speed (f).

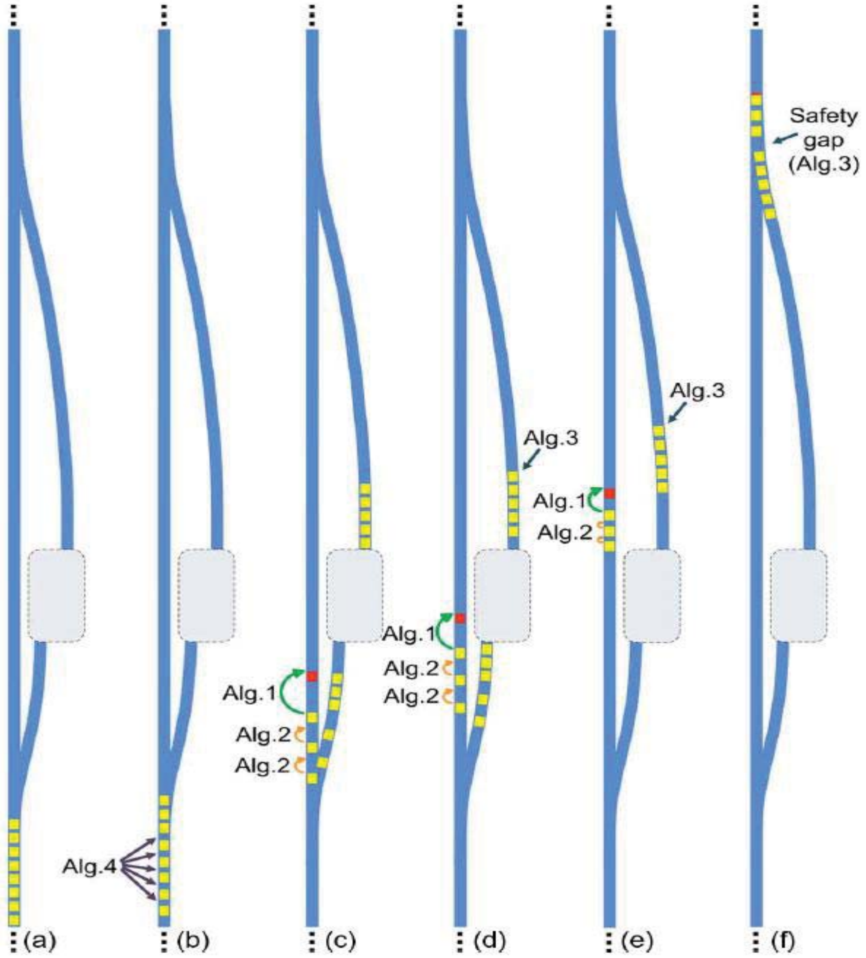


Figure 51. Illustration of dynamic vehicle and platoon movements in and out of platform loop (Fernandes & Nunes, 2015)

The distance between vehicles (d_{min}) in a platoon is based on Gehring and Fritz (1997) and calculated by the sum of the distance at rest (d_r) of 0.2 m, added to the product of velocity (V) and processing time (τ) of 100 ms, the length of the vehicle (L_v), and a margin for position error (e) assumed to be 10% of the length of the vehicle. The technical aspects of turnouts will be discussed in Section 5.6. The distance between vehicles can be calculated, in metres, as:

$$(20)\ d_{min} = D_r + V * \tau + L_v + e$$

It logically follows that the length of a platoon (L_{pn}) when moving at maximum speed (V) is calculated by the number of patterns in operation (N_{px}), the number of vehicles of each pattern in a platoon (N_{vx}), the length of each vehicle (L_v), and the minimum distance between vehicles (D_{min}):

$$(21)\ L_{pn} = N_{px} * N_{vx} * L_v + D_{min}(N_{px} * N_{vx} - 1)$$

The headway (H_{min}) is the space between platoons that operate under the assumption of moving block signalling. Starting from calculations by Takagi (2014), the minimum headway between platoons (in seconds) need to take into account the maximum speed in m/s (V), braking rate in m/s² (β), platoon length in metres (L_{pn}), and switch clearance in metres (L_c). Considering that H_{min} is longer than dwell time (T_d) and that each pattern (P_x) has its own station loop, then it is not necessary to include dwell time in its calculations because by the time a vehicle is arriving at the platform, the preceding vehicle has already departed.

$$(22)\ H_{min} = \frac{\frac{V^2}{\beta} + L_{pn} + 2L_c}{V}$$

The length of the vehicles also influences the design of platforms and the respective platform sections as explained on Figure 49. The length of each platform section (L_{px}) is dependent on the length of the vehicles (L_v), the number of vehicles per pattern (N_{vx}), the distance between vehicles at rest, and a certain extra length for vehicles to manoeuvre without crashing with those on the adjacent sections as illustrated in Figure 52.

$$(23)\ L_{px} = L_v * N_{vx} + D_r(N_{vx} - 1) + 2L_{pc}$$

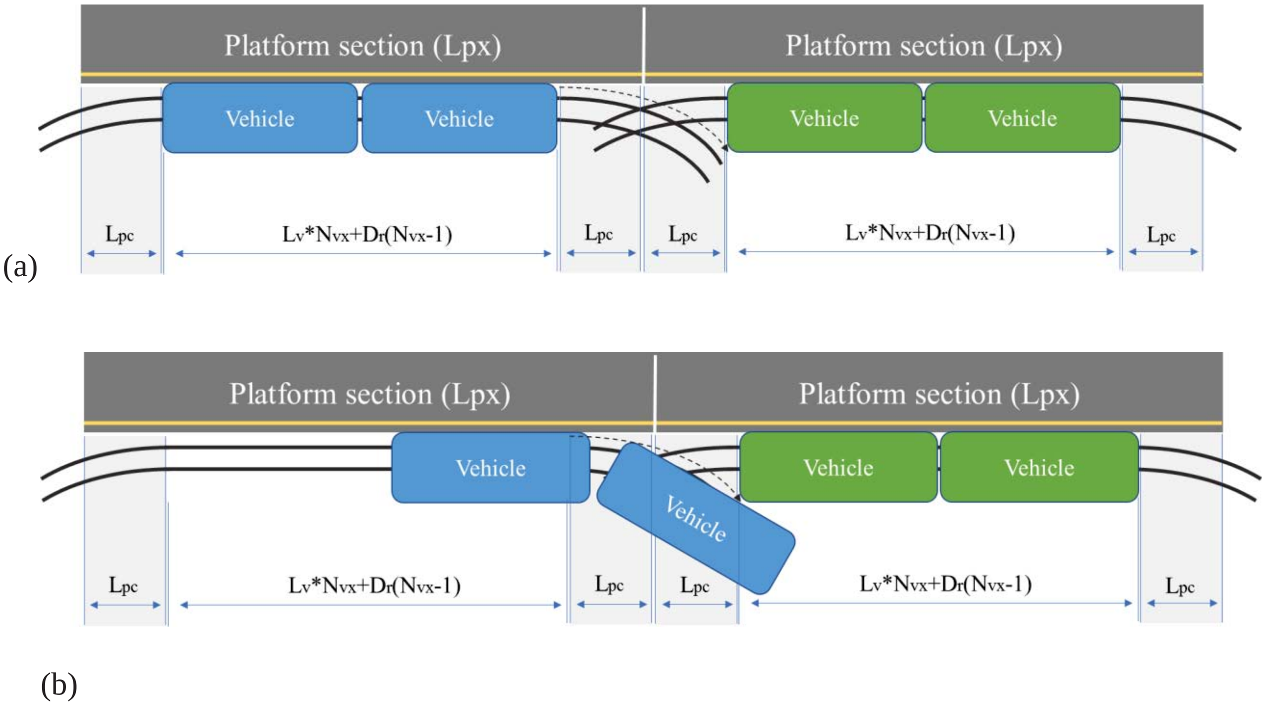


Figure 52. Illustration of platform sections as a function of rolling stock parameters

Each section needs to contain an additional clear area for entering and leaving the platform in a safely manner, where vehicles turn 3 metres from platforms given the assumed width of trains of 2.6 m (Figure 53).

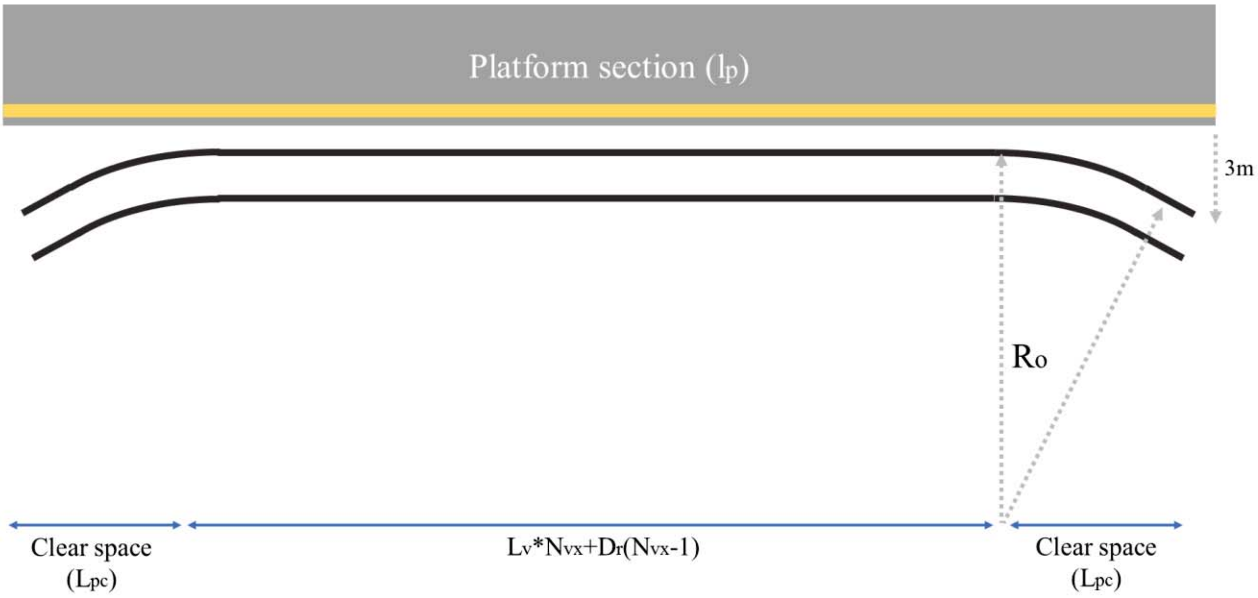


Figure 53. Illustration of platform section layout

The clear space area (L_{pc}) is a function of the outer radius of the curve profile (R_o), which is in turn dependent on the minimum turning radius (R_t), wheelbase of the vehicle (L_w), vehicle width (W), and track gauge (G). The concept relies on underlying assumption is that each vehicle has four axles, and is not articulated, as a means to simulate a baseline reality.

As shown in Figure 54, the radius that needs to clear at a distance of 3 m of the platform is the outer radius (R_o). Based on Vuchic (2007), there are four different radii to be taken into consideration: turning radius from the centre of the bogies (R_t), R’ as the radius of the centre of gravity, and the inner (R_i) and outer radii (R_o) based on the wider curve profile.

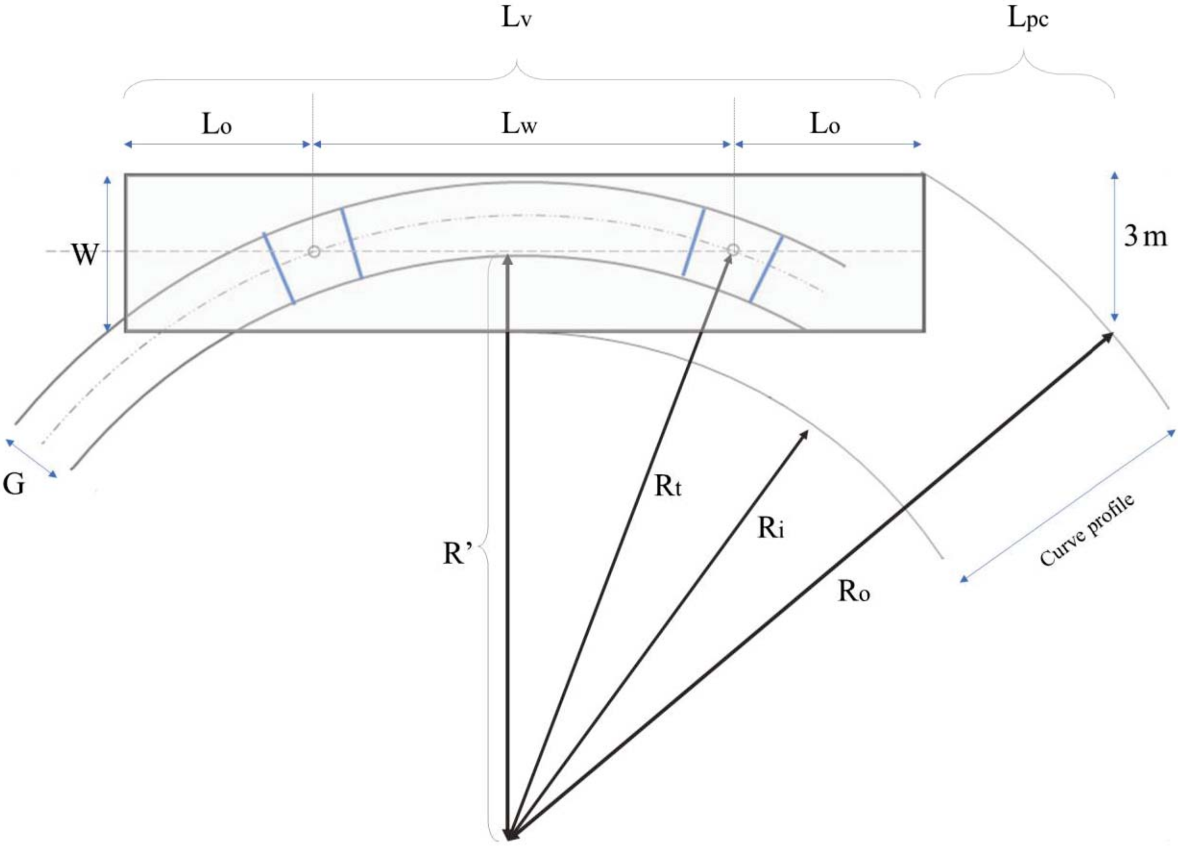


Figure 54. Calculation of the clear space area for platform section (based on Vuchic (2007))

Firstly, there is an assumption that R’ cannot be smaller than R_t minus half the track gauge in order to avoid overturning. Therefore, as the concept aims at finding the minimum radius possible, R’ will be considered to be:

$$(24)\ R' = R_t - \frac{G}{2}$$

To find L_{pc}, it is first necessary to find the outer radius (R_o), which is based on the turning radius (R_t), following this sequence of equations:

$$(25)\ R_t = \frac{{L_w}^2 + G^2}{4}$$

Where L_w is the wheelbase, assumed to be 60% of the vehicle length, based on the proportions found on the London Underground fleet (London Underground, 2007). G is the track gauge, assumed to meet the UK standard of 1,435 mm. From these equations, the inner (R_i) and outer (R_o) radii can be calculated respectively with these equations:

$$(26)\ R_i = R' - \frac{W}{2}$$

$$(27)\ R_o = \sqrt{\left(R_t + \frac{W}{2}\right)^2 + (\frac{L_w}{2} + L_o)^2}$$

Where L_o is the overhang between the centre of the bogies and the front or back of the vehicle. With these variables in hand, it is possible to calculate L_{pc} (Figure 55).

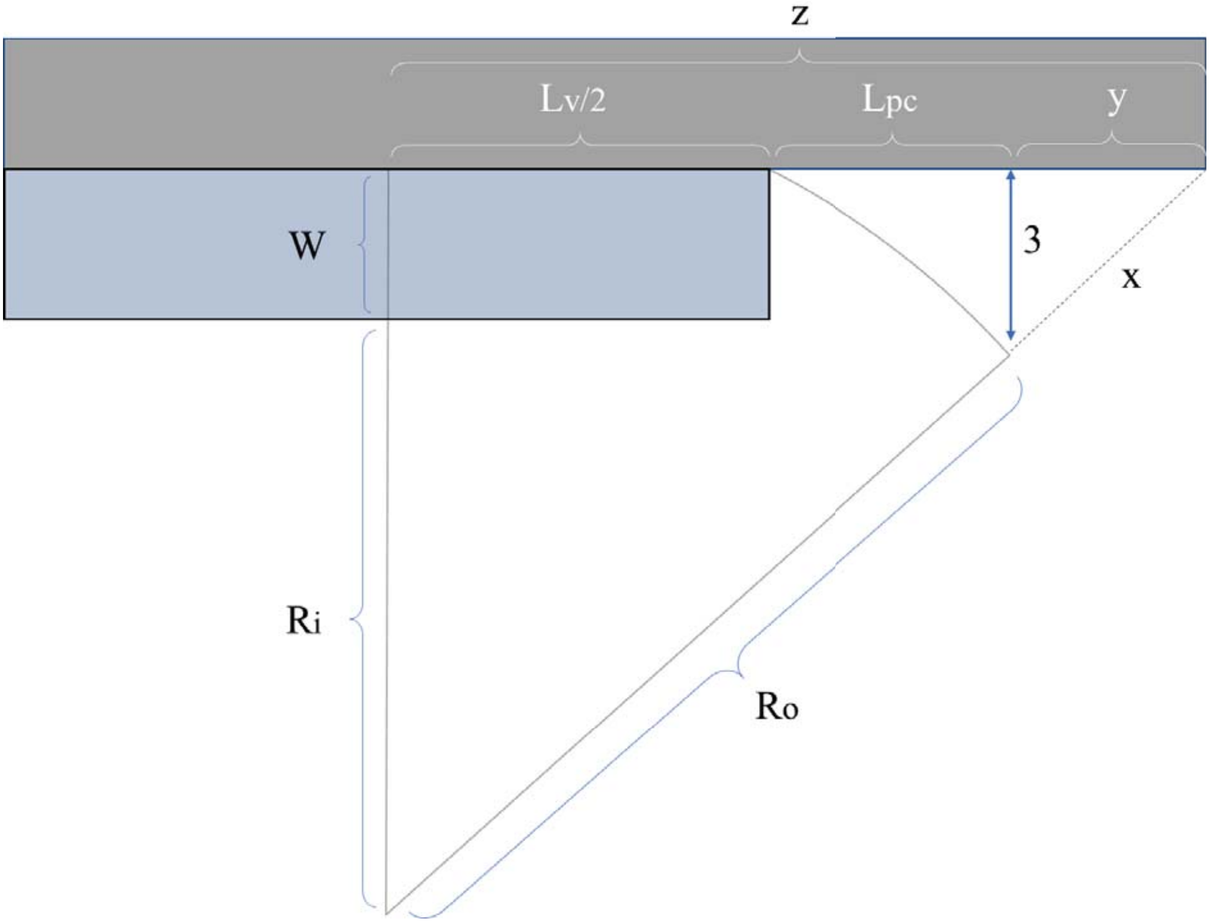


Figure 55. Platform clearance margin in relationship to rolling stock variables

The process is based on basic geometry and follow these equations:

(28) $L_{pc} = z - \frac{L_v}{2} - y$

(29) $y = \frac{3z}{R_i + W}$

(30) $(R_o + x)^2 = (R_i + W)^2 + z^2$

(31) $x = \frac{3(R_o + x)}{R_i + W}$

At this point, some assumptions over physical specifications need to be made as functions become circular. It is evident that the capacity of the system depends on the capacity of each vehicle, the number of vehicles composing a platoon, and the frequency with which vehicles attend the station. It is also evident that the capacity of each vehicle depends on its dimensions. On the other hand, longer vehicles require longer platforms and, as door provision is non-linear, may increase dwell time. Longer vehicles also increase platform time (T_p) as passengers will need to walk longer distances between vehicles when necessary. In addition, longer platforms increase the distance between stations and can impact access time (T_a).

For instance, if one uses a top-down perspective, then calculations of rolling stock specification would begin from the assumption of 100,000 passengers per hour per direction as stated in the earlier sections. To do so, vehicle dimensions (L_v and W) can be deduced by the necessary vehicle capacity (C_v), which is the total capacity (C) divided by frequency (F). However, frequency is based on the headway between platoons (H_{min}), which in turn depends on the vehicle length.

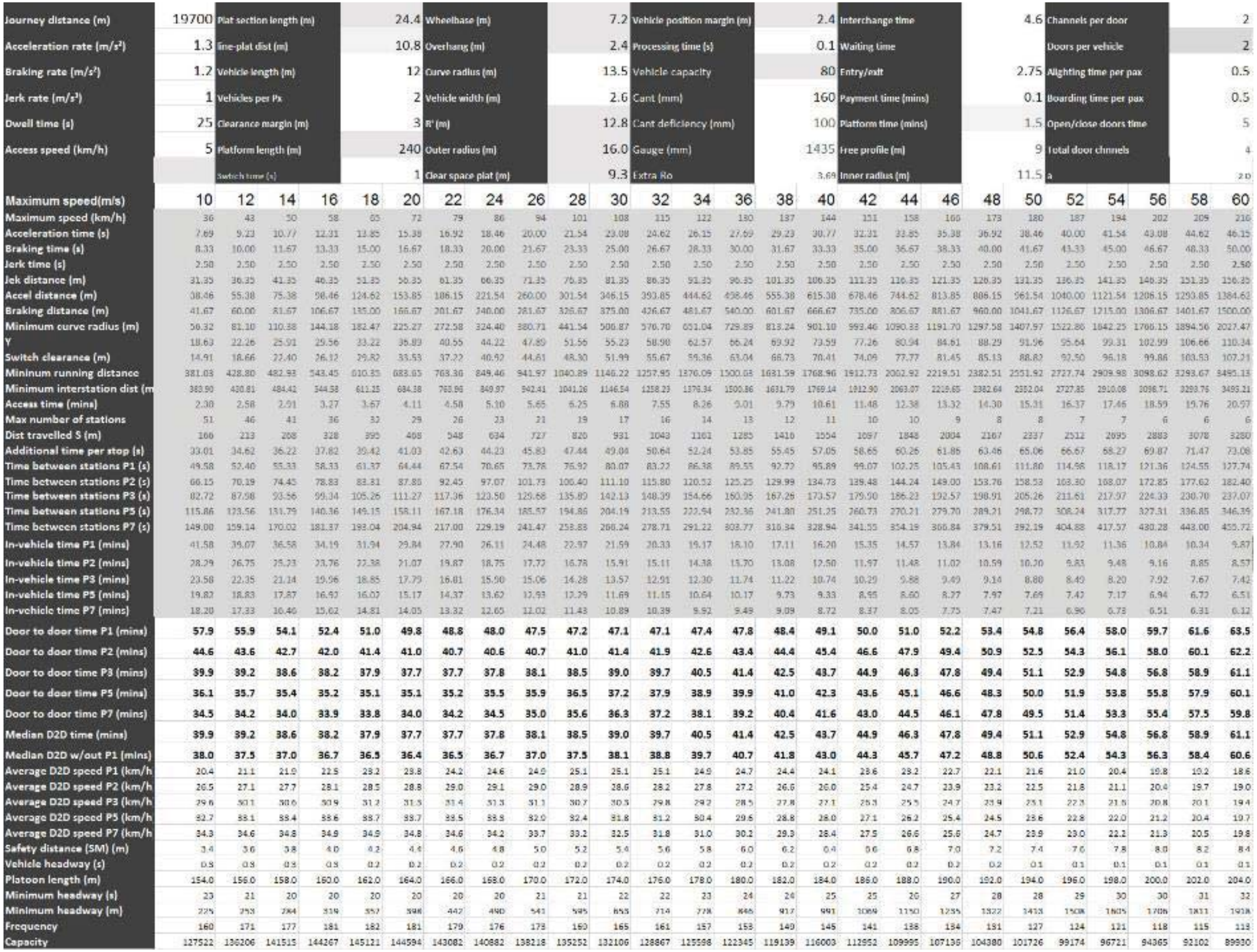
Therefore, the circular functions require one or some variables to be regarded as input variables. For the model produced, vehicle dimensions were chosen as input variables because they are intrinsically related to detailed design which is already uncertain at that level of abstraction. The width of the vehicle (W) was set at 2.6 m, based on the London Underground fleets (London Underground, 2007), and the length of the vehicle (L_v) at 12 m for testing the impact on curve radii. For vehicle capacity (C_v), it was assumed that 80% of the total area was available for passengers, with a loading factor of 4 passengers/m².

Using the inputs in vehicle length, it is possible to determine dwell time based on the number of doors (n_d), the number of channels per door (n'), and the average time per passenger to board or alight (t'). This time is estimated at 0.5 seconds per passenger using one channel (Vuchic, 2005; Thoreau, et al., 2016). On top of that, dwell time must also account for the time needed to safely open and close doors (t_{oc}), estimated at 2.5 seconds each. For calculation purposes, an extreme scenario will be assumed where all passengers in a full vehicle alight at the platform, being then replaced by another group of passengers boarding that fully occupy the vehicle. It logically follows that dwell time can be calculated by:

$$(32)\ T_d = \left(\frac{2C_v}{n_d * n' * \frac{1}{t'}} \right) + t_{oc}$$

5.4.4. Model results

All variables and functions of the operational concept were then aggregated in an Excel-based model and subsequently simulated in MatLab. Journey distance was takes as the main input, followed by technical parameters such as acceleration, braking, jerk, and vehicle dimensions. As Figure 56 shows, for every combination of inputs, there is an optimal combination of maximum line speed and access time that results in the minimum door-to-door journey time. Model results indicate that the closer station spacing maintains the access time at its minimum while the optimised operational model increases the average in-vehicle speeds.



Journey distance (m)	19700	Plat section length (m)	24.4	Wheelbase (m)	7.2	Vehicle position margin (m)	2.4	Interchange time	4.6	Channels per door	2
Acceleration rate (m/s^2)	1.3	line-plat dist (m)	10.8	Overhang (m)	2.4	Processing time (s)	0.1	Waiting time		Doors per vehicle	2
Braking rate (m/s^2)	1.2	Vehicle length (m)	12	Curve radius (m)	13.5	Vehicle capacity	80	Entry/exit	2.75	Alighting time per pax	0.5
Jerk rate (m/s^3)	1	Vehicles per Px	2	Vehicle width (m)	2.6	Cant (mm)	160	Payment time (mins)	0.1	Boarding time per pax	0.5
Dwell time (s)	25	Clearance margin (m)	3	R' (m)	12.8	Cant deficiency (mm)	100	Platform time (mins)	1.5	Open/close doors time	5
Access speed (km/h)	5	Platform length (m)	240	Outer radius (m)	16.0	Gauge (mm)	1435	Free profile (m)	9	Total door chnnels	4
		Switch time (s)	1	Clear space plat (m)	9.3	Extra Ro	3.69	Inner radius (m)	11.5	a	2.0

Maximum speed(m/s)	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58	60
Maximum speed (km/h)	36	43	50	58	65	72	79	86	94	101	108	115	122	130	137	144	151	158	166	173	180	187	194	202	209	216
Acceleration time (s)	7.69	9.23	10.77	12.31	13.85	15.38	16.92	18.46	20.00	21.54	23.08	24.62	26.15	27.69	29.23	30.77	32.31	33.85	35.38	36.92	38.46	40.00	41.54	43.08	44.62	46.15
Braking time (s)	8.33	10.00	11.67	13.33	15.00	16.67	18.33	20.00	21.67	23.33	25.00	26.67	28.33	30.00	31.67	33.33	35.00	36.67	38.33	40.00	41.67	43.33	45.00	46.67	48.33	50.00
Jerk time (s)	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50
Jek distance (m)	31.35	36.35	41.35	46.35	51.35	56.35	61.35	66.35	71.35	76.35	81.35	86.35	91.35	96.35	101.35	106.35	111.35	116.35	121.35	126.35	131.35	136.35	141.35	146.35	151.35	156.35
Accel distance (m)	38.46	55.38	75.38	98.46	124.62	153.85	186.15	221.54	260.00	301.54	346.15	393.85	444.62	498.46	555.38	615.38	678.46	744.62	813.85	886.15	961.54	1040.00	1121.54	1206.15	1293.85	1384.62
Braking distance (m)	41.67	60.00	81.67	106.67	135.00	166.67	201.67	240.00	281.67	326.67	375.00	426.67	481.67	540.00	601.67	666.67	735.00	806.67	881.67	960.00	1041.67	1126.67	1215.00	1306.67	1401.67	1500.00
Minimum curve radius (m)	56.32	81.10	110.38	144.18	182.47	225.27	272.58	324.40	380.71	441.54	506.87	576.70	651.04	729.89	813.24	901.10	993.46	1090.33	1191.70	1297.58	1407.97	1522.86	1642.25	1766.15	1894.56	2027.47
Y	18.63	22.26	25.91	29.56	33.22	36.89	40.55	44.22	47.89	51.56	55.23	58.90	62.57	66.24	69.92	73.59	77.26	80.94	84.61	88.29	91.96	95.64	99.31	102.99	106.66	110.34
Switch clearance (m)	14.91	18.66	22.40	26.12	29.82	33.53	37.22	40.92	44.61	48.30	51.99	55.67	59.36	63.04	66.73	70.41	74.09	77.77	81.45	85.13	88.82	92.50	96.18	99.86	103.53	107.21
Mininum running distance	381.03	428.80	482.93	543.45	610.35	683.65	763.36	849.46	941.97	1040.89	1146.22	1257.95	1376.09	1500.68	1631.59	1768.96	1912.73	2062.92	2219.51	2382.51	2551.92	2727.74	2909.98	3098.62	3293.67	3495.13
Minimum interstation dist (m	383.90	430.81	484.42	544.58	611.25	684.38	763.96	849.97	942.41	1041.26	1146.54	1258.23	1376.34	1500.86	1631.79	1769.14	1912.90	2063.07	2219.65	2382.64	2552.04	2727.85	2910.08	3098.71	3293.76	3495.21
Access time (mins)	2.30	2.58	2.91	3.27	3.67	4.11	4.58	5.10	5.65	6.25	6.88	7.55	8.26	9.01	9.79	10.61	11.48	12.38	13.32	14.30	15.31	16.37	17.46	18.59	19.76	20.97
Max number of stations	51	46	41	36	32	29	26	23	21	19	17	16	14	13	12	11	10	10	9	8	8	7	7	6	6	6
Dist travelled S (m)	166	213	268	328	395	468	548	634	727	826	931	1043	1161	1285	1416	1554	1697	1848	2004	2167	2337	2512	2695	2883	3078	3280
Additional time per stop (s)	33.01	34.62	36.22	37.82	39.42	41.03	42.63	44.23	45.83	47.44	49.04	50.64	52.24	53.85	55.45	57.05	58.65	60.26	61.86	63.46	65.06	66.67	68.27	69.87	71.47	73.08
Time between stations P1 (s)	49.58	52.40	55.33	58.33	61.37	64.44	67.54	70.65	73.78	76.92	80.07	83.22	86.38	89.55	92.72	95.89	99.07	102.25	105.43	108.61	111.80	114.98	118.17	121.36	124.55	127.74
Time between stations P2 (s)	66.15	70.19	74.45	78.83	83.31	87.85	92.45	97.07	101.73	106.40	111.10	115.80	120.52	125.25	129.99	134.73	139.48	144.24	149.00	153.76	158.53	163.30	168.07	172.85	177.62	182.40
Time between stations P3 (s)	82.72	87.58	93.56	99.34	105.26	111.27	117.36	123.50	129.68	135.89	142.13	148.39	154.66	160.95	167.26	173.57	179.90	186.23	192.57	198.91	205.26	211.61	217.97	224.33	230.70	237.07
Time between stations P5 (s)	115.86	123.56	131.79	140.36	149.15	158.11	167.18	176.34	185.57	194.86	204.19	213.55	222.94	232.36	241.80	251.25	260.73	270.21	279.70	289.21	298.72	308.24	317.77	327.31	336.85	346.39
Time between stations P7 (s)	149.00	159.14	170.02	181.37	193.04	204.94	217.00	229.19	241.47	253.83	266.24	278.71	291.22	303.77	316.34	328.94	341.55	354.19	366.84	379.51	392.19	404.88	417.57	430.28	443.00	455.72
In-vehicle time P1 (mins)	41.58	39.07	36.58	34.19	31.94	29.84	27.90	26.11	24.48	22.97	21.59	20.33	19.17	18.10	17.11	16.20	15.35	14.57	13.84	13.16	12.52	11.92	11.36	10.84	10.34	9.87
In-vehicle time P2 (mins)	28.29	26.75	25.23	23.76	22.38	21.07	19.87	18.75	17.72	16.78	15.91	15.11	14.38	13.70	13.08	12.50	11.97	11.48	11.02	10.59	10.20	9.83	9.48	9.16	8.85	8.57
In-vehicle time P3 (mins)	23.58	22.35	21.14	19.96	18.85	17.79	16.81	15.90	15.06	14.28	13.57	12.91	12.30	11.74	11.22	10.74	10.29	9.88	9.49	9.14	8.80	8.49	8.20	7.92	7.67	7.42
In-vehicle time P5 (mins)	19.82	18.83	17.87	16.92	16.02	15.17	14.37	13.62	12.93	12.29	11.69	11.15	10.64	10.17	9.73	9.33	8.95	8.60	8.27	7.97	7.69	7.42	7.17	6.94	6.72	6.51
In-vehicle time P7 (mins)	18.20	17.33	16.46	15.62	14.81	14.05	13.32	12.65	12.02	11.43	10.89	10.39	9.92	9.49	9.09	8.72	8.37	8.05	7.75	7.47	7.21	6.96	6.73	6.51	6.31	6.12
Door to door time P1 (mins)	**57.9**	**55.9**	**54.1**	**52.4**	**51.0**	**49.8**	**48.8**	**48.0**	**47.5**	**47.2**	**47.1**	**47.1**	**47.4**	**47.8**	**48.4**	**49.1**	**50.0**	**51.0**	**52.2**	**53.4**	**54.8**	**56.4**	**58.0**	**59.7**	**61.6**	**63.5**
Door to door time P2 (mins)	**44.6**	**43.6**	**42.7**	**42.0**	**41.4**	**41.0**	**40.7**	**40.6**	**40.7**	**41.0**	**41.4**	**41.9**	**42.6**	**43.4**	**44.4**	**45.4**	**46.6**	**47.9**	**49.4**	**50.9**	**52.5**	**54.3**	**56.1**	**58.0**	**60.1**	**62.2**
Door to door time P3 (mins)	**39.9**	**39.2**	**38.6**	**38.2**	**37.9**	**37.7**	**37.7**	**37.8**	**38.1**	**38.5**	**39.0**	**39.7**	**40.5**	**41.4**	**42.5**	**43.7**	**44.9**	**46.3**	**47.8**	**49.4**	**51.1**	**52.9**	**54.8**	**56.8**	**58.9**	**61.1**
Door to door time P5 (mins)	**36.1**	**35.7**	**35.4**	**35.2**	**35.1**	**35.1**	**35.2**	**35.5**	**35.9**	**36.5**	**37.2**	**37.9**	**38.9**	**39.9**	**41.0**	**42.3**	**43.6**	**45.1**	**46.6**	**48.3**	**50.0**	**51.9**	**53.8**	**55.8**	**57.9**	**60.1**
Door to door time P7 (mins)	**34.5**	**34.2**	**34.0**	**33.9**	**33.8**	**34.0**	**34.2**	**34.5**	**35.0**	**35.6**	**36.3**	**37.2**	**38.1**	**39.2**	**40.4**	**41.6**	**43.0**	**44.5**	**46.1**	**47.8**	**49.5**	**51.4**	**53.3**	**55.4**	**57.5**	**59.8**
Median D2D time (mins)	**39.9**	**39.2**	**38.6**	**38.2**	**37.9**	**37.7**	**37.7**	**37.8**	**38.1**	**38.5**	**39.0**	**39.7**	**40.5**	**41.4**	**42.5**	**43.7**	**44.9**	**46.3**	**47.8**	**49.4**	**51.1**	**52.9**	**54.8**	**56.8**	**58.9**	**61.1**
Median D2D w/out P1 (mins)	**38.0**	**37.5**	**37.0**	**36.7**	**36.5**	**36.4**	**36.5**	**36.7**	**37.0**	**37.5**	**38.1**	**38.8**	**39.7**	**40.7**	**41.8**	**43.0**	**44.3**	**45.7**	**47.2**	**48.8**	**50.6**	**52.4**	**54.3**	**56.3**	**58.4**	**60.6**
Average D2D speed P1 (km/h	20.4	21.1	21.9	22.5	23.2	23.8	24.2	24.6	24.9	25.1	25.1	25.1	24.9	24.7	24.4	24.1	23.6	23.2	22.7	22.1	21.6	21.0	20.4	19.8	19.2	18.6
Average D2D speed P2 (km/h	26.5	27.1	27.7	28.1	28.5	28.8	29.0	29.1	29.0	28.9	28.6	28.2	27.8	27.2	26.6	26.0	25.4	24.7	23.9	23.2	22.5	21.8	21.1	20.4	19.7	19.0
Average D2D speed P3 (km/h	29.6	30.1	30.6	30.9	31.2	31.5	31.4	31.3	31.1	30.7	30.3	29.8	29.2	28.5	27.8	27.1	26.3	25.5	24.7	23.9	23.1	22.3	21.5	20.8	20.1	19.4
Average D2D speed P5 (km/h	32.7	33.1	33.4	33.6	33.7	33.7	33.5	33.3	32.9	32.4	31.8	31.2	30.4	29.6	28.8	28.0	27.1	26.2	25.4	24.5	23.6	22.8	22.0	21.2	20.4	19.7
Average D2D speed P7 (km/h	34.3	34.6	34.8	34.9	34.9	34.8	34.6	34.2	33.7	33.2	32.5	31.8	31.0	30.2	29.3	28.4	27.5	26.6	25.6	24.7	23.9	23.0	22.2	21.3	20.5	19.8
Safety distance (SM) (m)	3.4	3.6	3.8	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0	6.2	6.4	6.6	6.8	7.0	7.2	7.4	7.6	7.8	8.0	8.2	8.4
Vehicle headway (s)	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
Platoon length (m)	154.0	156.0	158.0	160.0	162.0	164.0	166.0	168.0	170.0	172.0	174.0	176.0	178.0	180.0	182.0	184.0	186.0	188.0	190.0	192.0	194.0	196.0	198.0	200.0	202.0	204.0
Minimum headway (s)	23	21	20	20	20	20	20	20	21	21	22	22	23	24	24	25	25	26	27	28	28	29	30	30	31	32
Minimum headway (m)	225	253	284	319	357	398	442	490	541	595	653	714	778	846	917	991	1069	1150	1235	1322	1413	1508	1605	1706	1811	1918
Frequency	160	171	177	181	182	181	179	176	173	169	165	161	157	153	149	145	141	138	134	131	127	124	121	118	115	113
Capacity	127522	136206	141515	144267	145121	144594	143082	140882	138218	135252	132106	128867	125598	122345	119139	116003	112952	109995	107136	104380	101726	99174	96721	94365	92102	89930

Figure 56. Visual illustration of the model

The model calculates average door-to-door time for each pattern (P_x) as a function of the maximum speed on the line (V), and the combination of physical specifications as invariable parameters. Maximum speed is set between 36 km/h (10 m/s) and 216 km/h (60 m/s), ascending in 2 m/s increments. illustrates the average door-to-door times for the same 19.7 km journey using five different patterns: stopping at every station, and stopping at every second, third, fifth and seventh stations ($P_{1,2,3,5,7}$) respectively.

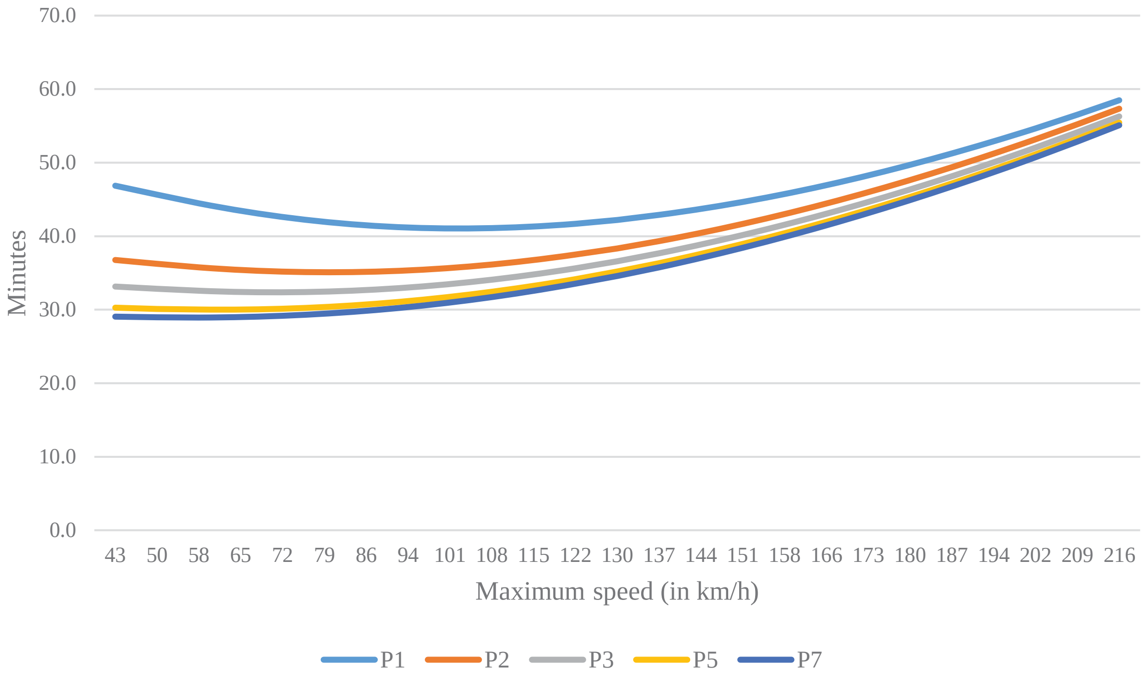


Figure 57. Door-to-door journey times for a 19.7 km journey and $P_{1,2,3,5,7}$

It can be observed that passengers that would travel only using a P_7 service would experience travel times close to the goal of 30 minutes. More specifically, door-to-door travel times for that pattern are below 34 minutes when maximum speed on the line is only 65 km/h, indicating an efficient combination. Nonetheless, as the curves also show, that benefit comes at the cost of sub-optimal door-to-door journey times for all other services. Therefore, it is better to evaluate the model using the median door-to-door journey times (Figure 58). With that, the door-to-door journey time curve follows that of P3, indicating an optimal maximum line speed of 72 km/h, leading to a median door-to-door journey time of 37.7 minutes.

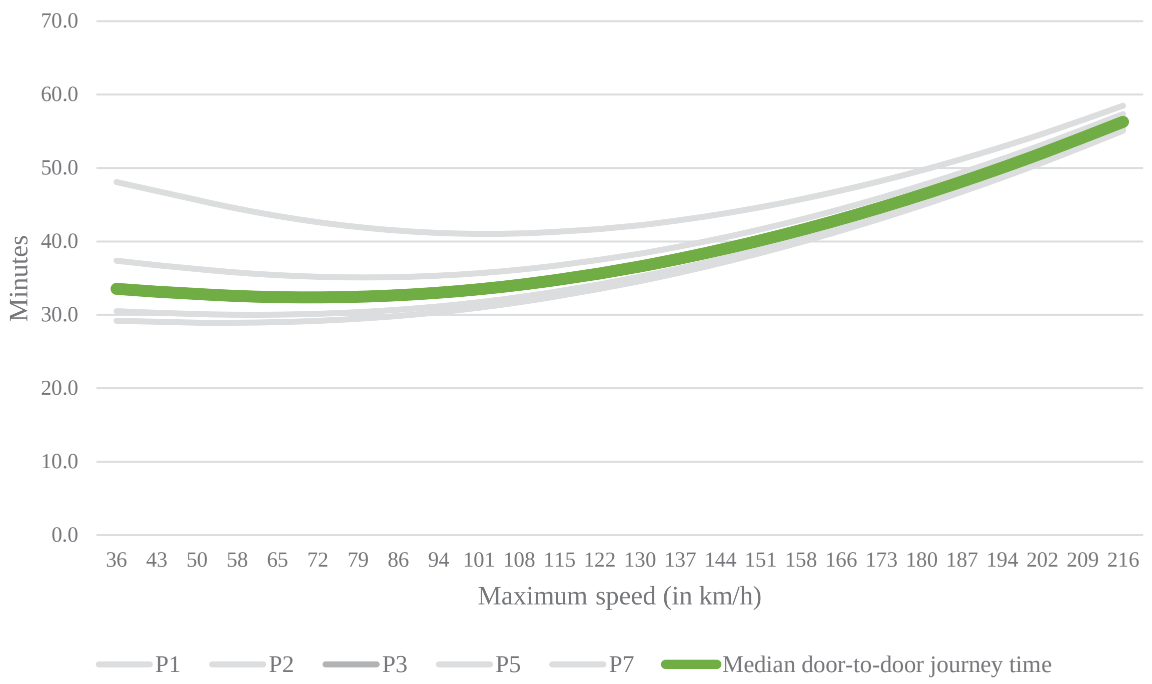


Figure 58. Median door-to-door journey times for a 19.7 km journey and $P_{1,2,3,5,7}$

On the other hand, such calculations may be skewed towards the lower side as most passengers would not opt to travel on a service that stops at every station when they have the choice of greater speeds offered by other patterns. Therefore, the median calculation should exclude P_1, because its use is seen as sparser than other services. Based on the evidence from TfL (2017b) that passengers choose the route with the smallest number of changes, it can be inferred that users will prefer the services that do not require any changes along the same line. For instance, a passenger that wants to travel for 8 stops can choose between various options but will choose travelling on a P2 service for four stops. Alternatives such as P7 + P1 may be considerable, but incur a change, extra platform time (T_p), waiting time (T_w), and dwell time (T_d).

In this operational concept, a change in services along the same line is only required to complete journeys where the number of stations between origin and destination is a prime number. Considering that the operational concept comprises a line of 29 stations, and that users will choose the option with the least number of changes, only 4 pairs of origin-destination do not count with a non-stop option and thus require one change along the time (11, 13, 17, 23).

Of those, alternative combinations using P_2 to complete the journey might be quicker than using P_1.

It is premature to affirm which percentage of users will use each service because the exact demand for each pair of origin-destination is unknown. In addition, the results found by TfL (2017b) show that some users might take sub-optimal routes. Nonetheless, the calculations above illustrate that P_1 services are likely to be the least used as most of the origin-destination pairs can be reached without any change along the line.

When P_1 is excluded, median door-to-door journey time is 36.4 minutes, when the maximum speed on the line is 72 km/h (Figure 59). That represents a 25% reduction in travel times compared to the current capability model. Moreover, it must be noted that the usage of certain patterns may increase over distances. For example, passengers travelling around 20 km will most likely use P_5 and P_7 services, reaching average door-to-door travel times of 35.1 and 34 minutes, respectively.

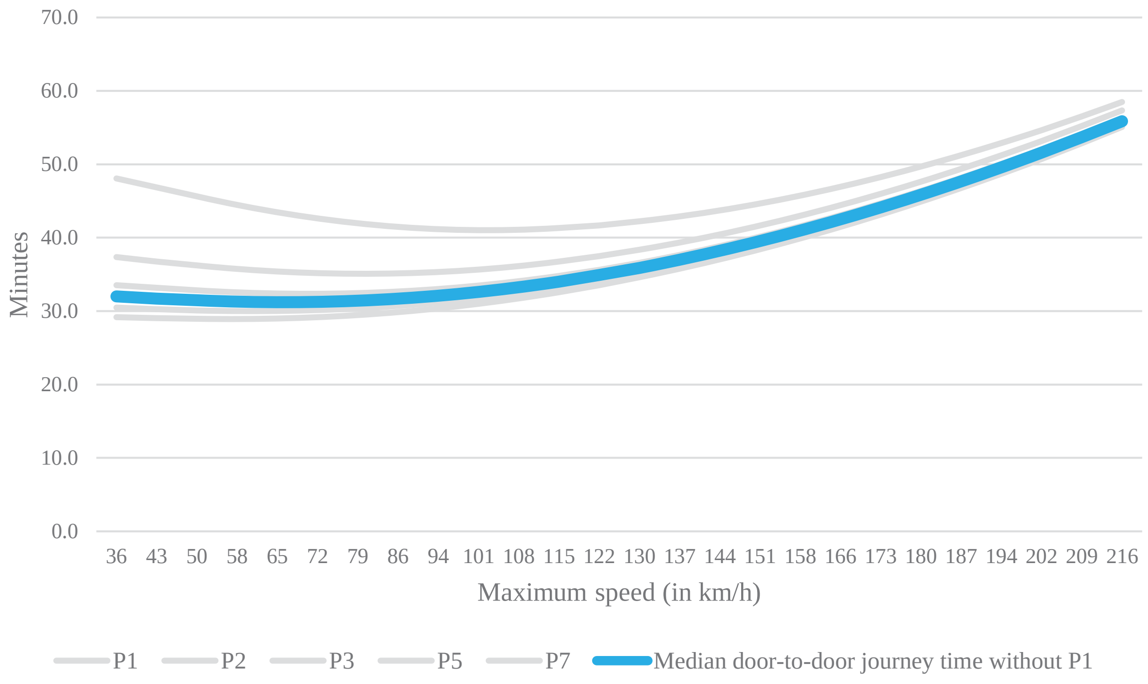


Figure 59. Median door-to-door journey times for a 19.7 km journey and $P_{2,3,5,7}$

Finally, if station access is improved, then the operational concept has the potential to achieve even shorter door-to-door travel times when compared to drivers. Assuming that entry and exit are reduced to 2 minutes each, and that interchange time is improved to also 2 minutes, then the median door-to-door travel time (excluding P_1) becomes 32.3 minutes (Figure 60). For those travelling mostly on P_5 and P_7 services, door-to-door travel time reaches the Marchetti constant of 30 minutes.

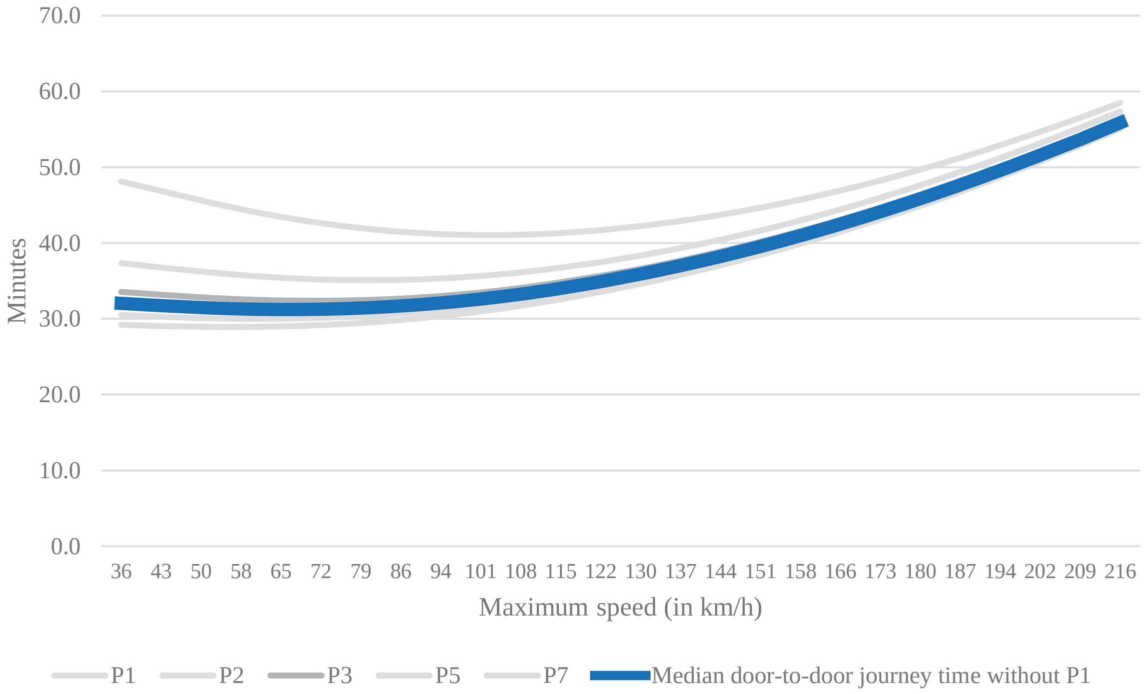


Figure 60. Median door-to-door journey times for a 19.7 km journey and $P_{2,3,5,7}$ with improved station access

In terms of capacity, the initial model shows a theoretical capacity of up to approximately 111,000 passengers per hour per direction. The peak capacity is found when the maximum line speed is 86 km/h, although it is only 2% smaller for the speed where journey times are lowest. Travelling at that speed, the minimum distance between vehicles (d_{min}) in a platoon is 24.4 metres under an ambitious assumption of 0.8 seconds between vehicles. This means platoons are 364 metres long when there are two vehicles of each P_x in each platoon. Consequently, the minimum headway (H_{min}) is approximately 30 seconds, which makes waiting time (T_w)

practically negligible. The technical aspects of its feasibility will be discussed in the next sections.

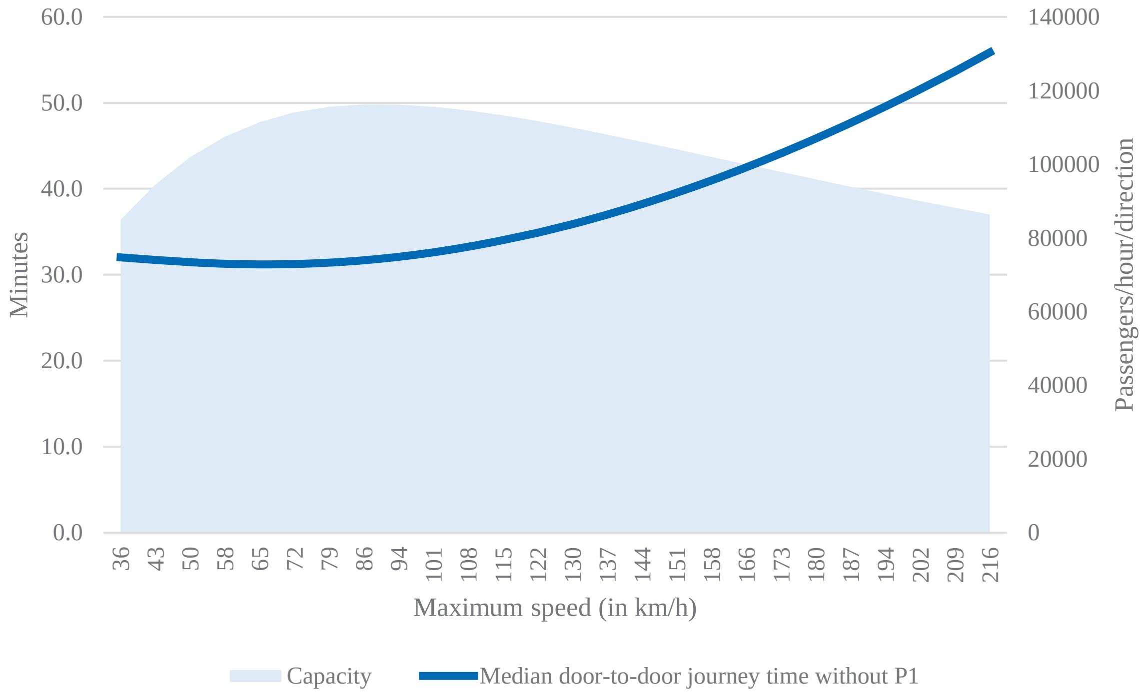


Figure 61. Median door-to-door journey times and theoretical capacity for a 19.7 km journey and P2,3,5,7 with improved station access

When vehicles are 12 m long and 2.6 m wide, each platform section (L_{px}) needs to have 9.3 metres of free space on each end, meaning that they are 43 metres long. Since the first and last platform sections only need one free space, total platform length (L_p) is then 240 metres.

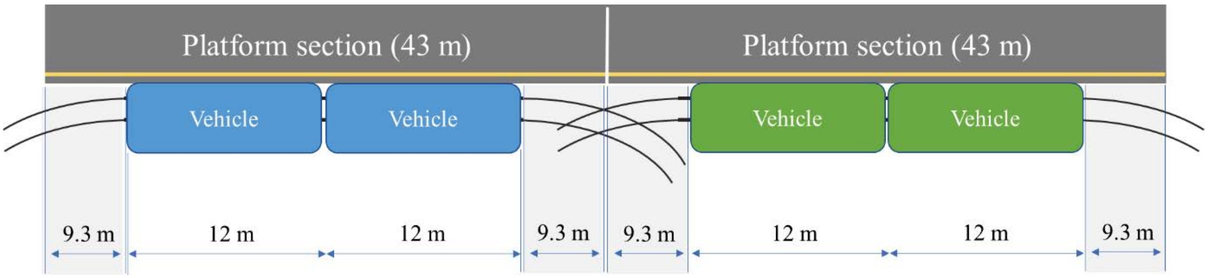


Figure 62. Dimensions of vehicles and platform sections

Under the previous assumptions for rolling stock dimensions (80% of the space occupied by passengers at a loading factor of 4 passengers/m²), each vehicle has the capacity (C_v) to carry

80 passengers. Assuming that (1) vehicles have one double channel door per every 5 metres of body length, (2) doors open and close in 5 seconds combined, and (3) each passenger takes 0.5 seconds to board or alight per channel, then the minimum dwell time (T_d) is 25 seconds.

5.4.5. Simulation results

Using the variables from the model, the operational concept was then run on a MatLab simulator to evaluate its dynamic robustness and stability. The focus of the simulations was less on the outcomes in terms of speed and capacity, and more on the dynamic interactions between vehicles following distinct patterns in a complex setting. The simulator starts from the main assumption that vehicles can autonomously communicate with each other and adjust speed curves according to trajectory projections, in accordance to Grade of Automation 4 (UITP, 2016). In addition, the simulator does not account for any necessary speed reduction between stations, in that the speed curve consists solely of acceleration, jerk, and braking (set at 1.3 m/s^2, 1 m/s^3, and 1.2 m/s^2 respectively). When the maximum speed on the line is 72 km/h, the minimum running distance between stations (D_{min}) is 685 metres. It logically follows that, for a trip of 19.7 km, there will be 29 stations between the origin and the destination. The simulator assumes the journey as the line length, simulated as a circle line for the evaluation of the interaction between all vehicles in continuous but asynchronous laps. The parameters used in the simulation are those from the model and listed on Table 4.

Table 4. Operational parameters used in the simulations

Parameter	**Value**
Maximum speed (V)	72 km/h (20 m/s)
Acceleration (α)	1.3 m/s^2
Jerk (j)	1 m/s^3
Braking (β)	1.2 m/s^2
Distance between stations (D)	685 m
Transport unit length	48.4 m
Dwell time (T_d)	25 s
Total length of line	19,700 m
Headway between vehicles	1 s
Headway between platoons (H_{min})	30 s

5.4.5.1. Normal operations

The line simulated is circular, meaning that vehicles do not change trajectory when reaching the 29th station, but rather start over at 0 m. The simulator treats vehicles as unit entities, meaning that if there are two vehicles of each P_x in a platoon, they will be counted as one transport unit. The reason for that is the simplification of the train lines in the output graphs and allows for different combinations of number of P_x in each platoon. Figure 63 illustrates the outputs of the simulator (train graphs of distance over time) by running one transport unit of each Px, for only one lap on the circular line, under normal and undisturbed operations.

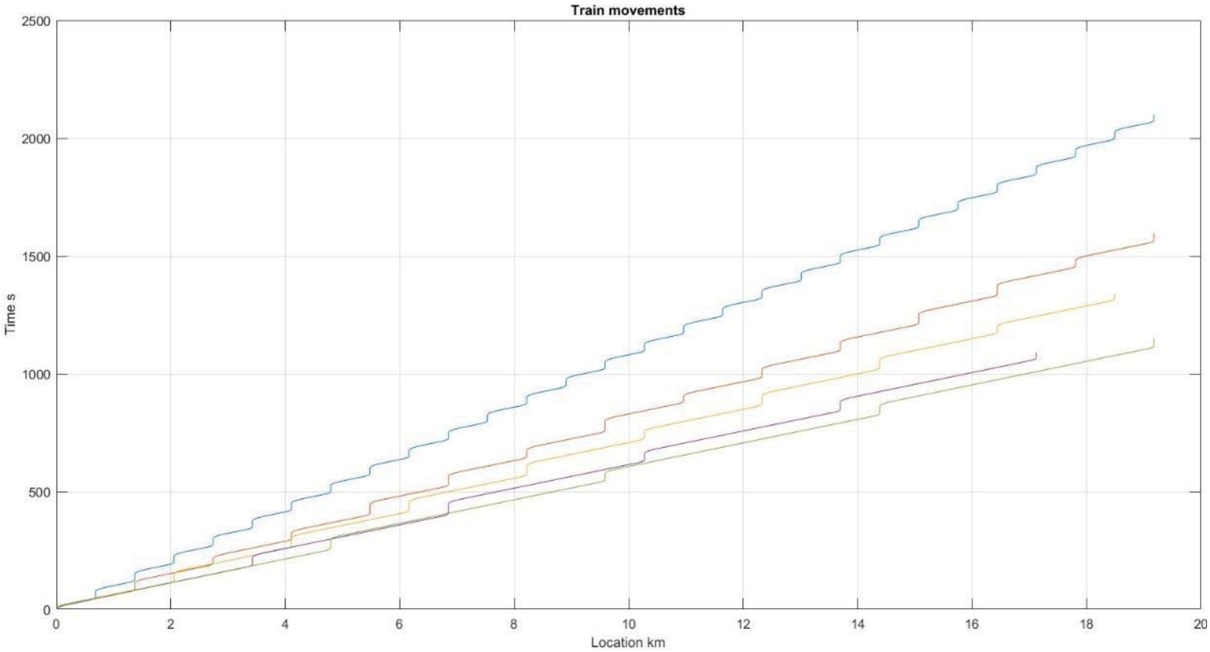


Figure 63. Illustration of simulator with one transport unit for each P_x

The second base assumption is that there will be one platoon per station, meaning that every station will be served by all patterns within the minimum headway. Thus, when there are 29 platoons, there will be 145 transport units when five patterns are in use ($P_{1,2,3,5,7}$). Headway is 30 seconds and the safety distance between vehicles within a platoon was set at 1 second between vehicles for simplification purposes. Dwell time is initially fixed at 25 seconds following the operational concept described in the model.

Firstly, the 145 transport units are simulated over only one lap. Results show that the system behaves as predicted, in that there are no conflicts between train paths nor vehicles must wait at the platform to return to the main line (Figure 64).

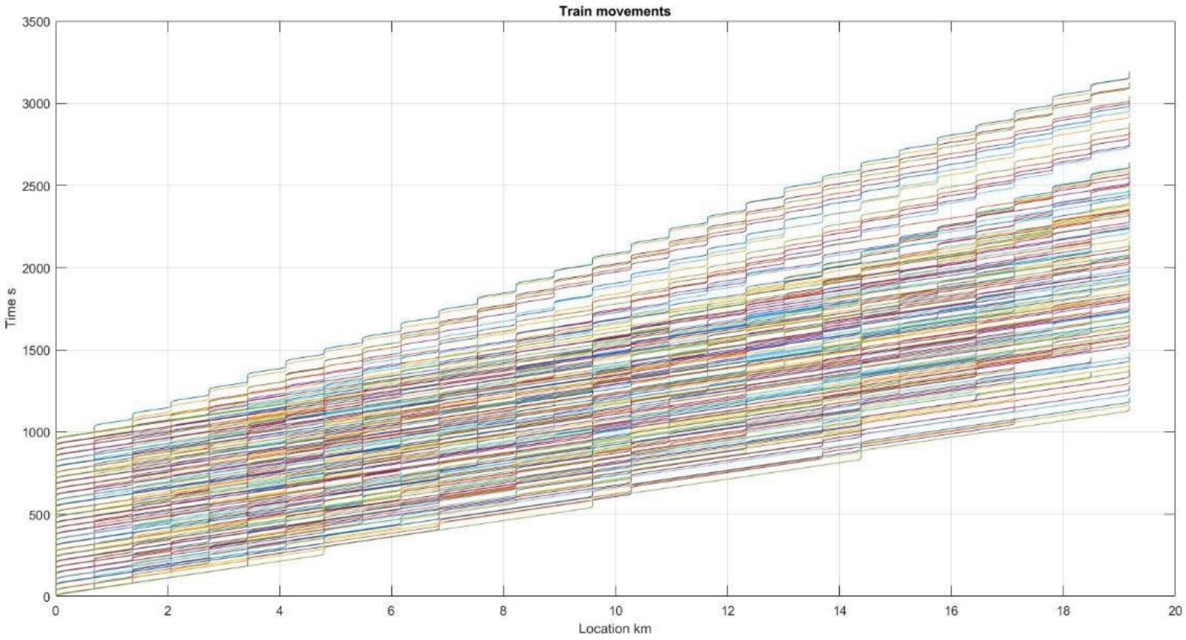


Figure 64. Simulation of 145 transport units performing one lap on a circular line with 29 stations

One observation is that once the system is running, platoons assemble and disassemble dynamically, as vehicles prioritise following their pattern over platoon formation. Some transport units may travel part of the journey separately as the simulator does not include commands to change speeds for platoon assembly. Figure 65 provides a closer look in the train lines for clarity, also illustrating the stability under complex operations.

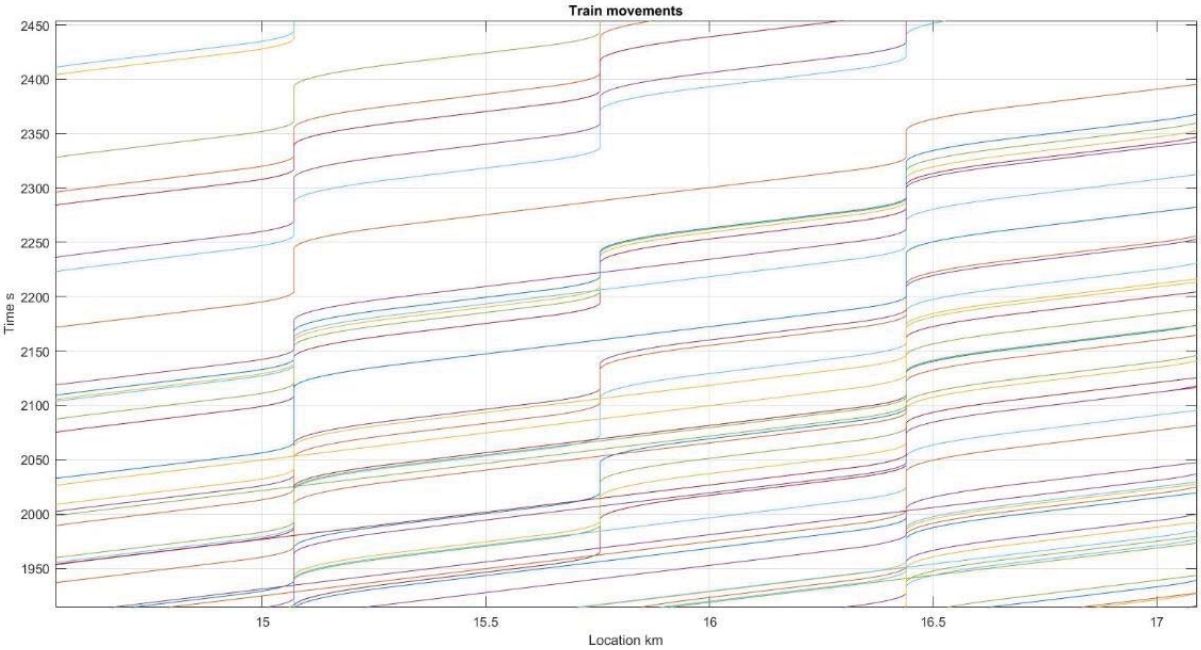


Figure 65. Close up of train lines of 145 transport units over one lap on a circular line with 29 stations

When the simulation includes more laps, the graph becomes difficult to read, highlighting some limitations of the simulator (Figure 66). However, the main point is that the system tends to maintains its stability most of the time, in that dwell times are not affected and no two transport units occupy the same track at the same time. The headway between vehicles, even though not clearly readable on the graph, is kept at or above 1 second.

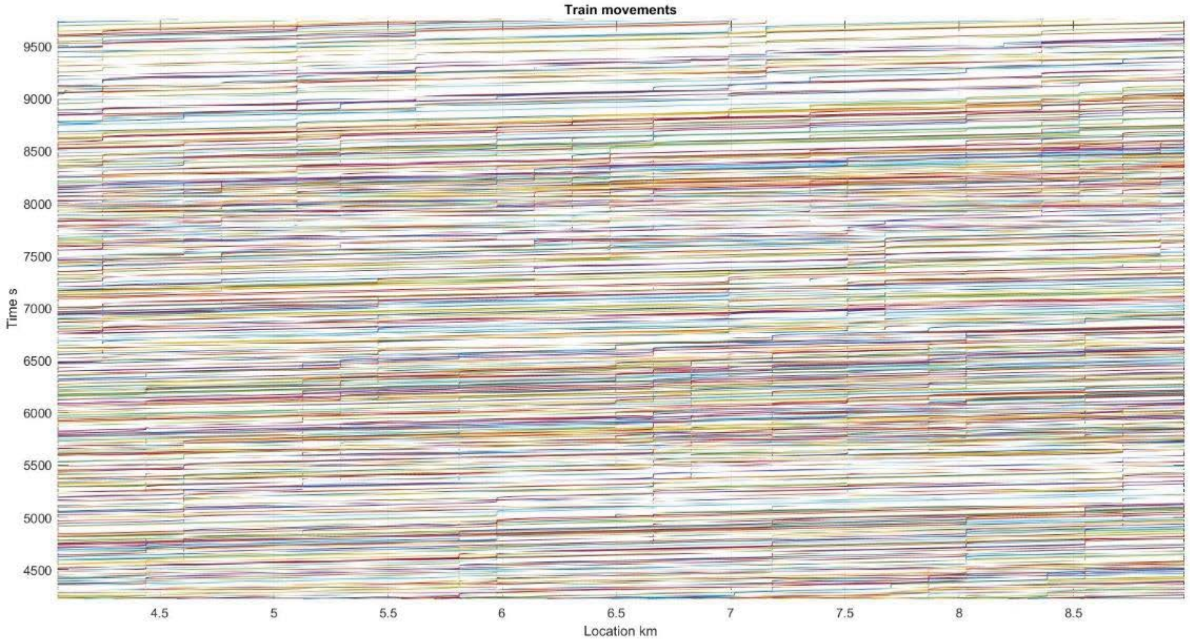


Figure 66. Simulation of 145 transport units over 7 laps on a circular line with 29 stations

Figure 67 shows the difficulties in simulation outputs to cope with the individual train paths. Nonetheless, the system maintains some robustness in headway and dwell times, in that vehicles of different patterns eventually end up reassembling in a platoon. Most of the times, vehicles travel at different distances from each other, which can complicate signalling and communication. More importantly, one area has been highlighted to show an instance when the transport unit leaves the platform to accelerate and join a transport unit which did not stop. Although of very close proximity, the distance between vehicles during this moment is of one second, and would thus require a robust train control system for the manoeuvre.

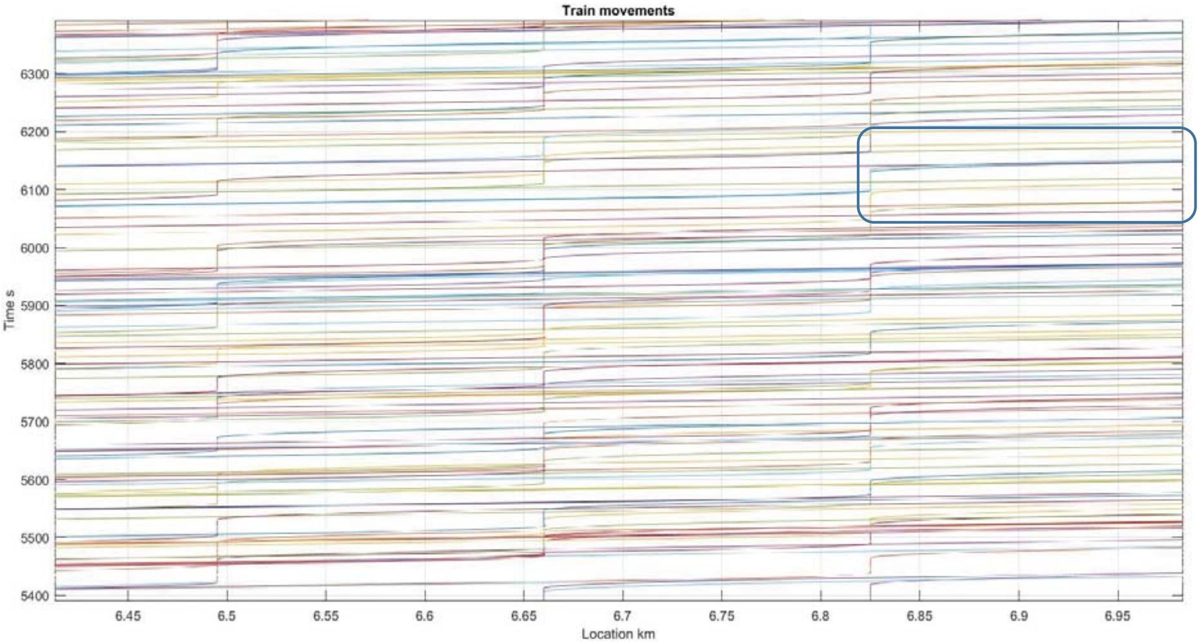


Figure 67. Zoom view of simulation of 145 transport units over 7 laps on a circular line with 29 stations

5.4.5.2. Disrupted operations

The second round of simulation looked at increasing complexity with a more realistic setting. Rather than fixed at 25 seconds, dwell times were set randomly between 25 and 60 seconds. More specifically, each time a transport unit stops at a station, it will pick a dwell time at random, simulating real-life situations that rely on passenger behaviour. Although automated systems may experience smaller deviations, the purpose of the simulation is to push variance to a greater extent to investigate its robustness. As a result, figures Figure 68 and Figure 69 show that the system remains stable under severe variation in dwell times, and that vehicles can recover their path once they attend the platform. Moreover, platoon formation seems to be maintained throughout operations. However, at certain points, the headway between vehicles drops below 1 second, indicating instances of potential conflict.

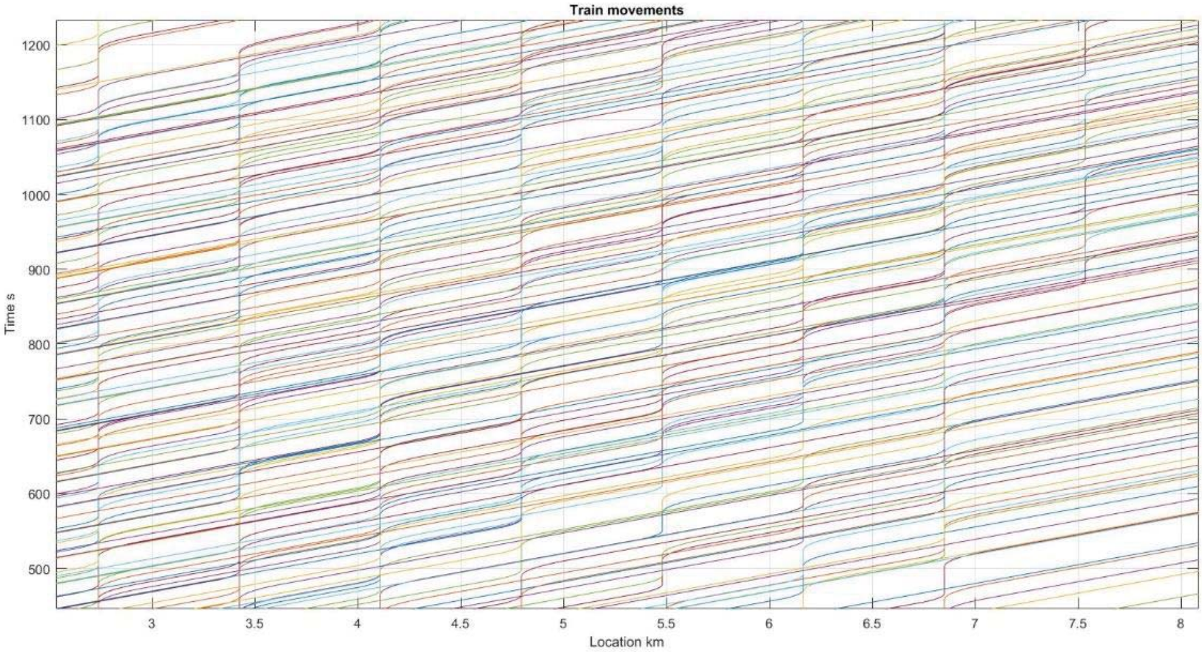


Figure 68. Simulation of 145 transport units over one lap on a line with 29 stations, with random dwell times between 25 and 60 seconds

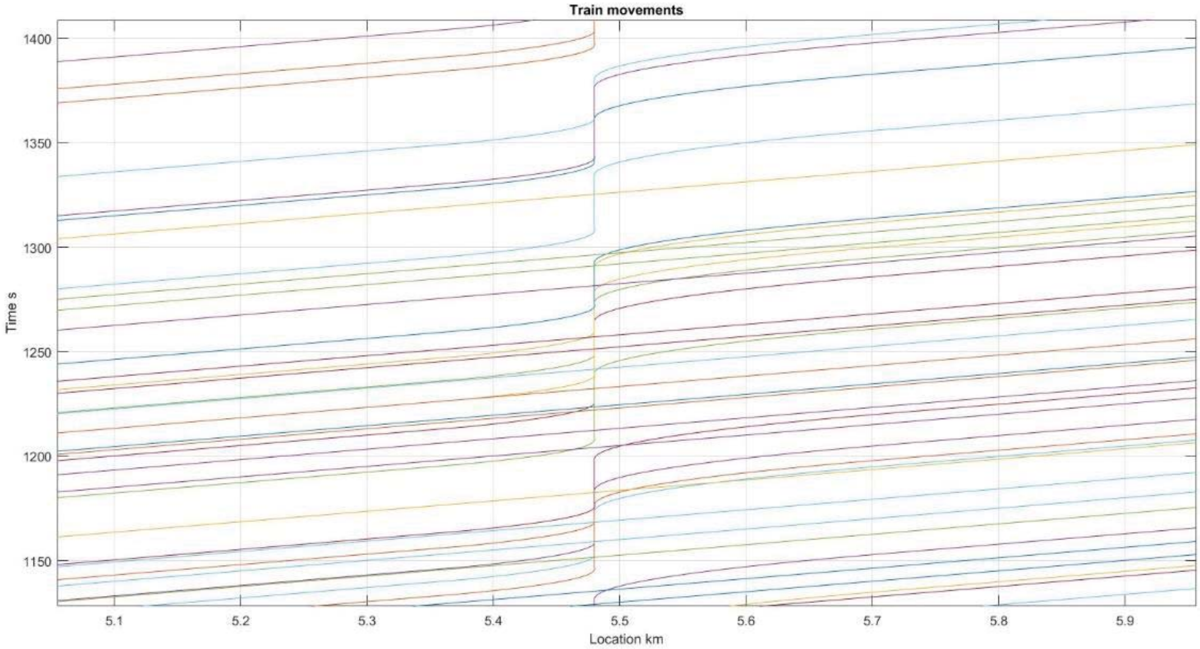


Figure 69. Zoom view of simulation of 145 transport units over one lap on a line with 29 stations, with random dwell times between 25 and 60 seconds

5.5. Definition of functional requirements

From the description of a working operational concept, it is possible to derive functional requirements. As mentioned before, this stage is not yet concerned with physical elements and specifications, in that only functional aspects are to be analysed. They will in turn lead to the selection of technologies,

Given the high level of abstraction of the operational concept dealt with in this thesis, it is unrealistic to develop a complete set of requirements. In fact, as it will be discussed in the next chapter, the heuristic nature of the thesis means that the aim is not on finalising the operational concept, but rather to develop an operational concept that kickstarts further investigations for radical change.

Therefore, this section will focus on the functional requirements that are seen as critical to the operational concept. Their criticality is selected based on two criteria: (1) essentiality for the operational concept to function as described, or (2) difference to the operations of current systems. A more comprehensive list of the functional requirements that arise from the operational concept can be found in Appendix B.

5.5.1. Pattern operation

The operation of different services attending stations according to various patterns is not necessarily new as explained before. Yet, the particular arrangement of patterns and number of stations creates some functional requirements under specific circumstances.

The main requirement already stated is that the number of stations cannot be a multiple of any of the patterns (P_x). When operating on a fully circular line, there is no need for adaptation and all patterns attend all stations as predicted, taking a number of laps equal to its own P_x to do so.

When operating on straight lines, a few adjustments must be made for P_2 and P_3 because otherwise these services will only attend part of the station in repeating cycles. Therefore, P_2 and P_3 services must be deployed in alternating initial stations in the beginning of daily operations. First service shall start from station 1, second service from station 2, then and third service from station 3 in the case of P_3.

5.5.2. Changing tracks

The most critical functional requirement concerns vehicles changing tracks, from the main line to a station and vice-versa. With the calculations from the model (equation 20), the spacing between vehicles within a platoon (d_{min}) is 0.8 seconds. That means that if vehicles change directions passively, track equipment must be able to move tracks out, lock, move tracks back, and lock again within this time. It logically follows that under that circumstances, a switch must be able to change directions and lock its position in no more than 0.3 seconds, assuming 100 ms for processing. The technical and technological decisions for this aspect will be further discussed in the next section.

5.5.3. Virtual coupling and reaction time

The notion of platooning inherently encompasses virtual coupling, where vehicles operate as in train formation but not physically connected. The minimum distance between vehicles depends on processing time to allow for vehicle-to-vehicle data transfer, and the definition of a safety margin to allow for inaccuracies in reading and transfer. To achieve the results of the model, processing time is required to be 100 milliseconds. In addition, safety margin must be no less than 20% of the vehicle length to allow for safe operations in the event of errors in data transmission.

5.6. Selection of candidate technologies

The operational concept provides specific requirements for its fulfilment. So far specific design requirements have not been addressed because the objective is to devise an optimal solution regardless of present technical and technological feasibility. This section addresses the technological gaps by measuring the necessary performance of the system against the performance of off-the-shelf counterparts. Since the operational concept is, as the name suggests, still conceptual, the measures of performance (MoP) focus mostly on the overarching functional aspects, which are listed and compared on Table 5.

Table 5. Measures of performance of overarching functional aspects

Function	**Requirement**	**Capability**	**Source**
Switch actuation time	<= 0.3 s	0.75 s	(Bemment, 2017)
Maximum speed	>= 72 km/h	90 km/h	(Siemens, 2012)
Acceleration	1.3 m/s²	1.33 m/s²	(Siemens, 2012)
Deceleration	1.2 m/s²	1.2 m/s²	(Siemens, 2012)
Jerk	1 m/s³	0.98 m/s³	(Powell & Palacín, 2015)
Open vehicle doors	<= 2.5 s	2.5 s	(FERSIL, 2008)
Close vehicle doors	<= 2.5 s	2.5 s	(FERSIL, 2008)
Grade of automation	4	4	(UITP, 2016)
Data transmission for platooning	<= 100 ms	40 ms	(Nardini, et al., 2018)
Error margin	20%	20%	(Takagi, 2014)
Platform length	240 metres	220 metres	(Mizutani, 2016)
Cant	160 mm	180 mm	(RSSB, 2007)
Cant deficiency	100 mm	110 mm	(RSSB, 2007)
Entry times	2 minutes	2.75 minutes	(Transport for London, 2013)
Interchange times	2 minutes	4.6 minutes	(Transport for London, 2013)

From the table, it can be seen that most of the capabilities for line design and rolling stock performance are already met by industry or research. This means that they populate region A in Figure 27 and do not pose risks to system development. Therefore, the critical technological gaps involve mainly the platooning aspect of the operational concept, in the form of signalling and track switching.

The case for platooning capabilities requires a more thorough investigation that involves the backcasting method more closely. The next steps aim at drawing a technological bridge between current capabilities and the desired state in the future. Its fulfilment, however, depends on the projects that emerge from this research. Headways in rail involve distinct assumptions than on roads. Firstly, rubber-tarmac adhesion is considerably greater than steel wheel-steel track. Secondly, there is a limit in deceleration and jerk that standing passengers can physically endure. The distance between vehicles in a platoon depends on their ability to decelerate safely in the event of an emergency.

5.6.1. Signalling

In that, the operational concept relies on the assumption that platoons can operate using relative moving block signalling as opposed to ‘brick wall’ moving block. While the latter calculates the minimum headway based on the preceding vehicle being stationary, the former adopts a more dynamic and perhaps realistic assumption that the preceding vehicle will move a certain distance even in the event of an emergency (Nakamura, 1998; Takagi, 2014). While moving block signalling is commonly used in contemporary metro systems, the adoption of relative speeds in headway calculation is still theoretical in the railways, thus inhabiting the space between B and C in Figure 27.

Relative moving block signalling is an intrinsic requirement for the operational concept to perform as necessary. Without it, the minimum headway between vehicles in a platoon would jump from 0,8 seconds to 12.8 seconds, and consequently platoons would be over 2,000m long, making the whole operation inviable. Notwithstanding this need, there are technological decisions to be made regarding turnouts that are directly related to the minimum headway between vehicles.

5.6.2. Switch actuation

When the headway between vehicles (D_{min}) is 0.8 seconds, the use of traditional passive turnouts becomes an issue. Firstly, the switch must be able to move out and back in a time much shorter than the headway between vehicles. In British practice at Technology Readiness Level (TRL) 9, a time of 8 seconds is used for safe operations (Bemment, et al., 2017). The current capability achieved in research is 0.75 seconds, which comes from the REPOINT project that stands at TRL 6 (Bemment, 2017). Secondly, considering that the switch clearance distance (L_c) is 33.5 metres when maximum speed is 72 km/h, it logically follows that just the time for the vehicle to clear the switch time would surpass 0.8 seconds regardless of the actuation time.

A potential solution to such considerable technological gap is that vehicles actively switch between tracks, similarly to road transport. However, this would require a complete change in the system, from normal wheel-rail interface to magnetic levitation. On the other hand, Mattos et al. (2016) have successfully developed a prototype of a superconducting maglev metro system in which vehicles can actively change direction following the polarisation of the magnets on the tracks. The research is currently also on TRL 5, although the track switching has not yet been fully tested. Therefore, a TRL 2 will be adopted in this specific circumstance. Assuming that vehicles can switch between tracks actively, the headway between vehicles in a platoon can be further reduced with a finer calibration of processing time and safety margins. The next section will analyse the impact of the range of D_{min} on door-to-door travel times against other parameters.

A more realistic approach however, in terms of Technology Readiness Levels (TRL), is to adjust d_{min} to an amount within which the track can move the switch blade, lock, vehicle clears L_c, actuation back to neutral position, and lock. Even with this adjustment, the expected headways are still short for railway operations. Nonetheless, platooning has been successfully demonstrated in automotive environments at a TRL 6. The fastest actuation possible is currently at 0.75 seconds (Bemment, et al., 2017), adding 1.5 seconds to the headway between vehicles (d_{min}). To that 200 ms of actuator processing time are added for control and error margins altogether (100 ms in each movement), totalling an extra 1.7 seconds to the original d_{min}. Therefore, the equation for the headway between vehicles adds the extra switch time to equation (20). D_r, τ , L_v, and e are assumed to have the same values of 0.2 m, 100 ms, 12 m, and 2.4 m respectively. In metres, d_{min} becomes:

$$(33)\ d_{min} = D_r + V * \tau + L_v + e + 1.7V$$

The change in d_{min} does not affect door-to-door travel times, but expectedly reduces the theoretical capacity of the system. Figure 70 shows the theoretical capacity of the model according to the maximum line speed (V). When the maximum speed on the line (V) is 20 m/s (72 km/h), the minimum headway between vehicles (d_{min}) is approximately 50 metres, equivalent to 2.5 seconds (which is 0.8 seconds from the original calculation added by 1.5 seconds of extra switching time and 0.2 seconds for control and error). The resulting theoretical capacity with this arrangement has a maximum value at 76,221 passengers per hour per direction.

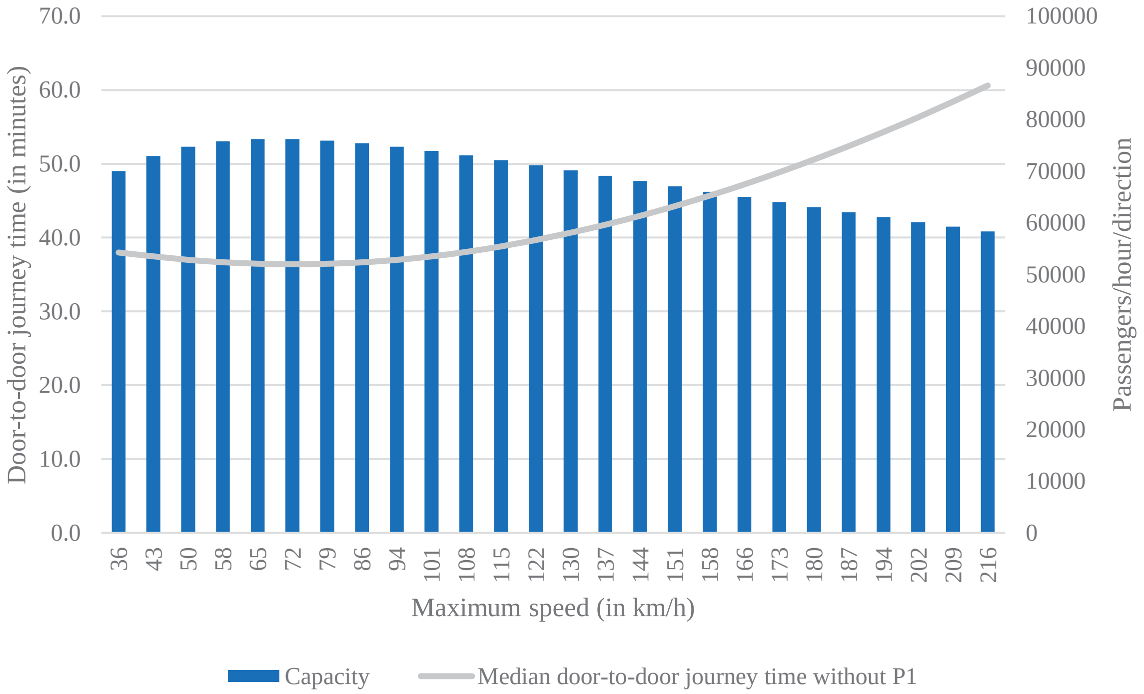


Figure 70. Theoretical capacity and median door-to-door travel times for $P_{2,3,5,7}$ for a 19.7km journey assuming passive track switching

Each technology represents a letter in Figure 27. The more conservative approach using passive track switching represents B-type technologies, while active switching is a C-type technology. These will be taken into account in the next section.

5.7. Selection of optimised solution

This section draws from the technologies screened in the previous section to analyse their benefits against costs and risks. Each of them will be assessed using sensitivity analysis over a range of inputs to measure the impact on the outputs of median door-to-door travel time and theoretical capacity.

For critical parameters, a simplified Pugh decision matrix was used to calculate the benefits in face of risks, uncertainty, and other parameters. The process of optimised solution selection is open-ended in the sense that the level of detail in the calculations can go as far as feasible. On the other hand, populating parameters and weightings for conceptual systems has proven difficult especially concerning technologies which are not in the market yet. The Pugh decision matrix addresses that with relative weightings that can be reviewed and iteratively refined.

Maximum line speed (V) has been used as the main input throughout the development of the operational concept so will not be addressed. For the calculations of sensitivity, entry and interchange times (T_e and T_i) were reset to their initial values of 2.75 and 4.6 minutes respectively. This will allow for a separate analysis of individual impacts on the overall outputs of the system. In addition, sensitivity analysis will be calculated based on the original operational concept which does not take passive track switching into account. For all calculations, the output considered for door-to-door travel times is the median value excluding P_1.

5.7.1. Non-critical parameters

Technologies classified as A-type (according to Figure 27) are those off-the-shelf that already achieve the required performance, already in use, and thus do not pose any risk to system deployment. Nonetheless, sensitivity analysis was also performed in their inputs for two reasons: (1) discovery of ‘over-engineering; (2) discovery of alternative off-the-shelf technologies that can provide benefits with no increase in risks. Table 6 summarises the findings. Individual graphs and more detailed analysis of each parameter can be found in Appendix C. The impacts on travel time and capacity illustrate the percentage of change in the output (travel time / capacity) for each percent of change in the input (row).

Table 6. Sensitivity analysis of non-critical parameters

Function	**Initial input value**	**Final input value**	**Impact on travel time**	**Impact on capacity**
Acceleration	1.3 m/s²	0.9 m/s²	0.098	--
Braking	1.2 m/s²	0.8 m/s²	0.11	0.41
Jerk	1 m/s³	0.6 m/s³	0.036	--
Opening/closing doors	2.5 seconds	2 seconds	0.018	--
Number of doors per vehicle	2	3	0.047	--
Data transmission	100 ms	40 ms	--	0.039
Position margin error	20%	10%	--	0.047
Cant	160 mm	180 mm	0.006	0.042
Cant deficiency	100 mm	110 mm	0.005	0.027
Track gauge	1435 mm	1067 mm	0.013	0.078

From the results observed, it can be highlighted that:

- Entry and exit times (T_e) were found to be the overall component with the greatest potential impact on door-to-door travel times. As opposed to the other components which indirectly influence travel times and/or capacity, every extra minute of T_e relates to an extra minute on door-to-door journey time (T_t). When reduced to 2 minutes each, median door-to-door travel time for $P_{2,3,5,7}$ is 32.3 minutes. The explanation for excluding P_1 from calculations is found on section 5.4.4. Nonetheless, entry and interchange times (T_e and T_i) are dependent on station design, which involves an entirely and dedicated new process of system engineering. It is difficult to perform measures of performance appropriately as various technologies are involved and little data is available.

- Parameters such as acceleration and braking should be kept at the highest levels that capability and passenger comfort and safety can afford. The impact of lower inputs on the system are significant. The system has shown to be moderately sensitive to jerk, in that small variations in input will not significantly affect the outputs of the system in terms of travel times.

- Indirect inputs such as door times, error margin, and data transmission offer marginal gains when pushed beyond the current capabilities. Moreover, they directly impact passenger safety and system reliability. Therefore, they should not be changed from the original values assigned in the operational concept.

- Track infrastructure elements such as cant, cant deficiency, and gauge have small influence on median door-to-door travel times and theoretical capacity. On the other hand, the final values considered are also on TRL 9, thus offering no risk for the marginal gains. Moreover, a simple change in track gauge from the standard 1,435 mm to a Cape gauge (1,067 mm) used in Japanese systems has shown to increase capacity. When track gauge was changed in the model, theoretical capacity increased by 2%. When the model was changed to combine gauge with cant and cant deficiency values of 180 mm and 110 mm respectively, there is an increase of 2.7% in theoretical capacity.

- The number of doors per vehicle and the number of channels per door have shown to have a considerable impact on travel times as they reduce dwell times. An increase to 3 doors per vehicle (with 2 channels per door), or 3 channels per door (with 2 doors per vehicle) can reduce median travel times by 2.34%. As there is no technological gap for this case, the marginal gains are achieved with no increase in risk or uncertainty. This particular parameter was not included in Appendix C because the number of doors or channels needs to be a natural value.

5.7.2. Critical parameters

Firstly, one technology has shown to be crucial for the operational concept to achieve its objectives. While moving block technologies are already in place in current systems, relative moving block is still in experimental stages, currently at a TRL of 4. Nonetheless, relative moving block signalling must be an inherent component of the system so it can perform accordingly. Therefore, this technological gap is not addressed in this section but in the next section with an overall discussion of results.

Secondly, two technologies were identified in the previous section to enable the platooning component. Each one enables the system to perform differently, offering distinct risks. The first is active switching in which a superconducting maglev vehicle can change directions through changes in polarisation on track. This option offers a higher capacity at a lower TRL, whereas normal track-based switching with blades and rapid actuation offers lower capacity at a higher TRL but lower capacity.

The reason for using a Pugh decision matrix to evaluate critical technologies is that a simple benefit to risk analyses would bias the process against more innovative ideas. For the two possible technologies for changing tracks, four parameters were used in the Pugh decision matrix: theoretical capacity, TRL, maintenance costs, and building costs. For the first two parameters, there is objective reasoning based on set values. The latter two are conducted in a more subjective manner the absence of precise data. The set is limited to a small number of parameters for the purpose of illustrating the framework, and the lack of manpower needed to develop a set of accurate values.

The Pugh decision matrix begins with the definition of weightings of each parameter on the overall benefit of a technology. Each parameter is compared to the others on the list, using a three-level scale: if the parameter is seen as more important, then it is assigned a value of 1. If both are seen as equally important, then both are assigned 0.5. Finally, for the lower importance of the comparison, zero is assigned. The reading of each valuation is done by comparing the parameter on the horizontal axis to the parameters on each column.

Table 7 illustrates the process for this case study. As defined in the scope of the thesis, this research focuses on journey time and capacity as the main objectives of the case study. Therefore, capacity was considered as the most important parameter, followed by TRL, then maintenance costs. This does not illustrate the reality of decision making processes in the railways, however. Prioritising capacity at this stage means that technologies will be chosen with that in mind, but subsequent iterations where costs, safety, and reliability are taken into account are needed. Their weighting is assigned dividing their added importance by the total. Travel time was not included in the weighting as both technologies offer the same performance.

Table 7. Weighting calculations for technology parameters

	Capacity	TRL	Maintenance	Building costs	Total	Weighting
Capacity		1	1	1	3	0.5
TRL	0		0.5	1	1.5	0.25
Maintenance	0	0.5		0.5	1	0.166667
Building costs	0	0	0.5		0.5	0.083333

Subsequently, each technology is compared to the baseline technology in their impacts. The baseline is a generic metro system and the value for capacity used in the comparison is 80,000 passengers per hour per direction as found in the literature. A score of 5 is assigned to the baseline meaning that it is the average system. for the new technologies, a score higher than 5 means that it offers an improvement, whereas lower scores mean reduction in feasibility.

For instance, active switching scores higher than the baseline on maintenance because most parts are in-vehicle rather than laid on tracks. On the other hand, passive switching not only uses similar track infrastructure but also requires more effort with off-line tracks to stations. For that reason, both concepts score lower than baseline on building costs as they theoretically require more infrastructure with the different track layouts. Capacity is calculated as a proportion of the theoretical capacity over the baseline capacity, and TRL is calculated in a similar manner by adjusting the levels to a base of 5. E.g. TRL_a is 2*5/9.

Table 8. Pugh decision matrix of two technologies for track switching

	Baseline	Active switching	Passive switching
Capacity	5	6.94	4.76
TRL	5	1.11	3.33
Maintenance	5	7	4
Building costs	5	4	3
TOTAL	**5**	**5.25**	**4.13**

Results indicate that although more conceptual, active switching carries a higher weighted benefit when compared to faster passive switching. The higher weighting given to capacity has shown influence in the overall results. However, the faster passive switching also scored lower in other parameters. In terms of capacity, faster passive switching does not bring significant benefit as the risks are still high for the technology, and the theoretical capacity is actually lower than in the baseline system. Since both technologies carry lower TRLs than the baseline and thus higher uncertainty, it would be advisable to focus on that which provides the greater overall benefits. The feasibility of the system becomes a time function of achieving the technical and technological requirements of the operational concept.

Therefore, besides the pattern strategy, the operational concept to be adopted consists of autonomous vehicles that are 12 metres long and can actively switch between tracks. Each vehicle shall comprise three sets of double-channel doors. For that, alternatives to conventional wheel-track systems shall be sought. A Cape gauge of 1,067 mm is seen as beneficial to the system, and cant and cant deficiency should be increased to 180 mm and 110 mm respectively.

5.8. Discussion of results

5.8.1. Summary of the operational concept and main results

The operational concept presented in this research has shown initial potential to achieve its objectives within the scope of door-to-door travel times and theoretical capacity. Assuming that all requirements can be met, the system can offer a median door-to-door travel time of 31.2 minutes for a 19.7 km journey. This is lower than driver's average time found in the literature and close to the Marchetti constant of urban travel (Gyimesi, et al., 2011; Marchetti, 1994). The theoretical capacity achieved using active switching technologies is approximately 111,000 passengers per hour per direction. When compared to the maximum estimated for a rapid transit system (Vuchic, 2007), the operational concept offers a 38.75% increase in theoretical capacity.

The operational strategy using patterns for different services can significantly reduce in-vehicle time as time penalties for intermediate stops are minimised. On top of that, the specific algorithm for the definition of number and type of patterns according to the number of stations on the line improves skip-stop style operations. Contrary to the traditional strategies found in the real world, the strategy used in this research ensures that all passengers can board any service at any station.

The operational strategy requires specific line layouts to operate to ensure that all vehicles attend all stations. To do so in a linear arrangement, the line needs to be operated in a 'carousel' style where the vehicles continue to follow patterns despite the eventual change in directions. Three options were identified but the circular line was adopted because it provides a clearer illustration. All layouts can lead to similar results considering that vehicles keep following their patterns after the end of the line. However, adjustments are needed when operating P_2 and P_3 services on straight lines. These services require extra programming to avoid stopping only at part of the stations on the line.

Entry and exit times (T_e), as well as interchange times (T_i), have also shown to have significant impact on travel times and should be investigated further. As they are systems themselves, this research could not elaborate on specific station designs for an optimal solution. Nonetheless, station development can be achieved by running the framework once again where high-level objectives are the time requirements for T_e and T_i. This shows an important potential of the framework when used iteratively on different levels of abstraction of the system but also the need for trade-offs between requirements and the operational concept. The optimal design for reducing T_e and T_i to a minimum may also be the costliest and the one that demands the greatest land take. Therefore, informed compromises between the reduction in journey times and incurred construction costs are necessary to identify a point of equilibrium. If necessary, high-level objectives may need updating if expectations are impossible to meet in practical terms.

While the pattern strategy and the identification of station design requirements have a significant impact on door-to-door travel times, off-line stations have a crucial role on capacity increase. When services stop on secondary tracks, braking and dwell times need not be included in headway calculations, thus permitting shorter headways. Values for curve radii were derived from maximum line speed and vehicle dimensions. The distance between the passing loop and the main track was prioritised as an attempt to reduce land requirements. Similarly, the chosen platform layout divides a regular platform linearly in one section for each pattern instead of various platforms in parallel in order to minimise platform time (T_p) when users need to change between patterns. Each section is linked to the main line by its own secondary line so that preceding traffic does not interfere with attending stations. While the design reflect the choices that resulted in the best outputs in journey times and capacity, there are various challenges and risks to be observed and will be discussed in section 5.8.3.

The third aspect of the operational concept involves platooning autonomous vehicles of different services together in order to minimise waiting time and increase overall capacity. With this arrangement, it was possible to infer that every passenger would have a waiting time of half the headway regardless of the service used. On the other hand, a platform time (T_p) penalty had to be added to the calculation for situations where passengers need to move along the platform to board another service on the same line. The need to change services within the same line is only necessary when the number of stops between origin and destination is a prime number. As previously shown, only a small share of the passengers would need to change services so T_p is less prevalent than other penalties. The platform time penalty T_p depends on the length and number of vehicles starting operations as a platoon. The longer the length of vehicles, the longer the platform needs to be, and the longer passengers need to walk to board a different service at a different platform section.

The process of technology selection also devised an alternative solution, based on normal track switches with faster actuation (REPOINT) that was lower on the risk axis. As expected, the lower risk was followed by lower capability and considerably reduced overall performance. Under higher TRL technologies (yet not TRL 9), the alternative operational concept uses traditional wheel-rail interfaces and passive track switching. It can maintain the faster door-to-door journeys yet imposes a significant reduction in capacity. For this reason, despite the lower risk, the concept using the REPOINT switch scored lower in the risk benefit analysis for two reasons: (1) it still comprised technologies which are not TRL 9, and (2) capacity is one of the main objectives of the project and thus carries a high weighting in the analysis. Should a solution use only TRL 9 components, the benefits would not be realised and the operational concept would neither be viable nor desirable.

From there, the sensitivity analysis for technology selection focused mostly on non-critical aspects of rolling stock and infrastructure. Assumptions had to be made and are explained in the next section. Jumping into specific design and architecture models would limit the capabilities of the solution and hinder the extent of radical innovation. Instead, the process focused on the functional aspects in order to derive physical requirements. The ability to investigate the impact of different technologies on the operational concept helped understand the interfaces between components and how they affect each other.

The system has been found to be more efficient when vehicles are 12 metres long, 2.6 metres wide, with 3 sets of double doors. However, the greater area dedicated to door space will inevitably have an impact on the vehicle capacity and seat provision. Moreover, the structural feasibility of the requirements were not analysed. These factors should be investigated further to return more realistic requirements that may or may not require changes in the operational concept or other design choices.

Rolling stock properties such as maximum speed, acceleration and braking were found to produce optimal results at 72 km/h, 1.3 m/s^2, and 1.2 m/s^2 respectively. They should not be reduced from their required values, otherwise they would have a considerable impact on door-to-door travel times. The system was found to be less sensitive to jerk, indicating that a lower value may be adopted for improved passenger comfort without large impacts on the overall performance. All parameters are already in use in current systems, meaning that they do not impose any risk or uncertainty to feasibility.

Apart from turnouts, the technological requirements analysed through Measures of Performance (MoP) were found to be on TRL 9, thus indicating low risk. A narrower gauge has been found to reduce the minimum curve radius and consequently reduce in-vehicle time. While various values were analysed, it is more economical to limit options to standardised gauges, such as the 1,067 mm suggested. However, it must be noted that the optimum operational concept involves active steering, which could potentially change the track requirements according to the technology adopted. This requires further investigation with the addition of extra parameters.

5.8.2. Main assumptions and scope limitations

The operational concept was developed with the underlying assumption of no technical or technological limitations, leading to an investigation to find one possible solution and backcast requirements from it. Such assumptions may denote a fragility at the current stage of the operational concept, but they are precisely the main objective of the framework in identifying potential solutions at a first stage to then refine with subsequent iterations.

The operational concept presented is but one option, and focused on two main aspects of journey times and capacity. Other equally important aspects of cost, safety, and reliability have not been evaluated at this stage due to the scope of the project but are essential in subsequent stages to investigate how to overcome practical barriers to implementation in a real world context.

For instance, the operational concept was based on the assumption of zero failures, which are unlikely given the number of switches required to move vehicles from the main line to their respective. In the layout proposed for off-line stations, vehicles would cross the path of the following vehicles with very tight margins, which increase the chances of collisions. In addition, the project assumed that vehicles would not lose communication and data would be processed in the shortest time. In reality, vehicles are not able to travel that closely because of potential failures and losses which can cause severe accidents at 72 km/h. While the results indicate potential roadmaps for future solutions, results are understood as (1) but one option among various operational concepts, and (2) a first iteration of the framework to be followed by subsequent processes that focus on other operational aspects.

Weightings on values of travel time were deliberately overlooked in the calculations of the resulting door-to-door travel times but should not be taken as irrelevant. Weightings of time penalties compared to in-vehicle time are difficult to predict in a new system, yet econometric analyses may help infer values in further research. In addition, social perceptions of value of time change from place to place, meaning that bespoke measurements should be conducted. Therefore, the values achieved relate to their physical properties and not their perception by users.

Another main assumption concerns station spacing. The operational concept adopted equidistant stations along the line, when in reality spacing reflects density. In traditional geographies, stations are close together in the central area, and further apart in the outer regions of the city. Vuchic (2005) has conducted extensive research that shows that optimum station spacing should be inversely proportional to the decreasing density of demand in a line from the central area to the outskirts of a city. The higher the density in demand, the closer the stations should be. This is an important issue because each station built on a certain line can have an impact on journey times, capacity, and capital and operational expenditures. In that regard, the model should be expanded to analyse operations with multiple station distances. On the other hand, it is to be noted that the framework stands on a normative future process and so stations can be used to generate demand (Cervero & Sullivan, 2010). Vuchic (2005) also acknowledges that ridership can sometimes be hiding a larger potential demand.

On the operational side, passenger behaviour was assumed to be generally uniform and non-disruptive. Passengers were assumed to walk to the station and travel within the system with the lowest times for each trip component. They were also assumed to distribute evenly along platforms which simplifies the evidence from the real world. Another simplified assumption was that passengers distribute evenly inside the vehicle and boarding was never blocked so that maximum capacity could always be achieved with the minimum amount of time. Nonetheless, simulations accounted for a variable dwell time in order to mimic passenger disruption at platform. Passenger flow needs to be more accurately simulated to understand the achievable capacity of the system in a real context. Vehicles were assumed to be able to accelerate and brake at expected values at all times. Variances in adhesion or traction power should be investigated to refine the models further. Communication was assumed to be constant and always available between trains. These may simplify the operational concept but such assumptions were key to its development at the initial stage.

The selection of an optimal solution was simplified because the main focus of the thesis is on the broader methodological aspects rather than the solution itself. Whereas an illustration of the process was used in this thesis, it has been more comprehensively shown in the first phase of a Horizon 2020 project for the development of high level architectures for the next generation of switches and crossings (S-CODE, 2017). The operational concepts developed in the project sustain the expected level of radical innovation, added by the robustness of iterative technical analyses in terms of feasibility, maintenance, and costs. After initial workshops, work package leaders of the project agreed that the framework helped thinking ‘outside the box’, focusing on the functional aspects rather than on physical limitations.

For similar reasons, simulations were not conducted using a bespoke simulator. Therefore, there was little flexibility in adjusting the parameters and the algorithms for more realistic tests.

As a consequence, the results from the simulation are to be seen as illustrative rather than definitive. The simulations assumed no conflicts from a highly complex network of switches and crossings around each station area which would prove challenging in a real environment. In addition, headway parameters used in the models and simulations are highly conceptual as they are based on theoretical solutions. Within the conceptual context, simulations indicate some robustness to the model. With the introduction of randomised dwell times, the system maintained its stability on most of the test runs.

5.8.3. Challenges and risks to implementation

While the operational concept achieves its goals in journey times and theoretical capacity, the limitations of the scope result in important challenges and risks to implementation faced by the operational concept that must be acknowledged for further investigation. Vuchic (2005) offers a set of aspects from both passenger and operator perspective that sheds light on the challenges and risks that need to be addressed by the operational concept if it is to be realised. The order they are discussed does not reflect their level of importance.

5.8.3.1. Availability

Passengers require a high level of locational availability (proximity to station) and temporal availability (hours of service) from urban transport systems (Vuchic, 2005). The first requirement was in fact the focus of the operational concept during its development in order to reduce the distance between stations without impacting average in-vehicle speeds. The second has not been studied in relation to the maintenance required. Given the ambitious technology selection, it is important to assess their maintainability requirements during later stages.

5.8.3.2. Speed, frequency, and capacity

These aspects are important from both passenger and operator perspective. The solution devised from the operational concept focuses on door-to-door journey speeds and capacity, which can in theory be achieved with lower speeds than some operational metros. Within the solution proposed, capacity can be increased with greater frequency in the form of shorter headways. These aspects are crucial to achieve the results, but rely on a level of automation and availability of virtual coupling technologies that may question the timescale for implementation. In addition, running vehicles more frequently and closer to each other increases the risk of accidents and their severity. These risks have not been included in the research.

5.8.3.3. Reliability

Reliability is perhaps the aspect that requires the most consideration from the operational concept. The solution proposed to improve journey times and theoretical capacity involves a considerable degree of complexity in design and operational strategy. To begin with, the choice for platform sections to be laid linearly rather than in other arrangements reflects the mains objectives of journey times and capacity, but incurs important trade-offs. Off-line stations incur greater costs in construction and land take. These costs have not been added to the risk benefit calculations because they were outside the scope of the project, but are expected to require compromises in the operational concept to overcome practical barriers to implementation. In addition, linear stations facilitate user movement and improve interchange times, but they require a significantly more intricate network of switches and crossings that can affect safety and reliability. As Vuchic (2005) highlights, ‘greater complexity always increases frequency of breakdowns’. In the event of using traditional track turnouts, the complexity of the arrangement may render the solution unfeasible and the operations too risky.

With a layout that involves platform sections, there would be at least two switches for each platform section at each station, multiplied by the number of stations. With that, the likelihood of a failure increases significantly when compared to current metro systems that adopt fewer switches to avoid such issues. On the other hand, active switching based on levitation technologies may offer a less complex solution because switching would not require mechanical devices and would instead change directions with polarisation in magnets on tracks. However, the feasibility of such systems is still uncertain with no proven concept in real world operations, and faults in polarisation can still happen, meaning that the number of turnouts could still be an issue. It logically follows that these issues require further investigation that could not be addressed by the scope of this research.

The adoption of platooning in the operational concept also carries various uncertainties in its feasibility and technological requirements that are higher on the risk axis. By adapting convoying algorithms used for road transport, the system is able to deliver the high performance estimated in the objectives. Yet running vehicles very close to each other in virtual coupling requires greater accuracy and processing speed than currently achieved through existing control systems, which is challenging considering the low adhesion nature of steel wheel on steel track systems. Another challenge is the actuation time of turnouts. Track switches can only physically move at a certain speed, which in turn would require vehicles to be farther apart.

In that sense, adopting platooning strategies would most likely have to involve a different type of interface between vehicles and tracks on turnouts, leading to specific technological requirements which are still in lower TRLs. Vehicles must be able to change track within a very short time frame, meaning that they most likely need to involve active switching based on superconducting levitation technologies which are still to be fully developed. The operational concept therefore becomes dependent on the achievement of such technological performance to be entirely feasible. Nonetheless, investigations into future technologies are exactly the goal of the framework in achieving normative scenarios. The recognition of the need of a different interface can be either a challenge or a technological project. The functional requirements derived from the operational concept may lead to an iteration of the framework in order to design such capability. On the other hand, this endeavour adds considerable risks over feasibility and costs.

5.8.3.4. Costs

This question of the number of stations track layout requirements also raises questions on the feasibility of retrofitting current metro systems. Vuchic (2005) points out that incremental costs per station decreases only slightly when station spacing is shorter. This is an important discussion because it concerns the extent of the benefit to cost ratio that the operational concept can achieve. In addition, in areas where the network density is high such as London and other developed megalopoleis, the need for extra tunnels and tracks may be financially prohibitive. In addition, the interfaces between the system with conventional lines may increase the complication of interchanges and thus affect trip penalties that are critical to its overall performance. Furthermore, the complication of the station layout may prove a challenge to overcome before the system can be designed to be reliable or cost beneficial.

In areas where network density is lower such as in developing countries, there is an important trade-off between applicability and feasibility. While these areas may have reduced costs for land take and also provide a blank canvas with fewer interfaces with existing lines, the costs and scale of such project may still remain a considerable barrier to implementation in face of local budgets. The costs of such systems depend on more detailed analyses that go beyond the scope of the current research, but findings in that regard will provide fruitful information to complement the risk-benefit analysis.

Moreover, there are general life-cycle and operational aspects that require further analysis beyond the initial technical realm. The operational challenges in running vehicles very close to each other around stations on a passing loop and the pressing issue of complexity of switches and crossings are expected to require extensive operational control and maintenance. For now, it is uncertain whether this level of service can be economically feasible. On that matter, the boundaries of the scope of this research become evident, highlighting the conceptual nature of this stage of the framework that results in but one option for an operational concept. An ad hoc analysis of the trade-off between system performance and operational costs can identify the economic viability of the project.

5.8.3.5. Flexibility

The operational concept comprises a more varied stopping pattern than the operational strategies currently in use as a means to reduce door-to-door journey times and increase the theoretical capacity of urban rail systems. Adopting a platooning design permits more flexible operations. Nonetheless, the short headways in relation to dwell times and the proximity of vehicles while travelling at maximum speed may in fact create an opposite effect of rigidity in operations. With very tight operations, the system becomes more fragile in face of unpredicted events. Simulations analysed the impact of disruption of dwell times on overall traffic, but disruptions on the main line have not been investigated. A degree of flexibility needs to be ensured so that small alterations in a complex environment do not spiral out of control. In that, further simulations are necessary to refine optimal frequency and vehicle-to-vehicle interaction.

5.8.3.6. Summary

From the assumptions, limitations, and challenges that surround the operational concept developed, several main issues are highlighted as the most critical to the technical feasibility of the project and stand as the main areas for further work. This section summarises them with regards to the practical barriers to be overcome:

- The number of switches involved in the design chosen for off-line stations. Having a switch for each platform section means that the system will comprise at least the product of the number of stations and the number of patterns. This highlights the trade-offs of offering more varied services with the increasing likelihood of switch failures that can eventually put the system to a halt if it runs on a circular line.

- In that, alternative platform layouts with simpler designs should be studied to find trade-off points between operational performance and reliability.
- Vehicles travelling with a distance of less than one second between them raises concerns over the actuation speed of switches, and the robustness of vehicle-to-vehicle communication. Should these barriers not be overcome, the operational concept may not achieve a sufficiently greater theoretical capacity to justify the higher operational requirements and costs.
- The solution proposed in the operational concept for the required switch actuation time involves superconducting levitation technologies. These have been developed and tested in controlled environments recently but application in the real world is still uncertain.
- Vehicles entering and leaving the platform with very short gaps between them. Platform design was chosen with tight curves and lines that cross each other, meaning that a vehicle leaving the platform will cross the line of another entering the station area. These movements significantly increase the risk of collision if adequate control systems are not in place.
- Off-line stations may help in reducing door-to-door journey times but there is an obvious cost issue in building a four-track system with stations on a passing loop. The cost of operational maintenance for the complex strategy also needs to be analysed to identify trade-off points where the system may be economically feasible. Similarly, the study assumed that T_e and T_i could be reduced but the extent of the possible reduction depends on the costs that design choices would incur.
- The scope did not include more specific analysis on the feasibility of retrofitting existing lines in robust networks. This aspect requires investigation in order to assess the economic viability of the operational concept.

5.8.4. Conclusion

While the operational concept has shown to potentially meet the targets of journey times and theoretical capacity, it would be premature to assess the overall feasibility of the solution. While various parameters were found to be of low risk, optimum system performance is still dependent on low-TRL technologies for track switching and convoying. These need to be realised before the system can be designed and developed. There are important reliability and safety issues to be reckoned with for the complex and tight movement of trains around platforms. Platooning tests with road vehicles have been successful, but the reality in the railway realm has not yet been proven outside simulations. Active switching using magnetic levitation is also a highly conceptual capability. Therefore, while the technical feasibility of the operational concept depends on the issues being overcome, the feasibility of the study was shown by the use of the framework to devise new potential solutions.

The case study provides a high-level illustration of the framework for radical innovation, especially of its heuristic nature, which was the main aim of the research. At a conceptual level, the case study has shown potential to achieve its objectives in term of journey times and theoretical capacity. The solution achieved in the case study is not final, nor is intended to be. The framework is an iterative process in order to maintain system coherence in complex problems. By switching the focus from current capabilities to optimum requirements, it was possible to derive an operational concept that satisfies the objectives and overcomes the systemic limitations found, and to identify the further technological developments that would be needed to realise the achievement of the stated requirements.

The purpose of the operational concept is precisely that of providing a blueprint of the system with which system engineers can further develop each aspect at lower levels of abstraction in an iterative process of the framework. The models show consistency with the reality of systems, indicating that the solution can potentially achieve its theoretical performance. However, there are important assumptions made in the model, but these are in fact the core of the framework. Its main purpose is to overcome current limitations and adopt teleology rather than determinism to devise necessary technologies to achieve a future goal.

6. DISCUSSION

6.1. Discussion of framework

6.1.1. Overall view

There seems to be then an inherent conflict that grows over time between the future requirements of the system context and the incremental evolution of its current capabilities. These systems, which include urban railways, tend to encounter systemic limitations after a period of incremental evolution that make them inadequate in face of changing external conditions. Moreover, engineering socio-technical systems take various years to be finished, meaning that in the current fast-changing environment they might be outdated by the time they are deployed. The research has shown how the proposed approach can help introducing radical innovation in complex systems. The thesis provided an illustration within a limited scope, in that its full application would need to embrace all of the practical barriers to achieve the desired results in a real world environment.

It was expected that the operational concept developed in the case study would not achieve a final and detailed solution, because the main intention of the research concerned the methodological aspects of radical innovation as a whole. In that, there are various technical aspects that require further work, also due to the magnitude of the operational concept as a whole system. Examples include station design for faster entry, exit, and interchange times, and also innovative turnout solutions that can operate within the required timeframe. Moreover, specific research on values of travel time weightings and costs would further improve the accuracy of benefit to risk calculations.

More importantly, the rather narrow scope of the case study highlights the challenges in the feasibility of the framework at the full scale. While the operational concept developed could provide a potential solution for door-to-door journey times and theoretical capacity, similar approaches would need to be applied for aspects such as reliability, cost, and safety. The additional capability assessment would be expected to require changes in the operational concept that lead to subsequent iterations to elicit the requirements to guide technological development.

However, the framework aims exactly on the identification of such requirements, so that the process can be conducted iteratively at increasing levels of detail. Its first iteration is not supposed to provide detailed designs, but to offer a robust and comprehensive blueprint of the system for the following steps. Within that domain, the results of the case study illustrate the potential of the framework in identifying critical aspects, the technologies that need addressing, and the selection between alternative solutions based on risk and other parameters. As expected, the framework helped defining specific directions for the next steps in technological development in order to achieve the normative future scenario chosen.

6.1.2. Innovation

The main benefit of the framework is the focus on functional aspects without an initial concern over technical and technological feasibility of physical components. Based on the backcasting method, it involves a normative approach to the future, identifying optimal operational concepts to solve problems, and then looking at the necessary technologies that make it possible. It logically follows that without the constraints of current technological capability, engineers are able to devise more radically innovative solutions that are also future-proof, in the sense that they are adequate to future scenarios.

From the literature review, it was shown that concerns over the uncertainty of normative futures tend to be overestimated because technological advancements are not a matter of chance, but a project of choice. Therefore, there is little difference in the accuracy of deterministic or normative forecasts on man-made systems because they are projects and not random occurrences. The unimaginable acceleration in technological advancements of recent centuries have shown that technical and technological requirements in the modern age are not a matter of possibility but a matter of time. In that sense, a normative approach to future technologies becomes desirable because, considering that future technologies eventually become reality, it is more efficient to drive advancements in directions established beforehand.

One example of this perception is the famous work of Konstantin Tsiolkovsy. Long before the age of space exploration, Tsiolkovsky (1903) calculated the systems requirements of rocket engines for orbiting the Earth and for cosmic flights. At that time, the necessary technology was not available, but the work proved relevant decades later. When technologies were developed based on his blueprints, they gave way to new boundaries in scientific and engineering potential.

The case study successfully highlighted this endeavour. The framework focuses on future scenarios, therefore current limitations should not be taken as set in stone. As shown in the literature review, radical innovation requires a completely different approach to the traditional incremental processes, because the latter eventually encounter systemic paradoxes that prevent any further improvement without trade-offs. In addition, the ever-faster changes in systems environment renders them outdated increasingly quicker. The framework then is able to produce an operational concept which looks forward and devise a solution that drives technological capability advancement to meet future requirements and objectives.

6.1.3. Robustness

On the other hand, it does not mean that just a vision will suffice in order to achieve working solutions. While the vision provides guidance in the form of objectives, it is necessary to use systems engineering processes to elicit specific requirements to be met. With that, two outcomes are achieved: (1) identification of off-the-shelf technologies that already meet the required performance, or (2) specific performance definitions for new technologies to be developed. The sequence of technology identification and optimal solution selection is an iterative process, where the technologies can be assessed on their benefit to risk ratio. In that, engineers can compare the performance of the system with off-the-shelf technologies that partly fulfil the requirements against ideal technologies that do not yet exist. Doing so, it is possible to analyse whether it is better to compromise in performance or in uncertainty.

When decided, the solution provided by the operational concept is not the end of the project, nor is intended to be. The framework should be used as an iterative process over different levels of abstraction. For instance, the overarching operational concept defined specific objectives for station design in terms of entry/exit and interchange times. From there, the process should be applied again in order to devise specific innovative designs that meet the target performance. Once functional requirements within that domain were elicited and technologies identified, the process will be used again, until it reaches a detailed design.

This also highlights the need for a multidisciplinary approach, which benefits from systems engineering processes. By understanding the inherent paradoxes in interfaces, it is possible to find the functional aspects that require change in order to overcome systemic limitations. On that, the use of models and simulations add significant robustness compared to the more traditional normative and explorative scenario building. The various modelling processes and techniques are crucial for drawing a precise image of the operational concept that is useful for subsequent analysis. In fact, they are responsible to the added robustness to future scenarios in contrast to the more general descriptions of current methods.

As illustrated in the case study, they were responsible for transforming an idea into an operational concept, and an operational concept into a defined set of requirements for technological performance. The models and simulations help grounding the vision into a specific and better-defined project, which encounters functional and physical limitations that need addressing. In the example of the case study, trade-offs in line layout and actuation time have brought to attention important issues with reliability, safety, and cost that need solving before the proposed solution can be seen as such.

There is an important benefit from the recent improvements in modelling capabilities in the form of added precision. Modelling processes have long been part of innovation. There is a quote attributed to Nikola Tesla (2010) that summarises this synergy:

> “My method is different. I do not rush into actual work. When I get a new idea, I start at once building it up in my imagination and make improvements and operate the device in my mind. When I have gone so far as to embody everything in my invention, every possible improvement I can think of, and when I see no fault anywhere, I put into concrete form the final product of my brain.”

However, the recent advances in processing power enables models and simulations to be much more detailed in their depiction of complexity, meaning that it is possible to investigate operational concepts robustly. The notion of Model Based Systems Engineering (MBSE) lies on this increased capacity to simulate complex systems with greater level of detail, almost matching the properties of physical prototypes. With that in mind, the framework benefits greatly from improved capabilities in modelling and simulation. Firstly, because it is possible to investigate radically innovative solutions in greater detail, adding robustness to the solutions analysed. Secondly, because the greater processing power also enables the models to test different combination of technologies, adding to the risk benefit analysis. Thirdly, because the greater level of details leads to more specific design requirements for new technologies when needed, thus improving the potential feasibility of the operational concept.

This means that a normative approach to the future is not necessarily risky when systems engineering processes are included in the project. The backcasting perspective changes the role of technologies from being initial constraints to being final products, giving way to radically innovative solutions which would not be taken into account otherwise. It is the adoption of models and simulations, now more advanced than ever, that allow operational concepts to be tested under the assumption of technological feasibility and the reassessed in their risks and benefits. Even more so, the process identifies critical and non-critical technologies, which help determine the selection of technologies and the informed compromises between performance and uncertainty.

6.1.4. Application

Although the thesis and the case study focused on urban rail systems, it can be inferred that the framework can be used for other complex socio-technical systems. The limitation in scope aimed at clear explanation via analogy, but in no means the framework could only be applied in the railway realm. As mentioned in the literature review, most man-made systems follow a similar process in which incremental innovation gradually returns less benefits for each unit of engineering effort. Eventually, the internal trade-offs render the system incapable of adapting to constant changing conditions.

One of the benefits of a framework based on the logic of discovery is that it is open-ended and flexible enough to be applicable to various problems. The core of the framework lies on the inversion of engineering process to devise radically innovative solutions. Such approach can be used to any system, although benefits will be greater for complicated and complex socio-technical ones. These systems tend to encounter important systemic limitations due to: (1) their constantly changing conditions in their physical and social environments, and (2) their greater chances of internal trade-offs that arise from their complex structures.

In these systems, the large number of sub-systems and components challenges synergistic development, because each may experience distinct progress according to market forces that are not always aligned with each other or overall societal needs. With that, the framework can assist on technological development by highlighting the potential value of certain components that would otherwise be overlooked or of late adoption by the market. The framework identifies one or few possible solutions that may be taken forward for further analysis. On that, a robust benefit to risk analysis which includes safety, costs, and reliability, can identify those that are worth investigating in more detail.

6.1.5. Replicability

Even though the framework focuses on reducing the uncertainty and risk of innovative operational concepts, the main shortcoming of the framework lies on its own epistemological nature. For being based on the heuristics of the logic of discovery, it cannot be a method, even less so a process, so its adoption does not depend on the project but on the designers and engineers. As explained in Chapter 3, there is no recipe for invention. Consequently, the replication of the framework is not simply a matter of following steps. It requires engineers to adopt a different posture towards their work, which relies on self-control and no algorithm can guarantee that.

From the literature review, it became clear that the distinction between the logic of justification and the logic of discovery encompasses a choice of perspective. Logically, in the physical sciences there is little space for heuristic methods, yet they are significantly more relevant in the artificial realm. That is because in man-made systems, choice is inherent, meaning that the conservatism with which they are designed is mostly due to epistemological paradigms. In that, perhaps the biggest challenge of the framework is that it requires engineers to work in a different way than they are used to, and that can lead to two negative outcomes: (1) the potential for radical innovation in the development of an operational concept will be hindered, because of conservative choices; and (2) radically innovative solutions may be abandoned due to a conservative bias on uncertainty and risk.

In addition, the absence of a precisely defined set of steps means that the framework can be open to interpretation. With different interpretations, there is no standard set of outputs, and with no standard outputs it is difficult to define what the end result should look like. As observed in the S-CODE (2017) project where the framework was applied, this flexibility can lead to initial debates and divergences between stakeholders.

On the other hand, flexibility is what makes the framework most fruitful. As innovation is inherently a heuristic process, engineers need to be able to adapt processes to the specific characteristics of the problem in hand. The framework, although open-ended, provides an epistemological support to ensure that the order of prioritisation is kept. In other words, the main benefit of the framework is to maintain the focus on the functional aspects in the beginning of the project, leaving the technical and technological assessments for later stages.

Therefore, the framework can be applied to any problem that involves complex systems in the socio-technical realm. Logically, it does not apply to natural systems as normative futures are impossible to devise. The framework is more fruitful where radical innovation is required, which is usually the case in rapidly changing environments such as urban areas. It would be indeed beneficial to apply the framework in systems outside the spectrum of transport in order to collect results and measure the level of replicability.

6.1.6. Relevance

The main reason for this research is the current conflict between the need for radical innovation in changing environments, and the inherent limitations of incremental evolution in systems in that aspect. The framework is proposed in a time of rapid changes in the urban environment, especially in developing countries. Their cities are experiencing an explosion in urbanisation rates and infrastructure systems are failing to follow.

In such context, a framework that focuses on radical innovation is considerably relevant to leapfrog advancements and keep pace with changing demands and social and spatial structures. There is an extra advantage of places where infrastructure is poor as benefit to risk ratios tend to be higher. In addition, there is little overlapping with current systems in that new projects are more necessary than retrofitting existing ones. Finally, there is a logical aspect of learning from systems in place and avoiding repetition of inherent limitations where possible.

This is not to say that the framework only applies to developing countries. In developed regions, radical innovation also applies but robust infrastructure and costs may hinder the potential benefit to risk ratio. Nonetheless, the framework is generally aimed at making the best use of technological development for large socio-technical systems that take a long time to be completed. Its most valuable aspect is the balancing of the two sides of socio-technical systems, identifying potential technologies to match future societal demands. It is argued that a normative approach to technologies in a systems engineering process can help leapfrogging the distance between changing environments and incremental evolution.

6.2. Recommendations

Based on the discussion and the results of the application of the framework, it is recommended that:

1. The framework is understood as an iterative process in which each iteration is not final but another step in the direction of greater levels of detail.

2. A full application of the framework is necessary to understand whether a number of iterations in increasing levels of details can overcome all practical barriers to achieve expected results in a real world environment.

3. The framework can be applied to socio-technical systems that involve complex interactions and interfaces, being most useful in the context of rapidly changing environments.

4. The framework is not an algorithm, but an open-ended heuristic, meaning that its flexibility should be used for achieving a combination of innovation and robustness.

5. Engineers using the framework should adopt the perspective of radical innovation, leaving the assessment of technological to later stages to enable more innovative solutions.

6. Models and simulations are valuable assets in the application of the framework, as they can greatly increase the level of detail of the operational concept and lead to more robust risk benefit analyses.

7. The framework should be applied to socio-technical systems outside the spectrum of transport in order to assess its replicability.

7. CONCLUSION

The starting point of this research was the transformation in the urban landscape and the recent debate on how to reinstate the equilibrium in travel times in megalopoleis. Their expansion, as a result of the reduced cost of travel permitted by technological advancements, now challenges the sustainability of transport systems to provide access within reasonable travel times. As travel distances in these urban giants continue to grow, the needs to faster travel becomes a key issue for the century when the size and number of megalopoleis is only expected to increase.

The literature is very settled on the notion of an anthropological invariant time that people dedicate to travelling (travel time budgets), and that human settlements have been approximately one hour wide over centuries. Logically, the faster one can travel, the larger the city can grow. However, the emergence of megalopoleis as sprawled polycentric urban areas now challenge the ability of current modes of transport to maintain such balance. Furthermore, it created a systemic imbalance between where public transport users spend in average least 50% more time to cover similar distances to drivers in many of these cities. This contributes to the threats to the future social, economic, and environmental sustainability or urbanisation across the world.

One of the reasons for the disparity in travel times between private and public systems lies on inherent paradoxes between trip components in the latter. The first original contribution to knowledge of this research is an in-depth understanding of the inherent trade-offs between access and in-vehicle speeds observed in public transport systems. Such notion, although widely known, had not been yet linked to the disparity between the emergence of megalopoleis and the inequalities in door-to-door travel times between different modes of transport. From the models, it was realised that these paradoxes cannot be overcome by incremental evolution of systems in the context of ever longer travel distances, and thus requires radical change to adapt to new conditions.

Nonetheless, radical innovation is not an easy task, especially concerning complex systems such as the socio-technical ones found in the urban environment. Therefore, in face of the wider challenges of future sustainability, the main objective of this thesis turned into the development a framework with which engineers can design radically innovative solutions which are robust enough to become projects for future systems. This process is the core novel contribution to knowledge as it looks into new ways of introducing innovation to engineering. To do so, it was necessary to conduct an extensive and thorough review of the literature in three of the main epistemological dichotomies from the philosophy of science.

Firstly, the distinction between the logic of discovery and the logic of justification. Although the engineering field is concerned with deliberate efforts to create artificial systems, the perspective of work is still very much rooted in the methods of the natural sciences. These methods, considered the only valued scientific approach, require the separation between the observer and the object in order to permit complete falsifiability and replication. It becomes fairly obvious to see where the application of such logic in engineering can hinder progress. Engineering is inherently concerned with the active development of new machinery to fulfil a need. Therefore, it is illogical to expect engineers to adopt only an observational attitude to problems.

Secondly, the different approaches to the future. Large engineering projects such as those of socio-technical system are usually long in duration, taking several years to be completed. It logically follows that they are inevitably forward-looking, and in their context of rapid changing environments, radical innovation is a necessity for sustainability. This means that the deterministic approach rooted in the natural sciences does not suffice. It is based on forecasts and projects no significant changes in the system, which is essentially the requirement for radical innovation. The research then focuses on normative approaches as they not only envision change but rather promote it. Its most common method, namely backcasting, has been is use for decades in the field of planning and provides a good basis for radical change. It starts with a desired scenario and works backwards in order to identify the necessary measures and actions required to achieve that end result.

However, planning methods are known for the lack of technical precision in face of the complexity of scenarios and therefore have not since found a place in the engineering realm. For that reason, a few researchers have suggested its combination with systems engineering for more robust scenarios. Since no instance was found in the literature, the thesis used the gap to develop a framework based on the combination of both approaches. Systems engineering oppose to the traditional reductionist approach and focus on the entirety system, especially its interfaces. They add robustness to a normative scenario by devising an operational concept and eliciting requirements from it. With that, it is possible to analyse specific technological needs that will lead to the solution when available.

The framework that arises from these three pillars is a set of heuristics to guide engineers to achieve radical innovation. It cannot be a method due to the absence of a step-by-step process to be followed, which is perhaps its main shortcoming. However, the very nature of the logic of discovery is heuristic and not algorithmic, which explains why there cannot be a defined set of steps to achieve innovation. On the one hand, the flexible structure of the framework makes is more applicable to various problems. On the other hand, replication and effectiveness are less guaranteed because they depend on the posture adopted by those using it.

It is well anticipated that this heterodox perspective to engineering may be received with cynicism or some level of suspicion regarding its relevance. In fact, it is expected that it does so. The framework was developed from a theoretical perspective and has been used only in two occasions so far, both in which the author had an active role in the project. Thus, it is imperative that other projects, especially those outside the transport realm, apply the guidelines proposed in order to verify its replicability and robustness when conducted by different researchers in various socio-technical systems.

In this thesis, the framework was applied in an illustrative case study to the very initial problem encountered of travel times in megalopoleis. Based on the evidence that users choose the fastest mode they afford to travel, the first objective of the novel system in was to approximate door-to-door journey times to 30 minutes in the context of long commutes. Moreover, based on the demographic and geographical forecasts of megalopoleis, the objective for theoretical capacity is to carry 100,000 passengers per hour per direction.

When modelling the current capabilities of the system, it became clear that incremental evolution does not suffice because of the inherent conflicts between access times and in-vehicle speeds. Several operational strategies have been attempted, such as local/express and skip-stop services. However, they comprise important trade-offs in capacity or inequalities in travel times and accessibility for users. Using the essence of radical innovation promoted by the framework, it was possible to develop an operational concept that can fulfil the objectives without the trade-offs of previous solution.

Benefitting from the technical freedom of the initial steps of the framework, the operational concept consists of three main aspects. Firstly, an optimised skip-stop strategy where services follow pre-defined stopping patterns along the line (e.g. every station, or every two, three, five, or seven stations). When the number of stations on the line is not a multiple of the patterns, it ensures that all services will attend all stations eventually, taking a number of laps equal to the pattern to do so. With that, the system can offer significantly higher average in-vehicle speeds that normal operations, and the discrepancies in travel times for different users of previous strategies are solved.

Secondly, the operational concept relies on stations located on secondary tracks off the main line. This ensures that the services that stop more often do not hold others back and thus limit their in-vehicle speed gains. In addition, it prevents users from experiencing increased waiting times. The optimal arrangement encountered was a linear disposition of platform sections, each with its own set of tracks connecting to the main line. This facilitates user experience and saves on space consumption.

Thirdly, the operational concept investigates the use of platoons of vehicles rather than conventional trains in order to improve the efficiency of the system. Assuming fully autonomous vehicles convoying using algorithms similar to those in road transport, it is possible to include all stopping patterns in all platoons, which reduces headways at the station. In addition, such system is more economical as the number of vehicles of each pattern can be adapted to the demand.

Once the normative scenario was complete, the combination of backcasting and systems engineering made possible to delve in much further detail than previous uses of the methods separately. Models and simulations were used to identify the functional and physical requirements of the system that would allow the operational concept to achieve its objectives. Aspects from vehicle dimensions to cant were tuned in order to achieve an optimal result. Subsequently, technologies were identified in their feasibility and risk to increase the robustness of the solution.

Results show that, assuming all requirements possible, the operational concept can indeed fulfil its objectives and provide a median door-to-door time of 32 minutes for a 19.7 km journey, while offering a theoretical capacity of around 110,000 passengers per hour per direction. To make this possible, a few technological gaps identified need to be overcome. Most importantly, the functions for changing tracks require new technologies based on active switching. Using optimised version of current systems would not impact travel times but would considerably affect the capacity of the system. There is also further work required on station design which guidelines have been simplified in the thesis. Finally, the costs of retrofitting existing systems may change the benefit to risk ratio to an extent that the improvements in travel times and capacity cannot be justified. On the other hand, new lines would prove much more feasible, and sprawling developing megacities could benefit from future-proof systems.

Some of the technologies debated in this thesis are yet to be developed. While that may seem overly ambitious, it is exactly the point of the framework. Considering that uncertainty is a natural quality of futures studies, normative scenarios are no more uncertain than deterministic ones when it comes to engineering. Quoting Dennis Gabor (1963), ‘futures cannot be predicted, but they can be invented’.

In conclusion, the work presented in this thesis, both at the methodological and the technical level, is not final and it does not intend to be so. For being rooted in the logic of discovery, the process is not exhaustive but iterative. In contrary, the main purpose of this thesis is to set a starting point for a new epistemology of engineering which is more in tune with the needs and capabilities of the 21st century. Social and environmental changes are happening increasingly fast in the digital age or urbanisation, and the natural attractiveness of megalopoleis require systems to adapt increasingly fast. For large scale infrastructure systems that take years to complete, this means that a future oriented approach is more necessary than ever, especially in developing countries where they are less pervasive. Rather than chasing the tail of social changes and environmental concerns, systems are now bound to leapfrog the trace and aim at fulfilling these aspects in their future stages. To do so, however, requires a shift in the traditional processes, with a greater emphasis on the functional needs than current physical capabilities. Current scientific and technological prowess now enables normative scenarios to be more than simple visions, and refined models can assess the robustness and feasibility of solutions in future contexts before the first stone is even laid. Therefore, it is perhaps time to tackle engineering projects more openly, using methods that give space for radical innovation. With that, this thesis hopes to ignite a new perspective on developing complex urban systems in the future.

8. REFERENCES

Acheson, P., Dagli, C. & Kilicay-Ergin, N., 2013. Model Based Systems Engineering for System of Systems Using Agent-Based Modeling. Procedia Computer Science, Volume 16, pp. 11-19.

Alexander, I. & Stevens, R., 2002. Writing better requirements. London: Pearson Education Ltd.

Altshuller, G. S., 1984. Creativity as an exact science: the theory of the solution of inventive problems. Amsterdam: Gordon and Breach Publishers.

Altshuller, G. S., 1999. The innovation algorithm: TRIZ, systematic innovation, and technical creativity. Worcester: Technical Innovation Center, Inc.

Amer, M. & Daim, T. U., 2010. Application of technology roadmaps for renewable energy sector. Technological Forecasting and Social Change, 77, pp. 1355-1370.

Anderson, G., 1973. Methods in futures studies: a view from the theory of science. Technological Forecast and Social Change, 5(3), pp. 303-317.

Anderson, J. E., 2000. A review of the state of the art of personal rapid transit. Journal of Advanced Transportation, 34(1), pp. 3-29.

ATKINS. 2011. Rail value for money study: Whole System Programme Management Final Report. London: Department for Transport.

Ausubel, J. H., 2014. Cars and civilization. [Online]
Available at: https://phe.rockefeller.edu/docs/Cars and Civilization.pdf
[Accessed 06 November 2017].

Ausubel, J. & Marchetti, C., 2001. The evolution of transport. The Industrial Physicist, 7(2), pp. 20-24.

Baiceanu, R., 2015. A Decade of Home Price Changes in NYC Reveals a Very Divided City. [Online]
Available at: https://www.propertyshark.com/Real-Estate-Reports/2015/10/14/a-decade-of-home-price-changes-in-nyc-reveals-a-very-divided-city/
[Accessed 02 April 2018].

Banister, D., 2011. The trilogy of distance, speed and time. Journal of Transport Geography, 19, pp. 950-959.

Banister, D. & Hickman, R., 2005. Visioning and Backcasting for UK Transport Policy (VIBAT), London: Department for Transport.

Banister, D. & Hickman, R., 2007. Looking over the horizon: Transport and reduced CO2 emissions in the UK by 2030. Transport Policy, 14(5), pp. 377-38.

Banister, D. & Hickman, R., 2013. Transport futures: thinking the unthinkable. Transport Policy, 29, pp. 283-293.

Barbosa, B., 2016. Alívio na lotação de Metrô e CPTM deve ficar para 2017. [Online]
Available at: https://noticias.uol.com.br/cotidiano/ultimas-noticias/2016/01/11/alivio-na-lotacao-de-metro-e-cptm-deve-ficar-para-2017.htm
[Accessed 06 April 2018].

Batty, M., 2005. Cities and complexity: understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. Cambridge, Massachusetts: MIT Press.

Batty, M., Besussi, E. & Chin, N., 2003. Traffic, urban growth and suburban sprawl, London, UK: UCL Centre for Advanced Spatial Analysis.

Beirão, G. & Cabral, J. A., 2007. Understanding attitudes towards public transport and private car: a qualitative study. Transport Policy, 14(6), pp. 478-489.

Bemment, S., Ebinger, E, Goodall, R. M., Ward, C. P., Dixon, R. 2017. Rethinking rail track switches for fault tolerance and enhanced performance. Proceedings of the ImechE Part F: Journal of Rail and Rapid Transit. 231(9). pp. 1048-1065

Bemment, S. 2017. The REPOINT faul-tolerant switch. Singapore, IRSE ASPECT.

Berkhout, F. & Hertin, J., 2002. Foresight futures scenarios: Developing and applying a participative strategic planning tool. Greener Management International, 37, pp. 37-52.

Bettencourt, L. M., Lobo, J., Strumsky, D. & West, G. B., 2010. Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. PLoS ONE, 5(11), pp. 1-9.

Bieber, A., Massot, M. & Orfeui, J., 1994. Prospects for daily urban mobility. Transport Reviews, 14(4), pp. 321-339.

Blainey, S., Hickford, A. & Preston, J., 2012. Barriers to passenger rail Use: A review of the evidence. Transport Reviews, 32(6), pp. 675-696.

Blanchard, B. S. & Fabrycky, W. J., 2006. Systems engineering and analysis. New Jersey: Pearson Education.

Blumenfeld, M., Schmid, F. & Roberts, C., 2016b. A systems approach to developing future urban systems: creating a metro that fits the megalopolis. Proceedings of the International Conference on Railway Engineering, pp. 1-6.

Blumenfeld, M., Stewart, E. & Larios, E., 2016a. Using model based systems engineering to increase the robustness of monitoring systems in railways. Proceedings of the 7th IET Conference on Railway Condition Monitoring, pp. 1-6.

Blum, U., Haynes, K. E. & Karlsson, C., 1997. The regional and urban effects of high-speed trains. The Annals of Regional Science, 31(1), pp. 1-20.

Boehm, B. W., 1988. A spiral model of software development and enhancement. Computer, 21(5), pp. 61-72.

Braithwaite, R. B., 1953. Scientific explanation: a study of the function of theory, probability and law in science. Cambridge: Cambridge University Press.

Carmichael, R. D., 1930. The logic of discovery. Chicago: The Open Court Publishing.

Cervero, R., 1998. The transit metropolis: a global inquiry. Washington, DC: Island Press.

Cervero, R. & Sullivan, C., 2010. Towards green TODs. Working paper UCB-ITS-VWP-2010-7, Berkeley: UC Berkeley Center for Future Urban Transportation.

Chen, C. & Mokhtarian, P. L., 1999. A review and discussion of the literature on travel time and money expenditures, Davis: University of California.

Christensen, C. M., 2000. The innovator's dillema: when new technologies cause great firms to fail. Boston, MA: Harvard University Review Press.

Cintra, M., 2014. Os custos de congestionamento na cidade de São Paulo, São Paulo: FGV São Paulo School of Economics.

Coates, J. F., 2000. Scenario planning. Technological Forecasting and Social Change, 65(1), pp. 115-123.

Companhia de Engenharia de Tráfego de São Paulo - CET/SP, 2017. Mobilidade no Sistema Viário Principal, São Paulo: Companhia de Engenharia de Tráfego de São Paulo - CET/SP.

Connor, P., Harris, N. & Schmid, F., 2015. Designing and managing urban railways. London: A&N Harris.

Council for Science and Technology, 2009. A national infrastructure for the 21st century, London, UK: Council for Science and Technology.

Cox, W., 2017. World urban areas - 13th Annual Edition: April 2017, Belleville, IL: Wendel Cox Consultancy.

Crutchfield, J. & Wiesner, K., 2010. Simplicity and complexity. Physics World, 23(2), p. 36.

Dendrinos, D. S. & Mullally, H., 1985. Urban Evolution: Studies in the Mathematical Ecology of Cities. Oxford: Oxford University Press.

Department for Transport, 2015. Average time taken to travel to work by region of workplace and usual method of travel, Great Britain: October to December 2015 - Table TSGB0111, London: s.n.

Deweerdt, S., 2016. The urban downshift. Nature, 531, pp. s52-S53.

Dittmar, H. & Ohland, G., 2004. The new transit town: best practices in transit-oriented development. Washington, DC: Island Press.

Douglass, B., 2015. Agile systems engineering. Burlington: Morgan Kaufmann.

Dreborg, K. H., 1996. Essence of backcasting. Futures, 28(9), pp. 813-828.

Durance, P. & Godet, M., 2010. Scenario building: uses and abuses. Technological Forecasting and Social Change, 77(9), pp. 1488-1492.

EIA, 2016. International Energy Outlook 2016, Washington: U.S. Energy Information Administration .

ESRI, 2015. The age of megacities. [Online]
Available at: http://storymaps.esri.com/stories/2014/growth-of-cities/
[Accessed 25 August 2018].

Evans, R., 2015. I am because we are: a systems approach to strategy design, Birmingham: University of Birmingham.

Experience Tokyo, 2015. The new E235 JR Yamanote Line train on a test run. [Online]
Available at: http://experiencetokyo.net/spotted-the-new-e235-jr-yamanote-line-train-on-a-test-run/
[Accessed 02 April 2018].

Federal Aviation Administration, 2014. FAA Systems engineering manual. Washington, DC: Federal Aviation Administration.

Fernandes, P. & Nunes, U., 2015. Multiplatooning leaders positioning and cooperative behavior algorithms of communicant automated vehicles for high traffic capacity. IEEE Transactions on Intelligent Transportation Systems, 16(3), pp. 1172-1187.

FERSIL, 2008. The COPP system - device to open and close platform screen doors, Paris: FERSIL.

Forstall, R. L., Greene, R. P. & Pick, J. B., 2009. Which are the largest? Why lists of major urban areas vary so greatly. Journal of Economic and Social Geography, 100(3), pp. 277-297.

Fosberg, K. & Mooz, H., 1991. The relationship of System Engineering to the project cycle. Proceedings of the National Council for Systems Engineering First Annual Conference, pp. 57-61.

Freyss, M., Giesen, R. & Muñoz, J., 2013. Continuous approximation for skip-stop operation in rail transit. Transportation Research Part C, 36, pp. 419-433.

Fu, L., Liu, Q. & Calamai, P., 2003. Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record, 1857, pp. 48-55.

Furth, P. & Day, F., 1985. Transit routing and scheduling strategies for heavy demand corridors. Transportation Research Record, 1011, pp. 23-26.

Gabor, D., 1963. Inventing the future. Bungay: Secker & Warburg.

Gabor, D., 1969. Normative technological forecasting. Technological forecasting, 1(1), pp. 1-3.

Gallopin, G. A., Hammond, A., Raskin, P. & Swart, R., 1997. Branch points: global scenarios and human choice, Stockholm: Stockholm Environment Institute.

Galvin, R., 1998. Science roadmaps. Science, 280(5365), p. 803.

Gazibara, I., Goodman, J. & Madden, P., 2010. Megacities on the move: your guide to the future of sustainable urban mobility in 2040, London: Forum for the Future.

Geddes, P., 1915. Cities in evolution: an introduction to the town planning movement and to the study of civics. London: Williams & Norgate.

Geels, F. W., 2004. From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. Research Policy, 33(6-7), pp. 897-920.

Gehring, O. & Fritz, H., 1997. Practical results of a longitudinal control concept for truck platooning with vehicle to vehicle communication. Boston, Massachusetts.

Givoni, M. & Banister, D., 2012. Speed: the less important element of the high-speed train. Journal of Transport Geography, Volume 22, pp. 306-307.

Givoni, M. & Rietveld, P., 2014. Do cities deserve more railway stations? The choice of a departure railway station in a multiple-station region. Journal of Transport Geography, Volume 36, pp. 89-97.

Godet, M., 1979. The crisis in forecasting and the emergence of the "Prospective" approach. Oxford: Pergamon Press.

Godet, M., 1986. Introduction to la prospective: seven key ideas and one scenario method. Futures, 18(2), pp. 134-157.

Goodwin, P. B., 1981. The usefulness of travel budgets. Transportation Research Part A, 15(1), pp. 97-106.

Gottmann, J., 1961. Megalopolis: the urbanized northeastern seaboard of the United States. New York: The Twentieth Century Fund.

Green, D. R., 1988. Distance to work in Victorian London: a case study of Henry Poole, Bespoke Tailors. Business History, 30(2), pp. 179-194.

Gunn, H. F., 1981. Travel budgets - a review of evidence and modelling implications. Transportation Research Part A, 15(1), pp. 7-23.

Gu, W., 2015. Queues in Transportation Systems. [Online]
Available at: http://www.mypolyuweb.hk/w1981gu/res.htm
[Accessed 24 April 2018].

Gwilliam, K., 2002. Cities on the move: a World Bank urban transport strategy review, Washington, D.C.: The World Bank.

Gwilliam, K., 2003. Urban transport in developing countries. Transport Reviews, 23(2), pp. 197-216.

Gyimesi, K., Vincent, C. & Lamba, N., 2011. Frustration rising: IBM commuter pain survey, New York: IBM.

Hall, F. L., 1996. Traffic stream characteristics. In: Traffic Flow Theory. Washington, D.C.:US Federal Highway Administration.

Hall, P., 2009. Looking backward, looking forward: the city region of the mid-21st century. Regional Studies, 43(6), pp. 803-817.

Hall, P., 2014. Cities of Tomorrow: An Intellectual History of Urban Planning and Design Since 1880. 4th edition ed. New Jersey: Wiley-Blackwell.

Hall, P. & Pain, K., 2009. The Polycentric Metropolis: Learning from Mega-City Regions in Europe. London: Routhledge.

Handy, S., 2005. Smart growth and the transportation-land use connection: what does research tell us?. International Regional Science Review, 28(2), pp. 146-167.

Handy, S., Weston, L. & Mokhtarian, P. L., 2005. Driving by choice or necessity?. Transportation Research Part A, Volume 39, pp. 183-203.

Hanson, N. R., 1958. Patterns of Discovery. Cambridge: Cambridge University Press.

Hennig, B. & Dorling, D., 2014. Inequalities in London. [Online]
Available at: http://www.londonmapper.org.uk/analysis/inequality-in-london/
[Accessed 02 April 2018].

Heres, D., Jack, D. & Salon, D., 2013. Do public transport investments promote urban economic development? Evidence from Bogotá, Colombia. Transportation. 41(1). pp. 57-74.

Hickman, R. & Banister, D., 2007. Looking over the horizon: transport and reduced CO2 emissions in the UK by 2030. Transport Policy, Volume 14, pp. 377-387.

Höjer, M. & Mattson, L.-G., 2000. Determinism and backcasting in future studies. Futures, 32(7), pp. 613-634.

Holt, J. & Perry, S., 2014. SysML for Systems Engineering. London: IET.

Hoornweg, D. & Pope, K., 2016. Population predictions for the world’s largest cities in the 21st century. Environment and Urbanization, 29(1), pp. 195-216.

Howard, E., 1902. Garden cities of to-morrow. 2nd ed. London: S. Sonnenschein & co.

Hybertson, D. W., 2009. Model-oriented systems engineering science. Boca Raton, FL: Taylor and Francis Group, LLC.

INCOSE UK, 2009. Why do Systems Engineering? Manage complexity, reduce your risk, Ilminster: International Coucil of Systems Engineering - UK Chapter.

INCOSE UK, 2014. Capability systems engineering guide, Ilminster: International Council of Systems Engineering - UK Chapter.

INCOSE, 2011. Systems engineering handbook: a guide for system life-cycle processes and activities, San Diego: International Council on Systems Engineering.

Instituto Brasileiro de Geografia e Estatística, 2011. Indicadores sociais municipais: uma análise dos resultados do universo do censo demográfico 2010, Rio de Janeiro: IBGE.

Instituto Nacional de Estadística, Geografía e Informática, 2007. Encuesta Origen – Destino de los Viajes de los Residentes de la Zona Metropolitana del Valle de México 2007. Mexico City: INEGI.

International Energy Agency, 2017. Key world energy statistics, Paris: IEA Publications.

ITF, 2017. Transport demand and CO2 emissions to 2050, Paris: OECD Publishing.

IUGS, 2005. Megacities: our global urban future, Leiden: Earth Sciences for Society Foundation.

James, R., 1984. Solutioneering. New Scientist, 31 May, p. 34.

Jantsch, E., 1967. Technological forecasting in perspective, Paris: OECD.

Jara-Díaz, S. R. & Guevara, C. A., 2003. Behind the subjective value of travel time savings. Journal of Transport Economics and Policy, 37(1), pp. 29-46.

Jones, K. T. & Madsen, D. B., 1989. Calculating the cost of resource transportation: a Great Basin example. Current Anthropology, 30(4), pp. 529-534.

Jouvenel, B. d., 1967. The art of conjecture. New York: Basic Books.

Kahneman, D., 2011. Thinking fast and slow. London: Penguin Books.

Kahn, H. & Wiener, A. J., 1967. The year 2000: a framework for speculation on the next thirty-three years. London: Macmillan.

Kaplan, A., 1964. The conduct of inquiry: methodology for behavioral science. San Francisco: Chandler.

Karekla, X. & Tyler, N., 2012. Reduced dwell times resulting from train–platform improvements: the costs and benefits of improving passenger accessibility to metro trains. Transportation Planning and Technology, 35(5), pp. 525-543.

Kellett, J. R., 1969. The impact of railways on Victorian cities. Abingdon: Routledge.

Kelly, R. L., 1983. Hunter-gatherer mobility strategies. Journal of Anthropological Research, 39(3), pp. 277-306.

Kennedy, C. A. et al., 2015. Energy and material flows of megacities. Proceedings of the National Academy of Sciences of the United States of America, 112(19), pp. 5985-5990.

Kenworthy, J. R., 2003. Transport energy use and greenhouse gases in urban passenger transport systems: a study of 84 global cities. International Sustainability Conference, 17-19 September.pp. 1-28.

Kikuchi, D., 2017. Tokyo plans new effort to ease commuter hell on rush-hour trains. [Online]
Available at: https://www.japantimes.co.jp/news/2017/07/06/national/tokyo-plans-new-effort-ease-crowding-rush-hour-trains/
[Accessed 2 April 2018].

Kobayashi, T., Morofuji, E. & Watanabe, Y., 2011. Sleeping Time Keeps Decreasing, Male Housework Time is Increasing from the NHK Japanese Time Use Survey, Tokyo: NHK Broadcasting Culture Research Institute.

Koen, B. V., 1985. Definition of the engineering method. Washington, D.C.: American Society for Engineering Education.

Kossiakoff, A., Sweet, W. N., Seymour, S. J. & Biemer, S. M., 2011. Systems engineering principles and practice. 2nd ed. Hoboken, New Jersey: John Wiley & Sons.

Kuhn, T., 1962. The structure of scientific revolutions. Chicago: The University of Chicago Press.

Lakatos, I., 1976. Proofs and refutations: the logic of mathematical discovery. Cambridge: Cambridge University Press.

Lakatos, I., 1978. The methodology of scientific research programmes: philosophical papers - volume 1. Cambridge: Cambridge University Press.

Lang, R. & Knox, P., 2009. The New Metropolis: Rethinking Megalopolis. Regional Studies, 43(6), pp. 789-802.

Laube, F. B., Kenworthy, J. R. & Zeibots, M. E., 1999. Towards a science of cities: city observation and a formulation of a city theory. In: M. Hesse, ed. Settlement structures, spatial mobility and traffic: on the road to sustainability in urban regions. Erkner: s.n., pp. 99-118.

Lee, Y., Shariat, S. & Choi, K., 2014. Optimizing Skip-Stop Rail Transit Stopping Strategy using a Genetic Algorithm. Journal of Public Transportation, 17(2), pp. 135-164.

Ligtvoet, W. & Hilderink, H., 2014. Towards a world of cities in 2050 – an outlook on water-related challenges. Background report to the UN-Habitat Global Report, The Hague: PBL Netherlands Environmental Assessment Agency.

Lim, G., 2008. Passenger load on train: what's acceptable and what's not. [Online]
Available at: https://www.lta.gov.sg/apps/news/page.aspx?c=3&id=84
[Accessed 06 April 2018].

Lin, S. & Kernighan, B. W., 1973. An efective heuristic algorithm for the traveling-salesman problem. Operations Research, 21(2), pp. 498-516.

Linstone, H. A., 1969. When is a need a need? The problem of normative forecasting in a changing environment. Technological Forecasting, 1(1), pp. 55-71.

Linstone, H. A. & Turoff, M., 1975. The Delphi method: techniques and applications. London: Addison-Wesley.

London Underground. 2007. Rolling Stock Information Sheet. London: Transport for London.

Loo, B., Chen, C. & Chan, C., 2010. Rail-based transit-oriented development: lessons from New York City and Hong Kong. Landscape and Urban Planning, 97, pp. 202-212.

Lucas, K., Mattioli, G., Verlinghieri, E. & Guzman, A., 2016. Transport poverty and its adverse social conequences. Transport, 169(6), pp. 353-365.

Lyons, G., Jain, J. & Holley, D., 2007. The use of travel time by rail passengers in Great Britain. Transportation Research Part A: Policy and Practice, 41(1), pp. 107-120.

Lyons, G. & Urry, J., 2005. Travel time use in the information age. Transportation Research Part A: Policy and Practice, 39(2-3), pp. 257-276.

Mackie, P. J., Sara-Díaz, S. & Fowkes, A. S., 2001. The value of travel time savings in evaluation. Transportation Research Part E: Logistics and Transportation Review, 37(2-3), pp. 91-106.

Marchetti, C., 1994. Anthropological invariants in travel behaviour. Technological Forecasting and Social Change, Volume 47, pp. 75-88.

Marien, M., 2002. Futures studies in the 21st Century: a reality-based view. Futures, 34(3), pp. 261-281.

Martens, K., 2006. Basing transport planning on principles of social justice. Berkeley Planning Journal, 19(1), pp. 1-17.

Martino, J. P., 1969. Technological forecasting and the autonomy of technology. Technological Forecasting, 1(1), pp. 73-81.

Masini, E., 2006. Rethinking futures studies. Futures, 38, pp. 1158-1168.

Mattos, L. S. et al., 2016. MagLev-Cobra Operational Tests. IEEE Transactions on Applied Superconductivity, 26(3), pp. 1-4.

McCahill, C. & Garrick, N., 2012. Automobile use and land consumption: empirical evidence from 12 cities. Urban Design International, 17(3), pp. 221-227.

McDowall, W., 2012. Technology roadmaps for transition management: the case of hydrogen energy. Technological Forecasting and Social Change, 79, pp. 530-542.

McKenzie, B. & Rapino, M., 2011. Commuting in the United States: 2009, Washington, DC: U.S. Census Bureau.

McKinsey Global Institute, 2011. Urban world: mapping the economic power of cities, New York, USA: McKinsey & Company.

McNulty, R., 2011. Realising the potential of rail in Great Britain - final independent report of the rail value for money study, London: Department for Transport.

Meadows, D., 2008. Thinking in systems: a primer. Hartford, USA: Chelsea Green Publishing.

Meadows, D. H., Meadows, D., Randers, J. & Behrens III, W. W., 1972. The limits to growth: a report for the Club of Rome's project on the predicament of mankind. New York: Universe Books.

Melander, L., 2018. Scenario development in transport studies: methodological considerations and reflections on delphi studies. Futures, 96, pp. 68-78.

Metrô/SP, 2007. Pesquisa origem destino 2007, São Paulo, Brazil: Companhia do Metropolitano de São Paulo.

Metro/SP, 2013. Pesquisa de mobilidade 2012 - região metropolitana de São Paulo, São Paulo: Cia do Metropolitano de São Paulo.

Metro/SP, 2018. Portal de Transparência. [Online]
Available at: https://transparencia.metrosp.com.br
[Accessed 05 April 2018].

Michalewcz, Z. & Fogel, D. B., 2004. How to solve it: modern heuristics. Berlin: Springer-Verlag.

Mizutani, K., 2016. Series E235 Rolling Stock – new EMU commuter train in Tokyo area. Graz, Schienenfahrzeugtagung.

Moavenzadeh, F. & Markow, M. J., 2007. Moving Millions: Transport Strategies for Sustainable Development in Megacities. Dordrecht: Springer.

Mobilize Brasil, 2017. Lançado mapa das distâncias a pé entre estações do metrô de SP. [Online]
Available at: http://www.mobilize.org.br/noticias/10424/lancado-mapa-das-distancias-a-pe-entre-estacoes-do-metro-de-sp.html
[Accessed 09 April 2018].

Mohammad, S., Graham, D., Melo, P. & Anderson, R., 2013. A meta-analysis of the impact of rail projects on land and property values. Transportation Research Part A: Policy and Practice, 50, pp. 158-170.

Mokhtarian, P. L., 2003. Telecommunications and travel: the case for complementarity. Journal of Industrial Ecology, 6(2), pp. 43-57.

Mokhtarian, P. L. & Chen, C., 2004. TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. Transportation Research Part A, 38(9-10), pp. 643-675.

Morecroft, J. D., 1982. A critical review of diagraming tools for conceptualizing feedback system models. Dynamica, 8(1), pp. 20-29.

Moriarty, P. & Honnery, D., 2008. Low mobility: the future of transport. Futures, 40(10), pp. 865-872.

Morio, J., 2011. Global and local sensitivity analysis methods for a physical system. European Journal of Physics, 32(6), p. 1577–1583.

Moss, M. & Qing, C., 2012. The Emergence of the Super-Commuter, New York: Rudin Center for Transportation.

Müller, R., 2009. The notion of a model: a historical overview. In: A. Meijers, ed. Phylosophy of Technology and Engineering Sciences. Burlington, MA: Elsevier, pp. 637-664.

Mumford, L., 1961. The City in History: Its Origins, Its Transformations, and Its Prospects. San Diego: Harcourt, Brace and World.

Musk, E., 2013. Hyperloop Alpha, Hawthorne: SpaceX.

Nakamura, H., 1998. Analysis of minimum train headway on moving block system by genetic algorithm. Transactions on the Built Environment, 34, pp. 1013-1022.

Nardini, G. et al., 2018. Cellular-V2X communications for platooning: design and evaluation. Sensors, 18(5), pp. 1527-1549.

Neto, V. & Villac, M. I., 2013. Habitação de Interesse Social (HIS) como instrumento de construção da cidade e inclusão social. [Online]
Available at: http://www.vitruvius.com.br/revistas/read/arquitextos/14.162/4964
[Accessed 02 April 2018].

Neuvonen, A. et al., 2014. Low-carbon futures and sustainable lifestyles: a backcasting scenario approach. Futures, 58(2014), pp. 66-76.

Newman, P. & Kenworthy, J. R., 1991. Transport and urban form in thirty-two of the word's principal cities. Transport Reviews: A Transnational Transdisciplinary Journal, 11(3), pp. 249-272.

Niiniluoto, I., 2001. Futures studies: science or art?. Futures, 33(5), p. 371–377.

Norman, D. A. & Verganti, R., 2014. Incremental and radical innovation: design research vs. technology and meaning change. Design Issues, 30(1), pp. 78-96.

Office for National Statistics, 2014. Annual estimates of employees from the Business Register and Employment Survey (BRES) 2013, London: Office for National Statistics.

Office for National Statistics, 2014. Commuting and Personal Well-being, London: Office for National Statistics.

Orlikowski, W. J., 1993. CASE tools as organizational change: investigating incremental and radical changes in systems development. MIS Quarterly, 17(3), pp. 309-340.

Ormrod, J. E., 2007. Educational psychology: developing learners. London: Pearson.

Owen, A. & Levinson, D., 2014. Access across America: transit 2014, University of Minnesota: University of Minnesota.

Pachi, A. & Ji, T., 2005. Frequency and velocity of people walking. The Structural Engineer, 83(3), pp. 36-40.

Pagliara, F., Silva, J. A. & Chen, G., 2015. Investigating the role of High Speed Rail in shaping metro-regions. Lisbon, Portugal, European Regional Science Association.

Pai, M., 2009. What Is super dense crush load?. [Online]
Available at: http://thecityfix.com/blog/what-is-super-dense-crush-load/
[Accessed 06 April 2018].

Parkinson, T. & Fisher, I., 1996. Rail transit capacity. Washington, D.C.: Transportation Research Board.

Perez, C., 1983. Structural change and assimilation of new technologies in the economic and social systems. Futures, 15(5), pp. 357-375.

Phaal, R., Farrukh, C. J. & Probert, D. R., 2004. Technology roadmapping - a planning framework for evolution and revolution. Technological Forecasting and Social Change, 71, pp. 5-26.

Polya, G., 1945. How to solve it: a new aspect of mathematical method. New Jersey: Princeton.

Popper, K., 1959. Logic of scientific discovery. Camberley: Hutchinson & Co.

Popper, K. R., 1961. The poverty of historicism. London: Routledge.

Pourbaix, J., Steriu, M. & Saeidizand, P., 2015. Mobility in cities database: synthesis report, Brussels: International Association of Public Transport.

Powell, J. P. & Palacín, R., 2015. Passenger stability within moving railway vehicles: limits on maximum longitudinal acceleration. Urban Rail Transit, 1(2), pp. 95-103.

Pugh, S., 1991. Total design integrated methods for successful product engineering. Wokingham: Addison-Wesley Publishing Company.

Pugh, S., 1996. Creating innovative products using total design. Reading, Massachussetts: Addison-Wesley Publishing Company.

Quist, J., Thissen, W. & Vergragt, P. J., 2011. The impact and spin-off of participatory backcasting: from vision to niche. Technological Forecasting and Social Change, 78(5), pp. 883-897.

Rail Technology Magazine, 2013. Interior refurbishment for London Midland Class 150s. [Online]
Available at: http://www.railtechnologymagazine.com/Rolling-stock/interior-refurbishment-for-london-midland-class-150s
[Accessed 02 April 2018].

Reichenbach, H., 1938. Experience and prediction: an analysis of the foundations and the structure of knowledge. Chicago: University of Chicago Press.

Rigby, R., 2011. Business traveller: the rise of super-commuters. [Online]
Available at: http://www.ft.com/cms/s/0/2375423a-2c1d-11e1-b194-00144feabdc0.html#axzz4JNLZYFCf

Rinne, M., 2004. Technology roadmaps: infrastructure for innovation. Technological Forecasting and Social Change, Volume 71, pp. 67-80.

Roberts, E. B., 1969. Exploratory and normative technological forecasting: a critical appraisal. Technological Forecasting, 1(1), pp. 113-127.

Robinson, J. B., 1988. Unlearning and backcasting: rethinking some of the questions we ask about the future. Technological Forecasting and Social Change, 33(4), pp. 325-338.

Robinson, J. B., 1988. Unlearning and backcasting: rethinking some of the questions we ask about the future. Technological Forecasting and Social Change, 33(4), pp. 325-338.

Robinson, J. B., 1990. Futures under glass: a recipe for people who hate to predict. Futures, 22(8), pp. 820-842.

Robinson, J. et al., 2011. Envisioning sustainability: recent progress in the use of participatory backcasting approaches for sustainability research. Technological Forecasting and Social Change, 78(5), pp. 756-768.

Robinson, T., Chan, E. & Coelingh, E., 2010. Operating platoons on publc motorways: an introduction to the SARTRE platooning programme. Busan, South Korea.

Ropohl, G., 1999. Techné: Research in Philosophy and Technology, 4(3), pp. 186-194.

Roth, G. J. & Zahavi, Y., 1981. Travel time “budgets” in developing countries. Transportation Research Part A, 15(1), pp. 87-95.

Royal Society for Public Health, 2016. Health in a hurry: the impact of rush hour commuting on our health and wellbeing, London: Royal Society for Public Health.

Royce, W. W., 1970. Managing the development of large software systems: concepts and techniques. Proceedings of the IEEE WesCon, pp. 1-9.

RSSB, 2004. Guidance on gauging. London: Rail Safety and Standards Board.

RSSB, 2007. Track system requirements, London: Rail Safety and Standards Board.

Rutkin, A., 2016. Hyperloop's first flight. New Scientist, 230(3074), pp. 20-21.

Saltelli, A. et al., 2008. Global sensitivity analysis. the primer. Chichester: John Wiley and Sons.

Samaniego, H. & Moses, M. E., 2008. Cities as Organisms: Allometric Scaling of Urban Road Networks. Journal of Transport and Land Use, 1(1), pp. 21-39.

Saviotti, P. P., 1986. Systems theory and technological change. Futures, 18(6), pp. 773-786.

Schaeffer, K. H. & Sclar, E., 1980. Access for All: Transportation and Urban Growth. New York: Colmbia University Press.

Schäfer, A., 2000. Regularities in travel demand: an international perspective. Journal of Transportation and Statistics, 3(3), pp. 1-31.

Schickore, J., 2014. Scientific Discovery. [Online]
Available at: https://plato.stanford.edu/archives/sum2018/entries/scientific-discovery/
[Accessed 29 01 2018].

Schmidt, R. F., 1993. IEEE P1220-standard for system engineering-a commercial standard for improving competitiveness. Fort Worth, TX, USA, AIAA/IEEE.

S-CODE, 2017. High Level Architecture Design Document, Birmingham: European Comission.

Siemens, 2012. Siemens to supply new metro for Munich. Berlin, InnoTrans, September 18–21.

Simon, H., 1981. Sciences of the artificial. Cambridge: MIT Press.

Smith, S., 2017. Victoria line trains set to arrive every 100 seconds from May. [Online]
Available at: https://www.standard.co.uk/news/transport/victoria-line-trains-set-to-arrive-every-100-seconds-from-may-a3461636.html
[Accessed 05 07 2018].

Snowden, D., 2002. Complex acts of knowing: paradox and descriptive self-awareness. Journal of Knowledge Management, 6(2), pp. 100-111.

Sommer, R. M., 2012. Mobility as an infratructure of democracy. In: Sustainable urbanism and beyond: rethinking cities for the future. New York: Rizzoli.

Sorensen, A. & Okata, J., 2011. Megacities: urban form, governance and sustainability. Tokyo: Springer.

Spinoza, B., 1996. Ethics. London: Penguin Books.

Steer Davies Gleave, 2002. The Case for Rail, London: Steer Davies Gleave.

Steg, L., 2005. Car use: lust and must. Instrumental, symbolic, and affective motives for car use. Transportation Research Part A, Volume 39, pp. 147-162.

Strano, E., Shai, S., Dobson, S. & Barthelemy, M., 2015. Multiplex networks in metropolitan areas: generic features and local effects. Journal of the Royal Society Interface, 12(111), pp. 1-9.

Suddendorf, T., 2013. The gap: the science of what separates us from other animals. New York: Basic Books.

Suddendorf, T. & Corballis, M. C., 2007. The evolution of foresight: What is mental time travel, and is it unique to humans?. Beharioural and Brian Sciences, 30(3), pp. 299-313.

Takagi, R., 2014. High speed rail – a thought on capacity. Birmingham, University of Birmingham.

Tank, T. & Linnarts, J. M., 1997. Vehicle-to-vehicle communication for AVCS platooning. IEEE Transactions on Vehicular Technology, 46(2), pp. 528-536.

Taylor, S. & Green, O., 2001. The moving metropolis: a history of London’s transport since 1800. London: Laurence King.

Tero, A. et al., 2010. Rules for Biologically Inspired Adaptive Network Design. Science, 327(5964), pp. 439-442.

Tesla, N., 2010. My inventions: the autobiography of Nikola Tesla. New York: Cosimo Classics.

Tetlock, P. & Gardner, D., 2016. Superforecasting: the art and science of prediction. London: Random House Books.

Thoreau, R. et al., 2016. Train design features affecting boarding and alighting of passengers. Journal of Advanced Transportation, 50(8), pp. 2077-2088.

Timilsina, G. R. & Dulal, H. B., 2010. Urban road transportation externalities: costs and choice of policy instruments. The World Bank Research Observer, 26(1), p. 162–191.

Tirachini, A., Hensher, D. A. & Rose, J. M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transportation Research Part A, 53, pp. 36-52.

Tirachini, A., Sun, L., Erath, A. & Chakirov, A., 2016. Valuation of sitting and standing in metro trains using revealed preferences. Transport Policy, 47, pp. 94-104.

Tough, J. & O´Flaherty, C., 1971. Passenger conveyors: an innovative form of communal transport. Surrey: Ian Allan.

Transport for London, 2008. Distance between adjacent Underground stations - FOI 1001647942, London: Transport for London.

Transport for London, 2011. Journey demand and service supply breakdown. FOI-0611-1011, London: Transport for London.

Transport for London, 2013. Gate-to-platform and interchange walking times. FOI-1216-1112, London: Transport for London.

Transport for London, 2017a. Travel in London - report 10. London: Transport for London.

Transport for London, 2017b. Review of the TfL WiFi pilot. London: Transport for London.

Transport for London, 2018. Facts and figures. [Online]
Available at: https://tfl.gov.uk/corporate/about-tfl/what-we-do/london-underground/facts-and-figures
[Accessed 05 April 2018].

Trigg, T., 2015. Cities Where It's Faster to Walk than Drive. [Online]
Available at: https://blogs.scientificamerican.com/plugged-in/cities-where-it-s-faster-to-walk-than-drive/
[Accessed 22 01 2018].

Tsiolkovsky, K. E., 1903. The exploration of cosmic space by means of reaction devices. Scientific Review, 5, pp. 45-75.

Tversky, A. & Kahneman, D., 1974. Judgement under uncertainty: heuristics and biases. Science, 185(4157), pp. 1124-1131.

UITP, 2016. World report on metro automation - July 2016, Brussels: International Association of Public Transport (UITP).

UN-DESA, 2014. World urbanization prospects: the 2014 revision, New York: The United Nations Department of Economic and Social Affairs.

UN-HABITAT, 2012. Sustainable Urban Energy: A Sourcebook for Asia, Nairobi, Kenya: United Nations Human Settlements Programme (UN-HABITAT).

Van Wee, B., Rietveld, P. & Meurs, H., 2006. Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time. Journal of Transport Geography, 14, pp. 109-122.

von Bertalanffy, L., 1967. General System theory: foundations, development, applications. New York: George Braziller.

von Kármán, T., 1946. Future airborne armies. Dayton: Army Air Forces Scientific Advisory Group.

Vuchic, V., 2007a. Urban transit: operations, planning and economics. Hoboken: John Wiley & Sons.

Vuchic, V., 2007b. Urban Transit: systems and technology. Hoboken: John Wiley & Sons.

Walsh, C. L. et al., 2015. Are wildcard events on infrastructure systems opportunities for transformational change?. Futures, 67, pp. 1-10.

Wardman, M., 2004. Public transport values of time. Transport Policy, 11(4), pp. 363-377.

Weaver, W., 1948. Science and complexity. American Scientist, 36(536), pp. 536-544.

Weiss, P. A., 1969. The living system: determinism stratified. In: A. Koestler & J. R. Smythes, eds. Beyond reductionism. London: Hutchinson & Co, pp. 3-55.

Wener, R. E. & Evans, G. W., 2011. Comparing stress of car and train commuters. Transportation Research Part F: Traffic Psychology and Behaviour, 14(2), pp. 111-116.

Wertheimer, M., 1938. Laws of organization in perceptual forms. London: Kegan Paul, Trench, Trubner & Company.

West, G., 2017. Scale: the universal laws of life and death in organisms, cities and companies. London, UK: The Orion Publishing Group.

West, G. & Bettencourt, L., 2010. A unified theory of urban living. Nature, 467, pp. 912-913.

Whitehead, A. N. & Russell, B., 1927. Principia Mathematica: Volume 3. 2nd ed. Cambridge: Cambridge University Press.

White, S. M. & Rotton, J., 1998. Type of commute, behavioral aftereffects, and cardiovascular activity - a field experiment. Environment and Behavior, 30(6), pp. 763-780.

Winterhalder, B., 1981. Optimal foraging strategies and hunter-gatherer research in anthropology. In: B. Winterhalder & A. E. Smith, eds. Hunter-gatherer foraging strategies: Ethnographic and archaeological analyse. Chicago: University of Chicago Press, pp. 13-35.

World Energy Council, 2011. Global Transport Scenarios 2050, London: World Energy Council.

World Energy Council, 2013. World energy scenarios: composing energy futures to 2050, London: World Energy Council.

World Health Organisation, 2015. Registered vehicles - data by country. [Online]
Available at: http://apps.who.int/gho/data/node.main.A995
[Accessed 28 01 2018].

Yates, F. E., 1987. Self-organizing systems: the emergence of order. New York: Plenum Press.

Young, M. & Blainey, S. P., 2017. Development of railway station choice models to improve the representation of station catchments in rail demand models. Transportation Planning and Technology , 41(1), pp. 80-103.

Zahavi, Y., 1974. Traveltime budgets and mobility in urban areas, Washington, D.C.: US Department of Transport.

Zahavi, Y. & Talvitie, A., 1980. Regularities in travel time and money expenditures. Transportation Research Record, Volume 750, pp. 13-19.

Zheng, S. & Kahn, M., 2013. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceedings of the National Academy of Sciences, 110(14), pp. 1248-1253.

APPENDIX A

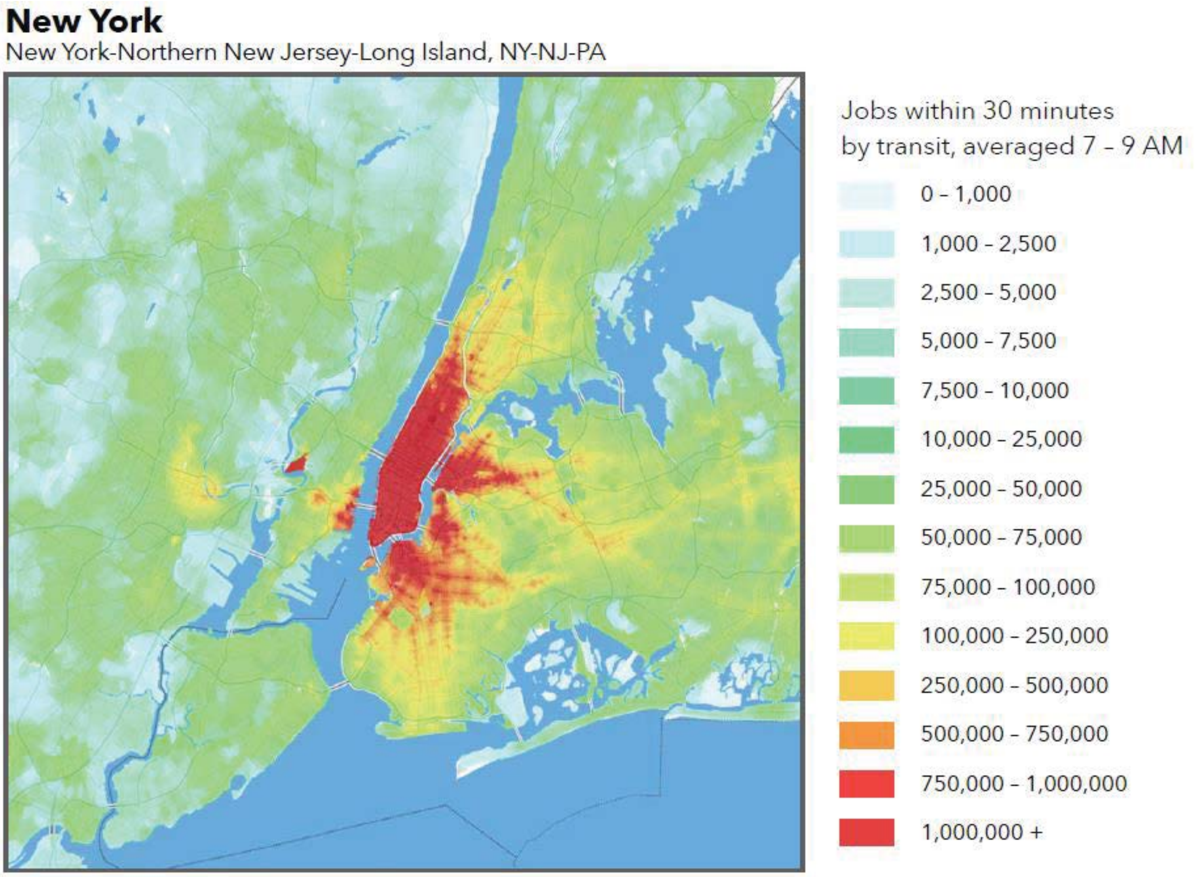


Figure 71. Jobs accessible within 30 minutes by public transport in New York

(Owen & Levinson, 2014)

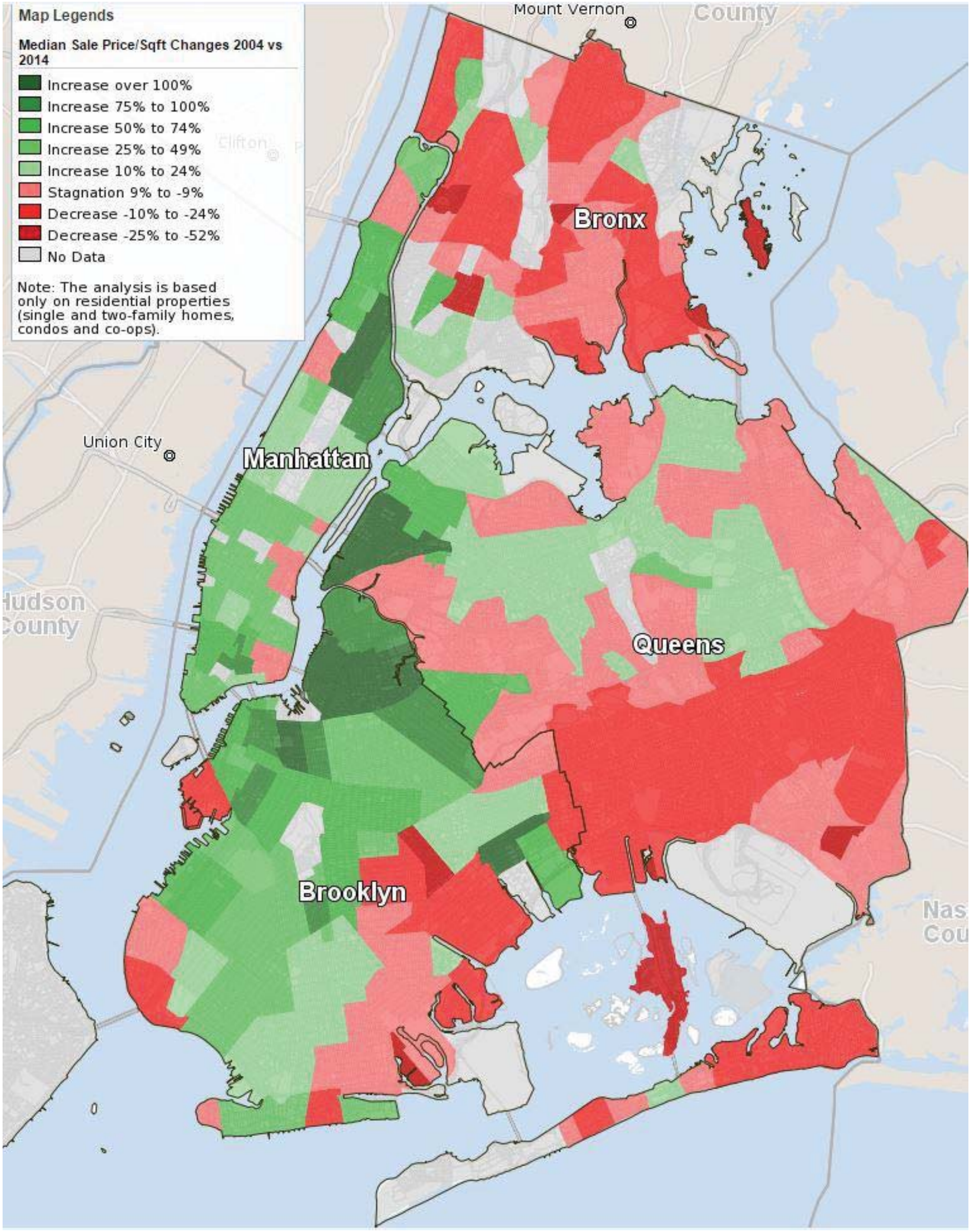


Figure 72. Residential price changes between 2004 and 2014 in New York (Baiceanu, 2015)

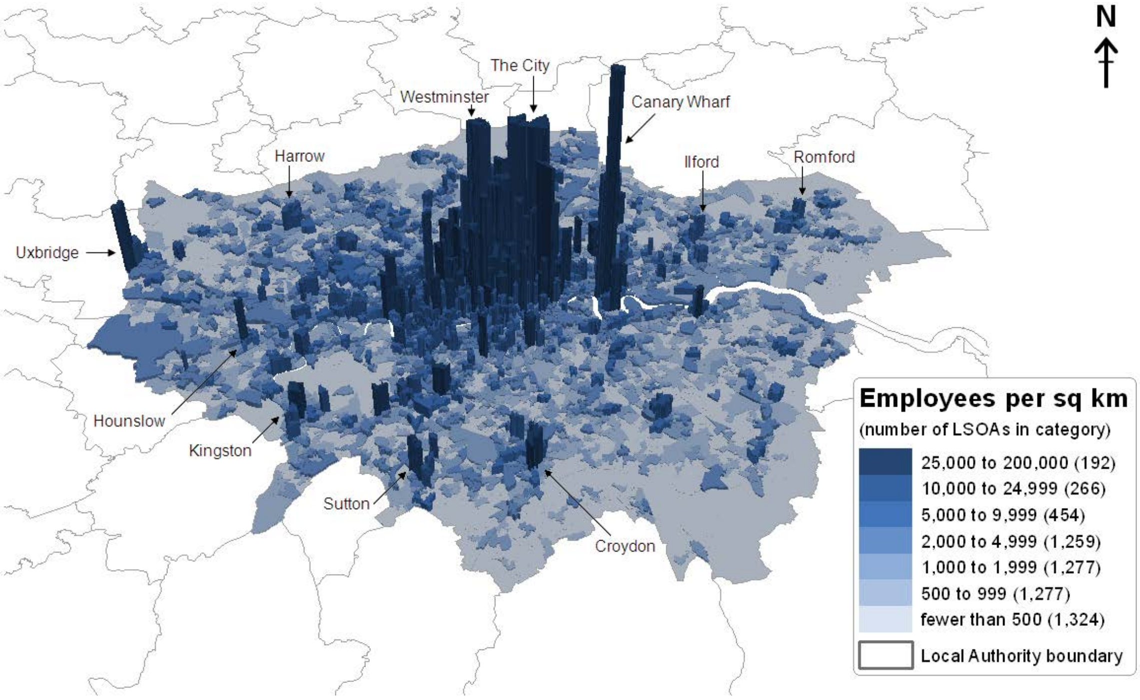


Figure 73. Number of employees per sq. km in London (Office for National Statistics, 2014)

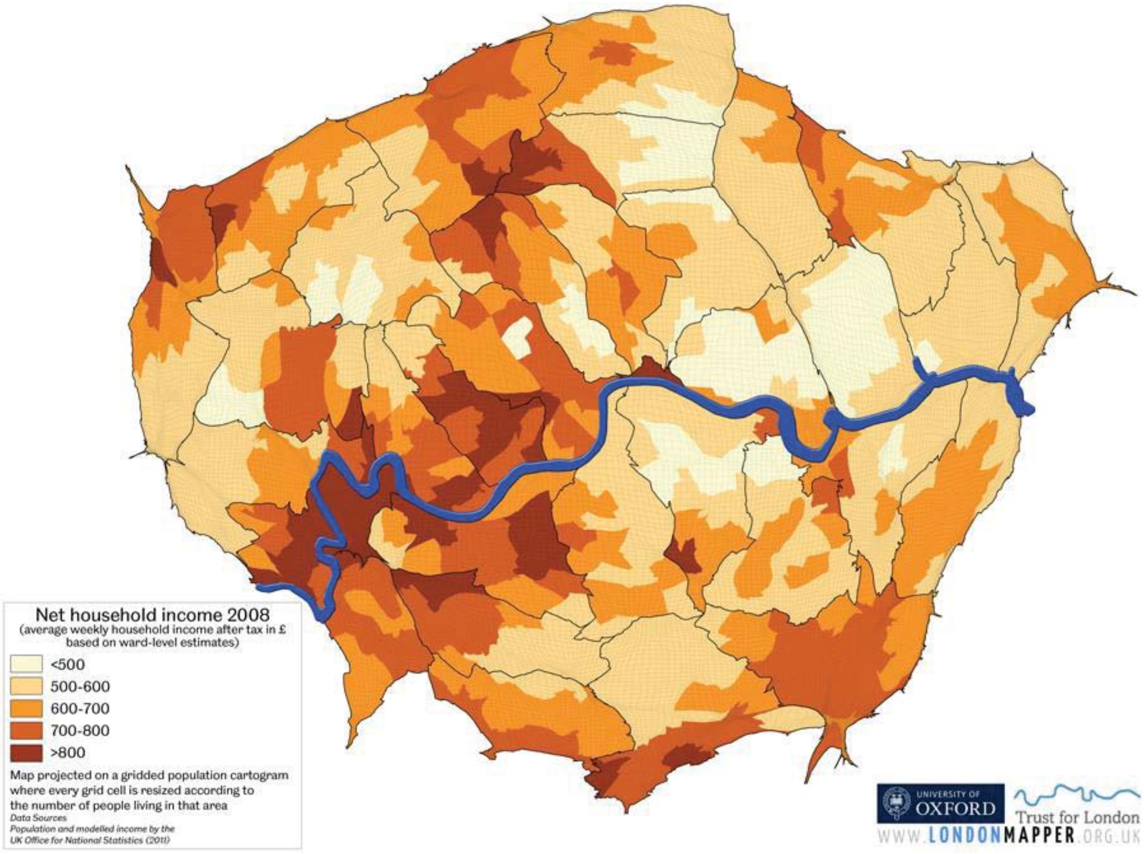


Figure 74. Average household income after tax (before housing costs) based on ward-level estimates for 2008 (Hennig & Dorling, 2014)

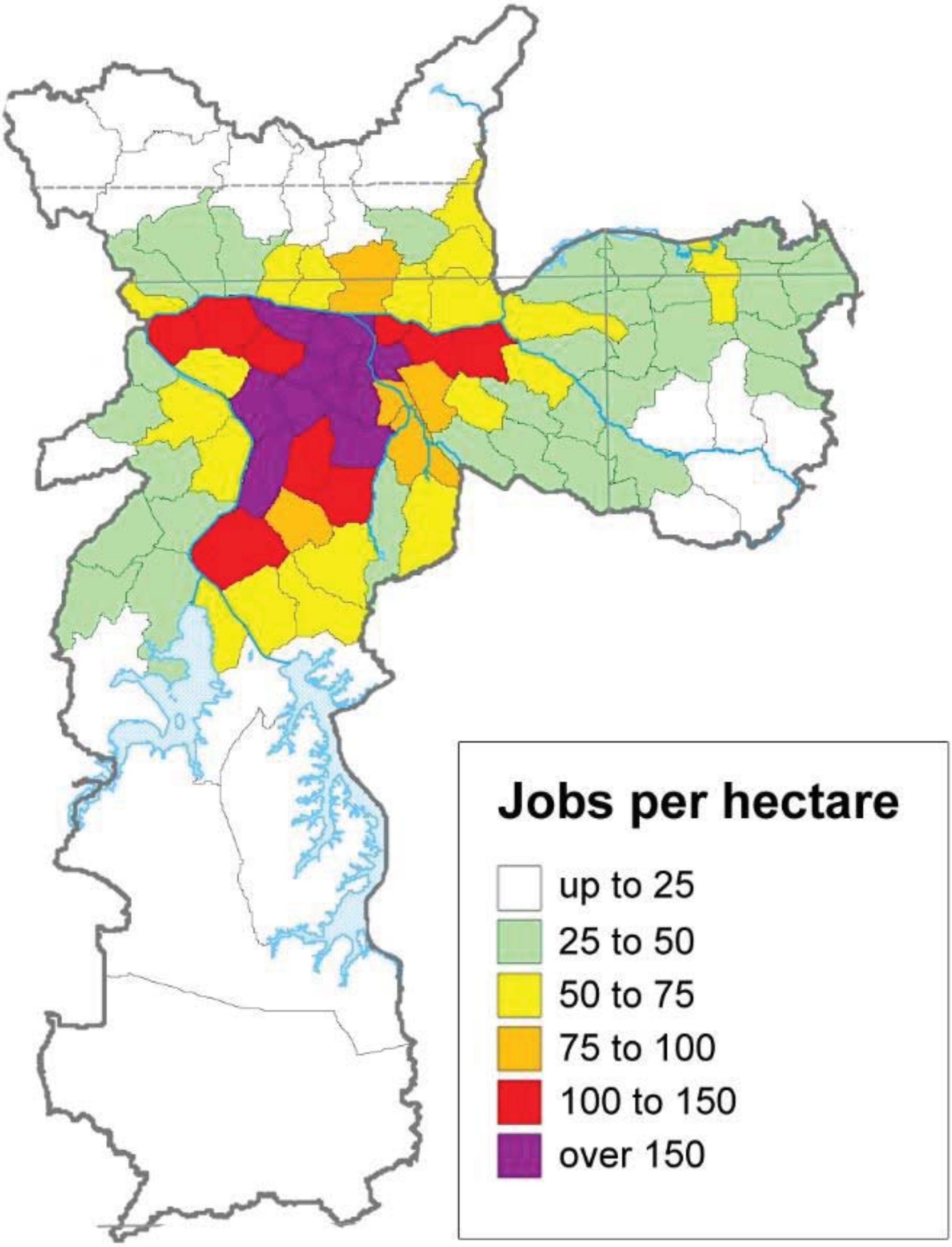


Figure 75. Density of jobs in São Paulo (Metrô/SP, 2007)

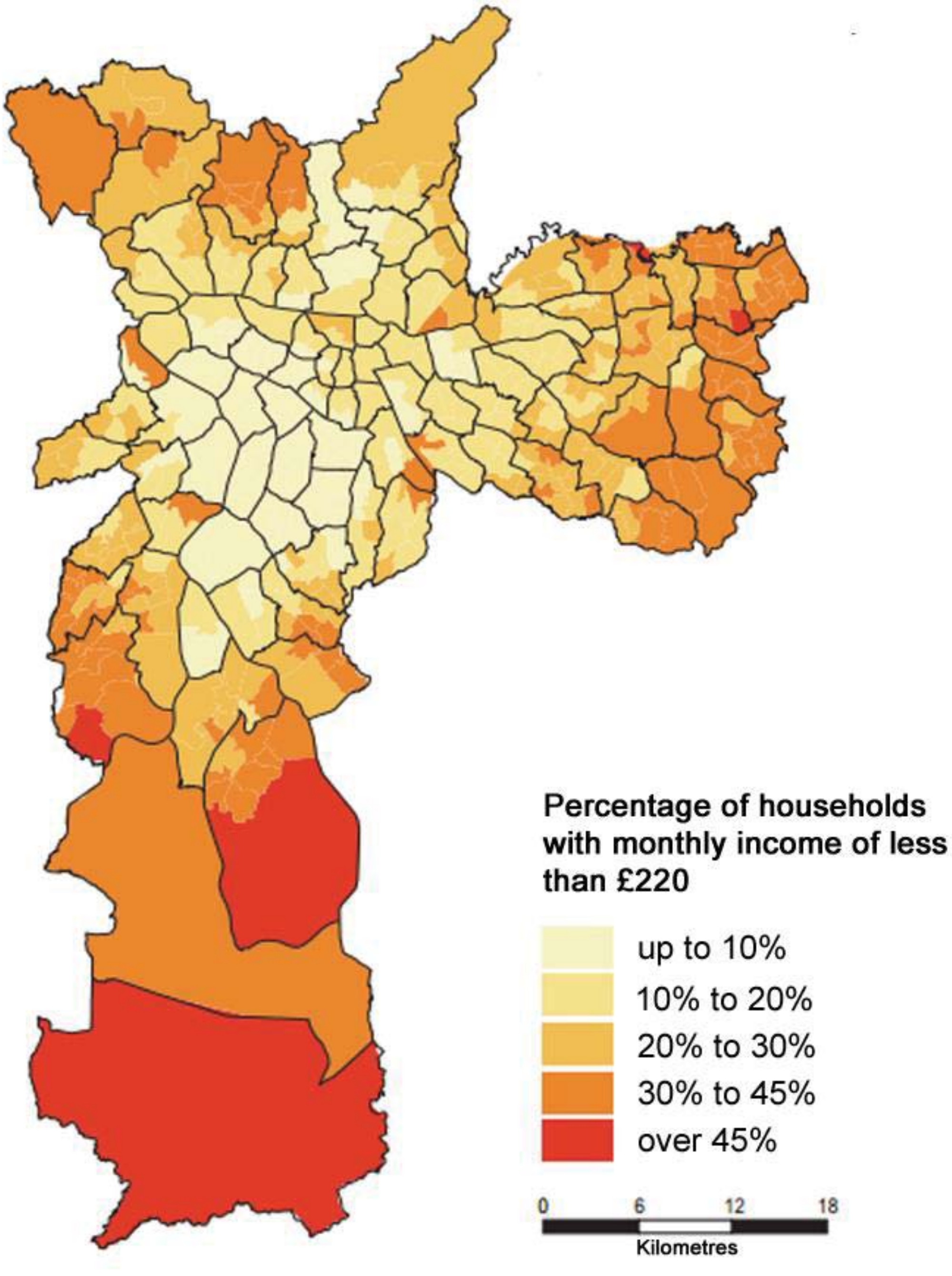


Figure 76. Percentage of households in São Paulo with a monthly salary of less than £220

(Neto & Villac, 2013)

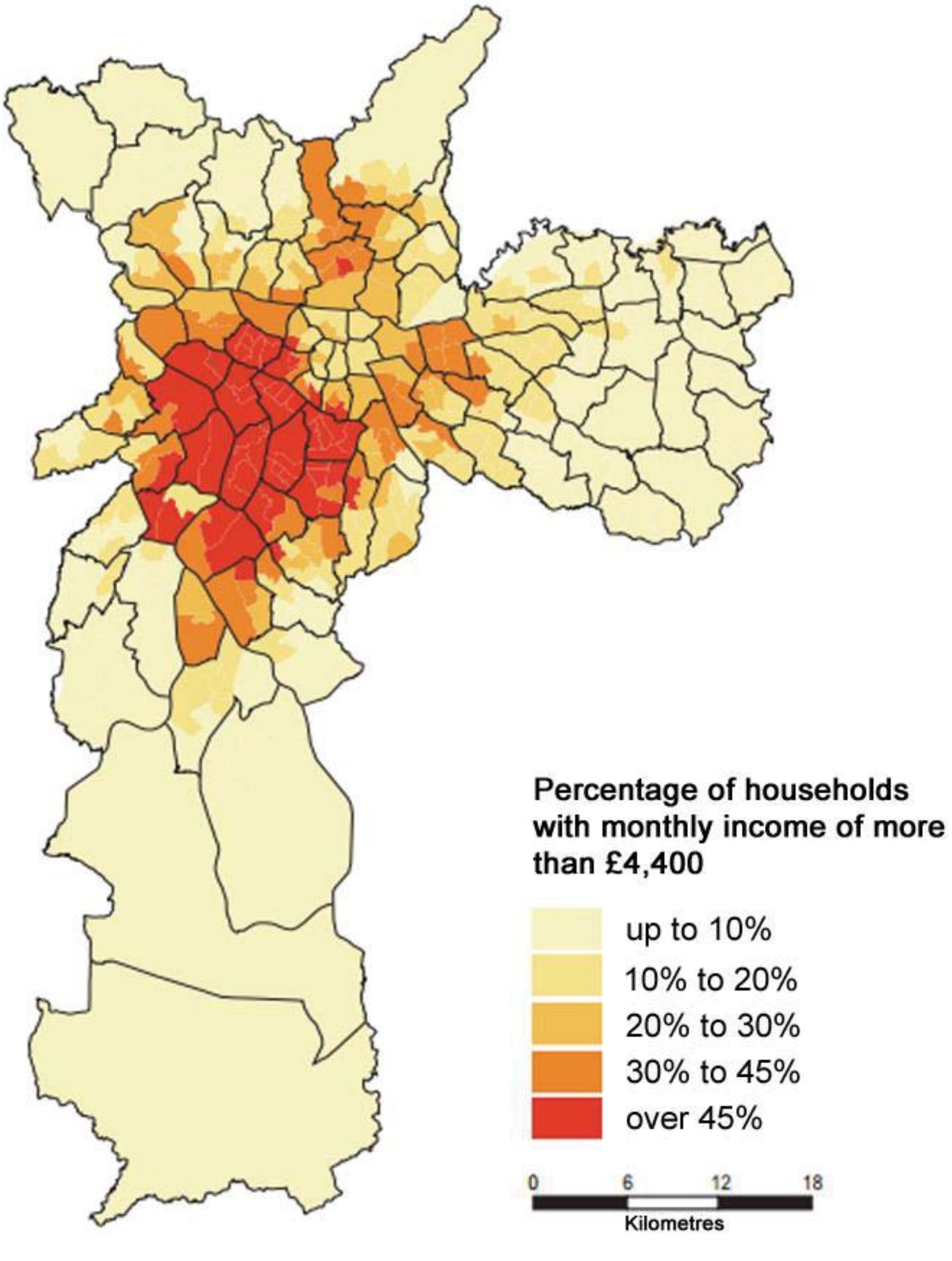


Figure 77. Percentage of households in São Paulo with a monthly salary of more than £4,400

(Neto & Villac, 2013)

APPENDIX B

ID	Requirement
1	All services shall operate only according to their assigned patterns
1.1	P1 vehicles shall stop at every station
1.2	P2 vehicles shall stop at every two stations
1.2.1	When the line is straight, P_2 vehicles shall be deployed in alternate stations when beginning daily operations
1.3	P3 vehicles shall stop at every three stations
1.4	P5 vehicles shall stop at every five stations
1.5	P7 vehicles shall stop at every seven stations
2	The number of stations shall be different to any multiple of any pattern in operation
3	Stations shall be located 685 m from each other
3.1	Stations shall be places on secondary lines at a distance of 10.8 m from the main line
3.2	Platforms shall be divided in six sections, being one for every Px in operation and one for emergencies
3.2.1	Each platform section shall be 43 m long
3.2.1.1	Each platform section shall include 9.3 m clearance area for manoeuvring on both ends, unless they are on platform ends where they shall include one clearance area on the open end.
3.2.2	Each platform section shall have its own track departing from and returning to the main line
3.3	Dwell time at platforms shall be no more than 25 seconds
3.3.1	Doors must open within 2.5 seconds
3.3.2	Doors must close within 2.5 seconds
4	Vehicles shall run autonomously at GoA 4
4 .1	Maximum speed on the line shall be limited to 72 km/h
4.2	Vehicles shall accelerate at a rate of 1.3 m/s^2
4.3	Vehicles shall brake at a rate of 1.2 m/s^2
4.4	Jerk rates shall be set at 1 m/s^3
4.5	Vehicles must maintain a safety distance of at least 24.4 metres from the preceding vehicle when they are part of the same platoon
4.5.1	On-board drive system must process and respond to data packets in 0.1 seconds

4.6 Platoon leaders must keep a minimum headway of 30.9 seconds from the preceding platoon leader

4.7 Error margin must be less or equal to 20% of vehicle length

4.8 Vehicles shall be 12 metres long

4.9 Vehicles shall be 2.6 metres wide

4.10 Vehicles shall offer capacity for at least 80 passengers at no more than 4 pax/m^2

5 Track gauge shall be standard UK of 1435 mm

5.1 Cant shall be 160 mm

5.2 Cant deficiency shall be 100 mm

APPENDIX C

1. Acceleration

Changes in acceleration do not result in changes in the theoretical capacity of the system. Nonetheless, the influence that acceleration rates have on door-to-door travel times is certainly non-negligible. Due to the quadratic nature of the equations, the sensitivity to acceleration inputs is hyperbolic. As Figure 78 shows, increasing the acceleration would not result in significant time savings, where accelerating at an impractical rate of 1.6 m/s² would only reduce travel time by 1.2%. On the other hand, reducing acceleration would have an increasingly more important influence on the median door-to-door travel time. For instance, a reduction to 0.9 m/s² would result in a 3% increase in travel times, a sensitivity of 0.098 (0.098% change in outputs for every 1% change in inputs). In a more extreme scenario, an acceleration input of 0.6 m/s² would have a sensitivity of 0.15. Acceleration rates of 1.3 m/s² are currently in use, indicating a TRL 9. Higher acceleration rates can jeopardise passenger safety and thus only reductions are in fact considered.

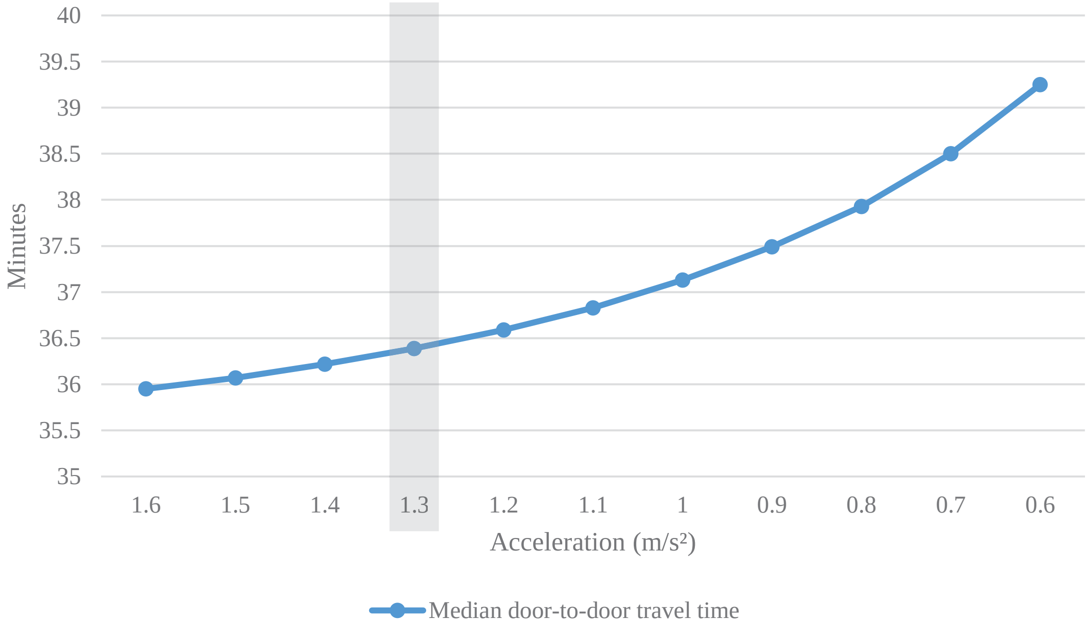


Figure 78. Influence of different acceleration inputs on median door-to-door travel times. Shaded area highlights the original value.

2. Braking

Braking inputs influence both the median door-to-door travel times and theoretical capacity of the system. Capacity is more significantly affected than travel times. In a similar way to acceleration, increases in braking rates result only in marginal gains in travel times in face of severe impacts on passenger comfort (Figure 79). A 33% increase in braking to 1.6 m/s² would result in 1.73% reduction in travel times, and 8.75% increase in capacity. Consequently, the sensitivity of the increase is 0.05 and 0.26 respectively.

On the other hand, reducing braking rates leads to a greater impact on the outputs of the system. Lowering the same 33% to 0.8 m/s² would result in a 3.68% increase in travel times and a 13.86% reduction in capacity. Sensitivity for these parameters are 0.11 and 0.41 respectively. Since high braking rates also poses risks to passenger safety, only reductions are considered.

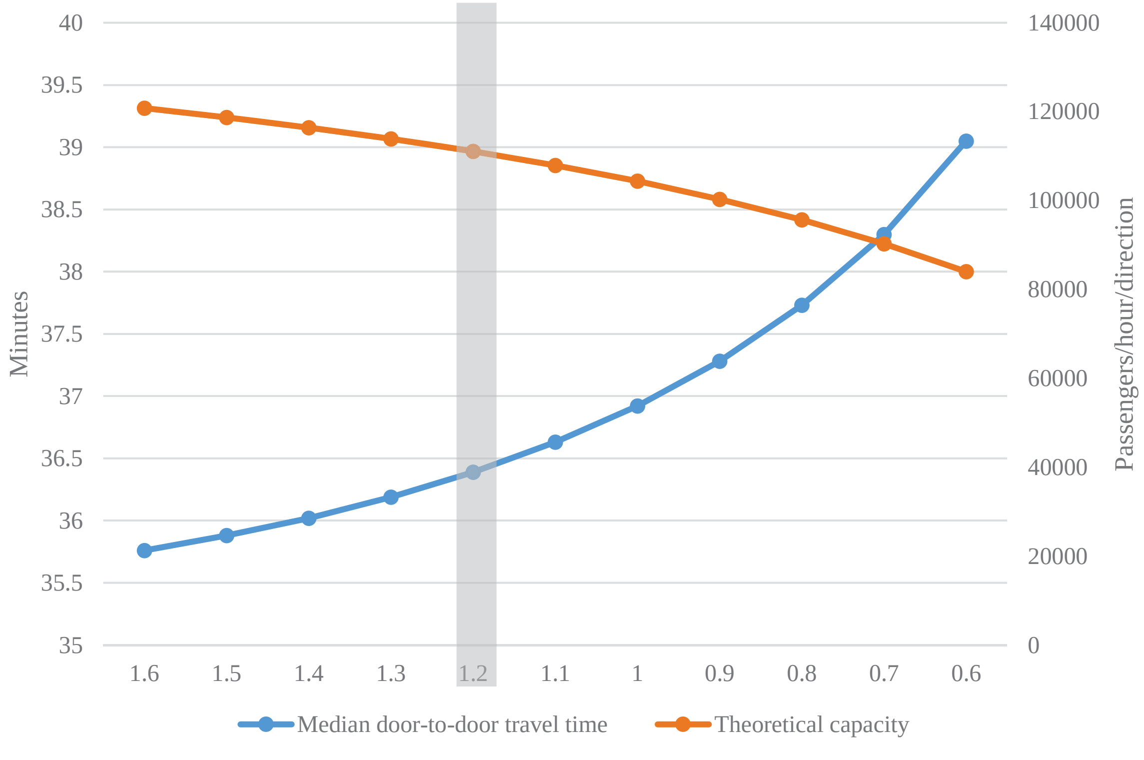


Figure 79. Influence of different braking inputs on median door-to-door travel time and capacity. Shaded area highlights the original value.

3. Jerk

The influence of different jerk inputs on system outputs are marginal, as shown on Figure 80. Increases in inputs over 1 m/s³ were not considered as they raise safety issues for standing passengers. The sensitivity of jerk inputs is of 0.03, and only influences travel times. Considering that rate of 1 m/s³ is already achievable, the choice for a lower jerk becomes mostly a decision concerning passenger comfort.

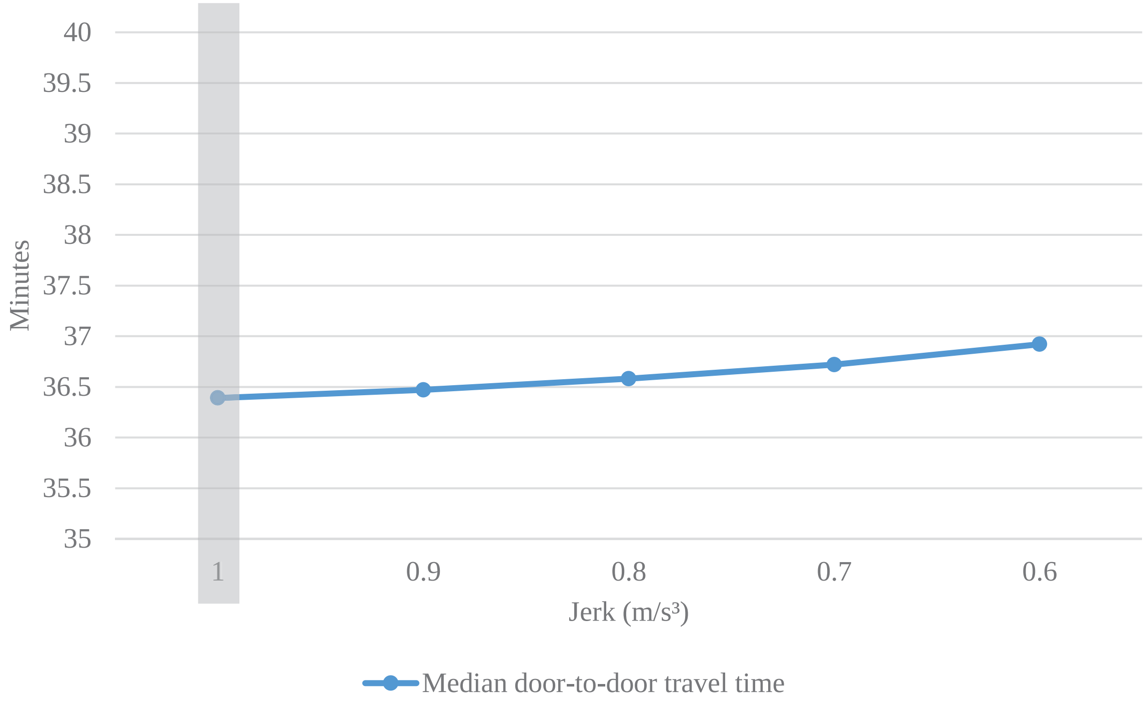


Figure 80. Influence of different jerk inputs on median door-to-door travel times. Shaded area highlights the original value.

4. Opening/closing door times

The sensitivity of door times is low on door-to-door travel times and does not apply to capacity. The reason for that is that since vehicles stop off the main line, dwell times are irrelevant to headway calculations. Consequently, the influence on door-to-door journey times is linear, where every extra second to open or close the doors would inflict two extra seconds per stop (Figure 81). A reduction of 20% in door opening/closing times to 2 seconds would result in an overall saving of 0.36% in travel times, meaning that the sensitivity of the input is 0.018.

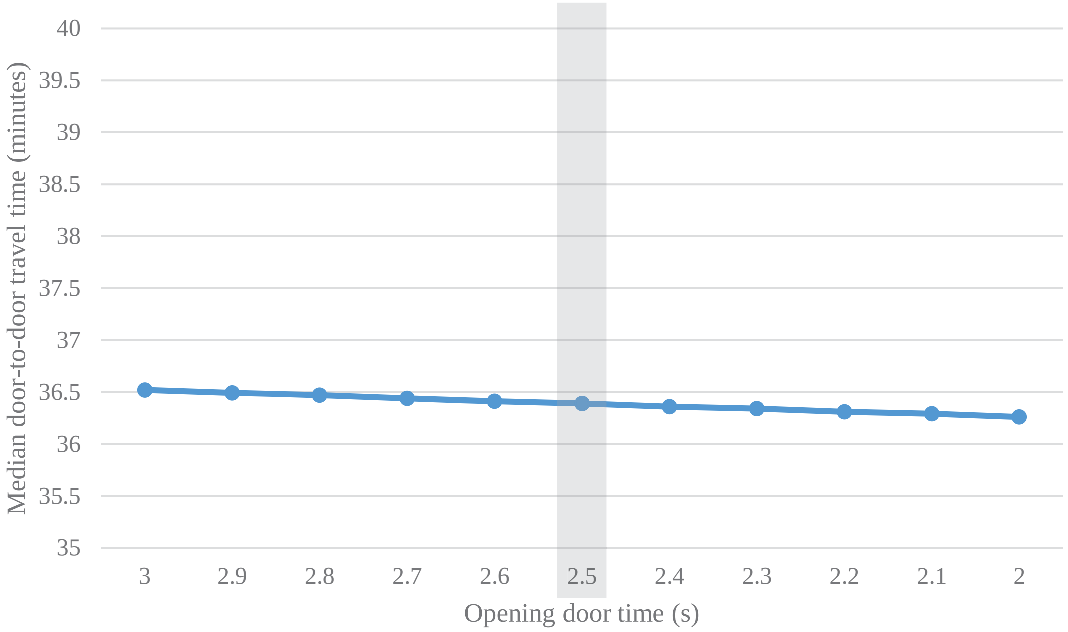


Figure 81. Influence of different door opening/closing time inputs on median door-to-door travel times. Shaded area highlights the original value.

5. Data transmission

Data transmission time has no impact on travel times but influences the theoretical capacity of the system because it affects the headway between vehicles and the headway between platoons. Sensitivity is moderate in comparison to the other non-critical parameters, yet small considering the engineering efforts required. A 60% reduction in data transmission time from 100 ms to 40 ms would result in 2.37% increase in theoretical capacity, revealing a sensitivity of 0.04 (Figure 82). However, reducing the assumption from 100 ms to 40 ms means also a change in TRL from 9 to 4.

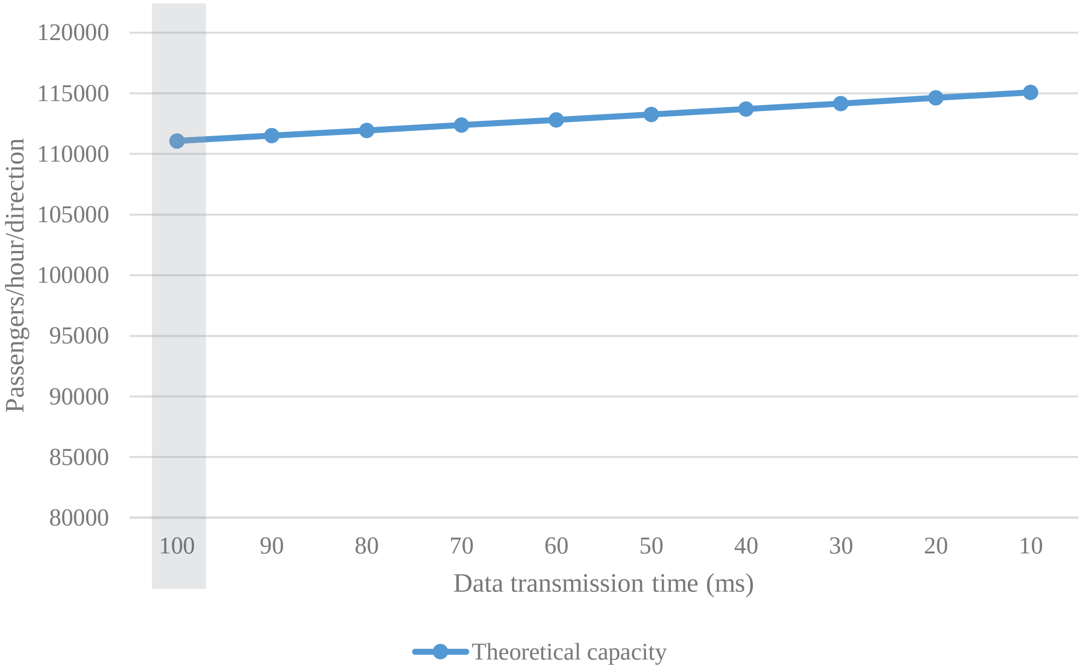


Figure 82. Influence of data transmission time inputs on theoretical capacity. Shaded area highlights the original value.

6. Vehicle position error margin

In a similar way to data transmission, error margin in vehicle position only influences the capacity as it impacts the headways used in the system. Sensitivity is moderate and linear, yet impact is low considering safety aspects and Technology Readiness Levels. For a 50% reduction in error margin (from 20% to 10% of vehicle length), there is a 2.37% increase in theoretical capacity. The resulting sensitivity is 0.047.

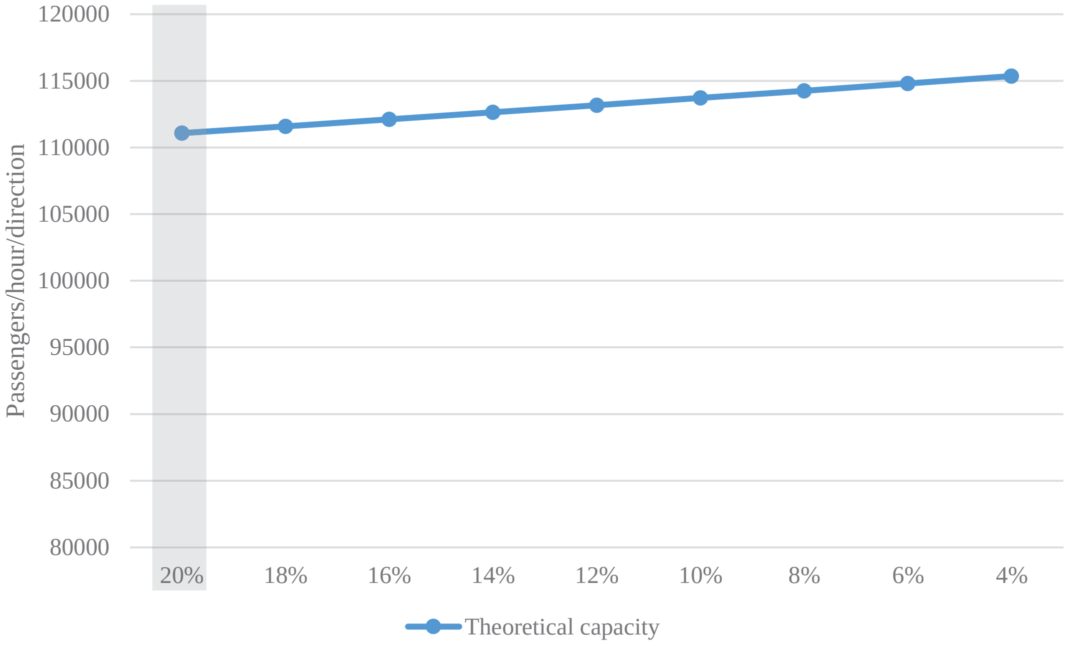


Figure 83. Influence of vehicle position error margin inputs on theoretical capacity. Shaded area highlights the original value.

7. Cant and cant deficiency

Changes in cant and cant deficiency have very little influence on travel times, and small to moderate impact on capacity. Since they influence curve radius, they influence the distance between stations and also the switch clearance distance, thus indirectly impacting travel times and capacity. Figures Figure 84 and Figure 85 show that the reduction is marginal when increasing cant and cant deficiency to 180mm and 110mm respectively. Sensitivity for cant is 0.006 for travel times and 0.042 for capacity. For cant deficiency, sensitivity is 0.005 for travel times and 0.027 for capacity. However, considering the higher values carry a TRL of 9, even marginal gains can be obtained without an increase in risk or uncertainty.

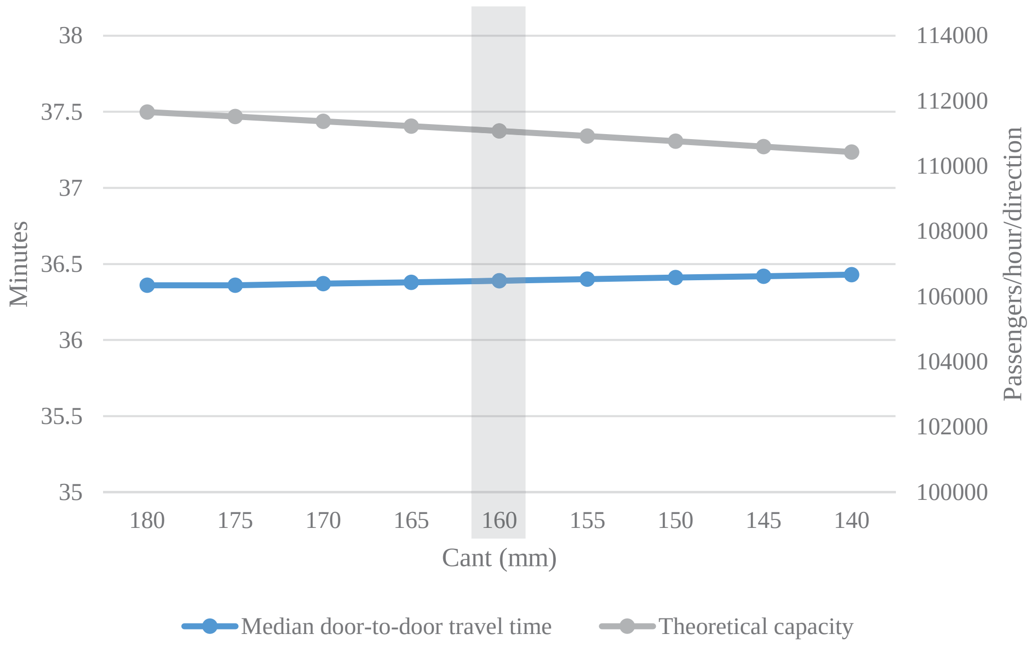


Figure 84. Influence of cant inputs on theoretical capacity. Shaded area highlights the original value.

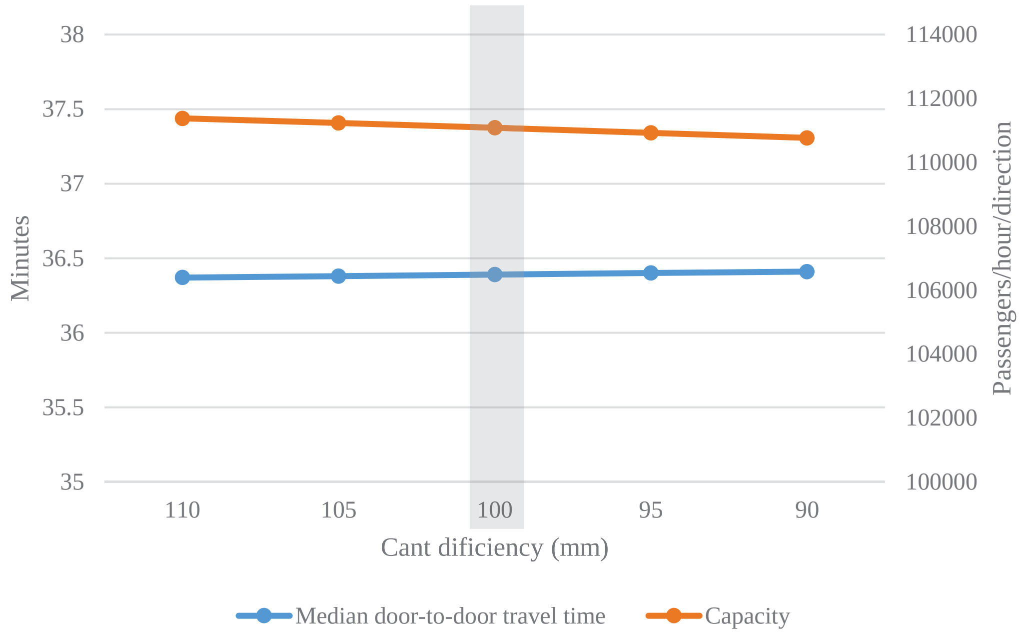


Figure 85. Influence of cant deficiency inputs on theoretical capacity. Shaded area highlights the original value.

8. Gauge

Sensitivity analysis on gauge is not to be taken as straightforward because the parameter is internationally standardised. Therefore, measuring the impact of every 1% change in distance will not necessarily produce a realistic assessment. Nonetheless, in sight of the conceptual nature of the system, it may be beneficial to study the impact of non-standard gauges in search of an optimal value that can be transformed into a requirement. Figure 86 illustrates the influence of track gauge on median door-to-door travel times and theoretical capacity. Sensitivities are 0.01 in travel times and 0.06 in capacity.

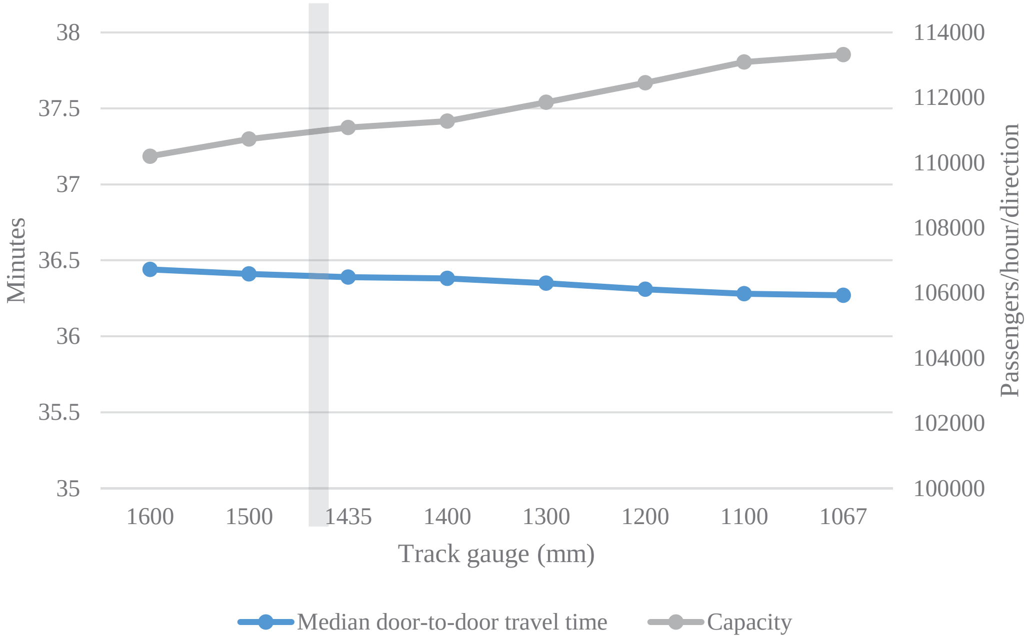


Figure 86. Influence of track gauge inputs on theoretical capacity. Shaded area highlights the original value.