TU Delft | Master Thesis

D.M.T. (Dion) Mol

The Potential Profitability of a European High-Speed Rail Network

The Potential Profitability of a European High-Speed Rail Network

by

Dion Matthew Thomas Mol

[4943694]

to obtain the degree of Master of Science (MSc)
in Transport, Infrastructure and Logistics (TIL)
at Delft University of Technology

Project duration: February 20, 2024 – November 4, 2024

Assessment Committee

Chair:	*prof. dr. O. (Oded) Cats* Professor at Delft University of Technology Faculty of Civil Engineering and Geosciences (CEG) Department of Transport and Planning
First Supervisor:	*dr. F. (Frederik) Schulte* Assistant professor at Delft University of Technology Faculty of Mechanical Engineering (ME) Department of Transport Engineering and Logistics
Second Supervisor:	*dr. A. (Alessandro) Bombelli* Assistant professor at Delft University of Technology Faculty of Aerospace Engineering (AE) Department of Air Transport and Operations
Additional Examiner:	*dr. Y. (Yongqiu) Zhu* Assistant professor at Delft University of Technology Faculty of Civil Engineering and Geosciences (CEG) Department of Transport and Planning

Front page image by Jawed Gfx (n.d.)

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Preface

Somewhere between Paris and Amsterdam, traveling at 300 km/h on a Eurostar train, I struck up a conversation with the conductor. I told him about this thesis and my aim to design an optimal international European high-speed rail network. His immediate reaction was sceptical, and for several minutes, he mentioned the various challenges I would face in achieving that goal, before wishing me good luck.

A few months later, I proudly write this preface with a completed report. I could have never done it alone. Many people and key events shaped me as a person, making sure I had the right mindset to go into this project, and the discipline to spend almost nine months and well over 1,200 hours to carefully shape my research and report. In this section, I would like to share my gratitude towards these people.

Starting with my supervisory team: Oded, thank you so much for being the chair of my committee. The fact that you agreed to take this role on long before the start of this project while helping me to correctly shape the scope, propelled me with confidence and discipline to work hard throughout the project. Frederik, your suggestions fixed the missing links in terms of scientific substantiation. Great examples of this are in the fields of literature gap analysis and benchmarking, but your ideas proved to be elevating the entire project. Alessandro, thank you for helping me out with the optimisation part especially. Also, your suggestion of web scraping was absolutely crucial for making the network realistic and enabled my optimisation model to be applied a network much larger than I could have ever wished for. As a whole, I want to thank my supervisory team for their time and trust in me, the valuable suggestions and meetings. Even though all of you helped in steering this project towards the end goal, within these bounds I felt a great sense of freedom to make this project truly my own. Thank you for that, as it was greatly beneficial for my productivity and motivation throughout these nine months.

To my family, and especially to my parents, thank you for allowing me to freely choose my path in life, supporting me in every single career decision I make, and providing me with a happy and safe living- and working environment. Without all of that, I would have never been able to write this thesis. I am thankful for my entire family. To my brother Jesse for his great interest along the way. To my girlfriend Wag, who is the most important person in my life. I can be quite a workaholic – you were crucial in finding the right balance between work and leisure. Without you, I could have never written this thesis. Thank you for helping me to pick colours for the final line design, which was quite a challenge for me. Also, thank you so much for proofreading this report with great attention to detail. To my grandmother Thea, who probably is my biggest fan and has followed my university career with great enthusiasm. I want to dedicate this report to the ones of my family watching from above – you know who you are. You are dearly missed, and I hope you are looking down proudly with a smile.

Lastly, I want to thank the friends I have made along the way during my time in Delft. All of you belong to the smartest people I have ever met, and you were crucial into making my time in Delft productive -but above all- an absolute pleasure to look back on. Siebe, Yuran, Lars, it is thanks to you that even though Covid ruined the second half of my time at Civil Engineering, I smile when looking at my bachelor's degree. We have made great memories together that I will carry with me for the rest of my life. Aik, besides being a great friend, you practically functioned as an additional supervisor. With our thesis projects running in parallel, I'm grateful for the many hours we spent working together and helping each other out during tough issues. You have played such a big part keeping me motivated and on track throughout this project.

Looking back at seven years of studying Civil Engineering and Transport, Infrastructure & Logistics at the Delft University of Technology, the six chapters of this report close a crucial chapter in my life at which I look back with happiness and gratitude.

I wish the reader of this report a pleasant reading.

Dion Mol
Delft, November 2024

Abstract

Despite its environmental benefits, the development of a European high-speed rail network has faced many challenges, such as cost overruns and disappointing demand, resulting in unprofitability. Countries tend to prioritise national interests over international cooperation, resulting in unconnected and smaller networks, being suboptimal. This study focuses on the optimal foundation of the network: it's topology, lines, and their operated frequencies. On the basis of extensive literature analysis, a demand forecasting model is built, in which trip generation and modal distribution are forecasted. Realistic data regarding travel alternatives, gathered through web scraping, serve as input. Optimal arc-specific HSR fares that maximise revenue are determined, while taking elastic demand into account.

Further literature analysis provides an overview of HSR cash-flows and their impacts on profitability, in which the demand forecasting model serves as input. The Transportation Network Design & Frequency Setting Problem (TNDFSP) is solved for optimal profitability over the network's lifetime. A new formulation of this problem is set up to optimise network and line design simultaneously for a large number of OD pairs while accounting for demand elasticity, addressing a crucial gap in current literature. The formulation is based on the Multi-Commodity Flow Problem (MCFP) but adapted in order to take OD pair flow routes into account. Our formulation is implemented, considering up to 111 cities and all their OD pairs, optimised within a few hours. The results show that high-speed rail can be profitable in Europe, having an optimal profit of €222.8 bn over a 40-year lifetime. The network is largely concentrated around a selected group of the largest cities in the north-western regions of the continent, and crucially relies on significant initial investments.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
2GLS	2-stage Generalised Least Squares
BNL	Binomial Logit Model
DID	Differences-In-Differences
FS(P)	Frequency Setting (Problem)
GDP	Gross Domestic Product
HS2	High Speed 2 (HSR construction project in the UK)
HSR	High-Speed Rail
IATA	International Air Transport Association (unique 3-letter airport identifying code: ISO-3)
IIA	Independence of Irrelevant Attributes
LP	Linear Program
MCFP	Multi-Commodity Flow Problem
MILP	Mixed Integer Linear Program
ML	Mixed Logit Model
MLE	Maximum Likelihood Estimation
MNL	Multinomial Logit Model
NL	Nested Logit Model
OD	Origin-Destination
OLS	Ordinary Least Squares
SERA	Single European Rail Area
TEN-T	Trans-European Travel Network
TND(P)	Transit/Transport Network Design (Problem)
TNDFS(P)	Transit/Transport Network Design & Frequency Setting (Problem)
VO(T)T	Value of (Travel) Time

Nomenclature

Sets and indices		
Symbol	**Description**	**Iter.**
$\boldsymbol{A}$	connections, arcs, edges, links	a
$\boldsymbol{A_p}$	set of valid arcs for OD pair p	a
$\boldsymbol{A^{cross}}$	set of crossing arc pairs	(i,j)
$\boldsymbol{A^{sel}}$	selected connections, arcs, edges, links	a
$\boldsymbol{A_i^{in}}$	arcs directed outwards from node $i \in \boldsymbol{N}$ (MCFP)	a
$\boldsymbol{A_i^{out}}$	arcs directed inwards to node $i \in \boldsymbol{N}$ (MCFP)	a
$\boldsymbol{D(k)}$	destination node of commodity $k \in \boldsymbol{K}$ (MCFP)	i
$\boldsymbol{G}$	graph $(\boldsymbol{N}, \boldsymbol{A})$	
$\boldsymbol{G_p}$	valid graph $(\boldsymbol{N_p}, \boldsymbol{A_p})$ for OD pair p	
$\boldsymbol{G^{sel}}$	selected graph $(\boldsymbol{N^{sel}}, \boldsymbol{A^{sel}})$	
$\boldsymbol{K}$	travel mode alternatives (MNL) commodities (MCFP)	k, j k
$\boldsymbol{L}$	lines	l
$\boldsymbol{L^{sel}}$	selected lines	l
$\boldsymbol{M}$	travel mode attributes (MNL)	m
$\boldsymbol{N}$	cities, nodes, vertices, UCs	i, j
$\boldsymbol{N_p}$	set of valid nodes for OD pair p	i
$\boldsymbol{N^{close}}$	city pairs that are located too closely (within l^{min})	(i,j)
$\boldsymbol{N^{sel}}$	selected cities, nodes, vertices, UCs	i, j
$\boldsymbol{O(k)}$	origin node of commodity $k \in \boldsymbol{K}$ (MCFP)	i
$\boldsymbol{P}$	OD pairs	p
$\boldsymbol{R}$	OD flow routes	r
$\boldsymbol{T^{life}}$	years of lifetime	t

Variables and parameters		
Symbol	**Description**	**Unit**
a	linear trendline, intercept coefficient	$[-]$
$AANP$	annual average number of passengers (Belal et al., 2020)	$[Mpeople]$
b	linear trendline, slope coefficient	$[-]$
c_{ak}	unit cost of transporting commodity k over arc a (MCFP)	[€]
c_a	travel cost of arc a	[€]
$C_{country(i)}^{surface}$	surface construction costs in country of city i per km	[€]
$C_{country(i)}^{tunnel}$	tunnel construction costs in country of city i per km	[€]
$C_{ij}^{T,infra}$	time-based infrastructure costs between city i and j	[€]
$C_{ij}^{T,train}$	time-based rolling stock costs between city i and j	[€]
$C_{ij}^{X,infra}$	fixed infrastructure costs between city i and j	[€]
$C_{ij}^{X,train}$	fixed rolling stock costs between city i and j	[€]
c_{nr}^{node}	whether OD flow route r covers node n	$[-]$
c_{ar}^{arc}	whether OD flow route r covers arc a	$[-]$
c_{pr}^{odpair}	whether OD flow route r covers OD pair p	$[-]$
m_{pr}^{odpair}	whether OD flow route r exactly matches OD pair p	$[-]$
c_{pl}	whether OD pair p is covered by line l	$[-]$
c_p	travel cost for OD pair p	$[-]$
c_r	travel time along route r	$[-]$
C_t^{TOT}	total costs in year t	[€]
D	demand/flow, operating days per year	$[pax\ /\ time]$ $[days]$
$d(a,b)$	distance between node a and b	$[km]$
$D_{HSR,ij}^{LT}$	average yearly HSR demand over lifetime, between city i and j	$[pax\ /\ time]$
$D_{m,p}$	demand for travel alternative k for OD pair p	$[pax\ /\ time]$
$D_{k,ij}$	travel mode alternative k demand/flow between city i and j	$[pax\ /\ time]$
D_{ij}	demand/flow between city i and j	$[pax\ /\ time]$
d_a	length of arc a	$[km]$

d_{ij}	distance between city i and j	$[km]$
D_p^{TOT}	total travel demand for OD pair p	$[pax\ /\ time]$
D_k	demand/flow of commodity k (MCFP)	$[-]$
d_r	length of route r	$[km]$
e	elasticity	$[-]$
e_{GDP}	GDP elasticity for HSR demand	$[-]$
f_a^{cost}	lifetime construction, operation and maintenance cost of arc a	$[€]$
f_r^{rev}	lifetime revenue of route r	$[€]$
f^{min}	minimum frequency	$[trains\ /\ h]$
f_{ij}	frequency of trains between city i and j	$[trains\ /\ h]$
f_l	frequency of trains between on line l	$[trains\ /\ h]$
GDP_i	GDP of city i	$[€]$
$GDPCAP_i$	GDP per capita of city i	$[€]$
H	operating hours per day	$[h]$
HI_i	height index for city i	$[m]$
h_l	headway of trains on line l	$[h]$
h_i	elevation of city i	$[m]$
h_{min}	minimum city elevation	$[m]$
h_{max}	maximum city elevation	$[m]$
k	gravity parameter (constant)	$[-]$
k^{detour}	detour factor	$[-]$
k_{ij}^{eco}	economic growth demand factor, between city i and j	$[-]$
k^{eff}	speed efficiency factor	$[-]$
k_{ij}^{ind}	induced demand factor, between city i and j	$[-]$
k^X	unit train acquisition cost	$[€\ /\ train]$
k^X	unit train operation and maintenance cost	$[€\ /\ seatkm]$
k^{peakhr}	peak hour factor	$[-]$
$k^{T,infra}$	unit time-based infrastructure costs	$[€\ /\ km/\ year]$
$k^{T,train}$	unit time-based rolling stock costs	$[€\ /\ seatkm/\ year]$
$k^{transfer}$	transfer penalty	$[€]$
$k^{X,infra}$	unit fixed infrastructure costs between	$[€\ /\ km]$
$k^{X,train}$	unit fixed rolling stock costs between city	$[€\ /\ train]$
L	line length (Belal et al., 2020)	$[km]$
l^{min}	minimum arc length, minimum distance between stops	$[km]$
l_r	Whether route r is a valid operating line	$[-]$
l_{ij}	line length between city i and j	$[km]$
M	arbitrary large constant	$[-]$
m_k	market share of travel mode alternative k	$[-]$
N	number of train seats (Belal et al., 2020)	$[pax]$
n_{ij}	number of trains between city i and j	$[trains]$
n_l	number of trains on line l	$[trains]$
$No.of\ Trains$	number of trains (Belal et al., 2020)	$[trains]$
n_r^{trains}	number of trains on route r	$[trains]$
n_p^{dest}	destination node of OD pair p	$[-]$
n_p^{org}	origin node of OD pair p	$[-]$
p_{ij}	minimum country-level economic growth rate, among city i and j	$[\%/year]$
P_i	population of city i	$[people]$
P_{ij}^T	HSR operational profitability between city i and j	$[€]$
P_{ij}^X	HSR justifiability between city i and j	$[€]$
P_{it}	population catchment of city i within t minutes	$[people]$
Q^{peak}	peak hour demand/flow	$[pax\ /\ time]$
Q_a	demand/flow for arc a	$[pax\ /\ time]$
q_r^{peak}	peak hour demand for route r	$[pax]$
q_r^{year}	yearly demand for route r	$[pax]$
R^2	R-squared correlation fit (OLS)	$[-]$
R_{ij}	revenue for connection between city i and j	$[€]$
s	travel distance, train seat capacity	$[km]$, $[pax]$
t	travel time	$[h]$
t^{access}	access time per station	$[h]$
t^{egress}	egress time per station	$[h]$
t^{dwell}	dwelling time per station	$[h]$

T^{life}	lifespan	$[years]$
t_a	travel time of arc a	$[h]$
t_{ij}	travel time between city i and j	$[h]$
t_l	travel time along length of line l	$[h]$
t_p	travel time along of OD pair p	$[h]$
t_r	travel time along route r	$[h]$
$TC_{k,ij}$	travel cost for travel alternative k between city i and j	$[€]$
$TT_{k,ij}$	travel time for travel alternative k between city i and j	$[h]$
U_k	utility of travel alternative k (MNL)	$[-]$
u_p	whether OD pair p is served	$[-]$
V	maximum operating speed (Belal et al., 2020)	$[km\ /\ h]$
$\bar{v}$	average operating speed	$[km\ /\ h]$
v^{max}	maximum operating speed	$[km\ /\ h]$
V_k	observed utility of travel alternative k (MNL)	$[-]$
$V_{k,ij}$	observed utility of travel alternative k between city i and j (MNL)	$[-]$
v_p	whether OD pair p is served without transfer	$[-]$
w_r	frequency of route r	$[-]$
W^{max}	Maximum allowed frequency	$[trains\ /\ h]$
x	slope of linear function (bias elimination)	$[-]$
x_{ak}	number of commodities k transported over arc a (MCFP)	$[-]$
x_{km}	value of travel attribute m for travel mode alternative k (MNL)	$[-]$
x_m	value of travel attribute m (MNL)	$[-]$
x_r	whether OD flow route r is selected	$[-]$
y	Intercept of linear function (bias elimination)	$[-]$
y_a	whether arc a is selected	$[-]$
y_{new}	new value of predicted demand (after bias elimination)	$[pax\ /\ time]$
y_{old}	old value of predicted demand (before bias elimination)	$[pax\ /\ time]$
Z	value of a graph	$[-]$
z_n	whether node n is selected	$[-]$
$z_{k,ij}$	whether travel alternative k is present between city i and j	$[-]$
α	parameter for population (gravity), intercept coefficient (OLS)	$[-]$
α_p, α_r	demand model linearisation coefficient: intercept	$[-]$
β	parameter for GDP (gravity), slope coefficient (OLS)	$[-]$
β^0	constant taste parameter (MNL)	$[-]$
β^{TC}	travel costs taste parameter (MNL)	$[utility\ /\ €]$
β^{TT}	travel time taste parameter (MNL)	$[utility\ /\ h]$
β_m	taste parameter of travel mode attribute m (MNL)	$[-]$
β_p, β_r	demand model linearisation coefficient: travel time	$[-]$
γ	parameter for distance (gravity)	$[-]$
γ_p, γ_r	demand model linearisation coefficient: travel cost	$[-]$
ε_k	unobserved utility of travel mode alternative k (MNL)	$[-]$

AI-generated high-speed train (Alam, n.d.)

1 Introduction

1 Introduction

This first chapter functions to introduce high-speed rail within its travel market, as well as the importance to shift more towards it (in section 1.1). It then proceeds to define the problems on the way to this shift (section 1.2) and the research questions that would arise when aiming to solve them (section 1.3). Section 1.4 provides a short methodological overview of the project, bridging to an outline for the rest of the report, presented in section 1.6. Section 1.5 provides the project's scope.

1.1 Background and Context

Breakfast in Stockholm, lunch in Berlin and dinner in Rome. For European travellers aware of the travel options between these cities, this will sound impracticable, as it would imply spending most of your day waiting inside an airport or transport vehicle. Even though the distances between large cities in Europe are on average much smaller than in other continents, it remains highly impractical to work in one city and live in another, a few hundred kilometres away. It is indicative for the lack of interconnectivity between European cities and the travel market potential that remains unserved. This section explores the possibilities and attempts of Europe to bring life to its long-distance travel market.

1.1.1 Long-Distance Travel Market

In Europe, generally speaking, four means of transport exist for travelling long-distance. In this paragraph, high-speed rail and its main competitors are addressed. Figure 1.1 below shows the popularity of each of the modes at various travel distances.

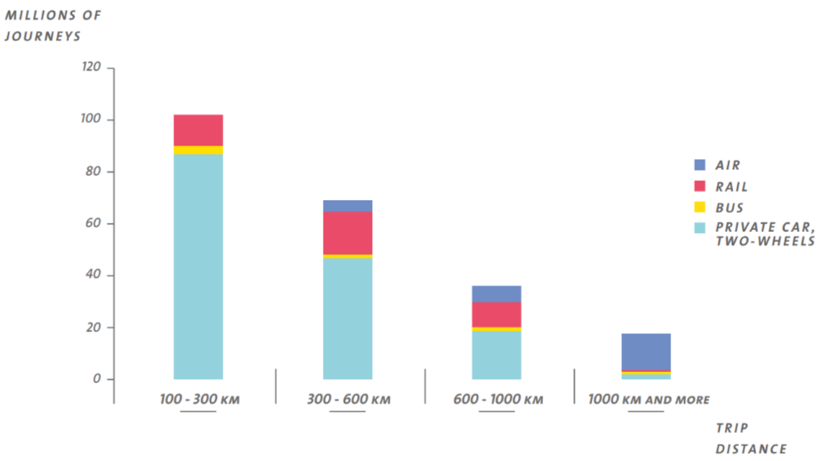


Figure 1.1 Modal shares in European long-distance travel (UIC, 2018)

Given the marginal contribution of bus travel at all distances, the mode is taken out of consideration here. The other three modes will be addressed below.

(1) Airplane

To this day, airplanes are able to fulfil long-distance travel demand the most, most often without a significant competitor. Airplanes dominate the European travel market for distances over 700 km. Increase the distance to 900 km, and in many cases their market share reaches near 100%, as travelling via air is most often the only practical option (Bleijenberg, 2020). Overall, travelling by air is a very popular travel alternative within the continent. In the European Union, aviation fuel is legally exempt from taxes (EU, 2024), which further benefits the aviation industry by allowing prices to remain lower. Up until a year before the pandemic, the EU member states showed continuous growth in air passenger numbers of 6.0% yearly on average, with the total number of passengers travelling intra-Europe exceeding 1.1 billion (Eurostat, 2019). This is indicative for the increased popularity of travelling long-distance. The airline industry has been recovering rapidly from the pandemic and is expected to do so fully by 2024. Projections indicate a continuous increase for the future as well (IATA, 2023), with passenger numbers expected to double within the next 20 years (Timperley, 2020), indicating the resilience and strength of the air travel market.

Despite being a popular travel option for many, flying comes with many downsides. Most prominent are environmental concerns: flights within Europe emit at least five to six times more CO_2 per passenger-kilometre than trains. Shorter flights emit higher levels of CO_2 per km (Bleijenberg, 2020) and per passenger-km (BBC, 2019; Figure 1.2) than longer flights. Even though the aviation industry has made significant strides regarding fuel efficiency (European Commission, 2021), the growth in air traffic has led to an increase in the contribution to climate change, expected to triple before 2050 (ICAO, 2019). Most airlines will rely upon traditional engines until then (Hepher & Frost, 2021), making it incompatible with the Paris climate agreements (Gössling & Humpe, 2020; Bleijenberg, 2020). As shown in Figure 1.2, travelling by air, car or bus is much more polluting than other modes of transport. Other disadvantages of air travel include accidents, aircraft noise, delays and congestion (Janić, 1999). These are all well-documented, but the growing concerns regarding emissions and climate change has forced the European Union to look for other alternatives and opting for High-Speed Rail (HSR) development (European Court of Auditors, 2018). Section 1.1.2 further elaborates on this.

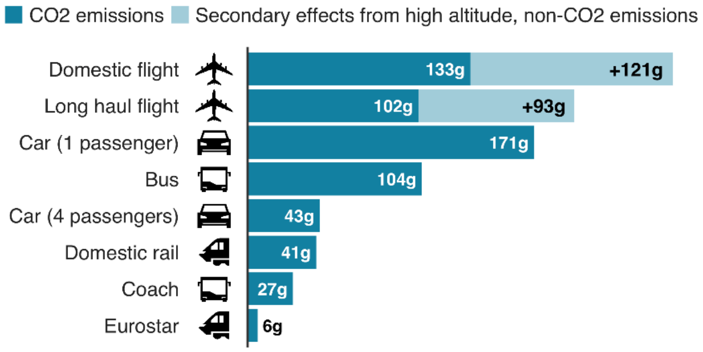


Figure 1.2 Total emissions per passenger-km for medium to long-distance intra-European travel (BBC, 2019)

(2) High-speed rail (HSR)

As shown in Figure 1.2, high-speed trains emit only a fraction of other air and road alternatives: on average seven times less per passenger-km (Strauss et al., 2021). Therefore, HSR is often raised as a more sustainable alternative for medium to long-distance travel between urban centres (López-Pita & Robusté, 2004). Trains are considered high-speed when exceeding 250 km per hour. Commercial speeds reach 350 km/h. For travel times up to two hours, HSR often is able to fully dominate the market, regularly leading to airlines giving up operating these trips. Even though the maximum speeds of high-speed trains are significantly lower than those of airplanes, it can be a competitive means of transport for travel times up to approximately four hours (UIC, 2018). This is due to smaller waiting times and the fact that HSR stations often can be placed in city centres, increasing its accessibility (Martín et al., 2014).

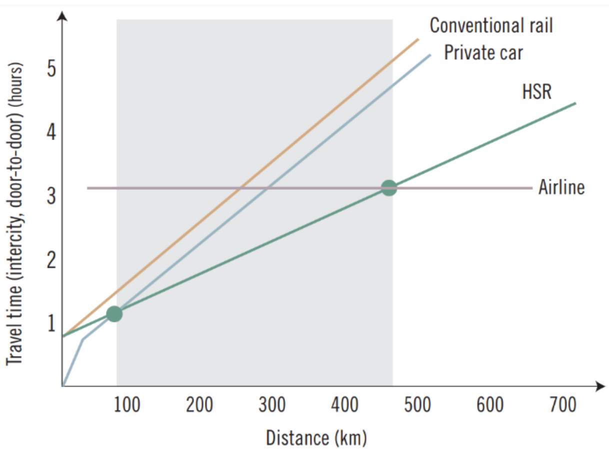


Figure 1.3 Competitive advantage of HSR over other modes (Barrón et al., 2012)

HSR technology has been in development and used in practice for decades. There are numerous advantages of high-speed trains over airplanes. Firstly, the use of clean and renewable energy, which commercial airplanes do not use at large scale - with no foreseeable use in the near future. Other advantages of HSR include much lower operating costs, potentially lower fares, better accessibility to urban and rural areas and more passenger comfort (Rajendran & Popfinger, 2022).

The benefits over air travel have made multiple countries in the world constructing high-speed rail lines. China can be seen as a world leader in this field, having a nation-wide network connecting all major cities, serving 74% of the country's inhabitants within a two-hour drive (Wang et al., 2015). In fact, two thirds of the world-wide HSR track length can be found in China's network (Chen, 2020). The fast growth of the network has been the consequence of heavy governmental subsidies and policy. High-speed rail has been very successful in China. The network served 2.4 billion passengers in 2019 (Zhang, 2024) and served twice the passenger number of domestic flights already back in 2013 (Bradsher, 2013).

(3) Car

In Europe, most long-distance trips are made by car (UIC, 2018). As can be seen in Figure 1.1, it remains the most popular way to travel for trips up to 1,000 km in general. Cars are more pollutive per passenger-km than all other competing modes (see Figure 1.2), contributing to 61% of all emitted CO_2 in the EU transport sector. Speeds reach far lower than those of (high-speed) rail and airplanes, but this is compensated by the ability of cars to also cover the distance directly to the origin and destination – the so-called first and last mile (Lu et al., 2023). Cars provide direct access to much more destinations than public transport, while having an increased level of comfort (Hiscock et al., 2002). Even though significant strides are made into making it a less polluting mode of travel, the EU aims to shift users more to rail, as it can play an important role into meeting set climate goals (European Parliament, 2019).

It is clear that high-speed rail is a viable alternative to environmentally unfriendly travel modes, and an urgent need for more sustainable medium- to long-distance travel options exists.

1.1.2 High-Speed Rail Policy

Following the example of China and all beforementioned benefits over air travel, it comes as no surprise that the EU has been pushing governments to develop international high-speed rail connections. Even though Europe has an extensive regular rail network, it was developed with national focus, complicating interoperability and efficiency when travelling internationally. This can be seen in Figure 1.4. Since the 1990s, the EU has set the end goal of developing an "efficient and competitive EU-wide railway network"; the so-called Single European Railway Area (SERA) (European Council, 2024). This vision has effected in policy by means of four railway packages, adapted by the European Council between 2001 and 2016, the last one coming into effect by 2021. These legislative packages are meant to allow and push for development of cross-border connections and interoperability among the different national rail systems. Sustainability of travelling rose to prominence with the development of plans for the Trans-European Travel Network (TEN-T) (European Commission, 2013), which supports the transition to more sustainable transport, connecting railway lines internationally with roads, ports and

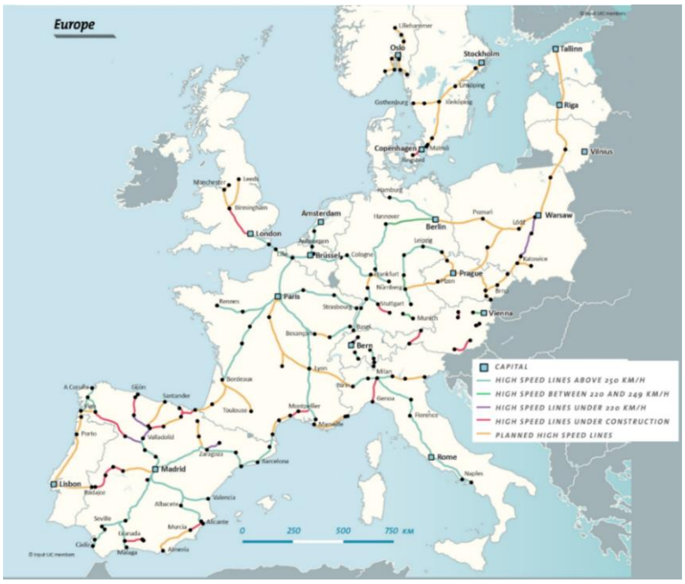


Figure 1.4 European HSR lines, either built or planned (UIC, 2018)
A larger version of this figure is provided in Appendix G

airports. As part of the European Green Deal, the EU has set the goal to double the length of HSR-lines by 2030, while also reserving the majority of a €25 billion budget for rail development (European Council, 2024). This is on top of the €23.7 billion already invested into high-speed rail infrastructure since 2000 (European Court of Auditors, 2018). In total, it aims to play a big part in further shifting travellers to climate-friendlier modes of transport, as the EU has set an objective to reduce the transport sector's greenhouse gas emissions by 90% between 1990 and 2050 (EEA, 2024).

1.1.3 European High-Speed Rail Network

As stated by European Court of Auditors (2018), there is no European high-speed rail network. It consists out of a set of rather smaller and nationally oriented networks, while cross-border connections are rare. This is further illustrated by Figure 1.4. Spain, France and Italy have national networks providing connections between their largest cities. Spain and France opted to have their capital serving as a HSR-hub, with radial lines traversing outwards to the other few largest cities. Italy's network more follows a line structure from Rome up to the northern cities. Germany operates a relatively large set of connections between the country's biggest cities but has no real nation-wide fully connected HSR network. Cross-border connections only exist via the Channel tunnel between France and the UK, and at the connections between the capitals of France, Belgium and The Netherlands.

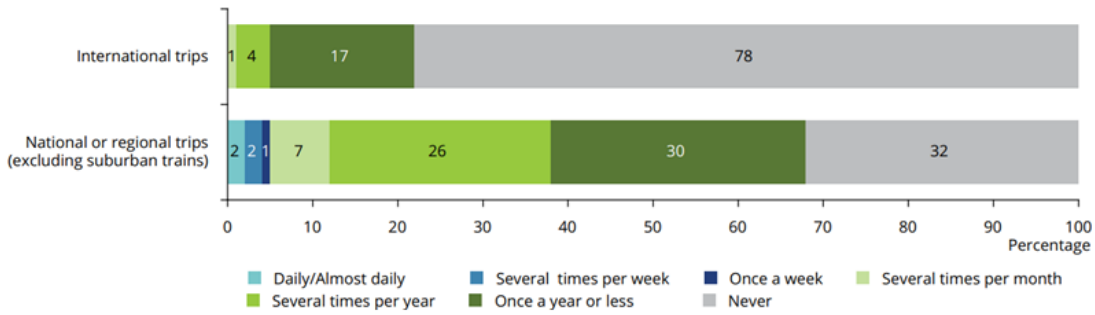


Figure 1.5 National and international HSR usage patterns, EU-28 excluding Malta and Cyprus (EC, 2018)

Thus, only three countries in Europe have a national fully-fledged network, and none of these countries are connected. There is no real EU-wide network, it mainly consists out of national-focussed isolated smaller networks. The result of this is also shown in passenger travel behaviour: only 1 % of Europeans travel internationally at least once a month. 78 % never travels internationally by HSR. For national or regional trips, these percentages are more HSR-favourable, but not by much. Only 5% of Europeans travel by HSR at least once a week. Figure 1.5 shows an overview.

1.2 Problem Definition

Building upon the context explained in section 1.1, three main problems can be identified, that complicate or prohibit the development of a Europe-wide high-speed rail network. In this paragraph, each of these are explained in detail.

1.2.1 Uncertainty of Justifiability and Operational Profitability

Even though the total Chinese HSR network has reported net profits (MarcoPolo, 2024), not all of its lines are operationally profitable. This mainly accounts for lines in the more sparsely populated mid-west, operating at lower speeds (Want China Times, 2013). The Chinese government solves this problem by subsidising heavily, which raises concerns about the justifiability among HSR experts.

This is not the only concern raised when it comes to high-speed rail. Critiques also exist regarding (construction) cost overruns, which can be substantial. This can be illustrated by the HS2-projects in the UK. Originally budgeted at £37.5 billion in 2009, the figure of expected construction costs has been increased multiple times, reaching £135 billion in 2020. These changes highly influence the profitability and justifiability of the project, even when adjusted for inflation. In the case of HS2, it has resulted in parts being cut off from the project entirely, while the benefit-cost ratio has decreased substantially. Concerns are now raised whether the line's revenue will ever justify the large initial investments (Tetlow & Pattison, 2023).

Even though the cost overruns of the HS2 project are exceptionally large, they are very common among (European) HSR projects. The same holds for delays in construction. These insecurities also occur in other HSR cash flows, and the expected demand. In Europe, three out of seven lines fail to have a sufficient number of passengers, while nine of out fourteen do not even have the potential number of passengers in the area to ever reach a sufficient amount (European Court of Auditors, 2018). The viability of these lines thus relies on subsidies. The uncertainties in (operational) profitability and justifiability are a main critique against HSR. These two terms concerns different parties and therefore they may not be used interchangeably:

- **Operational profitability** regards all operational revenues and costs, and therefore concerns the operator. If an HSR line is not operationally profitable, no company will be interested in operating the line, unless subsidies are able to compensate for this operational loss.
- **Justifiability** regards the extent in which the operational profitability is able to make up for the initial investments, and therefore concerns the party ordering construction of the line, which often are governmental bodies). If the construction of a HSR line is not justifiable, authorities will not be interested in building it.

An economically justifiable and operationally profitable Europe-wide HSR network as desired by the European Commission therefore remains a very complex research project. Thorough research on predicting justifiability and profitability of hypothetical lines accurately, with the beforementioned uncertainties in mind, is therefore a much-needed tool. It could then be used to design a European high-speed rail network, to demonstrate its profitability potential, while remaining economically justifiable. The result could function as a mean to check whether the main critique against high-speed rail is plausible. If not, it would show the conservatively calculated potential of HSR. This would potentially help pave the way for network development, changing the way continent is travelled across and significantly help battling climate change.

1.2.2 Ineffectiveness of European Policy

As already discussed in section 1.1.3 and displayed in Figure 1.4, there is still no European high-speed rail network. Also, it is expected that it will not be existent in the future soon. The most recent and prominent agreement concerns the European Green Deal, which aims to reduce transport-related emissions by 90% and triple HSR's traffic volume between 1990 and 2050 (European Council, 2024). As the total transport-related greenhouse gas emissions have only increased since 1990 (EEA, 2023), it can be concluded that the EU is not on track to meet its climate goals. Therefore, it can be concluded that the policies as designed by the EU have not reached their desired goals and effects. This is caused by multiple factors, as mentioned by European Court of Auditors (2018):

- There is no strategical EU-wide approach. EU member states still have a rather national focus when it comes to constructing high-speed rail lines. This is caused by the national rules that still exist and act as technical and administrative barriers when it comes to construction of cross-border connections.
- The European Union has no legal powers in forcing EU member states to construct the international rail connections as envisioned. This holds as well for an enormity of national rules still active in 2018, despite the European Commission attempting to mitigate them a decade ago.
- Countries might have adopted a more cautious position towards high-speed rail development, considering the uncertainties mentioned regarding profitability and justifiability, as well as recent inflation.

1.2.3 Knowledge Gap in HSR Network Design

As pointed out by Grolle et al. (2024), no HSR network design methods are currently available. This is a crucial research gap, since gaining information on selecting good or optimal configurations of connections directly influences the earlier mentioned justifiability and operational profitability of the networks as a whole. The problem needed to solve here can be generally captured as the "*Transit Network Design Problem (TNDP)*" and the *"Frequency Setting Problem (FSP)*", as they are described by Ibarra-Rojas et al. (2015), and applied to an European HSR network design by Grolle et al. (2024).

The **Transit Network Design Problem (TNDP)** is recognised as one of the hardest transportation problems to solve (Gao et al., 2005). It considers finding the optimal selection of arcs, able to accommodate the demand between then selected origin-destination (OD)-pairs (Costa, 2005). The inclusion of realistic travel behaviour following the choice of arcs selected, known as 'elastic demand' (Cascetta & Coppola, 2012), is the main contribution to the problem's complexity. Often, the goal is to maximise the flow served or minimise network costs. Pazour et al. (2010) are able to adopt a TNDP formulation to design a national high-speed rail network, for freight. The model itself is *uncapacitated*, meaning that the capacity of links is not taken into account. This is not constrained in the model itself but assessed by post-processing the results of the model. The problem has been studied as early as by Lampkin & Saalmans (1967), with increasing interest since the start of this millennium.

The **Frequency Setting Problem (FSP)** concerns the frequencies at which (specific) lines must be operated. It can be made more complicated by taking into account different planning periods. The line design has a large influence on the optimal frequencies, since it determines what trips require transfers and how demand flows over the network (Ibarra-Rojas et al., 2015). The author states the two problems mentioned above can be combined, into a **Transit Network Design Problem & Frequency Setting Problem (TNDFSP)**. The problem has been studied as early as by Patz (1925), also with increasing interest over the last two decades.

Extensive overviews of studies using different network design problems are provided by Guihaire & Hao (2008), Ibarra-Rojas et al. (2015), Chen et al. (2011) and Farahani et al. (2013). The authors conclude that while the problems are applicable to other modes, in practice they are applied to urban public transport only. In total, 21 TNDP and 23 TNDFSP- related works were found with some relation to this work, which are listed in Appendix A. These have been filtered to a selection of seven works with a closely related focus or approach and are shown in Table 1.1 below.

Table 1.1 Overview of strongly related network design problem solving works

Reference	Problem		Related focus	Solution method			Real case
	ND	FS		E	N	A	
Hasselström (1979, 1981)	✓	✓	Number of transfers	✓			
Bielli et al. (2002)	✓	✓	Pre-defined lines			✓	✓
Carrese & Gori (2002)	✓	✓	Demand, route length, travel time, operator cost	✓		✓	✓
Fusco et al. (2002)	✓	✓	Service level, demand, route length, overall cost	✓	✓	✓	
Chen et al. (2003, 2006)	✓		Expected profit	✓			
Chen & Subprasom (2007)	✓		Expected profit	✓			
Grolle et al. (2024)	✓	✓	Line length, frequency			✓	✓
This work	✓	✓	-	✓			✓

Problem: ND (Transit Network Design), FS (Frequency Setting)
Methods: E (Exact), N (Neighbourhood Search), A ((Evolutionary) Algorithm)

Even though a similar number of TNDP- and TNDFSP-related works were found, it is noticeable that closely related TNDP works are rarer. The goal of maximising profit was only found in works by Chen et al. (2003, 2006) and Chen & Subprasom (2007). In these works, the authors make use of a TNDP model *under* *uncertainty*, meaning that they take the variances of variables into account. Ordinary models would only take the expected value into account, and therefore do not deal with uncertainty. For both works, this variable is demand. While Chen & Subprasom (2007) only take the expected value of demand into account, Chen et al. (2003, 2006) also include its variance. This allows for realistic fluctuations, which increases the model's realism and scientific value of its outcome (Chen et al., 2011).

Studies on TNDFSP are increasingly prevalent and have been applied to real-life cases involving a wide variety of constraints and objectives. The approach of Bielli et al. (2002) is particularly interesting, since it allows the problem to be split in two; a pre-defined set of lines can be applied to a network resulting from solving a TNDP. Hasselström (1979, 1981) minimises for the number of transfers, while mainly the objective of costs minimisation is popular among TNDFSP applications – however in different fields. While Carrese & Gori (2002) minimise for the operator cost, Fusco et al. (2002) do so for the overall cost.

1.3 Research Goals

A high-speed rail masterplan is needed to fulfil its potential (Preston, 2023). This project aims to provide just that. The goal of this research can be formulated by means of a main research question. To answer this question, answers must first be found to a number of sub-questions.

1.3.1 Main Research Question

The main research question of this project is stated as:

"*Which European cities must be connected via High-Speed Rail, and how should these connections be served in order to lead to an (optimally) profitable network?****"***

The expected end result is a list of directly connected cities, and the different operating lines serving them, along with their frequency.

1.3.2 Sub-Questions

In order to answer the main research question, the project's methodology is split into three sections: demand forecasting, profitability estimation and network design. Section 1.4 provides more insights on them. The main research question can only be addressed after answering the following sub-questions.

Demand forecasting

1. What can be learned from completed high-speed rail projects, regarding their demand?
2. What models and impact factors can be used to forecast high-speed rail demand?

Profitability estimation

3. What can be learned from completed high-speed rail projects, regarding their profitability?
4. Into what cash-flows does high-speed rail profitability break down?
5. What factors influence these cash-flows and how can this relationship be captured?
6. What model can be developed to forecast high-speed profitability?
7. How do the cash-flows of 'profitability' translate into 'operational profitability' and 'justifiability'?

Network design

8. What can be learned from completed high-speed rail projects, regarding the requirements cities and connections need to fulfil?
9. How can linear programs be formulated in order to solve a realistic TNDFSP formulation to optimality within reasonable computation times?
10. What would an optimal high-speed rail network look like, with respect to its topology, operating lines and associated frequencies?
11. What are the policy implications from this optimal design?

1.4 Methodological Overview

Following the research goals and questions as listed in section 1.3, a short overview of the methodology can be formulated. This will be done in this paragraph. The general process followed in this project is split into three parts, as shown in Figure 1.6. Each of the parts mentioned here is a problem in itself, each requiring research and a methodology on its own, as they will be solved sequentially.

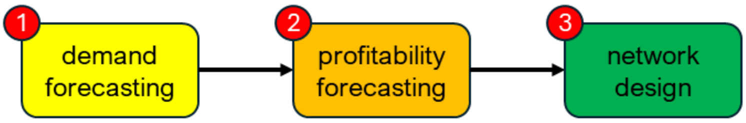


Figure 1.6 Schematic overview of methodology

It starts with a demand forecasting model, where demand is predicted for a hypothetical high-speed rail connection. The demand forecasting model functions as input for the profit forecasting model, which predicts the values of the different HSR cash-flows. From this, the justifiability and operational profitability can be assessed. The profit forecasting model is used to calculate the profitability of HSR connections. It also serves as the objective function for the network design model, since the goal is to maximise profit over the project's lifetime, while being justifiable. The network design problem considers optimisation, aiming to build and solve a problem-related TNDFSP (section 1.2.3). The expected end result is a list of built arcs, along with a list of operating lines serving them and their frequencies.

1.5 Research Scope

To have a solution of the highest scientifical value, the aim is to have all data in the models as close as possible to reality. The following scoping rules are set out.

Level of detail

The design encompasses decision making at the strategical level only. The two strategical decisions considered here are:

- What cities must be connected directly?
- What lines should operate these direct connections, and at what frequencies?

Time scope

The network should be designed to be built and completed as soon as possible, which means that it adheres to current demands and constraints. A logical building time frame would be in the order of 40 years, closely representing the lifetime of high-speed rail projects (European Court of Auditors, 2018). The aim of this project is to demonstrate how a optimally profitable network should be designed, and what it would look like. Forecasting of demand 40+ years into the future very accurately is a completely different science this project will not dive into. This way, it will be ensured that the scope will not become too large, sticking to the essence of the problem.

Area scope

Building and planning new tunnels takes decades. Their costs are hard to predict, and their benefits reach much further than only for passenger high-speed rail. This means that considering the building of tunnels to serve high-speed rail connections needs different expertise and its own research project. Therefore, for this project, the time frame enforces that connections can only be made on Continental Europe. This way, it will be ensured that the scope will not become too large, sticking to the essence of the problem to solve only.

Figure 1.7 provides the definition of the projects' area of interest.

Figure 1.7 Area of interest; own adaptation of Maix (2007)

In total, the scope encompasses 39 countries. For a full list and underlying rationale, see Appendix C.

1.6 Report Structure

Now that the logical steps towards the end goal of this research are explained, an outline of the rest of the report is presented here. It follows the methodological overview as described in section 1.4.

Chapter 2 collects all necessary knowledge regarding the defined three problems in the form of literature review. Founded on the garnered knowledge, a methodology to answer the main research question is determined in chapter 3. Thereafter, chapter 4 reports the results found after implementation of the methodology, while chapter 5 focuses on the policy implications resulting from this. At last, chapter 6 finalises the project by drawing conclusions, discussing the results and making recommendations.

2 Literature Review

AI-generated high-speed train (Alam, n.d.)

2 Literature Review

This chapter will function as a guideline, containing all necessary information needed to be able to formulate the project's methodology in chapter 3. The three problems defined in section 1.4 are addressed by sections 2.1-4, 2.5 and 2.6-7, respectively. Overviews are created of HSR demand-impacting factors (section 2.1), forecasting models (section 2.2) and their combinations (section 2.3). Demand dynamics are researched in section 2.4. Together, these sections aid the formulation of a forecasting model. Then, section 2.5 breaks down HSR profitability into cash-flows. Section 2.6 and 2.7 focus on network design and line design, respectively. Section 2.8 summarises the gaps found in literature, and bridges to the goals and methodology of this project.

2.1 Demand Impact Factors

Typically, demand for high-speed rail is forecasted by means of a specialised demand forecasting model. Castillo-Manzano et al. (2015) state that an accurate demand forecast is vital in order to verify the justification of proposed HSR lines. To be able to create such a model, this section identifies a complete list of factors influencing high-speed rail demand, according to academic sources. This serves as a starting point for understanding the concept of HSR demand, which will enhance the capability to build a demand forecasting model. The methodology building upon this can be found in section 3.2.

In order to collect a complete set of demand-influencing factors, a large literature review is required. The web was searched for papers attempting to forecast exclusively long-distance high-speed rail demand, involving high-speed rail as travel mode. Zhang et al. (2012), Börjesson (2014) and Nurhidayat et al. (2023) provide extensive overviews of long-distance travel demand studies, which serves as a starting point for our search. Further snowballing from these studies (combined with keywords "high-speed rail", "demand", "forecast" and "model") within academic platforms *ScienceDirect* and *Scopus*, and *Google Scholar* found 100 studies meeting the mentioned requirements. An overview of all these papers can be found in Appendix B.

2.1.1 Descriptive Statistics of Found Papers

Research was published in a span of forty years, between 1982 and 2023. Figure 2.1 shows the distribution of publication year among the studies. It can be noted that the interest for high-speed rail demand research and long-distance travel models has increased significantly over the years.

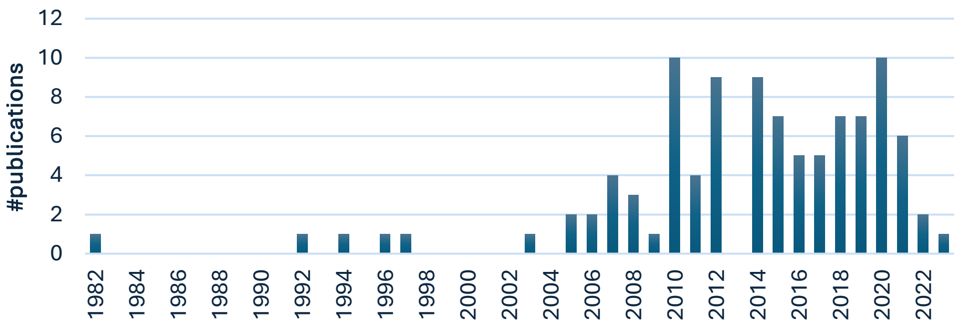


Figure 2.1 Publications of HSR demand factor papers, by year

All studies apply their models to case studies, considering a total of eighteen different countries separately, while some studies consider (parts of) continents such as Europe (6), East Asia (1) or the entire world (2). China is the most considered country by far (30 studies), which shows that the HSR culture also permeates the country's academic spheres. Mainly the western world is considered here, including almost every western European country. For each study, the assumed demand-impacting factors are listed. A full list per study can also be found in Appendix B. Factors with the same meaning (e.g. “fare” and “ticket price”) are treated as one. After this processing step, 38 factors of influence remain (a full list is provided in Table 2.1), which are divided in eight groups in Figure 2.2 below. Only a small number of factors are mentioned often, while 28 out of 38 are mentioned in less than 10% of the studies. It is indicative for the academic state-of-the-art of HSR demand forecasting: a small number of impact factors are obvious, while the set of non-obvious impact factors is large and requires attention in studies as well.

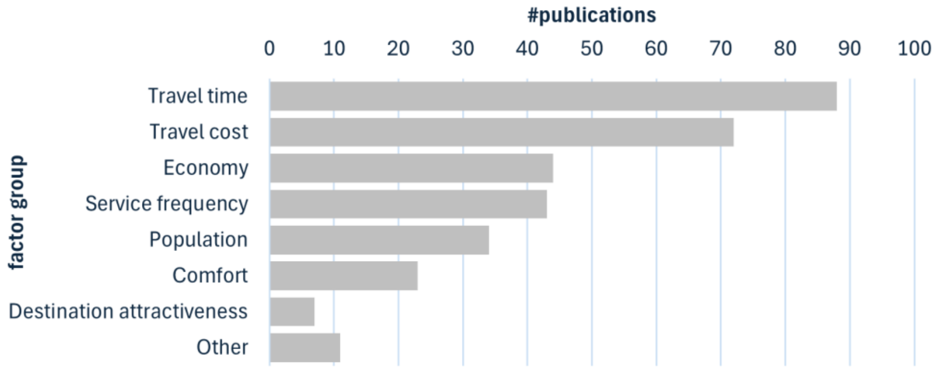


Figure 2.2 Most commonly mentioned HSR demand-impacting factors

The groups as mentioned in Figure 2.2 are presented one by one, each in their own paragraph, in order of their popularity in studies. The individual factors in each group are highlighted in bold.

2.1.2 Travel Time

For studies, this is the most obvious impact factor to think of, being thought of by 88% of them. Most of these studies use the notion of ‘total travel time’, which consists of multiple components.

Total travel time

For most travel modes, the **in-vehicle time** will represent the largest portion of the trip. Sometimes, a transfer between the transport vehicle is required, which would then add a **transfer time**, also related to the **number of transfers**. Some trips include **waiting time** at a transport hub before the transport vehicle departs, which applies to flight-based trips especially. For rail- and flight-based trips, passengers also have to travel to the airport or station, which adds an **access time**. The same holds when passengers exit the plane or train and have to travel to their actual destination (**egress time**). These last two are of great influence on determining the total door-to-door time (Moyano et al., 2018). For the same trip distance, it most often holds that in-vehicle times are shorter for flight-based trips. Waiting, access and egress times are most often shorter for rail-based trips, due to the closer proximity of rail stations to city centres. Brons et al. (2023) illustrate this in Figure 2.3.

For simplicity, most studies working with a demand forecasting model consider only the total travel time and do not break it down in each of its components. Some studies however, do (e.g. Cascetta & Cartenì (2014)) and find that all components on their own are statistically significant to be taken into account. Studies mostly focus on access time (22 studies), followed by egress time (13), in-vehicle time (8), waiting time (8), the number of transfers (4) and transfer time (1).

Relationship with HSR demand

All studies agree that higher travel times correspond with lower HSR demand. E.H. Michell (2024) provides the insights to this relationship. He compared train’s travel time with the market share of high-speed rail (on the HSR-air market), from actual reported passenger data. With this information, the author was able to produce Figure 2.4 with a reasonably good fit, highlighting the substantial influence of travel time on high-speed rail demand. Nelldal & Jansson (2010) and Jorritsma (2009) conclude that travel time is the most important impact factor of high-speed rail demand. The graph shows a non-linear relationship, while indicating HSR dominance for travel times up to four hours.

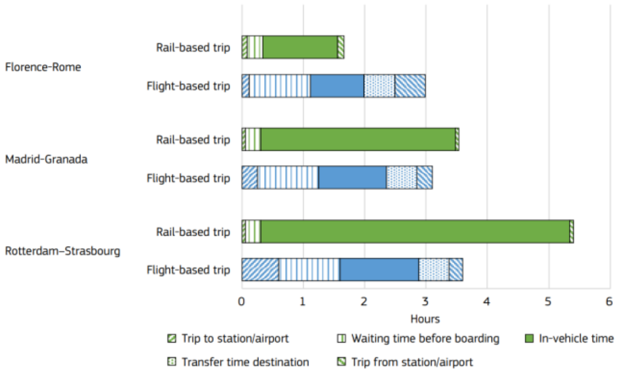


Figure 2.3 Composition of travel time components for different types of city-to-city trips within Europe (Brons et al., 2023)

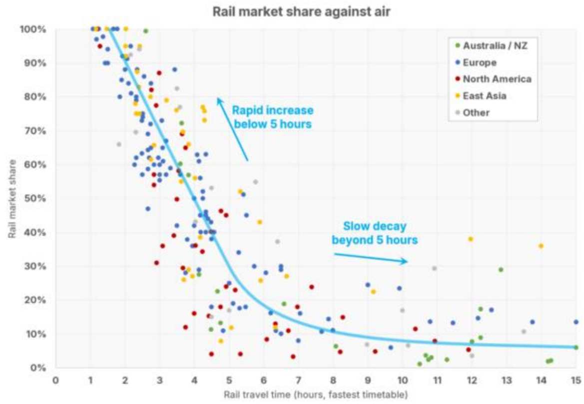


Figure 2.4 Relationship between travel time and HSR market share (E.H. Michell, 2024)

The variability around the figure's trendline also shows that total travel time is not the only explanatory variable determining high-speed rail demand. An important cause of this might be differences between how people from different countries *value* time; often indicated by the **Value of Time (VOT)**. This is taken into account by a limited number of only five studies (e.g. Hsu & Chung (1997), Outwater et al. (2010), Yang & Zhang (2012)). The authors show that it is a statistically significant parameter to consider. Also mentioned by studies are travel distance (by 34) and speed (5). Although these factors correlate with demand, they first and foremost directly influence the travel time.

2.1.3 Travel Cost

The second-most common thought of impact factor, considered by 72% of studies from this literature review. The total travel cost mostly comprises of the **ticket price** (in case of rail or air) but also fuel cost (car). For this reason, studies including car travel often opt for simply the **travel cost**, making comparisons easier. Recalling section 2.1.2, the total trip length often exists out of multiple parts, including access and egress. This pre- and post-travelling often add costs, which is also considered by travellers. For this reason and in order to make comparisons between different travel modes easier, the term "travel cost" will be used here.

In literature, the relationship between demand and its impact factor is often indicated by an elasticity value e, denoting the demand change with the value of the impact factor increases by 1% (in this context, a 1% increase of travel cost). However, it should be noted that these elasticities only hold for small changes in the pricing system, and can differ based on multiple other factors, such as the economic growth, travel time but also the VOT or price itself (de Bok et al., 2010). For this reason, elasticities should be treated as indicative.

2.1.4 Economy

It should be noted that the leap in popularity after travel time (88%) and travel cost (72%) is large: only 44% of studies include economic factors as demand-impacting factor. More studies look into the population's wealth, rather than the population figure itself (44% vs 34%), suggesting a greater importance. A city might have a large number of inhabitants, but that does not tell the complete story. They should also have a certain amount of wealth, since a wealthier population spends more money and is more likely to travel. This accounts for HSR as well. Cabanne (2003) shows that GDP elasticity in France lies somewhere between 0.3 and 0.5. The author also pointed out that the demand is much less sensitive to GDP rise than other long-distance travel modes such as air (elasticity 1.5) and car (1.3). Studies consider different ways to include the population's wealth into HSR demand-predicting models:

- **GDP (Gross Domestic Product)** is most commonly used (by 26% of studies), in different forms. The term stands for the total value of goods and services created within a certain period and is considered the standard way to measure an area's economic activity (OECD, 2022). For demand forecasting, most often the total GDP of the city is used (e.g. Albalate et al. (2015)), while a minority of studies (3%) choose to use **GDP/capita** (e.g. Couto & Graham (2007)). This term makes it easier to draw comparisons between cities. Barrón et al. (2012) estimated demand purely based on GDP growth and assumes an elasticity of 1.25. Travel data analysis over an 70-year period by Swedish operator Trafikverket (2021) concludes with a more conservative estimate of 0.70.
- **Average income** more accurately describes the actual spending power that the average inhabitant of a city has, and therefore is a more accurately demand impact factor than GDP. Data on income is much harder to find, which partly explains why it is less used in studies (19% vs 26%). Together with ticket fares, it is able to predict demand, as demonstrated by Bergantino & Capozza (2015). The income can also be aggregated to average levels per household, as executed by Ashiabor et al. (2007).
- **Welfare** as a whole is rarely found in studies, since it is rather hard to quantify in units, but Gu & Wan (2020) implement it in their demand model.

2.1.5 Service Frequency

In airline competition, the flight frequency share is the main market share determining factor when all other factors are equal (Hansen & Liu, 2015), thus a crucial factor to take into account. The frequency remains a prominent factor for studies, but less so for high-speed rail, which is illustrated by its appearance in HSR demand forecasting studies (only 43%).

The service frequency is determined by the number of departing high-speed trains per hour, within operating hours. Some of the reviewed studies call it "headway", which refers to the same concept, which is the average time between two consecutive departures.

Since the operated frequency has great influence on demand, it is often considered as a strategic design criterium. Recalling section 1.2.3: for this reason, Grolle et al. (2024) search for a European network by means of having the frequency per line as decision variable.

2.1.6 Population

This factor is mentioned by 34% of studies as demand-influencing factor. It is considered a constant value, as it cannot be changed due to design choices. Of course, higher population figures near HSR station locations generally correspond with higher HSR demand. Studies use different ways to include the population as a mean to predict high-speed rail demand:

- **Area population**, meaning they include the total population of the city, agglomeration or another administrative boundary. Since this data is often easy to find, it is a fast, cheap and popular way to include populations into demand models. Using the area's population has its flaws, since a HSR station in a city may also attract passengers from outside of the boundary or may not be accessible to all parts of the city uniformly.
- **Working population**¸ which is similar to the area population when considering its benefits and drawbacks. However, it is more an accurate representation of the number of people that would use high-speed rail regularly. It requires more extensive research, since data is harder to find. Yao & Morikawa (2005) manage to include it.
- **Number of households**, which is similar to the working population when considering its benefits and drawbacks, but data is much harder to find. Outwater et al. (2010) is an example of a study to include it, while also including the (local) unemployment rate.
- **Catchment area population** is by far the most accurate method to include population into the model, as it takes accessibility into account. HSR stations might attract travellers from outside the city in which it is located (Martínez et al., 2016). The size of the catchment area is determined by

a chosen origin-to-station travel time (Harnish, 2023). Martínez et al. (2016) showed that sizes of catchment areas vary among cities, while people living at closer proximity to stations have a higher chance of using the HSR service. The chance of people making use of the HSR line decreases with the distance: Martínez et al. (2016) find that 80% of HSR users lives in a catchment area of no more than 30 minutes. Informed by surveys, the authors apply catchment areas of 90-110 minutes in car travel time into their research.

The benefit of using catchment area population rather than city population can be illustrated by comparing the following two cities in Germany: Berlin (3.43 million inhabitants) and Wuppertal (0.36 million) (WPR, 2024), which have 60-minute catchments of 4.4 and 10.5 million people (Smappen, 2024). Even though Wuppertal is ten times smaller than Berlin, its population catchment is more than twice as large. This can be illustrated by Figure 2.5, which shows the catchment areas of both HSR stations over a population density map.

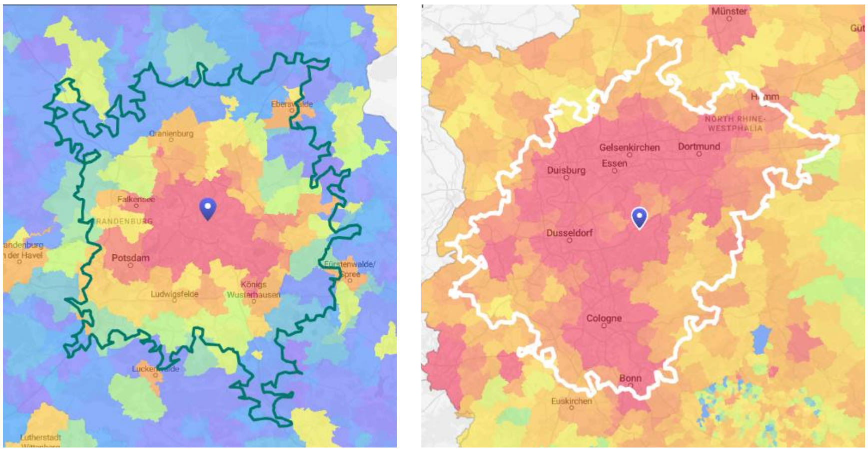


Figure 2.5 One-hour-drive population catchment areas for two German cities, plotted on a population density map, for Berlin (left) and Wuppertal (right) (Smappen, 2024)

While Berlin is a large and dense city, the population density directly outside of it is very low. On the other hand, Wuppertal lies in the middle of the Ruhr area, Europe's third most populous conurbation, behind only London and Paris (Eurocities, 2020). For that reason, density is included into the demand model by Clewlow et al. (2014), as denser cities generally lead to higher populations in catchment areas.

2.1.7 Travel Comfort

Passengers do not only want a frequent, fast and cheap service. When they travel, they except a certain level of comfort. 22% of reviewed studies take any of the following comfort-related factors into account:

- **Overall comfort** or **convenience** (10% of studies) of travelling, has proven to be a statistically significant dealbreaker to passengers when multiple long-distance travel modes are available, especially when it comes to high-speed rail (Pagliara et al., 2015).
- Passengers expect a reasonably **punctual** and **reliable** service (ARUP & OXERA, 2010). The authors find that, in the UK, a minute increase in delay decreases demand by 2%. However, punctuality and reliability is included by 3% and 6% of studies in this review, respectively.
- **Safety** (5% of studies) is taken for granted by most studies, and therefore one would expect that the influence on demand is not significant. It is indeed less significant than comfort, but as pointed out by Pagliara et al. (2015), it can be a dealbreaker for some, and higher safety levels do indeed correlate with higher demand.

- Passengers expect to be able to find a seat; this especially applies to rail modes. Busy routes where low **capacity** leads to **seat availability** issues decrease the popularity of that travel mode. For this reason, four studies take these impact factors into account.
- **Service** or **hospitality** is taken into account by three studies. It relates to operator personnel services in the transport vehicle itself.
- Most transport companies tend to make sure their **operating hours** are as customer-friendly as possible, as they achieve to maximise their passenger numbers and operating profit. But it remains an important factor to take into consideration. Yang & Zhang (2012) is the only study in this literature review to do just that.

2.1.8 Destination Attractiveness

The demand between two cities is determined by the willingness of people to bridge the distance. In models, travelling is seen as a disutility (as it requires time and money), thus it is important that the destination has something to offer, which makes the trip worthwhile (Mokhtarian & Salomon, 2001):

- **Work**. HSR-connected cities with fairly small travel times allow passengers to travel back and forth on the same day. In China, this has led to the development of a new commuting market (Ollivier et al., 2014). The number of commuting passengers logically depends on the **city attractiveness** and **industrial structure**. Only a few studies look into these factors when forecasting HSR demand: city attractiveness by Cartenì et al. (2017) and Chen (2017), industrial structure by Yu et al. (2021). Both are taken into account by Yao & Morikawa (2005).
- **Non-work** reasons usually refers to tourism or day trippers. For day trippers, again a relatively small travel time is required. Tourists are generally willing to spend more time travelling, which indicates that these are the kind of people travelling long distances. Therefore, Jiménez & Betancor (2012), Wang et al. (2018) and Zhang et al. (2020) include **tourism status.**

2.1.9 Other Impact Factors

The factors mentioned previously in this chapter fully cover 89% of all studies reviewed. The last five factors are described briefly in this section:

- **The number of air passengers** is a very direct way to predict the demand for high-speed rail. This data has great benefits: it is easily accessible, and an accurate representation of actual long-distance travel demand. For multiple studies, e.g. Grolle et al. (2024), it therefore functions as a starting point for forecasting high-speed rail demand.
- Li & Sheng (2016) and Nurhidayat et al. (2018) incorporate trip **purpose** as one of the factors to forecast demand. The latter hereby makes a distinction between 'travel reasons' as indicated in section 2.1.8.
- **Travel attitude** is incorporated in demand models by Pan & Truong (2020). Their reasoning is that travel demand might depend on the number of people actually willing to travel via a certain mode in the first place, which can be based on a general aversion or prejudices.
- Climate-aware passengers are more likely to choose climate-friendlier travel options. Therefore, **pollution** can influence travel demand. Implemented as a demand-impacting factor into a forecasting model by Chai et al. (2018).
- Cities with lower **internet usage** might generate less HSR demand. A model implementing this relationship is set up by Li et al. (2019).

2.1.10 Overview

An overview of all mentioned impact factors in demand forecasting studies is shown in Table 2.1 below.

Table 2.1 Overview of HSR demand-impacting factors reported in HSR demand studies, and the (expected) elastic relation

Impact group	Studies (%)	Factor	Unit	Relation	Studies
Travel time	88	Travel time	[h]	-	59
		Travel distance	[km]	-	34
		Access time	[h]	-	22
		Egress time	[h]	-	13
		In-vehicle time	[h]	-	8
		Wait time	[h]	-	8
		Value of Time (VOT)	[€/h]	-	5
		Speed	[km/h]	+	5
		Number of transfers	[-]	-	4
		Transfer time	[h]	-	1
Travel cost	72	Ticket price	[€]	-	43
		Travel cost	[€]	-	30
Economy	44	GDP	[€]	+	26
		Income	[€/yr]	+	19
		Welfare	[0-10]	+	1
Service frequency	43	Frequency	[h^{-1}]	+	43
Population	34	Population	[-]	+	33
		Unemployment rate	[%]	-	3
		Households	[-]	+	1
		Density	[km^{-2}]	+	2
Travel comfort	29	Overall comfort	[0-10]	+	8
		Reliability	[0-10]	+	6
		Safety	[0-10]	+	5
		Punctuality	[%]	+	3
		Convenience	[0-10]	+	2
		Capacity	[-]	+	2
		Seat availability	[%]	+	2
		Service	[0-10]	+	2
		Hospitality	[0-10]	+	1
		Operating hours	[h/day]	+	1
Destination attractiveness	7	General	[0-10]	+	3
		Touristic	[0-10]	+	3
		Industrial	[0-10]	+	2
Other	11	Air passengers	[-]	+	6
		Purpose	[-]	+	2
		Attitude	[-]	+	1
		Pollution	[CO_2]	-	1
		Internet usage	[%]	+	1

2.2 Demand Forecasting Models

As all 100 studies from the previous literature review implement their factors into forecasting models, their models will be analysed here. An overview of studies can be found in Appendix B. This literature review will provide a body of knowledge to choose the best suitable model in the methodology in section 3.2.

The studies use a relatively small number of model types, but within each type often a large variety of implementations exist. Only three different models were found in at least 5% of studies reviewed: logistic or linear regression, and gravity models. Demand forecasting models generally estimate either the total demand of all modes combined, or the market share of each single travel mode. The next four sections will provide insights to the usage of these demand-forecasting models in literature. After that, a more brief section is reserved for addressing model calibration (section 2.2.5) and an overview of models in section 2.2.6.

2.2.1 Logistic Regression Models

Logistic regression models, often called "logit models" are the most used model in demand forecasting, appear in 47% of studies from this review. In travel behaviour research, they are used to calculate market shares among different travel alternatives. Logit models are built upon the premise that to travellers, each travel alternative has a "utility", which states the satisfaction or value that a traveller derives from choosing that alternative (Arentze & Molin, 2013; Ben-Akiva & Lerman, 1985). In ordinary logit models, the utility of an alternative is a function of its attributes, as indicated by equation (2.1):

$$U_k = V_k + \varepsilon_k \tag{2.1}$$

Here, U_k is the utility of travel alternative k, which is the sum of the observed utility V_k and the unobserved utility ε_k. The latter term takes errors and deviations of behaviour into account, since the effective utility often does not match the observed utility. It is assumed to be independently and identically distributed (Ben-Akiva & Lerman, 1985). The observed utility for ordinary logit models is calculated by equation (2.2):

$$V_k = \sum_m \beta_m \cdot x_{km} \tag{2.2}$$

Here, x_{km} is the value of attribute m for travel alternative k. β_m is a 'taste'-parameter, representing the importance of attribute m when it comes to determining the utility of travel alternative k. The taste parameters are the values to be calibrated, as they are unknown at first (Mandel et al. 1994). Travellers are seen as "utility optimisers": the higher the utility of a travel alternative, the more likely they are to choose that option, thus the higher the market share m_k, which can be calculated by using equation (2.3):

$$m_k = \frac{\exp{(V_k)}}{\sum_j \exp{(V_j)}} \tag{2.3}$$

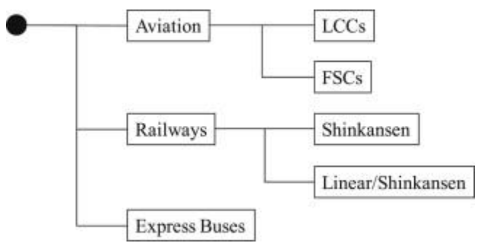
These market shares are then often calibrated on the basis of real-life data (most often large-scale survey outcomes), by calculating maximum-likelihood estimates for the taste parameters (Fiig et al., 2014). The logistic function ensures that the predicted probabilities fall between 0 and 1. Logit models therefore are often used to forecast market shares in widely varying business sectors (Geurts & Whitlark, 1992). A wide range of logit models exist, listed below. An overview is shown in Table 2.2.

- **Multinomial logit (MNL)** takes more than two travel modes into account, thus applicable to this case as indicated in section 1.1.1. Of the 47 studies reviewed, 19 used MNL. Their popularity can be attributed to their simplicity of implementation and ease of interpretation of results. The major drawback of MNL is the Independence of Irrelevant Alternatives (IIA) property, which means that the relative preference between two alternatives remains constant, regardless of the presence or absence of other alternatives (Lee et al., 2016). Therefore, it over- or underestimates market shares of alike travel modes (such as HSR and rail).

- **Binomial** or **binary logit (BNL)** is similar to MNL but only takes two options in account. Studies have implemented it in different fashions. Danapour et al. (2018), Nurhidayat et al. (2018) and Nurhidayat et al. (2019) only consider HSR and air transport as the two travel options. Another way to look at it, is the choice to travel (via HSR) or not (Cartenì et al., 2017) or choose between HSR or conventional rail (Cascetta & Cartenì, 2014); (Ren et al., 2020). In the reviewed studies, BNL is used half as frequent as MNL.

In literature, MNL and BNL are seen as ordinary logit models, in the sense that their observed utility function is linear (as in equation (2.2)). Numerous adaptations to this have been made to be able to produce more accurate, however more complex models:

- **Nested logit (NL)** takes alike travel modes such as HSR and conventional rail into account, to capture higher degrees of substitutions (Ben-Akiva et al., 2010). This eliminates the IIA property, making NL more accurate than MNL (de Palma et al., 2019). Together with MNL, NL is the most used logit model according to Fiig et al. (2014), which is reflected in the results of this literature review. The model is more complex than MNL, requiring more computing power. Its idea is to look at travelling as a multi-question decision-making process, making consecutive decisions about the travel mode and options within that class. For example, Outwater et al. (2010) and Inoue et al. (2015) put alike travel options in the same 'nest'. The nested structure of the latter is shown in Figure 2.6.



'

Figure 2.6 Nested decision-making structure in NL-model applied by Inoue et al. (2015)

- **Mixed logit (ML)** allows the previously defined taste parameters to be distributed randomly, just like the ε_k-values, which makes it a more generalised and flexible version of the ordinary logit model (Lee et al., 2016). Therefore, this model is more realistic, as a taste parameter usually is not assumed constant for an entire population. However, it requires more computation power to solve. Behrens & Pels (2012) applied both MNL and ML models in the London-Paris HSR/air market and statistically proved the ML model outperforms the MNL model. The ML model generally is more accurate than MNL, since the high flexibility tackles the IIA problem (Lee et al., 2016).
- **Box-Cox** allows for non-linearity in the model. In ordinary logit models, as can be viewed in equation (2.2), the observed value is calculated in a linear fashion. But in some cases, this might not be realistic. For this reason, Box-Cox logit models transform one or more variables into non-linear variables (Mandel et al. 1994). In transport demand modelling, Box-Cox logit models have shown to outperform ordinary logit models, as for example by last-mentioned study. This is mainly due to the fact that Box-Cox is able to capture the effects of diminishing returns: the impact of a price or travel time change decreases as the initial value is higher. The major drawback of Box-Cox is its complexity, as illustrated by Chirania (2012) and Gaudry (2008). Below, an overview of the different logistic model's characteristics is shown.

Table 2.2 Overview of different logit models in HSR demand forecasting literature

Model type	Properties			Benefits	Drawbacks	
	Alt.	Utility function	β-value		IIA	Complexity
Multinomial (MNL)	2+	linear	constant	high simplicity	yes	low
Binomial (BNL)	2	linear	constant	high simplicity	yes	low
Nested (NL)	2+	flexible	flexible	increased flexibility	no	medium
Mixed (ML)	2+	flexible	distribution	higher realism	no	high
Box-Cox	2+	non-linear	flexible	diminishing returns	yes	very high

Alt. = number of alternatives applicable to model

2.2.2 Linear Regression Models

These models represent the simplest type of demand forecasting, when it applies to the actual passenger numbers instead of market shares. The models assume that the relationship between demand and its impacting factors is linear. Equation (2.4) shows a general example of such a formulation for the demand D, which is a simplification of the version applied by Castillo-Manzano et al. (2015).

$$D = \beta^0 + \sum_m \beta_m \cdot x_m \tag{2.4}$$

It should be noted that the right-hand side of the equation above closely matches V_k in equation (2.2), as it still is a linear additive function of attribute values and attribute weights. A constant β^0 is added and represents the demand when the value of all demand-impacting factors equal zero.

As mentioned, linear regression models assume linear relationships between demand and its impact factors. In the case of demand forecasting this does not always lead to accurate forecasts, since some of the demand-impacting factors are not necessarily linear. This makes the model less popular than logistic regression (32% vs 47%), but its simplicity partly compensates for the mentioned drawbacks. Therefore, linear regression models remain relatively popular among demand forecasters. Of the 32 reviewed linear regression models, eleven were of the ordinary kind as described above.

Similar to logit models, multiple variations in linear regression models exist. In line with the literature reviewed, this section will discuss model types that have been mentioned multiple times. An overview of their distinguishing characteristics is provided in Table 2.3.

- **Panel linear regression** takes in data from multiple time instances over the same population. For example, Yu, Zhang, et al. (2021) use panel data from 2008-2015 to provide insights of HSR's influence on economic development. A panel linear regression model does not only forecast HSR demand, but also forecasts it for specific time instances. Therefore, it captures the trend as well as the relationship between HSR demand and its impact factors (Arellano & Honoré, 2001). The model is much more complex than ordinary linear regressions, since it requires estimating time-specific taste parameters. Panel linear regression models are used when the trend or development of HSR demand must be forecasted as well, and are performed by Chen (2017), Li et al. (2019) and Liu et al., 2019). The former writes the formula in a log-linear form, adding complexity while allowing to capture effects of diminishing returns.
- **DID (Differences-In-Differences) linear regression** also requires panel data but examines it differently. It allows to divide the population into groups, some exposed to a variable of interest and some not. By not changing the other variables among, the effect of the variable of interest can be examined. Mizutani & Sakai (2021) and Wan et al. (2016) investigate travellers using a certain flight route, while the variable of interest is the introduction of a HSR connection. Over a multi-year period, the effect of HSR on the air demand can be measured.
- **Multiple linear regression** is used when multiple dependent variables are of interest (Sinharay, 2010). In most reviewed studies, demand studies would only incorporate the passenger numbers of the mode of interest, therefore only having one dependent variable. Yang, Dobruszkes, et al.

(2018) break down the popularity of HSR links into two factors, city centrality and link connectivity, and incorporate both as dependent variables in their model. Zhang et al. (2017) estimate airlines passenger-kilometres in models with different independent variables.

- **Dynamic linear regression** is used when the relationships between HSR demand and its impact factors are assumed to be changing over time. This makes the model much more complex than panel linear regression, which is reflected in its popularity. Applications are scarce: both Dargay & Clark (2012) and Castillo-Manzano et al. (2015) use the model, accompanied by an ordinary linear regression model. This allows for verification of the dynamic relationship among variables. Applied in Spain, it finds that the dynamic model is superior.
- **Random effects models** include an error term ε_k. Recalling section 2.2.1, it was also used in multiple logit models. When calibrated, the error term captures parts of the unobserved relationships and variation in the data. Typically, the values of parameters are calculated by means of 2-stage Generalised Least Squares (2GLS), which calculates the error term's variance before the other parameters (Gerdtham & Jönsson, 2000). This makes the model computationally more difficult to solve, which partly explains the scarcity of studies using this model. Albalate et al. (2015) apply a 2GLS-random effects model with Spanish data, and finds that the HSR effect on airlines is a reduced number of offered seats, while the flight frequency is insensitive. Bergantino & Capozza (2015) use it to investigate the effect of HSR's presence on airline's pricing strategies and find that the effect strongly exists.

Below, Table 2.3 shows an overview of all mentioned regression models and their characteristics.

Table 2.3 Overview of different regression models in HSR demand-forecasting literature

Model type	Dependent variables	Temporal data	Grouping structure	Computation difficulty	HSR demand suitability
Ordinary	1	no	no	low	limited
Panel	1 (2D)	yes	flexible	high	good
DID	1	yes	yes (treatment & control)	moderate	interventions
Multiple	2 or more	no	no	manageable	limited
Dynamic	1 or more	yes	no	high	good
Random effects	1	yes	yes	very high	advanced

2.2.3 Gravity Models

These models are considered the most suitable method to estimate demand for new transit connections, especially when currently no direct service is available (Grosche et al., 2007). The model is built based on the concept of Newton's law of universal gravitation, which states the gravitational pull (the attractiveness of travelling) between two objects depends on their masses, distance and a scalar (Newton, 1687). In HSR demand forecasting studies, gravity models are used for calculating passenger flows rather than market shares. A standard gravity model takes only the populations of the two cities, the distance between them and a scalar into account (Wheeler, 2005). Its formula is presented by equation (2.5).

$$D_{ij} = k \cdot \frac{\left(P_i \cdot P_j\right)^{\alpha}}{\left(d_{ij}\right)^{\gamma}} \tag{2.5}$$

Here, D_{ij} is the flow between city i and j, separated by distance d_{ij}, with respective populations P_i and P_j. The distance is raised to a power distant-exponent γ to account for diminishing effects. Calibrating a gravity model requires estimating γ and scalar k. As common with other forecasting models, multiple adaptations to the original exist, seeking for a better model fit and better explanatory models. Sometimes, the distance d_{ij} between the cities is denoted by other inconveniences of travelling, such as travel time or travel costs (Wheeler, 2005).

In literature, the gravity model has been implemented in various adaptations to forecast high-speed rail demand. An overview of different types and their characteristics is provided in Table 2.4.

Table 2.4 Overview of different gravity models in HSR demand-forecasting literature

Model type	Dependent variables	Temporal data	Computation difficulty	HSR demand suitability
Ordinary	1	no	medium	good
Dynamic	multiple	yes	very high	very good

Shilton (1982) uses the basic model as stated in equation (2.5), but adds exponents to the population parameters, allowing for a better model fit. Martín & Nombela (2007) use the basic model but add capital stock (value of all transport infrastructure capital) as an measure of transport activity in the cities, along with the population. Leng et al. (2015) use it to forecast economic growth in a HSR-demand related study and includes GDP as an activity measure. It also raises the $P_i \cdot GDP_i$ -factors to exponents, which have to be estimated by calibration. Grolle et al. (2024) use a gravity model to forecast air demand, fitted to observed travel data. Yu et al. (2021) have made significant efforts in creating a ‘dynamic’ gravity model, which means that all parameters except γ and k attain an extra subscript for time t. For their research it was needed, since they wanted to account for changes in operating speeds over a four-year period. Instead of population numbers, the authors use GDP only.

2.2.4 Alternative Models

Vertical differentiation aims to understand how people weigh off different travel options, considering different characteristics that distinguish different transport modes, often in terms of quality of service attributes. It helps predicting the impact of a change in these attributes’ values on consumer preferences, and thus demand. For example, Xia & Zhang (2017) use the model to calculate competition effects on fares, traffic volumes and social welfare, resulting in interesting policy implications. Wang, Sun et al. (2020) investigate the effect of fare changes on the competition between long-distance travel modes and uses the result to find mode-specific optimal pricing policies. Both studies prove that the model can provide accurate results, but it remains unpopular, which is reflected in the scarcity of studies applying it. This is likely caused by its complexity to use and interpret results.

Different studies try to predict demand through logic: Fröidh (2005) does exactly that for the Svealand line in Sweden, Dobruszkes (2011) for five European city-pairs and Diez-Pisonero (2012) for Spanish HSR lines. These studies are not meant to invent the wheel but can be used to confirm or explain findings from statistical methods. Time series models to forecast demand are developed by Cabanne (2003) and Li & Schmöcker (2014). The model assumes that explanatory variables for a phenomenon change over time. A wide variety of time series exist (Profillidis & Botzoris, 2019), but the method is very scarcely applied in HSR demand forecasting. Cabanne estimates demand by taking data from over a 20-year period. However, as is pointed out by the author: during this time, GDP, transport network development and fares have significantly changed. The time series model is applied for this reason.

Some studies design their own (non-linear and non-logistic) formulas and fit these to data, as for example done by Ben-Akiva et al. (2010) and Miyoshi & Givoni (2012), resulting in reasonable model fits. Data analysis and other statistical procedures are performed by Martínez et al. (2016) and Zhong et al. (2014). Sweden has designed their own demand forecasting model named *Sampers*; it is applied to a case study by Nelldal & Jansson (2010). Stated preference choice models are used by Burge et al. (2010) and Hsu & Chung (1997), while the latter study invents its own rendition. Inventing new methods specifically for a study occurs regularly. Kroes & Savelberg (2019) design and use a new so-called ‘substitution model’, Yang & Zhang (2012) a ‘competition model’, Wang, Jiang et al. (2020) a ‘connectivity utility model and Sánchez-Borràs et al. (2010) a new econometric model. Building further upon economic influences, Zhang et al. (2014) use the Lerner index. This number was originally meant as a measure of market power but is here used to calibrate a panel linear regression model.

2.2.5 Model Calibration

In demand forecasting, typically two methods are used commonly to estimate the model parameters. Both will be briefly addressed here.

- **Ordinary Least Squares (OLS)** minimises the squared difference between the observed and predicted values by the model. It does so by changing the values of the model's parameters, until an optimal set is reached. It typically works well with non-complex models and is the most popular in use due to its simplicity, while still providing accurate results (Petropoulos et al., 2022). More about the method is explained in Appendix B.
- **Maximum Likelihood Estimation (MLE)** is much more complex, maximising a likelihood function, which measures the quality of the data fit. It works well with both linear and non-linear models but requires more computation power as it assumes the model parameters follow a probability distribution. Unique optimal solutions are not guaranteed (Petropoulos et al., 2022).

2.2.6 Overview

This paragraph functions to show an overview of all mentioned high-speed rail demand forecasting models, and their rate of occurrence in related demand studies. This is shown in Table 2.5 below.

Table 2.5 Overview of models applied in HSR demand studies

Model	Focus		Studies (%)	Model type	Studies
	MS	PF			
Logistic regression	✓		47	Multinomial (MNL)	19
				Nested (NL)	18
				Binomial (BNL)	9
				Mixed (ML)	5
				Box-Cox	3
Linear regression		✓	32	Ordinary	11
				Panel	7
				DID	4
				Multiple	2
				Dynamic	2
				Random effects	2
Gravity		✓	6	Ordinary	5
				Dynamic	1
Alternative		✓	21	New models	5
				Vertical differentiation	4
				Case studies	3
				Time series	2
				Other formula fitting	2
				Data analysis	2
				Choice model	2
				Sampers	1

Focus: MS = Market Share, PF = Passenger Flow

2.3 Model-Specific Demand Impact Factors

Now the HSR demand forecasting practice has been explored regarding applied impact factors (section 2.1) and models (section 2.2), their combinations can be analysed. The latter section showed that only three models are used commonly; this section will therefore consider only these models.

Only studies that use one type of model are included for this analysis, as to focus on how each factor is used specifically for each model. Then, for each factor, the occurrence was measured in each model type. The results of this analysis are presented in Figure 2.7. The goal of this analysis is to help choosing demand-impact factors once the demand forecasting model type is chosen (section 3.2.2).

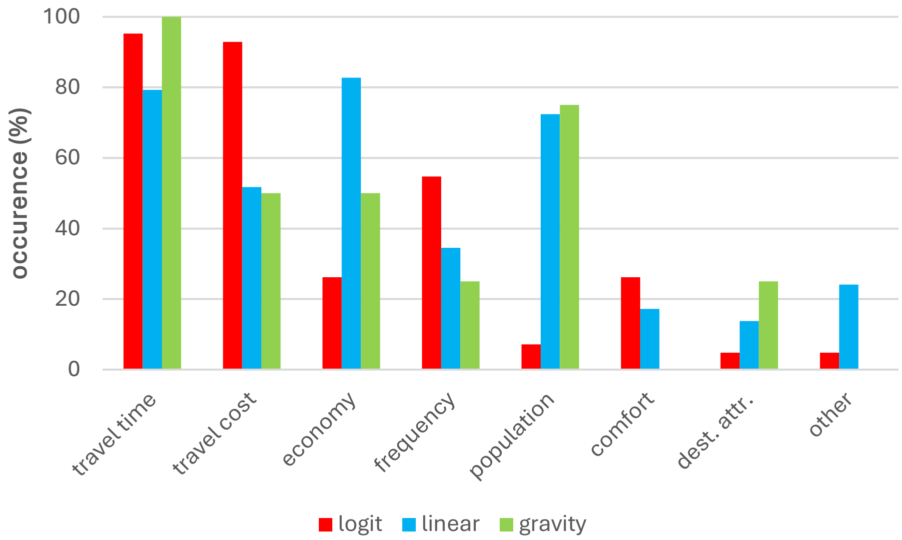


Figure 2.7 Impact factors usage in commonly used HSR demand forecasting models

The figure shows that some factors are very common to use some model types, but not in others. The best example for this is ‘population’, which is commonly used in linear regression and gravity models, but practically never in logit models. Travel time is the only factor popular among all model types. Travel costs are implemented in almost all logit models, but for other model types it is much less common. The ‘economy’ factors are mostly used in linear models. The analysis also shows that gravity models often are kept very simple: comfort-related and ‘other’ factors are not implemented.

2.4 Demand Evolution

The forecasting models evaluated in section 2.2 pointed out an extra interest of importance: the relationship demand and its impact factors could change over time. This dynamic dimension of demand will be addressed in this section - HSR demand forecasts must be accurate for the present and should retain this accuracy in the future.

High-speed rail demand can be divided into multiple parts, as is common practice in related literature (see works by e.g. Cascetta & Coppola (2011), Ben-Akiva et al. (2010), Russo et al. (2023)). This is presented in Table 2.6.

Table 2.6 HSR demand broken down into different flows (Cascetta & Coppola, 2011)

Type	Demand part		Cause example
Endogenous	**Diverted demand**	from other travel modes	e.g. shift from air / car / intercity
	Induced demand	direct	e.g. changes of travel characteristics
Exogenous		indirect	e.g. increase of mobility due to changes in lifestyles and land use
	Demand growth	economy-based	e.g. increase of mobility due to economic growth

2.4.1 Endogenous Effects

After opening of a new HSR line between two cities, it will immediately attract passengers that used to travel with other modes, such as flights, cars or other train services. This is called diverted demand. It will also attract new demand of people that were not travelling between the cities before. This phenomenon is called induced demand and can be caused by a change of any combination of factors reported in section 2.1. Well-reported in literature is a reduction in cost, increasing the demand. This follows the economic theory of supply and demand, reported from the perspective of travel demand by Noland (2001). The author provides a visualisation of this phenomenon, which is depicted here in Figure 2.8.

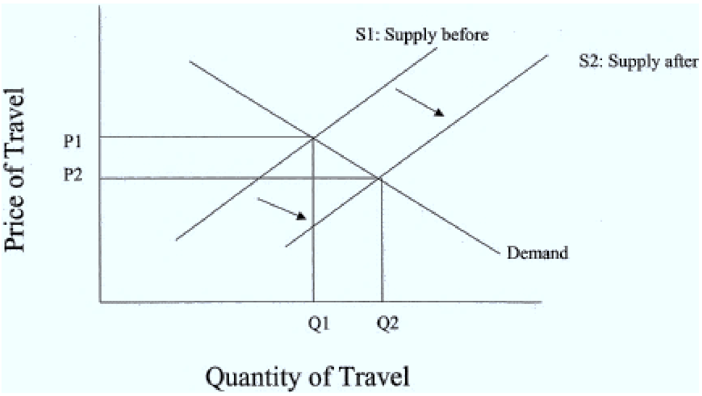


Figure 2.8 Induced demand (Noland, 2001)

Induced demand has been a reported often as an important factor of travel demand (Yao & Morikawa, 2005). It has a direct effect, but also indirect. Together, the diverted demand and direct induced demand determines the actual demand directly after opening a new HSR connection. Most demand models are not able to capture this kind of travel demand growth, since their estimated number of trips is inelastic to the quality of the travel environment (e.g. the level of service offered). According to Yao & Morikawa (2005), the level of induced demand is dependent on these factors especially.

2.4.2 Exogenous Effects

Induced demand, as mentioned in the previous section, also has an indirect effect since the opening of a HSR line can cause lifestyle and land use changes. This can be illustrated by the following example:

- A new high-speed rail station in a city makes the city more attractive to travel to, something which companies can respond to by settling there (land use change).
- A fast HSR connection to that city could then convince people to start working there (lifestyle changes), and thus increase the demand.

It explains why this part of induced demand is called 'indirect': these changes can take multiple years to gain momentum. But they can have great effects on the total travel demand (Yao & Morikawa, 2005). The authors report that indirect induced demand is caused by income and population growth, as well as differences in travel time, travel cost, access time and frequencies of modes offered.

Induced demand often is not considered by demand forecasting models, which typically underestimate demand by considerable margins. The level of induced demand has been reported to be significantly high; e.g. 35% of the original demand in France and Japan and even up to 40% in the UK (King, 1996).

The change of peoples travel behaviour is also caused by economic growth. As mentioned in section 2.1 and section 2.2, multiple models include this by means of GDP. A higher city GDP correlates with more generation of travel demand from that city, as it indicates a higher level of economic activity. In increase of GDP is considered a strong and popularly used indicator of economic growth and is typically expressed in a percentage point. For Europe, this percentage has been relatively stable up until the pandemic, indicating exponential growth (OECD, 2023).

The forecasting models evaluated in section 2.2 often only take the directly occurring effects in account, which are called endogenous. Cascetta & Coppola (2011) advise to also look at exogenous (indirect) factors, since they can have a large impact on (future) demand. This can be illustrated by the passenger numbers of the Eurostar service on the London-Brussels/Paris travel market over the years, in Figure 2.9.

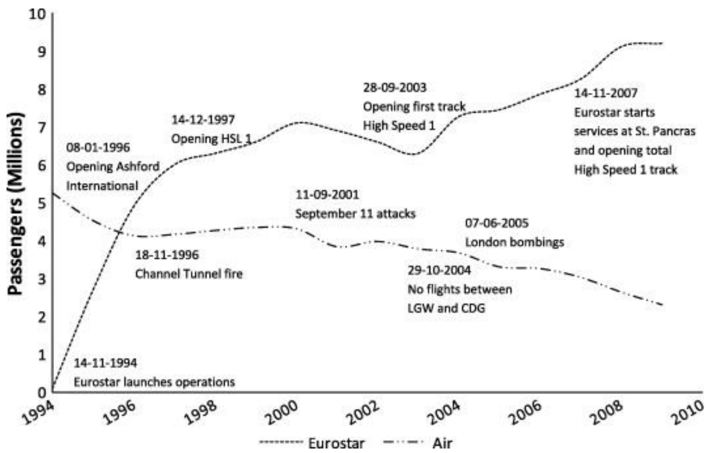


Figure 2.9 Passenger numbers for Eurostar and air services on the London-Brussels/Paris travel market, 1994-2010 (Behrens & Pels, 2012)

2.5 Profitability Breakdown

This paragraph dives into the second problem as mentioned in section 1.4: profitability forecasting. This will be the term used in this chapter, but it relates to the operational profitability and justification simultaneously. The next sections will each address one of the cash-flows, preceded by an overview.

2.5.1 Overview

Literature providing overviews of high-speed rail cash flows are scarce. Barrón et al. (2012) provide insights into this subject with great detail, applied to real European cases. This helps constructing a table of significant cash-flows that come into play during the lifetime of HSR projects. They are listed in Table 2.7. The obtained values are obtained from later-mentioned sources.

Table 2.7 Overview of HSR cash-flows (in 2024 euros) (Barrón et al., 2012)

Part	Types	Examples	Indicative value (low-high)
Infrastructure	Construction costs	Planning and land costs Infrastructure building costs Superstructure costs	€26.5 million / km (6.9 – 96.5)
	Operating & maintenance costs	Tracks Electrification Signalling Telecommunications	€100 thousand / km / year (45 – 110)
Rolling stock	Acquisition costs	Rolling stock	€37.5 million / train (20 – 107)
	Operating costs	Train operations Energy Sales & administration	€0.03 / seat-km
	Maintenance costs	Rolling stock & equipment	
Passengers	Revenue	Tickets	

From the table, it follows that the costs can be split in two parts, often regarded as track (infrastructure) and train (rolling stock). Each of these has a direct investment component (building or acquisition) and one running over the operation stage (operating and maintenance). It should be noted that with high-speed rail, also external costs emerge. Examples of these are land take, noise, air pollution, visual intrusion and barrier effects (Barrón et al., 2012). These costs cannot be measured and do not end up on the operator's balance sheet, and therefore do not count regarding operational profitability or (economical) justifiability. For this reason, these costs are here left out of the scope. The cash-flows in Table 2.7 can be attributed to two phases:

- **The construction phase**: here, all initial investments are made. It includes construction of the infrastructure and acquisition of the rolling stock.
- **The operational phase** relates to all costs regarding operation and maintenance, and revenue.

A thorough explanation of all cash-flows, an indication of their size and potential impact factors (if any) are provided in the next sections. Note: all monetary amounts have been adjusted for inflation, in order to be able to draw fair comparisons and predictions.

2.5.2 Infrastructure Costs

Infrastructure costs relate to any cost made in the lifetime of high-speed rail infrastructure. Initially when the infrastructure is built, there will be construction costs. After that, the infrastructure will be used, resulting in operating costs and maintenance costs, which are often referred to under the same heading in literature.

Construction costs

Among all initial expenses, these generally are by far the largest and therefore immediately have a large impact on the justifiability of the entire project. Barrón et al. (2012) investigated 45 constructed European projects and found that the total cost depends on the length and location. The authors report varying costs between 6.9 and 96.5 million euros, with an average of 26.5 million euros per kilometre. The results of their study are displayed in Figure 2.10 below.

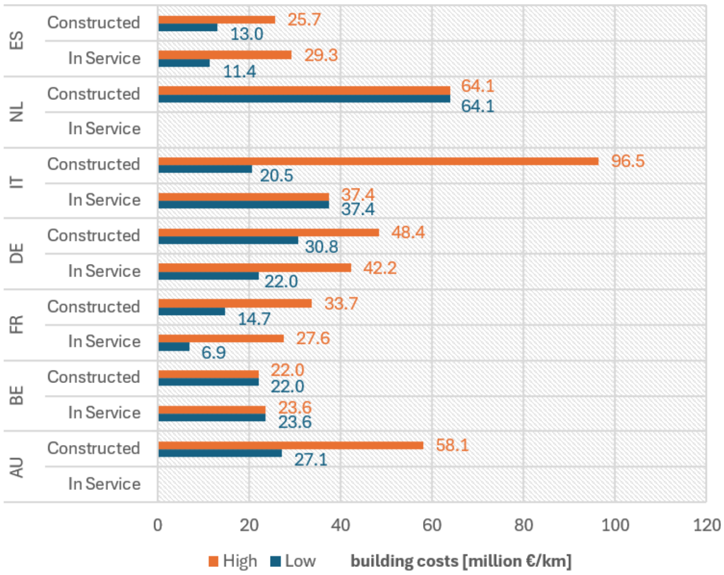


Figure 2.10 Building costs per kilometre for European HSR projects (adaptation of Barrón et al. (2012))

From the figure, it can be seen that construction costs vary wildly between countries, and less within the same country. This implies that the construction costs are heavily influenced by the country considered. UIC (2018) found that, in European practice, 1 km of new high-speed rail line costs between €18 and 48 million, a range that Belal et al. (2020) also find themselves in. Nash (2010) finds values between €18 and 60 million. France and Spain have the lowest construction costs, but these countries are also renowned for their very low population density in between cities. Trabo et al. (2013) indeed list population density as a main cost driver, along with elevation differences in terrain and the economic price level. UIC (2002) investigates the same subject and list four main cost drivers:

1. **Maximum design speed**. Faster high-speed rail connections require larger curve radii and therefore the amount of earthworks and civil structures such as bridges and tunnels increases. This effects massively in terms of costs, but not in terms of profitability. Higher design speeds result in more profitable HSR connections, with diminishing returns (Barrón et al. (2012), Belal et al. (2020) and Zhang (2024)).
2. **Infrastructure capacity.** By this, the purpose of the line is meant. HSR lines can also serve conventional rail of freight trains. More differences in use purposes and speeds among trains increase the construction costs, as mixed lines require extra safety measures. Normally, lines will be built double-track. Construction costs are likely to increase for mixed usage, along with the need for more tracks. For these reasons, the cost of HSR dedicated solely to passenger traffic is 20% lower than mixed lines (UIC, 2002).
3. **Power supply.** Most often, high-speed trains are powered electrically. Compared to alternative ways (e.g. diesel trains), this adds roughly 10% of costs due to investments in catenary and power supply (UIC, 2002).
4. **Exceeding scheduled time** also adds extra costs to the project. The construction time largely depends on the length of the longest tunnel and the environment (geology and population density).

The construction costs can be split into three components, following the definitions of UIC (2005): planning/land costs, infrastructure building costs and superstructure costs.

- **Planning and land costs** involve expenses made in the time frame between the birth of the projects' idea and the start of its construction. These costs thus include feasibility studies and technical designs but also land acquisition and other legal paperwork such as licenses and permits. The costs made here are represent a small but relatively constant 5-10% share of the total initial investment. The total planning and land costs typically depend on the size of the project (the length of line). More densely populated areas are generally harder to acquire as the land is more valuable, thus increasing the costs of the land acquisition part. In Europe, the earth works represent roughly 25% of total construction costs (Trabo et al., 2013).
- **Infrastructure building costs:** building a HSR line starts with terrain preparation and platform building. For clarity, this does not include the tracks, as they are part of superstructure. The infrastructure as described here is everything beneath the tracks. The costs of this part largely depend on the length of the line. In most cases, it represents 10-25% of the total building costs. The difficulty of terrain (see Figure 2.11) are known to increase costs, easily doubling the contribution to 40-50%. The civil structures contribute to 25% of the total construction costs (Trabo et al., 2013).
- **Superstructure costs** include the rails, and everything else built upon the prepared foundations: sidings, signalling systems, electrification mechanisms but also less conspicuous components such as communication and safety systems. In total, these costs usually represent 5-10% of the total costs (Barrón et al., 2012). The previously mentioned projects analysed by Trabo et al. (2013) accounted for €0.3 to 3.1 million in signalling costs (2-3%). Track construction costs between €1.3 and 6.2 million (8-10%).

Building new stations logically also induces costs. However, for this project it is taken as premise that new HSR infrastructure will be connected to existing stations in urban areas. Construction costs related to stations are therefore taken out of consideration for this project.

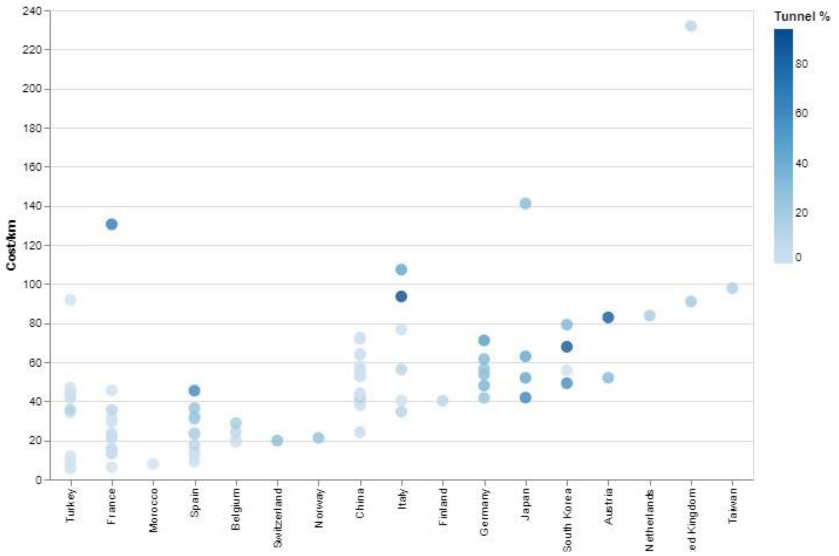


Figure 2.11 Construction costs of projects in different countries, showing their tunnelling percentage (Transit Costs Project, 2024)

The works by Trabo et al. (2013) and UIC (2002) show that the infrastructure costs largely depend on the project's length and environmental factors, but also point out that a number of design choices must be made up front. For this reason, the methodology in chapter 3 will start with making these decisions.

Operating & maintenance costs

Several of items mentioned in the previous section have to be operated: tracks and the electrification, signalling and communication systems. As opposed to the construction costs, the costs described here are not one-time investments, but expenses made on a day-to-day basis during the entire life span of the HSR line. All mentioned monetary values mentioned in this paragraph are yearly costs.

Most of the previously mentioned HSR components operated also require maintenance. Barrón et al. (2012) split these costs into five parts: maintenance of tracks, electrification, signalling, communication systems and other maintenance costs. The authors assess the costs in each of these categories for four European countries with an (extensive) high-speed rail network, as they are also mentioned in section 1.1.3: Spain, France, Italy and Belgium. Most of the operating and maintenance spending is done on tracks (40-67%), followed by the electrification system (8-19%), signalling system (10-35%) and telecommunications (4-17%). The most important takeaway from the authors research is that the total sum of operating and maintenance costs (per km) is very predictable. Nash (2010), UIC (2018), de Rus et al. (2020) and Railtech (2024) independently find values ranging between €100,000 and €110,000 per year, providing a more accurate representation of all costs.

Unlike construction costs, operating & maintenance costs are much less dependent on the location, or other factors of influence mentioned in the previous paragraph. This makes forecasting much easier.

2.5.3 Rolling Stock Costs

These costs represent a smaller share of the initial investment. Initially, the rolling stock must be bought (acquisition costs). During their lifetime, these trains must be operated and maintained, leading to extra costs. Both acquisition and operating/maintenance costs are addressed in the next two sections.

Acquisition costs

These are the most straightforward to determine, as they do not depend on numerous external factors and uncertainties such as found in construction costs. The acquisition costs simply depend on the choice of rolling stock. Janić (2017) makes an educated estimate of €55,000 per seat, which is a preferable unit to be able to compare different projects. Almujibah & Preston (2019) conclude that anywhere between €55,000 and €65,000 per seat is a realistic estimate to work with.

Even though faster trains are generally more expensive to buy, their acquisition cost is determined by numerous other factors (de Rus et al., 2020). These estimated costs per train are in line with estimations by last-mentioned work, who set a value of roughly €35 to €40 million, for a 350-seat train with an "economic life" of 30 years, based upon empirical research by UIC (2018). Trains reaching operating speeds of 350 km/h are even more expensive. Belal et al. (2020) found a range between €49 and €107 million per train set, based on different scenarios.

An important note to make here, is that the choice of rolling stock depends on the desired design speed and capacity, which relates back to the points made about these in section 2.5.2. This again stresses the fact that thoughtful decision making should be performed regarding this at the beginning of the related methodology part in chapter 3. Belal et al. (2020) dived deeper into the subject and found that a higher design speed increases the profitability of the high-speed rail line, despite the higher acquisition costs. This is in line with the findings of a HSR costs prediction model made by Barrón et al. (2012).

Another obvious indicator of the total acquisition costs is the number of trains bought. Seeking to find a good prediction of this, Belal et al. (2020) found a formula that fits real life data remarkably accurately. During validation, the maximum deviation from observed data was only one train, reaching a determination factor (R^2) of 98%. The formula is displayed in equation (2.6):

$$No. of\ Trains = \frac{63.287 \cdot AANP \cdot (6L + V)}{N \cdot V} \tag{2.6}$$

Here, $AANP$ is the annual average number of passengers (in million) for both directions combined, L is the line length (in km), V is the maximum operating speed (in km/h) and N is the number of train seats. Their model was validated for a wide range of its parameters.

It should be noted that equation (2.6) is statistically fitted to data regarding single lines instead of lines in a network. Also, it allows for less control since it simply takes the maximum operating speed, instead of an average, which tells much more about how the network is operated. Lastly, the number of passengers on a line is not easy to calculate in a network, as it encompasses flows from a great variety of OD pairs, while it remains unknown how many people use a certain line if multiple lines serve the same connection. In cases like these, it is much more convenient and straightforward to calculate the number of trains based on the operated frequency, as it leaves out complex estimations of demand and (average) operating speeds. Therefore, equation (2.6) is rewritten and simplified:

$$n_l \geq 2 \cdot f_l \cdot t_l \tag{2.7}$$

In Appendix E, the steps required to deduce this are specified. In equation (2.7), f_l is the operated frequency of the line in [trains / hour] on line l, and t_l is the travel time along the entire length of line l in [hours]. The value is multiplied by two in order to capture the full round-trip travel time as trains travel back and forth along the entire length of the line. The outcome is the number of trains n_l for line l.

Operating & maintenance costs

The total costs regarding this cash-flow depend on the number of trains bought. In literature, boundaries between operating and maintenance costs are not always defined clearly, resulting in great differences when comparing different sources directly. For example, Railtech (2024) uses 'operating cost' as an umbrella term that also include maintenance costs. Another research for a Californian high-speed railway estimated the full operating and maintenance costs per trainset per year at €11.3 million (Levinson et al., 1997).

Operating costs of trains largely depends on the energy required, which is specified for each train (Barrón et al., 2012). The authors find yearly operation costs in Europe varying between €20.5 -67 million, and yearly maintenance costs between €2.5-6.3 million, per train. In a more optimistic estimate, UIC (2018) estimates €1.2 million as yearly maintenance costs per train, based on the assumption that it travels 500,000 km each year. Nash (2010) finds values close to that of approximately €1.35 million.

As operating & maintenance costs for rolling stock mainly depend on the level of usage, a common unit of expression is the 'cost per seat-km', which Fröidh (2006) assess with great depth, finding strong relationships with operating speed and the number of seats.

Together, operating and maintenance costs directly depends on the number of seats displaced and the distance over which this is done, and is therefore typically estimated in units of euros per seat-km (Doomernik, 2017; Janić, 2017). Economies of speed apply: for faster operating speeds, the unit cost becomes lower (Kanafani et al., 2012). The same is true for economies of scale regarding the number of seats or seat density (Fröidh, 2006), who shows the dependency of unit costs for various factors of influence.

2.5.4 Ticket Revenue

Due to the complexity of predicting this figure, the literature review in section 2.1 provides special attention to this matter. For the sake of completeness to section 2.5, it is shortly addressed here. Revenue in high-speed rail is collected through ticket sales, simply depending on the demand and the price of a ticket (e.g. the 'fare'). As literature review pointed out these two factors are interrelated, operators generally set a price that maximises their passenger revenue; at least when the operator is not a governmental organisation (Qin et al., 2019).

Demand is inversely proportional to fares, as can be recalled from section 2.1.3. This means that setting a higher fare results in lower demand. However, the people that would still travel despite the higher price, pay more. Since revenue is determined as demand multiplied by the fare, this could result in higher or lower revenue. As mentioned, the primary goal of a rail operating company is to make and maximise profit. It is evident that fares and demand are two counteracting forces that need to be quantified in order to understand how the revenue can be optimised. Figure 2.12 describes the inner workings of this mechanism. For the sake of simplicity, the relationship between demand and price in the figure is considered linear. In the lower graph, the price can be set, influencing the demand. The revenue can be visualised; as it is the product of demand and price, it is represented as the area of this square. When plotting all price-revenue combinations, the upper graph will be attained, showing a parabola with optimum. This optimum represents the price-revenue combination with the highest revenue and is indicated with a red dot. The values of the optimal price and revenue can then be read from the axes.

2.5.5 Dynamic Component

As already indicated by section 2.4, it is crucial to take dynamic components into account – every factor or variable can change over time. For this reason, works by Belal et al. (2020) and Barrón et al. (2012) take evolving costs over the entire lifespan of a high-speed rail project into account. This lifespan takes the evolution of HSR technology into account; after a certain amount of time, the technology used in a built line will be outdated, and then there will be a demand for other, new lines that are up-to-date with the contemporary technology at that time. For example, the oldest HSR line in the world, Japan's Shinkansen lines, were built in 1964 and have been largely improved at least twice (Hood, 2006). Thus, when assessing the profitability or justifiability of a HSR line, it should only look at cash flows within its lifespan. Therefore, a project's planned lifespan must be set prior to any feasibility study.

The exact length of this is debated in literature, since it is hard to predict the evolution of HSR technology. Here, Belal et al. (2020) are inspired by Barrón et al. (2012) and both use a lifespan of 40 years. This roughly matches the life expectancy of rolling stock, which lies between 30 and 50 years (EEA, 2020).

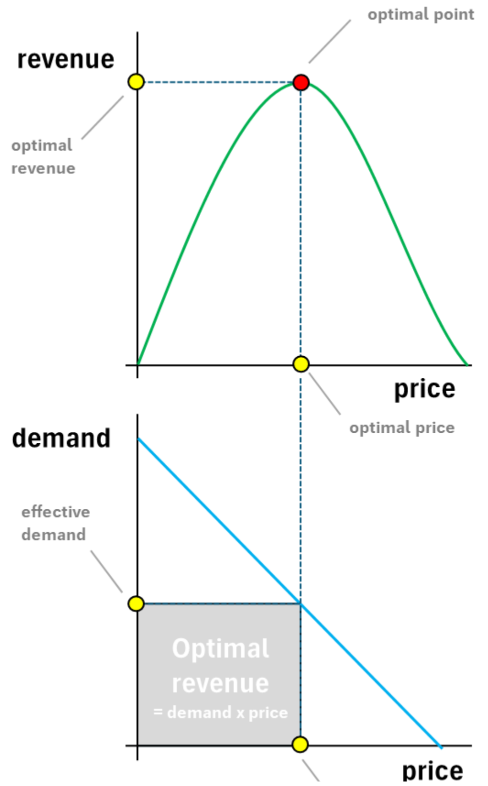


Figure 2.12 Maximisation of revenue by changing fares (adaptation of May et al. (2022))

2.6 Transport Network Design Problem

This literature review will provide insights into the Transport Network Design Problem (TNDP) and its related state-of-the-art. It can be seen as a more in-depth version of section 1.2.3, which already partly looked at related studies regarding this very problem. The general philosophy and an introduction of the TNDP will be given in section 2.6.1, after which a look is taken into the different ways that related problem formulations are applied in literature.

2.6.1 Introduction

The definition of the Network Design Problem (NDP) is provided by Feremans et al. (2003): it comes down to "finding the optimal subgraph of a graph, subject to side constraints". The meaning of these terms will be clarified later on in this section. Well-known variations of the problem are finding the shortest path from A to B (Shortest Path Problem) or finding the shortest round-trip along a pre-defined set of locations (Travelling Salesman Problem). The problem is applied in a wide range of scientific fields, ranging from biology to telecommunications. In the field of transportation, this concept is well-studied and is referred to as the Transport Network Design Problem (TNDP). It is applied to solve different transport-related problems such as optimal allocation of increased link capacity or optimal scheduling of maintenance in combination with link closures. Applications vary from road networks to public transportation (Dixit & Niu, 2023). In this project, the focus lies on one certain application of the TNDP: finding the optimal configuration of (new) links. Figure 2.13 illustrates the nature of this design problem.

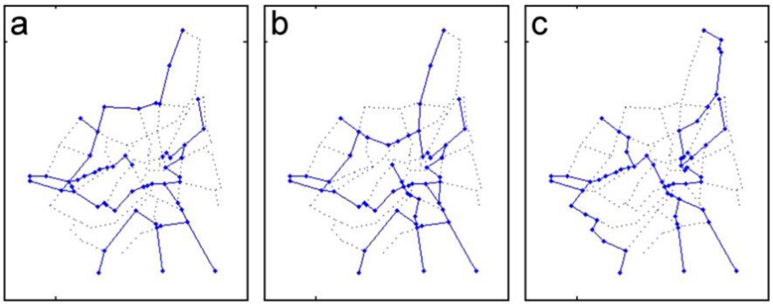


Figure 2.13 Three different solutions of a Transport Network Design Problem (Erkut & Gzara, 2008)

The figure above shows three different solutions of a problem with 35 city-pairs, each arc represented by a line between two dots. Referring to the original definition of the NDP by Feremans et al. (2003): the *subgraph* is formed by the set of selected city-pairs (solid lines) and associated cities (solid dots), while the *graph* encompasses all city-pairs (dotted and solid lines combined) and all cities.

It can be observed that there is a large number of possible subgraphs. The supply is determined by the choice of links, which on its term influences the demand. For example, compared to a direct connection, an indirect connection between two cities will still serve a certain level of demand, but most likely lower due to increased travel time and costs. In the most extreme case, a city-pair may not be connected at all, thus setting its effective demand to zero. The feedback between demand and service level is another well-known factor ('elastic demand', section 1.2.3), extremely complicating the problem (Dixit & Niu, 2023).

2.6.2 Mathematical Definition

The TNDP related to this project considers an undirected graph $G = (N, A)$, where N represents the set of cities and A represents the set of connections between them. Each arc $a \in A$ has certain attributes, such as the travel time or cost related to traversing it. For each arc, the decision can be made about whether it must be built or not. For this reason, a binary decision variable y_a is included. If the arc is built, y_a is set to 1 and zero otherwise. Most often, the selection of arcs is not an entirely free choice – it usually comes with a number of constraints. Section 2.6.4 provides a broader focus on this. One may think of a maximum number (or length) of selected arcs.

The selection of arcs a form a new set $A^{sel} \subseteq A$. The nodes associated with the arcs $a \in A^{sel}$ also form a new set $N^{sel} \subseteq N$. Together, they form a new graph $G^{sel} = (N^{sel}, A^{sel})$, which then logically is a subgraph of the original graph $G = (N, A)$. Each possible subgraph $G^{sel} \subseteq G$ has an associated 'value' Z which is driven by specific objective function of the mathematical model. The goal of the problem is to find the optimal subgraph, while adhering to all constraints. In this project, the goal is to maximise the network's lifetime profit, but numerous other objectives can be used. Section 2.6.3 provides a broader focus on this.

Generally, the input of the model is a demand matrix stating the demand from every node to every other node and an original graph G, where it is assumed that every passenger chooses the shortest route while only taking in-vehicle times into account (Durán-Micco & Vansteenwegen, 2021).

A promising recent development regarding optimal TNDP solving is the adaption of the Multi-Commodity Flow Problem (MCFP) formulation. Marín & García-Ródenas (2009) already observed that this approach is required when considering passenger flows for each OD pair. The OD pairs are then seen as commodities, which leads to an "efficient formulation" to handle "city-scale transit networks" (Ng et al., 2024). However with current practice, the formulation suits small instances of the problem only, due to the added complexity by OD pair specific flows (Gutiérrez-Jarpa et al., 2017), a problem overcome by last-mentioned authors by splitting the problem in two. First, they optimise for the network's topology only, based on the OD pair flows. Then, they optimise for the design of lines over this topology.

2.6.3 Objectives

Objective functions take various forms in applications of the TNDP. Based on their perspective, they can be split into two groups: user or operator. They both value the quality of a solution differently. Extensive reviews on TNDP studies are provided by Kepaptsoglou & Karlaftis (2009), encompassing 62 studies published between 1967 and 2007, and by Durán-Micco & Vansteenwegen (2021), for 30 studies ranging from 2009 to 2021. Analysis of these review papers shows that more recently, the focus has shifted towards optimisation from a user perspective rather than the operator's point of view. The following two sections will briefly address each of them.

User perspective

As found in literature review in section 2.1, travel time is the most important demand-influencing factor. This is reflected in the choice of objective. Durán-Micco & Vansteenwegen (2021) investigated thirty studies where TNDP was applied. Among them, twenty-seven (90%) opt for *total travel time*, which is the travel time summed over every individual passenger, in line with the user-oriented focus that TNDP studies generally have currently. A much alike metric used is total travel time savings, for example by Gutiérrez-Jarpa et al. (2017). Adapting the operators cost-focus to users, Iliopoulou et al. (2019) minimise the total user costs. Cadarso & Marín (2016) and Cadarso et al. (2017) minimise the maximum negative effects caused by a single disruption, thus finding the best worst-case scenario. In an attempt to maximise user-friendliness, Suman & Bolia (2019) maximise directness (in their context: the share of passengers not having to transfer). While most studies choose to serve all demand as a constraint, Yoon & Chow (2020) try to maximise the demand served.

Operator perspective

Among the studies investigated by last-mentioned authors, an operator perspective is much less popular. When chosen, the objective from this perspective is to maximise profit (Guihaire & Hao, 2008). This is in line with the findings of literature review in section 2.5: operators generally want to satisfy the demand while spending a minimum amount. The line length was found to be the most influential factor affecting costs. Speaking of which, multiple studies choose to minimise costs (Iliopoulou et al., 2019; Heyken Soares et al., 2020), construction costs (Gutiérrez-Jarpa et al., 2017) or operating costs (Yoon & Chow, 2020).

2.6.4 Constraints

Constraints are related to either the network's performance or limited resources (Fan & Machemehl, 2006a); (Guihaire & Hao, 2008). Due to the complexity of the problem, studies attempting to find solutions generally include a low number of non-complex constraints. For this same reason, most studies assign a fixed number of passengers to the shortest route (the so-called 'All-Or-Nothing-approach'). This leads to overestimation of passenger numbers as in reality, demand depends heavily on the level of service (e.g. travel time, travel costs, see section 2.1). Recall that this is 'elastic demand'. There are two types, as defined by Lee & Vuchic (2005); either the total demand for all modes together is kept constant while the demand per mode is elastic, or a varying total demand is considered.

Including any of the two types into TNDP solving further complexifies an already very complex problem, thus the vast majority of studies choose to work with a fixed demand ((Kepaptsoglou & Karlaftis, 2009); (Durán-Micco & Vansteenwegen, 2021)).

Some TNDP formulations already consider lines (a series of connected arcs and nodes) and logically set a minimum and maximum bound regarding what the number of connected stations in one line may be (Durán-Micco & Vansteenwegen, 2021). The authors note that literature makes use of varying constraints, which makes it hard to compare the quality of solutions. In the most basic formulation of the TNDP, it is constrained that lines can visit each node at most once and that all demand must be served (Kepaptsoglou & Karlaftis, 2009); (Guihaire & Hao, 2008). As stated before, the user-perspective has become more popular, meaning that operator-related metrics are used as constraints. Examples are related to fleet size, operator costs and other operator-related budgets, maximum line length (for one connection or the full network), capacity on lines and other constraints on the network's topology (Durán-Micco & Vansteenwegen, 2021).

An older, more extensive review by last-mentioned authors finds even more variations, such as the number of connections and lines, load factors, fleet availability, demand characteristics and related patterns, travel times for individual lines or the total travel time of all passengers and connectivity of nodes.

2.6.5 Solving Methods

As mentioned before, the TNDP is considered to be a very complex problem. Therefore, it is typically solved by means of (meta)heuristic methods (Durán-Micco & Vansteenwegen, 2021), which do not necessarily find the optimal solution, but an acceptable solution in relatively short time.

Exact methods

Only 10% of studies reviewed by the authors use an exact method. This method provides the optimal solution but can be very time-consuming. These studies often simplify the problem significantly by consider only a limited number of cities and connections and leaving out demand elasticity. Cadarso & Marín (2016) consider a very small, fictive network of 9 nodes, 30 arcs and 1,044 passengers. They define a TNDP formulation which minimises the maximum negative effects caused by a single disruption, finding the best worst-case scenario. They are able to produce exact results, which allows them to compare the results of different model formulations as their findings are optimal. The work of Cadarso et al. (2017) is a follow-up of this, based upon the same network, also focussing on risk management in a finite set of possible disruptions. They however compare even more exact models to calculate the optimal value, while their objective function considers a sum of different weighted costs. Their method finds optimal solutions within several seconds on an average personal computer. Remarkably, the studies working with exact methods do not use popular factors such as 'total travel time' or 'network length' as objective. Apparently these standard objectives do not work well with exact solving methods.

Compared to these first two studies, the work of Gutiérrez-Jarpa et al. (2017) is much more complex. It solves for three different objectives simultaneously: it minimises construction cost (an operator perspective), maximises time savings of the passenger flows (user perspective) and maximises patronage. The network is more complex: it encompasses 108 nodes, 3,789 arcs and 360,000 passengers. Solving the problem requires much more time but remains executable: several hours. Borndörfer et al. (2007) and a very recent work by Ng et al. (2024) transform a multi-commodity flow problem formulation into a TNDP, which allows them to efficiently optimise the choice of added links, while ensuring that all demand flows optimally over the network's links, while all demand is satisfied. Their model incorporates TNDP with the choice of operated lines (train routes) and solves to optimality fast, for relatively small problems. With some major simplifications, Borndörfer et al. (2007) are able to solve for a network with 410 nodes.

(Meta)heuristic methods

Other methods found by the authors in studies are most often metaheuristics: these are known for finding a reasonably good solution relatively fast, while not being designed for a specific model - hence their popularity over exact approaches. Guihaire & Hao (2008) classify applied (meta)heuristics into four 'big families': neighbourhood search, evolutionary search, hybrid search and greedy metaheuristics. The most popular are evolutionary algorithms, applied by 37% of studies. Even within this group, the exact approach varies from study to study. Feng et al. (2019), Heyken Soares et al. (2019) and Yang & Jiang (2020) for example, make use of a genetic algorithm where the bus routes are viewed as 'chromosomes'. The algorithm iteratively mutates and swaps the chromosomes, seeking for the lowest total travel time. For the latter work, the method is adapted to take two objectives into account simultaneously: optimising from both the operators' and user's perspective. Oliker & Bekhor (2020) pre-define a set of passenger routes for each OD pair, allowing to solve a network of 903 nodes, 2975 arcs and 5394 OD pairs with an bi-level optimisation algorithm, solving for the same network much faster than existing formulations.

Other studies employ a wide variety of solving methods; there is more experimentation with solving methods than objectives. Some studies use population-based algorithms inspired by natural phenomena, which are becoming increasingly more popular in TNDP solving (Iliopoulou et al., 2019). For example, Nikolić & Teodorović (2013) find the best selection of arcs by Bee Colony Optimisation (BCO), a technique that converges to the best solution by iteratively trying new solutions while making choices based on the increasing knowledge of the population. It owes its name to the process of bees searching for food. The authors show that the solutions can be of high quality and competitive with other algorithms. Kechagiopoulos & Beligiannis (2014) use the Particle Swarn Optimisation (PSO) technique, which is based upon the behaviour of social behaviour of bird flocks or fish schools. Both BCO and PSO are algorithms inspired by natural phenomena, but while a BCO solution learns from every decision made in the past, a PSO solution only does so from its own and nearby solutions. Fan et al. (2019) make use of an improved Flower Pollination Algorithm (FPA). Unlike BCO and PSO, FPA does not involve communication among solutions, but more attractive solutions are more likely to exchange information to improve. This is based upon the natural phenomena that flowers with attractive features are more likely to attract pollinators.

Metaheuristics applied in studies vary substantially. Other examples are: simulated annealing and the hill-climbing algorithm (Fan & Mumford, 2010), tabuu search algorithm (Yao et al., 2014), hyper-heuristics (Ahmed et al., 2019; Heyken Soares et al., 2020), stochastic beam algorithm (Islam et al., 2019) and learning-based route generation (Yoon & Chow, 2020).

2.7 Frequency Setting Problem

This literature review section will provide insights into the Frequency Setting Problem (FSP) and its related state-of-the-art. It can be seen as a more in-depth version of section 1.2.3, which already partly looked at related studies regarding this very problem. This section starts off by introducing the lines and frequencies, before stating their definition in section 2.7.2 in order to understand the nature of the problem the FSP solves. Section 2.7.3 then uses the definitions to mathematically define the problem. The next three sections then provide insights into how the problem is solved in literature; stating objectives (section 2.7.4), constraints (section 2.7.5) and solving methods (section 2.7.6).

2.7.1 Introduction

The TNDP introduced in section 2.6 only looks at which direct connections to be built between cities. However, this does not address how the network must be *operated* in order to sufficiently handle the demand. Generally speaking, transport networks are served by *lines*, each having their own designed *frequency*. The optimal incorporation these two factors into a transit network design is known as the Frequency Setting Problem. As denoted by Ceder & Wilson (1986) and Guihaire & Hao (2008), setting frequencies forms the second step of the overall public transport planning process. As they crucially impact operator costs, it is evident that they must be taken into account when evaluating candidate networks. For this reason, the FSP and TNDP are often solved simultaneously, in a so-called Transport Network Design & Frequency Setting Problem (TNDFSP) (Durán-Micco & Vansteenwegen, 2021). Below, Figure 2.14 illustrates the difference between ordinary network design, and when line design is added.

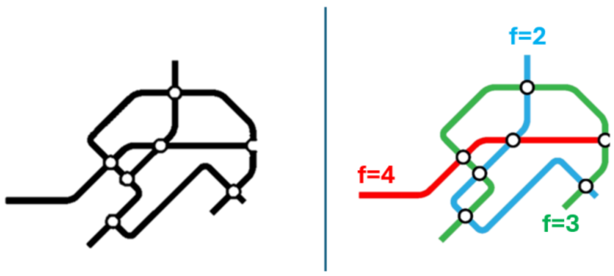


Figure 2.14 Visualised results of a TNDP (left) and TNDFSP with frequencies in terms of trains per hour (right) for a simple network

2.7.2 Lines and Frequencies

As explained, the Frequency Setting Problem involves making decisions about two aspects: lines and their operated frequencies. Both are defined as follows:

Line

A transit line is a set of connections and stations, which are visited in order by one specific train. Thus, they describe the routes trains will travel over when serving the network. On metro maps, they are generally indicated by a coloured line running through a number of stations. A station can be served by multiple lines, which often is the case for stations at central or other important locations. However, all stations of the network must be served by at least one line, otherwise they cannot be served. On a line, trains generally run from one terminus to the other before returning. In some cases, they run a circular path. Whether passengers need to transfer depends on whether their origin and destination are both on a transit line.

Line frequency

Each transit line comes with a designed frequency. This is an integer number, in this case often denoted as the number of trains departing from a station as part of that line, per hour. The interval between these passing trains most often is constant and an integer amount of minutes, called the *headway*. For this reason, not all possible frequencies are used in practice (Gallo et al., 2011). The frequency f_l in trains per hour thus depends on the headway h_l in hours and is specific for line l.

$$f_l = \frac{1}{h_l} \tag{2.8}$$

Like any transit network, a high-speed rail system is designed to be able to serve all demand. Since the peak hour is normative for the system load, the line frequencies are calculated accordingly. Therefore, the minimum needed frequency f^{min} in trains per hour is determined by the peak hour passenger flow during the day Q^{peak} and a train's passenger capacity s following the equation:

$$f^{min} = \left\lceil \frac{Q^{peak}}{s} \right\rceil \tag{2.9}$$

Here, both Q^{peak} and f^{min} consider passenger traffic in one direction only. Note that the right-hand side has to be rounded up, as frequencies are integer numbers and a lower frequency would not serve all demand. Hence the ceiling brackets on the right-hand side of the equation.

While the number of seats is simply dependent on the choice of train set, the peak hour passenger flow is harder to determine. Literature often makes use of a conversion ratio between peak hour and average flow. This ratio depends on numerous factors, but generally lies between 1.6 and 2.0 (Yue et al., 2023).

2.7.3 Mathematical Definition

The definition of the FSP is provided by Martínez et al. (2014). Similar to the TNDP definition, the frequency setting problem considers a directed graph $G = (N, A)$, where N represents the set of nodes (or vertices, in this context: cities) and A represents the set of arcs (or links, in this context: HSR connections) connecting these nodes. These schematically represent the potential movement of trains.

Now, the frequency setting problem adds a set of lines L to this. Each line $l \in L$ consists of a set of adjacent arcs. Each arc has a passenger flow Q_a, defined for all $a \in A$, originating from the solution of the TNDP. The set L encompasses many lines, much more than eventually chosen, as the goal is to select the optimal subset $L^{sel} \subset L$. For this reason, an integer decision variable f_l is included which denotes the frequency, defined for each line $l \in L$. If the line is not chosen, f_l is simply set to zero.

2.7.4 Objectives

An extensive FSP literature review is provided by Durán-Micco & Vansteenwegen (2021), for thirty studies ranging from 2009 to 2021. Analysis of these show objectives vary much less for FSPs than for TNDPs. Most commonly, frequency setting problems optimise operator's and/or user's costs, as they are highly influenced by the set frequency (Durán-Micco & Vansteenwegen, 2021). Higher frequencies indicate a greater fleet size, which this on its turn directly imposes operating and maintenance costs for both infrastructure and rolling stock, thus increasing operator's costs significantly. On the other hand, higher frequencies indicate lower waiting and transfer times for passengers, which reduces user's costs. Thus, user's and operator's costs are in conflict (Kepaptsoglou & Karlaftis, 2009). For this reason, most studies formulate an objective function that includes factors representing both sides and call it 'social welfare', as a one-sided optimal solution is not socially desirable. The few studies that choose one side in the objective function, generally address the wishes of the other in the formulation's constraints.

User's costs are typically addressed in the form of (total) travel time - thus including in-vehicle, waiting, transfer, egress and access time (see section 2.1.2). In some studies, (the number of) transfers are also penalised. Studies implement this into their objective function by expressing it in a monetary value. Their reasoning is that a transfer adds travel time, as perceived by the passenger. Then, the Value of Time (section 2.1.2) allows to attach a monetary value to penalise for the increase in perceived travel time. Based on a stated preference analysis, de Keizer et al. (2015) conclude this increase depends on numerous factors.

They set a value of 22.63 minutes in extra perceived travel time if a trip is not without transfer. The real transfer time is included in this value as well, allowing for calculation of monetary valued penalties for every passenger who has to transfer.

The operator's costs are usually represented by simply the number of train sets needed to cover the designed network, along with its lines and their frequencies.

2.7.5 Constraints

In literature, frequency setting problem formulations impose a wide variation of constraints. But most prominently, all demand on the network must be served, as otherwise it would not be regarded as a valuable solution (Canca et al., 2018). The design of lines and setting of their associated frequencies is namely guided by the flow of passengers over the network (Kepaptsoglou & Karlaftis, 2009).

Lines

The selection of lines generally is the first step. As the number of potential lines would otherwise become very large, most studies predetermine lines (e.g. Asadi Bagloee & Ceder (2011)), and subject them to numerous constraints to decrease the size of the set. The whole set of lines could be predetermined, or in some cases only the terminal nodes, or only the total number of lines (Durán-Micco & Vansteenwegen, 2021). The terminal nodes of the line can be constrained to be located on end points or transfer points within the topology of the network only (e.g. the left side of Figure 2.14), which last-mentioned author considers a more innovative approach, and this is implemented by Gutiérrez-Jarpa et al. (2017). Ordinary models cannot solve for realistically large networks due to the complexity of accounting for competition for every OD-flow. However, last-mentioned authors overcome this by first optimising the network's topology, then optimising the design of lines and their frequencies on this topology.

Constraints can also be imposed on the shape of the route in between the terminal nodes, for example:

- a minimum and/or maximum length/duration of the line (practical guideline),
- a minimum and/or maximum number of stops (practical guideline),
- a certain level of route directness, meaning that the routes are as straight or short as possible between the terminal nodes,
- and the degree of overlapping with other lines.

Frequencies

To constraint that the network must serve all demand, minimum frequencies are imposed for each connection, depending on the peak hour flow on that connection. Another constraint in this regard often imposed is a minimum frequency for all arcs, as it could be desirable in case the peak flow is relatively low, since waiting times would otherwise become intolerable. Equation (2.9) already addressed this interaction. Frequencies have upper boundaries as well, which depends on the minimum headway between successive trains allowable in operation (see equation (2.8)). As already discussed in section 2.7.2, line frequencies are constrained to be integer and nonnegative.

2.7.6 Solving Methods

In literature, TNDFSPs are solved by either exact methods or metaheuristics. Similarly to TNDPs, the exact approach is uncommon and used in 1 out of every 10 studies. Durán-Micco & Vansteenwegen (2021) analysed 48 studies; only five solve the problem by using an exact approach. The reason for this unpopularity remains the complexity of the problem, which is even more pronounced for TNDFSPs than TNDPs. Nowadays, metaheuristics are perfectly capable of finding a good solution fast. This section will show some examples of studies opting for either one of the solution approaches.

Exact

Even though the number of exact approaches in literature is small, they differ in approach. Zhang, Yang, Wu, et al. (2014) optimise a multi-modal network that optimises an integrated bus- and car network. To do so, they formulate the problem as a TNDFSP while optimising for both the operator and user's costs. They solve it efficiently by implementation of the Genetic Algorithm (see section 2.6.5) within a few seconds, attaining improved solutions compared to previous methods. Cancela et al. (2015) finds a new Mixed-Integer Linear Programming (MILP) formulation for a bus network and solves the problem successfully for relatively small networks. The authors emphasise that research on larger networks would benefit from solving with algorithms instead, as their network of 84 nodes and 143 arcs takes over four hours to solve. Liang et al. (2019) optimise a model under uncertainty (for travel preferences, travel times and demand) in two steps: first, they apply a column generation algorithm to identify passenger paths and transit lines. Then, they optimise for line frequencies and passenger flows by use of a linear program. Ranjbari et al. (2020) solve a full TNDFSP by use of pre-defining candidate lines and stations. Their research involves a sensitivity analysis, which finds that the ratio in which demand is met impacts the objective function's value the most. Zhou et al. (2021) use the very uncommon approach of formulating a non-linear model for an already complex problem, but linearise it by approximation. Despite this, they are able to produce sensible results with room for improvement.

In conclusion, it can be said that exact approaches can solve the problem, as long as the formulation is kept linear and the size of the problem is kept relatively small.

(Meta)heuristic methods

The same metaheuristic methods used for solving TNDPs can be found in TNDFSP-related literature. Durán-Micco & Vansteenwegen (2021) analysed 48 studies. More than half of these (54%) used evolutionary algorithms to solve. The most commonly used type falling under this is the Genetic Algorithm, as used by Asadi Bagloee & Ceder (2011) and Bourbonnais et al. (2019), who were able to optimise for medium sized urban transport networks. Building upon this, Cipriani et al. (2012) use a Parallel Genetic Algorithm (PGA), meaning that they implement two genetic algorithms at the same time, making optimal use of computation power. The authors conclude that the method is "robust and effective in producing reasonable solutions", with room for further decrease of computational times.

Nature-inspired metaheuristics are just as popular for solving TNDFSPs as they are for TNDPs. The same kinds can be found in TNDFSP literature. Common examples are ant colony optimisation (Yu et al., 2012), bee colony optimisation (Nikolić & Teodorović, 2014), fish swarm optimisation (Liu et al., 2020), swarm optimisation without (Iliopoulou, Tassopoulos, et al., 2019) or with multiple search strategies (Jha et al., 2019), intelligent water drops (Capali & Ceylan, 2020) and cuckoo search (Sadeghi et al., 2020).

2.8 Literature Gaps

This section will summarise the main research gaps found in literature, which need to be bridged for this project. This will be done in the shape of three paragraphs, one for every problem listed in section 1.3.2. These paragraphs summarise the findings in the respective sections.

Demand forecasting (section 2.1, 2.2, 2.3 and 2.4)

This field is well-researched. In order to predict the level of demand, a wide variety of models and factors are implemented. Due to the fact that opinions are so much divided about accurate demand forecasting, it can be considered a gap in literature on what method works best in general, but this is a whole thesis topic on its own and is therefore considered out of this project's scope. This project therefore aims to attain reasonably accurate demand forecasts, as it does lie within the core of what this project hopes to offer scientifically.

Profitability estimation (section 2.5)

As pointed out by section 2.5, the profitability of a HSR connection depends on numerous cash-flows, with each their own influence factors. Some of their values rely on simple rules of thumb, with accurate results (e.g. acquisition costs). Other factors, such as infrastructure construction costs, depend on numerous other factors and can be unpredictable (see section 1.2.1). This can be considered an important gap in literature, as the reporting on this is scarce, similarly to the accurate prediction of HSR profitability. For this reason, profitability is only scarcely included into network design LP formulations. Network optimisation while taking profitability into account is therefore considered an important research gap. This project will pay attention to by breaking down profitability, in order to fill this gap. Also, it aims to attain reasonably accurate profitability assessments, as it does lie within the core of what this project hopes to offer scientifically.

Network design (section 2.6 and 2.7)

Regarded as one of the most complex transportation problems to solve, the TNDP and TNDFSP are typically solved using heuristic methods (Guihaire & Hao, 2008), finding good but not optimal solutions within a reasonable time frame (section 1.2.3). Due to the problem's complexity, exact optimisation techniques are rarely used (Reeves, 1995). Studies that despite this still opt to face the challenge, make various compromises to keep their solving times within reasonable bounds. Section 2.6.5 pointed out they keep the number of nodes, arcs, passengers and/or OD pairs low. Also, they assume a fixed (inelastic) demand, despite it being identified as a crucial factor in order to design sound transport networks (Guihaire & Hao, 2008). Demand elasticity is considered one of the main factors complexifying the problem (Jiang et al., 2014), and therefore not accounted for by studies when applying an exact mathematical approach to a medium-to-large sized network (Kepaptsoglou & Karlaftis, 2009; Durán-Micco & Vansteenwegen, 2021). This puts the scientific value of optimal transport network designs, found through mathematical optimisation, under scrutiny.

In recent developments by Yan & Chen (2002) and Borndörfer et al. (2007), the Multi-Commodity Flow Problem (MCFP) formulation was able to produce optimal solutions to larger instances of the problem. It was found that a MCFP-based formulation is required to assess the values of individual OD pair flows, and to incorporate elastic demand (Marín & García-Ródenas, 2009). This is seen as inspiration and starting point for the formulation this project will build to bridge the literature gap.

It is evident that with current practices in literature, medium- to large-sized realistic and optimal transport networks cannot be designed with exact methods (Murray, 2003); (Guan et al., 2003), which is also evidenced by the literature review of Guihaire & Hao (2008). This project will attempt to bridge this gap, and account for demand elasticity in linear programming in larger instances of the problem. This can be linked back to the goals of this research as described in section 1.2.3, and is considered as the core of what this project hopes to offer scientifically.

3 Methodology

AI-generated high-speed train (Alam, n.d.)

3 Methodology

This chapter provides details the three problems mentioned in section 1.4, and therefore paves the road to provide answers to the research questions mentioned in section 1.3.2. As recommended by the literature review in section 2.5, it should start off by making some important design decisions, which have a great impacts on the design's nature. This is done in section 3.1. Then, each of the problems are addressed here separately in their own section, chronologically: demand forecasting (section 3.2), profitability estimation (section 3.3) and network design (section 3.4).

3.1 Foundational Premises

Decisions regarding any of the key points that arose in chapter 1 will be made here, accompanied by the necessary substantiation. These are foundational decisions that shape the overall framework and nature of the methodology. They serve as guiding principles or initial conditions upon which the design is based.

Lifetime

As indicated by section 2.5.5, the exact length is debated in literature, since it is hard to predict the evolution of HSR technology. Here, Belal et al. (2020) are inspired by Barrón et al. (2012) and both use a lifespan of **40 years**. This roughly matches the life expectancy of railway vehicles, which is between 30 and 50 years (EEA, 2020).

Design speed

Section 2.5.2 indicated higher design speeds result in more profitable HSR connection, with diminishing effects. This has been shown in theory by Barrón et al. (2012) and Belal et al. (2020) as well as in real life data (Zhang, 2024; Want China Times, 2013). For these reasons, researchers opt for a design speed of **350 km/h**, which matches the current maximum speed achieved by high-speed trains in commercial operation (Belal et al., 2020) and thus reflects the true profitability potential of high-speed rail.

Tracks and infrastructure capacity

Most high-speed rail lines in the world are **double-track**, which has the main benefit of being able to serve two directions simultaneously without conflicts. This project will follow the standard from practice, but it must be checked afterwards whether this is enough to cope with the train traffic needed to serve all demand.

Traffic type served

Literature review in section 2.5.2 found out that costs HSR dedicated solely to passenger traffic on average is 20% lower than mixed lines. As this project aims to find the potential of high-speed rail, the proposed network will be designed for **solely passenger traffic**. This also helps with verification of the infrastructure capacity, as explained in the previous paragraph.

Power supply

Even though literature review in section 2.5.2 found out that it adds roughly 10% to the total costs, the network designed will follow the standard of **electric power supply**, as it is common with most high-speed rail lines. This choice is made deliberately following the environmental context described in section 1.1.1.

Rolling stock

A train set fulfilling the design criteria must be chosen. Following from the earlier mentioned points in this section, these are a maximum speed of at least 350 km/h and electrical power supply. UIC High-Speed (2018) provides an overview of all active high-speed rolling stock in commercial operation, allowing for compilation of a complete list of trainsets fulfilling the design criteria, in Table 3.1 below.

Table 3.1 Candidate train sets

Name	Manufacturer	In service	V_{max} (km/h)	Cars	Seats	Operating country	Acquisition Cost (2024)	Source
CRH380A CRH380AL	CRRC Sifang	2010	350	8 16	480 1,061	China	€ 37,866,000 € 75,773,000	IRJ (2010)
CRH380B CRH380BL CRH380CL	Siemens CNR Tangshan CNR Changchun	2011	350	8 16 16	551 1,043 1,053	China	€ 39,350,000 € 78,700,000 € 62,536,000	CD (2009)
CRH380D	Bombardier CSR Sifang	2012	350	8	518	China	€ 31,483,000	BBD (2015)
CR400AF-A CR400AF-B	CRRC Sifang	2017	350	16 17	1,193 1,283	China	€ 70,633,000 € 88,292,000	GRR (2021)
CR400BF-A CR400BF-B	CRRC Tangshan CRRC Changchun	2017	350	16 17	1,193 1,283	China	€ 70,633,000 € 88,292,000	CD (2017)
KCIC400AF	CRRC Sifang	2023	350	8	576	Indonesia	€ 39,756,000	RSW (2022)

These trains are mostly produced by Chinese companies and/or active in China only, within the last fifteen years. The Jakarta-Bandung HSR line (Indonesia, opened in 2023) is currently the only place outside of China where 350 km/h trainsets are in operation. Their trains again are of Chinese manufacture. Trainsets are produced most often in 8-car (approx. 550 seats, €35 million per train) or 16-car configuration (approx. 1100 seats, €70 million per train). As for high-speed trains, the acquisition is considered the normative expense, the choice of trainset will purely be based on the acquisition costs. It can be derived that type **CRH380CL** provides the best cost-per-seat ratio. To reduce the complexity of the problem, the premise is to consider CRH380CL **homogeneously** operating the network.

Stations

As mentioned in literature review section 2.5.2, it is taken as premise that new HSR infrastructure will be connected to existing stations in urban areas. Construction costs related to stations are therefore taken out of consideration for this project. Thus, it is assumed that all stations can already meet regulations regarding e.g. platform lengths in order to serve the chosen rolling stock.

Separation of network design and frequency setting

The scientific environment of these two design stages, whether integrated or not, is explained by (Guihaire & Hao, 2008) in the figure on the right:

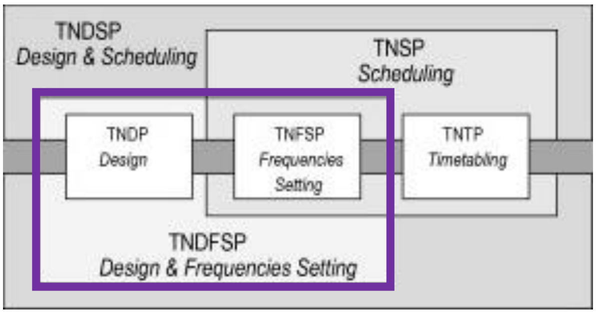


Figure 3.1 Schematic representation of integrated and separated approaches for network and line design (Guihaire & Hao, 2008)

Even though Gutiérrez-Jarpa et al. (2017) innovatively split two problem in two parts, network design and frequency setting commonly are handled integrated by means of a TNDFSP. An integrated formulation however produces higher-valued solutions, as it considers the interaction between network- and line design. Therefore, it is opted for to aim for an integrated solution.

3.2 Demand Forecasting

This section aims to outline the methodology for forecasting the demand of a hypothetical line connecting two cities.

3.2.1 Problem Definition

To be able to design an operationally profitable and economically justifiable high-speed rail network, demand is a crucial factor of influence. HSR connections with too low demand will never be profitable. High-speed rail requires substantial investments (see section 2.5). Therefore, adequate forecasts of demand are vital into guaranteeing that these investments are well-spent. These statements can be underpinned with the work of European Court of Auditors (2018). The authors concluded that nine out of fourteen assessed European HSR lines failed to generate sufficient demand in order to be successful. The total construction costs of these lines were €10.6 billion. These examples emphasise the importance of accurate demand forecasting.

The problem is much bigger and long-standing than this example alone. Flyvbjerg et al. (2005) investigated 210 transportation projects in 14 countries and concluded demand is overestimated for 90% of rail projects, with the average overestimation equalling 106%. Despite claims of forecasters that their methods have become more accurate, the author is able to show forecasting knowledge has grown, but not the accuracy of forecasting models. This can be attributed to political causes, which have a substantial influence on rail projects: decision-makers generally ignore or downplay financial risks under the guise of social welfare or other variables that are impossible to measure accurately (Flyvbjerg et al., 2005). Passenger forecasts for rail projects can be accurate if done in a scientific and independent manner (Börjesson, 2014). The author concludes that models assuming linear relationships can predict HSR demand 'reasonably well', while identifying three arising problems that complexifies HSR passenger forecasting which must be addressed in order to produce realistic forecasts:

1. It depends on more factors than regional travel models.
2. The non-linear relationship between demand and travel time.
3. HSR models are harder to calibrate than regional travel models, due to scarcer data.

3.2.2 Model Definition

Demand for high-speed rail is forecasted by means of dedicated models (section 2.2), including multiple impact factors (section 2.1), in which the combination of model and impact factors (section 2.3) is also important. As shown in section 2.2 and Table 2.5, established models most often do not estimate HSR demand directly. In practice, they focus on estimating either market shares or total passenger flow. These models are referred to as modal share models and trip generation models, respectively. This project exclusively relies on established models due to their well-documented benefits and drawbacks, providing greater insight into the functioning and reliability of the model.

Preliminary modelling choices

As the most popular model in literature, a logit model (section 2.2.1) is used in our model. A direct consequence of this choice is that the HSR demand for OD pair p; $D_{HSR,p}$ has to be estimated via its market share $m_{HSR,p}$. The most straightforward way to achieve this, is formulated in equation (3.1). The inspiration for this is drawn by the work of Sánchez-Borràs et al. (2010) and Leng et al. (2015).

$$D_{HSR,p} = m_{HSR,p} \cdot D_p^{TOT} \tag{3.1}$$

The consequence of previous choices suggests a model must be found to forecast the total travel demand between the two cities D_p^{TOT}. Section 2.2 showed two established models dedicated to this purpose exist: linear regression and gravity models. The latter is considered the most suitable method to estimate demand for new transit connections, especially when currently no direct service is available (Grosche et al., 2007), which is a key point for this project. For this reason, multiple long-distance travel demand related studies (e.g. Boelrijk, 2019; Grolle et al., 2024) opted for the gravity model as well.

Some factors impact market share but not the total demand, or vice versa (see the literature review of section 2.2). Literature summarised in Figure 2.7 showed that gravity models and logit models complement each other in terms of demand-impacting factors considered, allowing for an all-encompassing forecasting method when these two models are combined. The logit model will act as modal share model, the gravity model as trip generation model.

Model definition

The goal is not to understand the variation of preferences among the population, only to find the general preferences, which is needed to forecast demand. Therefore, an MNL is considered sufficient. Altogether, the model should consider a connection between city i and city j, incorporating prominently used and quantifiable impact factors. An overview of commonly used model-factor combinations was provided by Figure 2.7. Based on that figure, it is decided that the logit part covers travel time $TT_{k,ij}$ and travel cost $TC_{k,ij}$, while the gravity part takes in the city's population P_i, GDP values GDP_i, and a distance d_{ij} in between, raised to parameter γ, to take distance decay and diminishing returns into account. The same is done for the populations and GDP measures by parameters α and β, respectively, inspired by the work of Leng et al. (2015). Finally, k represents a scaling factor. This formulation provides a solution to all three complexing factors in HSR demand forecasting (see section 3.2.1): it encompasses more factors of influence than analysed studies, the non-linearity between demand and travel time is addressed by the logit part and γ-parameter, and it is calibrated on widely available flight demand data rather than scarce HSR data. Also, the model includes the five most popular demand impacting factors from studies (Table 2.1, Figure 2.7), excluding 'frequency', which is addressed separately as the 'frequency setting problem' in section 3.4.6.

The complete, to be calibrated model is presented as:

$$D_{AIR,ij} = \frac{\exp\left(V_{plane,ij}\right)}{\sum_{k\in \boldsymbol{K}} z_{k,ij} \cdot \exp\left(V_{k,ij}\right)} \cdot k \cdot \frac{\left(P_i \cdot P_j\right)^{\alpha} \cdot \left(GDP_i \cdot GDP_j\right)^{\beta}}{\left(d_{ij}\right)^{\gamma}} \tag{3.2}$$

In equation (3.2), $\boldsymbol{K}$ is the set of available travel modes; $k \in \{HSR, plane, train, car\}$. The logit part calculates the air market share, following the explanation in section 2.2.1. $z_{k,ij}$ is a binary value referring to the presence of travel mode k for city pair ij (consisting out of city i and j, where $i \neq j$). The utility functions $V_{k,ij}$ are defined for each mode $k \in \boldsymbol{K}$, and depend on the city pair ij of interest:

$$V_{k,ij} = \beta^{TT} \cdot TT_{k,ij} + \beta^{TC} \cdot TC_{k,ij} \tag{3.3}$$

where:

- The coefficients β^{TT} and β^{TC} represent the influence of travel time and travel costs. Their units are [utility / h] and [utility / €], respectively.
- $TT_{k,ij}$ and $TC_{k,ij}$ are the travel time and travel costs, respectively for mode k and city pair ij.

3.2.3 Data Collection

The model as defined in section 3.2.2 requires multiple parameters as input. How each of these will be gathered is explained below. Referring to the different demand flows described by Table 2.6, the model addresses diverted demand only, which is the estimated demand for the first year. Later, section 3.2.6 will leverage the model even further to also account for demand estimations in later years, which accounts for economic demand growth and induced demand (again, see Table 2.6). Inspiration for this is drawn from Leng et al. (2015).

Airports & air demand

Like all forecasting models, the one produced in this paragraph has to be calibrated to a significant amount of observed data, which enhances the model accuracy and precision (Morgan et al., 2003). The reporting of data on observed high-speed rail demand is limited. Therefore, a HSR demand forecasting model cannot be calibrated with observed HSR demand; here, it is calibrated with air passenger data.

Arguably the most comprehensive data set on long-distance travel passenger data currently accessible is provided by Eurostat, the statistical office of the European Union. Eurostat operates monthly-updated country-based databases, named *'avia_par_xx: detailed air passenger transport between the main airports of [country] and their main partner airports'* (Eurostat, 2024a). Here, 'xx' refers to the alpha-2 country code. An overview of databases is provided in Appendix C. From the scope as defined in section 1.5, six countries are not represented by the Eurostat databases, of which only Albania is the only country with an airport (ICAO, 2023). These passengers are captured from the other end of their routes.

Seasonal trends will be neglected, since this is considered outside of the scope. 2019 is the chosen year of interest, being the latest available year of normal operations before the pandemic. Airports will be identified by their respective unique IATA code, gathered through FlightsFrom (2024), their coordinates through ODS (2024).

The following five steps should be undertaken chronologically to preprocess the data, in order to create an eligible set of airports and associated pairs for model calibration, accompanied by air passenger data:

1. **Combining:** all country-specific data (Appendix C) is combined into one large data set.
2. **Time scoping:** filtering for only 2019 data.
3. **Area scoping:** only airports located within the scope (section 1.5) were kept.
4. **Airport combining:** In order to capture 'city-pair' demand instead of 'airport-pair' demand, airports serving the same city (full list: Appendix C) are treated as one airport.
5. **Directional combining:** To find the total travel potential between two cities, the number of passengers from the outward and return flights have to be combined. If the return flight is not present in the data, the total travel potential is calculated as twice the outward passenger number. It has to be checked with the data set whether this assumption can be made.

The five steps above form the air demand data set. However, for model calibration, the next step must be taken as well:

6. **Adaption to competition:** As the size of the catchment area is largely influenced by airport and city size (Lieshout, 2012; Grosche et al., 2007), forecasting studies opt to only include a sample of medium-to-large international airports that have non-overlapping catchment areas. As airports serving less than 1 million passengers per year are classified as 'small' by the currently active legal act of the European Commission (2014), they are excluded from the calibration data set. Also, only flight distances of over 1000 km should be considered here, as gravity models do not take competition from other modes into account (see Figure 1.1). This project's methodology will follow the philosophy of the author's work: only the largest airports per country, as long as their 2h-driving catchment area does not overlap with other airports, ensuring that each considered airport does not capture passengers from other large airports considered (Marcucci & Gatta, 2011). The size of airports is calculated by their total passenger movement (arrivals and departures) from the full data set considering all flights.

Population

Following from literature review, catchment area population is deemed the most accurate representation of population. Its size is determined by the chosen travel time, of which the figure has been debated in literature. The related literature review in section 2.1.6 showed maximum catchment areas of 90-110 minutes in travel time by car, with variations among cities. This research opts for to gather data for a maximum of 120 minutes, matching the maximum allowed access time of most studies (Marcucci & Gatta, 2011). The model will be estimated for every increment of 15 minutes in travel time catchment area at first, seeking the best model fit. The catchment area populations are gathered with use of Smappen (2024), which allows to choose a starting point and a travel time of various travel modes. Since section 2.1.6 showed that car travel time captures the distance decay well, this methodology will be followed here as well. The catchment area for each airport found by air demand data collection will be centred around the city centre.

Gross Domestic Product (GDP)

As population data was attained in the previous section, the model will be calibrated on GDP and GDP/capita, as we seek for the best fit. For model calibration, the data will be acquired by the following three databases from Eurostat:

1. met_10r_3gdp: *Gross domestic product (GDP) at current market prices by metropolitan regions* (Eurostat, 2024b)
2. tgs00003: *Regional gross domestic product by NUTS 2 regions - million EUR* (Eurostat, 2024d)
3. demo_r_d2jan: *Population on 1 January by age, sex and NUTS 2 region* (Eurostat, 2024c)

If the city is included, data set 1 finds us the city's total GDP directly. If not, is deducted by scaling the region's GDP (data set 2) down, based on the population of the region (data set 3) and the city itself. To find GDP per capita, population data from a data set by Florczyk et al. (2020) is used, which was created in name of the European Commission and reports the population for all European urban centres.

Distance

It is not allowed to use the distance travelled in any of the possible travel modes, as it would favour and/or disfavour the other travel modes. Instead, the great circle distance is used, which is calculated by the Haversine formula (Agramanisti Azdy & Darnis, 2020), allowing us to estimate with an error of 0.5% at maximum (Great Britain Navy Dept., 1987). Using this formula is much less time-consuming than manually looking up distances, without sacrificing much in terms of accuracy.

Travel cost & travel time (non-HSR)

Rome2Rio (2024) provides great insights into all travel options from origin to destination, for 'plane, train, bus, ferry and car'. For each travel mode (no matter how likely it is to be chosen by passengers), it provides the total door-to-door travel time and travel costs. For the latter, it also looks at minimum and maximum values for travelling the route indicated, taking into account periodical price variations. The benefit of using this source is the fact it shows realistic data that people weighing off their travel options would also use, preventing us from making any generalising assumptions. The major drawback: collecting data manually takes a long time. Therefore, the data will be collected via web scraping. By *Python* code, aided by web scraping package *selenium*, a computer can be assigned to surf the web, continuously inputting thousands of search queries and copying the data into an Microsoft Excel spreadsheet, for further analysis. Coordinates of the city centres are used as origin and destination for each search query, to represent realistic door-to-door trips.

The following processing steps will be undertaken before the travel costs and travel times of various travel alternatives are ready to serve as input for the demand forecasting model:

- **Step 1: mode simplification**; Road2Rio considers ten different travel modes: bus, bus ferry, car, car ferry, ferry, plane, rideshare, taxi, town car and train. Often, a combination of these travel modes are used in a sequence. The following mode changes were applied for simplification of analysis in later steps:
    - The modes 'car ferry', 'taxi' and 'town car' are changed to 'car', as they are in essence all car-based.
    - The mode 'bus ferry' is changed to 'bus', as it is assumed to be only accessible by bus.
    - The mode 'rideshare' is deleted, as all trips should already have a 'car' alternative.

  This brings the number of different modes down to five: plane, train, car, bus and ferry.

- **Step 2: mode sequence characterisation;** Each sequence should be translated to a characterisation: a base-mode. The scope of this project only considers plane-, train- and car-based trips. The following rules will be followed in order:

Table 3.2 Mode characterisation rules

#	Characterisation	Rules applicable to sequence	In scope
1	Bus-based	'bus' only	no
2	Ferry-based	'ferry' only	no
3	Car-based	'car' only	yes
4	Unusual car use	'car' (in)directly after non-car mode	no
5	Plane-based	'plane' uninterrupted	yes
6	Train-based	'train' uninterrupted	yes
7	-	none of the above	no

- **Step 3: travel time & costs;** Travel times are rewritten to the number of hours. The source provides lower and upper limits for the travel costs. Since the model defined in section 3.2.1 allows only one travel cost input per city pair, the travels cost acquired from the data is averaged. In case a city pair has multiple reported travel options for the same base, the one with the shortest travel time is reported. This is done for multiple reasons:
    - Section 2.1 pointed out it is the most important decision factor among travellers.
    - The model defined in section 3.2.1 allows only one travel time input per city pair.

- **Step 4: handling of missing data;** Rome2Rio (2024) always reports travel times but in some cases, if made use of special operators, travel costs are missing. Missing travel costs will be attained using an average cost per km for trips of only that mode, from the data set used for model calibration. The results of this are reported in section 4.1.1.

HSR travel cost

This is a design choice and refers to the fare setting. Section 3.3.1 will explain the related methodology.

HSR travel time

For new infrastructure, Rome2Rio (2024) does not provide any information. The maximum design speed v^{max} of 350 km/h set in section 3.1 cannot be reached for the entire duration – in Europe, the average speed is only 45% of the design speed (European Court of Auditors, 2018). Similarly to design speeds, the current peak efficiency of 90% is reached in China (Zhang & Zhang, 2021), considering connections without intermediate stops. As this project will try to find the potential of HSR in Europe, the maximum potential of current HSR practice is assumed. Therefore, this 90% efficiency k^{eff} will be used in calculating travel times:

$$TT_{HSR,ij} = \frac{l_{ij}}{k^{eff} \cdot v^{max}} + t^{dwell} + t^{access} + t^{egress} \quad (3.4)$$

Here, l_{ij} is the line length between station i and j, not to be confused with the (straight-line) distance between them, d_{ij}, which is generally shorter (Brons et al., 2023). k^{eff} is the earlier mentioned efficiency, thus set to 0.9 here. t^{dwell} is set to $1/12^{\text{th}}$ of an hour (5 minutes, based on Grolle et al. (2024)). The value of l_{ij} should be written in terms of [km] and v^{max} in [km/h]. The travel time $TT_{HSR,ij}$ will then be calculated in terms of [h]. The line length is acquired by finding the shortest distance over land. As Rome2Rio (2024) provides no option to avoid ferries (thus stay on land), Google Maps is web scraped instead.

Since all other travel data is considered door-to-door, and HSR is not yet, $t^{access} + t^{egress}$ represents the total access and egress time (section 2.1.2), and is set to 30 minutes, which is also used by Prov. N-Holland & Hardt (2020) and Sane (2020). It will be left out of equation (3.4) when calculating travel distances between HSR stations but left in when it considers OD travel times in order to calculate the travel alternatives' market shares and demand.

3.2.4 Model Calibration

For calibration, the model will be split in two. First, solely the gravity part of the model will be calibrated on the Eurostat data.

Gravity part

To calibrate our gravity model (eq. (3.2)), it is usually written in a log-linear form (Grosche et al., 2007).

$$\log(D_{HSR,ij}) = \log(k) + \alpha \log(P_i P_j) + \beta \log(GDP_i GDP_j) - \gamma \log(d_{ij}) \quad (3.5)$$

Calibration methods are briefly explained in section 2.2.5. The complexity of using MLE for this model has led to the decision to initially adapt the model to be fitted for the OLS solving technique. OLS is a well-established way in literature which can be used to calibrate (log-)linear functions. For this reason, it will be applied for calibration of the model. Detailed explanation of the method's workings and means to display results can be found in Appendix B.

Bias elimination

Even though Lieshout (2012) and Grosche et al. (2007) showed catchment area size varies along with city and airport size, it is kept constant in our model for simplicity. This might induce a bias in the calibrated model, as the number of potential passengers is overestimated for small airports and underestimated for large airports. For this reason, it will be investigated whether the calibrated model shows these biases.

When plotting the log of predicted demand y versus the log of observed demand x, its trendline is described by $y = ax + b$. For an unbiased model, $a = 1$ and $b = 0$, which simplifies to $y = x$. In any other case, the model is biased and therefore lacks accuracy. This bias can be eliminated by transforming the y_{old} value, following the following formula:

$$y_{new} = \frac{y_{old} - b}{a} \quad (3.6)$$

The benefits of an unbiased model are increased accuracy, while not harming the statistical significance of its parameters.

Logit part

Logit models usually are calibrated to large scale surveys data (see section 2.2.1). As conducting a survey is considered far outside the scope of this project, being time-expensive and performed in literature many times in the past. As an extensive overview of long-distance travel studies was compiled already in the literature reviews of section 2.1 and 2.2, it is opted for here to recall these studies, but now only considering the ones that use our model type MNL. The β^{TT} and β^{TC}-parameters can be estimated easily by statistical analysis of the results from these studies.

As was already shown in Table 2.5, 19 MNL-studies related to this field were found. Fifteen of them publish parameter estimates fitting to model defined in section 3.2.1. Their methodologies underline the necessity of choosing not to calibrate this project's MNL model to actual data: these studies use extensive surveys to estimate parameters, with the number of reported observations summing up to almost half a million. It is indicative of the complexity involved that estimating these parameters constitutes a distinct scientific discipline in its own right. Studies were applied in the UK, France, Germany, Spain, Portugal, Sweden and a few other countries outside of Europe, with the number of observations ranging from 40 to 63,000, and the r-squared model fit ranging between 0.075 and 0.822.

As the majority of these studies forecast demand by means of multiple model definitions, the number of models considered in this literature analysis reaches 57. An overview of them is shown in Appendix B. It is important to note that some studies calibrate models, each addressing different travel groups as they value travel time differently. These are commuting, business and leisure travellers.

Travel time

Almost all models defined by the studies included total travel time as a parameter. Some opt to split this term into separate parameters to quantify the influence of access and egress times, which is not considered here. Both hours and minutes are commonly used as units of travel time with equal frequency. For comparison, all reported estimates are converted to utility per hour. All reported parameter estimates were of the expected negative sign and are valued between -3.6 and -0.0054. The median is used here, it equals **-0.4606**.

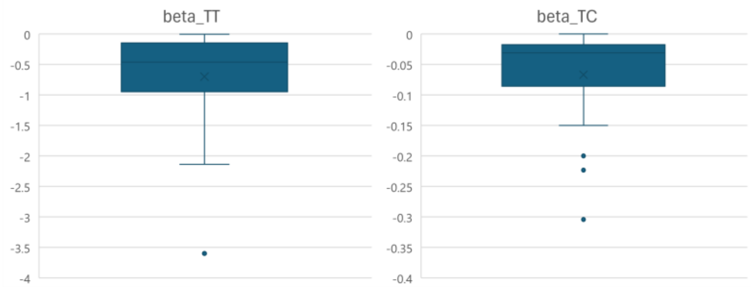


Figure 3.2 Reported values for travel time (N=53) and travel cost (N=47) parameters in MNL models

Travel cost

Also estimated travel cost parameters were found in almost all investigated studies. The unit reported varies much more, as it considers different currencies from different years. The value of the currency at the time of study is converted to 2024 €, to be able to make valid comparisons. The boxplot's shape closely in Figure 3.2 resembles the travel times, however vertically approximately ten times smaller. Again, all reported parameter estimates were of the expected sign and lie between -0.30413 and -0.000072. The median is used here, it equals **-0.03111**. Boxplots of both parameters are plotted in Figure 3.2 above.

3.2.5 Model Validation

To validate our demand model, we need to ensure it is able to accurately predict air demand using data that differs from the calibration data. In order to do so, we use the data points that were removed during the last step of data collection in section 3.2.3 for the sake of erasing effects of competition among airports, which was needed to correctly calibrate the model's parameters. This selection is chosen for as it encompasses data from the same scope (area, time), though being different from the calibration data. Recall that the calibrated part of the demand model serves the purpose of predicting the total passenger flow, not that of solely high-speed rail (section 3.2.4). Therefore, for validation, again a minimum threshold of 1,000 km in travel distance was imposed to eliminate the effects of competition among travel modes. The validation approach as taken by Belal et al. (2020) will be used here: we take five interesting connections and calculate how closely the model approaches the observed demand.

3.2.6 Demand Evolution

As mentioned before in sections 2.4 and 3.2.2 and Table 2.6, estimating demand for later years depends on the level of economic growth and induced demand. These factors are left out for calibration on purpose, as the demand model is calibrated to today's level of medium- to long distance travel demand. The following adaption to the forecasted first-year demand $D_{HSR,ij}$ by equation (3.2) is made to forecast future demand:

$$D_{HSR,ij}^{LT} = D_{HSR,ij} \cdot k_{ij}^{eco} \cdot k_{ij}^{ind} \tag{3.7}$$

Here, $D_{HSR,ij}^{LT}$ equals the average yearly HSR demand between city i and j over the project's lifetime. The scaling between these two factors is determined by factors k_{ij}^{eco} and k_{ij}^{ind}, representing economy-related demand growth and induced demand growth, respectively. These depend on the chosen city pair ij. Below, the computation of each of these two factors will be explained.

Economic growth

Trafikverket (2021) estimated a GDP elasticity with respect to demand of 0.7, which is a conservative estimate when drawing comparisons with other figures mentioned in section 2.1.4. To limit the chances of demand overestimation, this value is chosen to work with. World Bank (2023) reports yearly GDP growth for every country between 1961 and 2023. For this project, the mean yearly growth rate p from the last 40 years is taken into account for each country within the scope, as the data is not available at city-level. For a connection ij between two countries, the minimum yearly growth rate among them is considered normative for economy-based demand growth. This value p is used to compute the factor k_{ij}^{eco} as follows:

$$k_{ij}^{eco} = \frac{1}{T^{life}+1} \sum_{t=0}^{T^{life}} \left(1 + \frac{p_{ij} \cdot e_{GDP}}{100}\right)^{t} \tag{3.8}$$

Here, T^{life} is the project's lifetime in years, with $t \in \{0,1,2,\dots,T^{life}\}$. p_{ij} is the minimum economic growth rate among the two countries related to city pair ij as a percentage, and e_{GDP} represents the demand-related GDP elasticity.

Induced demand

Preston (2013) mentions the level of induced demand for multiple European high-speed rail projects which are in operation. From these numbers, it can be concluded that the amount varies between 11% and 50% of the total demand. However, most of the reported values are around and above 20%. To mitigate the risk of overestimation (which is something that more forecasting studies should do, see section 3.2.1), k_{ij}^{ind} is set to 1.20. As the level of induced HSR demand depends on the quality of the current rail connection (Leng et al., 2015), it will be halved if the new HSR infrastructure is an upgrade of an existing line. The rationale behind this, and the scenario itself will be further explained in section 3.3.2.

3.2.7 List of Assumptions

The following table will list the assumptions made during the demand forecasting methodology.

Table 3.3 Assumptions made in demand forecasting

Assumption	Impact	Reasoning
Bus not considered as travel mode	Very small	Bus market share is marginally small for full spectrum of long-distance trip lengths Excluded for model simplicity
Symmetrical passenger flow	Very small	Return flight passenger data is not always available, data set shows that assumption can be made safely as errors are very small. Also: it provides more data
All passengers travel via air for trips longer than 1000 km	Small	Gravity model must be calibrated to total passenger flow but the only data available is air passenger flow. Therefore, a lower distance bound must be chosen.
No preference variability among population	Medium	MNL implied simplicity
Mode choice is based on travel time and costs only, no base preference for modes	Medium	MNL simplicity
2024 travel attributes can be coupled to 2019 travel flows in demand forecasting	Medium	No workaround since annual 2024 passenger flow data and 2019 travel attribute data is not readily available Also, 2024 air demand is at approx. 2019 level
Conventional rail and HSR uncorrelated in demand forecasting (IIA property)	Medium	MNL implied simplicity
Flight passengers do not transfer	Medium	No workaround method found as flight data is captured based on flight legs, not OD pair
Constant catchment area size	Large	Gravity model implied simplicity

These assumptions are discussed in section 6.2.1.

3.3 Profitability Estimation

This section aims to outline the methodology for estimating the probability and justifiability of an high-speed rail line. Two important inputs are required here: the foundational premises in section 3.1 as they form the base environment for profitability estimations, and the demand forecasting in section 3.2 for allowing revenue to be estimated (in section 3.3.1). Section 3.3.2 provides estimations for the total costs, by summation of its earlier mentioned components. The third section put the found formulas into the context of operational profitability and justifiability. The last section lists the assumptions made.

3.3.1 Revenue & Fare Setting

Fares are set based on maximisation of revenue. Therefore, this section first deduces the revenue function, then optimises it to find the optimal fare.

Deducing the revenue function

Section 2.5.4 mentions that operators generally set a price that maximises their passenger revenue, and provides the general interplay existing between price, demand and revenue. This general approach will here be applied to the demand forecasting model defined in section 3.2.1.

While the number of air passengers was the centre of attention for demand model calibration, it must be centred around the number of HSR passengers to be able to calculate HSR revenue. From the model formulation, it can be seen that fare setting directly affects the travel cost of high-speed rail $TC_{HSR,ij}$. The gravity part of the formulation is unaffected in this formulation. The value of $TC_{HSR,ij}$ however does affect the logit part – the utility of HSR decreases, while the utilities of all other modes remain unaffected. As the utility of HSR is divided by those of all modes together, the market share of HSR decreases. This matches realistic effects.

However, as the revenue equals the demand multiplied with the fare, increasing fares also can work in favour of the total revenue R_{ij}. To find the optimal fare setting, that maximises the revenue, the maximum value of this equation must be found, by setting the derivate of the formula with respect to $TC_{HSR,ij}$ to zero:

$$TC_{HSR,ij} = \max_{TC_{HSR,ij}} \left(D_{HSR,ij} \cdot TC_{HSR,ij}\right) = \max_{TC_{HSR,ij}} \left(R_{ij}\right) \Rightarrow \frac{d}{dTC_{HSR,ij}}\left[R_{ij}\right] = 0 \tag{3.9}$$

This equation always has one unique optimal solution in our field of interest, corresponding with maximum revenue. The mathematical proof of this can be found in Appendix D.

Implementation

Python offers packages to find optima of sophisticated formulas. For this project, the *fsolve* package from the *scipy.optimize* library will be used, which is instructed to not only calculate the optimal fare setting, but also the related revenue and demand.

3.3.2 Costs

As stated in section 2.5, high-speed rail comes with a number of cash-flows related to cost. The results of Table 2.7 are split into the initial investments (construction stage) and recurring payments (operational stage). Both will be mentioned here with their impact factors and relationships, setting up a formula to estimate the total costs related to this cash-flow. The formulas of all cash-flows individually will be combined to create a formula to estimate the total costs. The literature review in section 2.5 will serve as a starting point.

Construction stage

In literature, these are split into fixed costs regarding infrastructure (construction costs) and rolling stock (acquisition costs). These expenses are made once, in the initial stages before operation.

Infrastructure

The base cost driver of infrastructure was found to be the length. For this reason, construction costs generally are reported in terms of [€ / km]. The base equation to calculate $C_{ij}^{X,infra}$ can therefore be written as:

$$C_{ij}^{X,infra} = k^{X,infra} \cdot l_{ij} \tag{3.10}$$

Here, $C_{ij}^{X,infra}$ represents the total fixed infrastructure cost between city pair ij, the length of the line is l_{ij} and $k^{X,infra}$ is the construction cost per km. In section 2.5.2 (Figure 2.10), it was found that unit construction costs vary wildly from location to location, as it depends on the local population density, GDP and the economic price level. Borgogno (2023) quantifies these relationships and produces unit construction costs per km, for surface ($C^{surface}$) and tunnelling (C^{tunnel}) separately, for 28 European countries. If data is missing regarding countries for this project, we calculate it as the average of its reported neighbours. It is assumed that when the difference in elevation between cities is greater, the connection between them requires more tunnelling, resulting in higher construction costs. Firstly, a height-index HI_i will be attributed to every city. Linear interpolation between the lowest and highest city elevations (h_{min} and h_{max}, respectively) will be used to attribute height-indexes to all other cities:

$$HI_i = (h_i - h_{min})/(h_{max} - h_{min}) \tag{3.11}$$

The choice of data set to determine the elevations per city will be made in section 3.4.8, as the selection of cities is to be determined there as well. The construction cost per km will be calculated as a linear interpolation between surface and tunnelling costs, averaging the involved cities and countries:

$$k_{ij}^{X,infra} = |HI_i - HI_j| \cdot \left(\frac{C_{country(i)}^{tunnel} + C_{country(j)}^{tunnel}}{2}\right) + \left(1 - |HI_i - HI_j|\right) \cdot \left(\frac{C_{country(i)}^{surface} + C_{country(j)}^{surface}}{2}\right) \tag{3.12}$$

The monetary values reported by Borgogno (2023) are 2017-based and will be adjusted for inflation. When rail infrastructure is already present between a city pair, one may also opt to upgrade the existing line rather than building a new one from scratch, as it could reduce costs significantly (European Court of Auditors, 2018). Conventional rail lines can only be upgraded to a design speed of 220 km/h at maximum (UIC, 2018). This means they are not interesting for this project. Instead, the authors state that only dedicated lines having maximum design speeds of at least 250 km/h could be upgraded. The construction of 350 km/h HSR infrastructure is roughly twice as expensive as for 250 km/h (Preston, 2013). Therefore, a 50% construction cost discount is applied, if the connection already provides a rail connection with an average travel speed exceeding 200 km/h. This can be calculated through Rome2Rio (2024) web scraping, performed in section 3.2.3. The threshold is lowered to account for stopping times, and to accurately capture the current network's performance limitations. A 250 km/h HSR line is not truly 250 km/h if the realised operating speeds are significantly lower. Recalling European Court of Auditors (2018), this problem is common among HSR lines. If this scenario applies, the current fastest rail alternative becomes the new HSR alternative, while the current second fastest rail alternative (if it exists) becomes the new rail alternative in the model of equation (3.2).

Rolling stock
The total acquisition costs is perhaps the most straightforward cash-flow to calculate. It simply amounts to the unit cost of a train, multiplied by the number of trains needed. The base equation to calculate $C_{ij}^{X,train}$ can therefore be written as:

$$C_{ij}^{X,train} = k^{X,train} \cdot n_{ij} \tag{3.13}$$

Here, $C_{ij}^{X,train}$ represents the total fixed rolling stock cost between city pair ij, the number of trains needed to be ordered to serve this city pair equals n_{ij}. Parameter $k^{X,train}$ is the acquisition cost per train set. As reported in section 3.1, it equals € 62.54 million. The number of trains is calculated by the earlier introduced equation (2.7):

$$n_{ij} = 2 \cdot f_{ij} \cdot t_{ij} \tag{3.14}$$

Here, f_{ij} is the frequency considered in one way (in trains per hour) and t_{ij} is the travel time required to cross the entire length of the line (in hours). Equation (3.13) and (3.14) can be combined:

$$C_{ij}^{X,train} = k^{X,train} \cdot \lceil 2 \cdot f_{ij} \cdot t_{ij} \rceil \tag{3.15}$$

Note that the formula for n_{ij} now stands between ceiling brackets. This means that the number between these brackets is rounded up, as acquisition costs are based on an integer number of trains.

Recurring payments
As followed from Table 2.7, these costs refer to expenses made during operation and maintenance. These are costs made after the initial stages during the actual operation stages, continuously across the years. The costs listed here are based on yearly amounts.

Infrastructure
In literature, the operation and maintenance costs are commonly taken together as some consider the latter term as covered by the former. Often, they are reported in terms of [€ / km], indicating the line length as the most important impact factor. Therefore, their nature is alike the infrastructure construction costs:

$$C_{ij}^{T,infra} = k^{T,infra} \cdot l_{ij} \tag{3.16}$$

Here, the yearly infrastructure time-based costs $C_{ij}^{T,infra}$ are calculated by a simple multiplication of the unit cost per kilometre $k^{T,infra}$ and the total line length l_{ij}, all specific for the connection between city pair ij. Literature review in section 2.5 showed that the dedicated HSR lines designed for speeds above 300 km/h require around €100,000 in maintenance and operation costs, yearly and per kilometre.

Rolling stock
The yearly rolling stock operation and maintenance costs cannot be expressed in a simple [€ / km], as it directly depends on the level of usage. Therefore, most educated guesses are presented in units of [€ / seat-km]. Therefore, the yearly time-based rolling stock costs $C_{ij}^{T,train}$ will be directly calculated from that unit:

$$C_{ij}^{T,train} = k^{T,train} \cdot s \cdot H \cdot D \cdot \frac{n_{ij} \cdot l_{ij}}{t_{ij}} \tag{3.17}$$

The hourly travelled distance by all trains on city pair ij is calculated by $\left(n_{ij} \cdot d_{ij}\right)/t_{ij}$. This then is multiplied with the number of operating hours per year $(H \cdot D)$, in order to find the total travelled distance per year, and multiplied by the number of seats s in order to find the yearly total seat-km. This is multiplied with the unit cost $k^{T,train}$, in order to find the yearly rolling stock-related operation and maintenance costs.

Analysis of the work of Fröidh (2006), introduced in section 2.5.3, yields a value of € 0.03 per seat-km for an operating speed of 350 km/h and train capacity of 1,053 seats, matching the nature of our problem.

In the equation presented above, t_{ij} is the travel distance between city pair ij. In this example $H = 18$ and $D = 365$.

Total costs

The formula for the total costs over lifetime T^{life} is expressed as:

$$C^{TOT} = C_{ij}^{X,infra} + C_{ij}^{X,train} + T^{life} \cdot \left(C_{ij}^{T,infra} + C_{ij}^{Ttrain}\right) \tag{3.18}$$

3.3.3 Operational Profitability & Justifiability

Recalling the meaning of these terms when first introduced in section 1.2.1; a high-speed rail connection should only be built if it meets both of these two conditions. Their meanings can now be illustrated even more clearly, by using the formulas set up in section 3.3.1 and 3.3.2.

Operational profitability

In order for a high-speed rail line to be operationally profitable, all the operational costs should be lower than the operational benefits. The first term concerns the operation and maintenance costs regarding infrastructure $C_{ij}^{T,infra}$ and rolling stock $C_{ij}^{T,train}$. Operational benefits concerns the ticket revenue R_{ij}. Thus, the operational profitability P_{ij}^{T} for a connection between city pair ij can be written as follows:

$$P_{ij}^{T} = R_{ij} - \left(C_{ij}^{T,infra} + C_{ij}^{T,train}\right) \tag{3.19}$$

Justifiability

This concerns whether the operational profitability is able to pay back the initial investments, within its lifetime. The first term was illustrated in equation (3.19), the latter in equation (3.10). The justifiability for a connection between city pair ij is called P_{ij}^{X} and can then be written as follows:

$$P_{ij}^{X} = T^{life} \cdot \left(R_{ij} - C_{ij}^{T,infra} - C_{ij}^{T,train}\right) - \left(C_{ij}^{X,infra} + C_{ij}^{X,train}\right) \tag{3.20}$$

3.3.4 List of Assumptions

The following table will list the assumptions made during the profitability estimating methodology.

Table 3.4 Assumptions made during profitability estimation

Assumption	Impact	Reasoning
No subsidy (scenarios)	Large	Simplicity
No change in travel characteristics of competing modes	Large	Complexity of forecasting is out of project's scope
No inflation or deflation	Large	Complexity of forecasting is out of project's scope
No variability / risk management for all	Medium	Simplicity

These assumptions are discussed in section 6.2.2.

3.4 Network Design

This section will provide a step-by-step manual on how the Transport Network Design and Frequency Setting Problem (TNDFSP) associated with this project will be solved. The network design was first introduced as the third problem in section 1.4 and related literature was reviewed in section 2.6 and 2.7, to provide a foundation of knowledge. Based on that, the solving method and related assumptions are presented in section 3.4.1 and section 3.4.2. Then, the model's source of inspiration is introduced (section 3.4.3), which forms the basis of the objective (section 3.4.4), leading up to the addition of TNDP (section 3.4.5) and FSP (section 3.4.6) elements. Together they form the model's formulation (section 3.4.7), after which the data sources are listed (section 3.4.8). Section 3.4.9 explains the model's pre-processing. Section 3.4.10 and 3.4.11 address how the linear program will be optimised and validated, respectively.

3.4.1 Solving Approach

Kepaptsoglou & Karlaftis (2009) classify the methodologies used to formulate and solve network design problems to optimality: analytical methods, which finds relationships between components of a small transport network, and mathematical programming, which finds an optimal solution for a greater variability of network sizes. The latter method therefore matches the goal of finding the potential of HSR, for a network potentially matching the size of Europe. Ceder (2001) summarises it nicely by stating that analytical models are not suited for network design, while mathematical programming models are. Therefore, the latter method will be used here.

3.4.2 General Assumptions

Section 2.6 listed a few general aspects of the TNDFSP which should be considered when implemented. These regard to assumptions that determine the nature of the formulation, and functions as a starting point when doing so. In this section, the choices regarding these aspects are made and substantiated.

Demand elasticity

As found in section 2.6, most studies assume fixed (inelastic) demand, ignoring how travel time and cost affect demand, despite its inherent elasticity. Since the goal of this study is to find the realistic potential of HSR, it becomes evident that the elasticity of demand cannot be neglected. Either the total demand flow between a city pair is kept constant, or it varies along with the performance level of all modes combined. To keep the model realistic, the latter option is opted for here. As the demand here is defined by a forecasting model, it is clear that it needs to be linearised in order to be implemented into the objective function.

Passenger path assignment

Literature review in section 2.6 found that most studies assign the passengers to the shortest path, as the model's complexity would otherwise severely increase (Kepaptsoglou & Karlaftis (2009). The authors also report that deviating passenger assignments have their own underlying extensive literature, which is deemed outside of this project's scope. Therefore, the common practice of assigning passengers to the shortest path will be followed here as well.

3.4.3 Model Inspiration: Multi-Commodity Flow Problem

The presence of model formulations in literature matching the nature of this project is limited – however, section 2.6 found a new development in TNDP formulation; the adaptation of the Multi-Commodity Flow Problem (MCFP), leading to an “efficient formulation” to handle “city-scale transit networks” (Ng et al., 2024). The philosophy behind the MCFP applied to transit network design is explained by Magnanti & Wong (1984). The problem concerns finding the lowest cost of sending commodities (goods or people) through a network. The multi-commodity problem's formulation will be adapted to the nature of this project, firstly to incorporate elements of network design (section 2.6), then of line design (section 2.7).

A general formulation is provided in the equations below:

$$\min \quad \sum_{a \in \boldsymbol{A}} \sum_{k \in \boldsymbol{K}} c_{ak} \cdot x_{ak} \tag{3.21}$$

$$\text{s.t.:} \quad \sum_{a \in \boldsymbol{A}_i^{out}} x_{ak} - \sum_{a \in \boldsymbol{A}_i^{in}} x_{ak} = \begin{cases} D_k & if\ i \in \boldsymbol{O}(\boldsymbol{k}) \\ -D_k & if\ i \in \boldsymbol{D}(\boldsymbol{k}) \\ 0 & otherwise\ \forall i \in \boldsymbol{N}, a \in \boldsymbol{A} \end{cases} \tag{3.22}$$

The demand for commodity k is named D_k. The unit cost and number of commodities k transported over arc a are defined as c_{ak} and x_{ak}, respectively. The set of all arcs $\boldsymbol{A}$ is split into $\boldsymbol{A}_i^{in}$ and $\boldsymbol{A}_i^{out}$, representing the arcs going in and out from node $i \in \boldsymbol{N}$. Equation (3.22) guarantees continuity at the nodes, which requires the definition of sets representing the origin and destination of commodity k, $\boldsymbol{O}(\boldsymbol{k})$ and $\boldsymbol{D}(\boldsymbol{k})$ respectively. The same in the other direction holds for the destination of the commodity, defined by set $\boldsymbol{D}(\boldsymbol{k})$. Numerous extra constraints can be added to this model – e.g. the capacity constraint for each arc (Magnanti & Wong, 1984).

3.4.4 Model Objective

As defined in section 1.4, this is the network's lifetime profitability, which is in line with the general approach in literature from the operator's perspective (Guihaire & Hao, 2008). In section 2.5.1, profitability was dissected into seven cash flows, with each given their own equation in section 3.3.1 and 3.3.2. Here, they are recalled and divided across the network and line design part of our formulation.

Table 3.5 Cash flow allocation to problems

Problem	Cash flow	Symbol	Related equation
Network Design (TNDP)	Ticket revenue	R_{ij}	(3.9)
	Infrastructure construction	$C_{ij}^{X,infra}$	(3.10)
	Infrastructure operation & maintenance	$C_{ij}^{T,infra}$	(3.16)
Line Design (FSP)	Rolling stock acquisition	$C_{ij}^{X,train}$	(3.13)
	Rolling stock operation & maintenance	$C_{ij}^{T,train}$	(3.17)

Following the equations defined listed in Table 3.5, it can be seen that rolling stock costs depend on the line design only, as they depend on the line's frequencies and length. Infrastructure-related costs depend on the length of selected arcs only, and therefore fit the network design part. The same holds for ticket revenue. It can be seen that all found cash flows can be associated with one of the two problems, which the model will integrate by formulation of a TNDFSP. This allows for an all-inclusive evaluation of profitability for a HSR network. Below, section 3.4.5 will focus on the TNDP (network design) elements to be included in the TNDFSP formulation. After this, the model's formulation will be extended by addition of FSP (line design) elements in section 3.4.6.

3.4.5 Network Design Elements

This section will reason its way towards implementation of TNDP elements into the optimisation model.

- **Nodes:** We define the set of nodes $\boldsymbol{N}$, equivalent to the MCFP set of nodes. To indicate if node n is selected for our final network, we define a binary decision variable z_n.
- **Arcs:** We also define the set of undirected arcs $\boldsymbol{A}$. Binary decision variables y_a are defined to indicate if arc a is selected for the final network. Each arc has been attributed with travel time t_a, travel cost c_a, length d_a, lifetime construction, operating and maintenance cost f_a^{cost}.
- **OD pairs:** The original MCFP set of commodities is translated to a set of OD pairs $\boldsymbol{P}$, as their meaning is equivalent for this problem.

OD flow routes

The main problem found with exact approaches was that they can only solve within reasonable running times for small to medium-sized networks (section 2.6.5 and 2.7.6). In the MCFP formulation, for each commodity k, decision variables are considered whether it is transported along arc a. Inspired by the earlier-mentioned work of Oliker & Bekhor (2020) and Liang et al. (2019), this can be reformulated without changing the intrinsic nature and optimal solution: for each commodity k, we can define a set of potential routes across the network instead. A route would then be defined as a valid sequence of arcs from origin to destination. The same is done for this project's model; to do so, we define the set of OD flow routes $\boldsymbol{R}$, inspired by the work of Arbex & da Cunha (2015). Each OD flow route r can be selected (or not) for our final network, which can be indicated by binary decision variable x_r. An algorithm must be written in order generate all possible OD flow routes. Details regarding this are presented in section 3.4.9. This idea roughly matches literature review in section 2.7.5, where it was advised to pre-determine potential lines instead in order to decrease model complexity. Parameters are introduced in order to couple each OD flow route r to nodes, arcs and OD pairs:

- c_{nr}^{node}: whether OD flow route r covers node n (1) or not (0)
- c_{ar}^{arc}: whether OD flow route r covers arc a (1) or not (0)
- c_{pr}^{odpair}: whether OD flow route r covers OD pair p (1) or not (0)
- m_{pr}^{odpair}: whether OD flow route r exactly matches OD pair p (1) or not (0)

Now, a number of attributes specific to each OD flow route r can be defined:

- The travel time $t_r = \sum_{a \in \boldsymbol{A}} (t_a \cdot c_{ar}^{arc})$
- The travel cost $c_r = \sum_{a \in \boldsymbol{A}} (c_a \cdot c_{ar}^{arc})$
- The length $d_r = \sum_{a \in \boldsymbol{A}} (d_a \cdot c_{ar}^{arc})$

Demand

Since the relationship between travel time and demand was found to be non-linear by our model definition in equation (3.2), it has to be linearised for this mathematical programming model:

$$D_{HSR,p} = \alpha_p + \beta_p \cdot t_p + \gamma_p \cdot c_p \tag{3.23}$$

Here, the demand $D_{HSR,p}$ is estimated by a 'maximum demand' α_p for OD pair $p \in \boldsymbol{P}$ (when travel time and costs are zero), an associated 'time decay' β_p and 'cost decay' γ_p, indicating by how much the demand would reduce for respectively one hour extra travel time t_p, or one euro in extra travel cost c_p. The coefficients α_p, β_p and γ_p can be obtained by linearisation of the relationship between demand (from the demand forecasting model), travel time and costs. Making use of travel time t_r and travel cost c_r, a yearly passenger flow q_r^{year} specific to route r can be calculated. In order to do this, the equation mentioned above must be rewritten in order to match the route-oriented nature of the problem. Therefore, we define the following:

- Intercept $\alpha_r = \sum_{p \in \boldsymbol{P}} \left(\alpha_p \cdot m_{pr}^{odpair}\right)$
- Travel time coefficient $\beta_r = \sum_{p \in \boldsymbol{P}} \left(\beta_p \cdot m_{pr}^{odpair}\right)$
- Travel cost coefficient $\gamma_r = \sum_{p \in \boldsymbol{P}} \left(\gamma_p \cdot m_{pr}^{odpair}\right)$

The values of travel time and travel cost itself retain same meaning: $t_r = t_p$ and $c_r = c_p$.
We can now present the redefined linearised demand function:

$$q_r^{year} = \alpha_r + \beta_r \cdot t_r + \gamma_r \cdot c_r \tag{3.24}$$

The travel time t_r is fixed for each route, while demand between origins and destinations depends on the effective travel time and cost, which are influenced by the network structure. To maintain a linear model, OD pair demand will be captured by coefficients from a trendline fitted to the demand forecasting model, based on travel time t_r and costs c_r. The trendline is fitted for $t_r \in [t_r,\ k^{detour} \cdot t_r];\ c_r \in\ [c_r,\ k^{detour} \cdot c_r]$, with $k^{detour} = 1.25$, the maximum allowed detour factor, as used by Grolle et al. (2024). The undirected graph $G = (N, A)$ sums demand in both directions to simplify the model, without changing its intrinsic logic.

Ticket revenue

Recall that the objective of the formulation is a maximisation of lifetime profitability, and that revenue was found to be a component of that. We define the lifetime revenue f_r^{rev} for route r as the product of its demand per year q_r^{year}, fare price c_r and the lifetime of the project T^{life} in years. The fare price is determined by its own optimisation process in section 3.3.1.

Constraints

Now all variables and parameters are defined, the constraints can be constructed.

(1) **Node selection**: A node n is selected if a selected OD flow route r flows over it (3.25). Also, a node n cannot be selected if no selected OD flow route r flows over it (3.26).

$$z_n \geq x_r \cdot c_{nr}^{node} \qquad \forall n \in \boldsymbol{N}, \forall r \in \boldsymbol{R} \tag{3.25}$$
$$z_n \leq \sum_{r \in \boldsymbol{R}} (x_r \cdot c_{nr}^{node}) \qquad \forall n \in \boldsymbol{N} \tag{3.26}$$

(2) **Arc selection**: An arc a must be selected if a selected OD flow route r flows over it (3.27). As arc selection automatically induces costs, a constraint oriented to the opposite such as (3.26) can be disregarded.

$$y_a \geq x_r \cdot c_{ar}^{arc} \qquad \forall a \in \boldsymbol{A}, \forall r \in \boldsymbol{R} \tag{3.27}$$

(3) **OD pair selection**: An OD pair p may be served by at most one selected OD flow route r (3.28).

$$\sum_{r \in \boldsymbol{R}} (x_r \cdot m_{pr}^{ODpair}) \leq 1 \qquad \forall p \in \boldsymbol{P} \tag{3.28}$$

(4) **Minimum node separation**: Having too short distances between selected nodes will undermine HSR's rationale (Rodrigue, 2017). Therefore, a minimum distance l^{min} between nodes is introduced, and a set of all node pairs $(i, j) \in \boldsymbol{N}$ that would violate l^{min} is introduced: $\boldsymbol{N^{close}}$. We define the following constraints to ensure the minimum distance between nodes is respected:

$$z_i + z_j \leq 1 \qquad \forall (i, j) \in \boldsymbol{N^{close}} \tag{3.29}$$

(5) **Non-crossing arcs**: Selected arcs are not allowed to cross each other. To constrain this, we define the set of crossing arc pairs (i, j) as $\boldsymbol{A^{cross}}$, with $i, j \in \boldsymbol{A}$ and $i \neq j$. The following constraints should hold:

$$y_i + y_j \leq 1 \qquad \forall (i, j) \in \boldsymbol{A^{cross}} \tag{3.30}$$

(6) **Decision variables**: As stated before: x_r, z_n and y_a are all binary:

$$x_r \in \{0, 1\} \qquad \forall r \in \boldsymbol{R} \tag{3.31}$$
$$y_a \in \{0, 1\} \qquad \forall a \in \boldsymbol{A} \tag{3.32}$$
$$z_n \in \{0, 1\} \qquad \forall n \in \boldsymbol{N} \tag{3.33}$$

Objective function

As stated in section 3.4.4, cash flows considered from the Network Design perspective are ticket revenue and infrastructure-related costs. The total ticket revenue can be calculated as the sum of f_r^{rev}, for selected routes r. For arc a, the total infrastructure-related costs were earlier defined in this section as f_a^{cost}. In order to calculate the network's total infrastructure-related costs, it must be summed over selected arcs a. The 'network design'-related profitability can be displayed as:

$$\sum_{r\in \boldsymbol{R}}(x_r \cdot f_r^{rev}) - \sum_{a\in \boldsymbol{A}}(y_a \cdot f_a^{cost}) \tag{3.34}$$

The introduction of line design elements in the next chapter will further extend this objective function, as well as the rest of the model's formulation.

3.4.6 Line Design Elements

This section will add line design elements to the formulation built in the previous section, completing the Transport Network Design & Frequency Setting Problem (TNDFSP). The line design parts assigns lines and frequencies to the network. Regarding the formulation of constraints, inspiration is drawn from the work of Baaj & Mahmassani (1991), who represent a very basic TNDFSP formulation.

Operating frame

We assume 18 operating hours per day (H) and 365 operating days per year (D).

Valid operating lines

Each route r will be attributed a binary value l_r, which equals 1 if it is a valid operating line and 0 if not. An OD flow route r is considered an invalid operating line if the travel time between both end points is more than 9 hours, since then a train cannot make a full roundtrip within an operating day, resulting in planning difficulties. By imposing this limit, trains can always be returned to at least one of the line's termini overnight. An OD flow route r is considered a valid and selected operating line if $x_r \cdot l_r = 1$.

Line frequencies

Each route r will be attributed an operating frequency in trains per hour, equalling zero for invalid operating lines (to be constrained later). A nonnegative integer decision variable w_r is introduced, along with a universal maximum frequency W^{max}, here set to 12 trains per hour, matching minimum headway rules for the same 350 km/h trains in the Chinese network (Tian & Zhang, 2024).

Number of trains acquired

This is related to the operated frequency w_r on route r. The associated nonnegative integer decision variable is defined as n_r^{train}.

Seat capacity

This is defined by parameter s, equalling the number of passengers one train can carry simultaneously.

Peak hour demand per direction

This is calculated by calculating the average demand per operating hour throughout the year, while multiplying with a peak hour factor k^{peakhr}, set to 1.25 (see section 2.7.2), which is a conversion factor between peak hour flow and average hourly flow. As the flow q_r^{year} combines flow in both directions, it is divided by two. The peak hour flow q_r^{peak} is then calculated as follows:

$$q_r^{peak} = \frac{k^{peakhr}}{2 \cdot D \cdot H} \cdot q_r^{year} \tag{3.35}$$

Transfer penalties

In order to address the user's perspective in the objective function as well, as monetary penalty is added for every passenger who has to transfer. Literature review in section 2.7.4 provides an added perceived travel time for a transfer of 22.63 minutes, while Wardman et al. (2012) provides a common European VoT of €14.80 per hour, adjusted for inflation. The penalty transferring passenger equals $(22.63/60) \cdot €14.80 \approx$ €5.58. Given the definitions of this model, the following situations must both occur for an OD pair p in order for a transfer penalty to be imposed:

1. Passengers of OD pair p must be travelling across the network. This means that any OD flow route r **exactly matching** the OD pair p **must be selected**. We therefore introduce a binary decision variable u_p indicating whether this statement is true (1) or false (0).
2. The OD pair p may not be served directly. This means that any OD flow route r **covering** that OD pair p **may not be selected**. We therefore introduce a binary decision variable v_p indicating whether this statement is true (0) or false (1).

We can verify the outcome of $u_p - v_p$ only equals 1 if an OD pair p is served, but not directly. As non-active OD pairs p cannot be served by selected operating lines it is constrained that $u_p \geq v_p, \forall p \in \boldsymbol{P}$**.**

Constraints

Now all variables and parameters are defined, the constraints can be constructed.

(1) **Frequency setting**: An operating line r cannot be selected if the corresponding OD flow route r is not selected, or when it is not considered a valid operating line. Also, maximum operating frequencies W^{max} universally apply.

$$w_r \leq W^{max} \cdot l_r \cdot x_r \qquad \forall r \in \boldsymbol{R} \tag{3.36}$$

(2) **Serve all demand**: The selected valid operating lines r must serve all demand.

$$s \cdot \sum_{r \in \boldsymbol{R}} (w_r \cdot c_{ar}^{arc}) \geq \sum_{r \in \boldsymbol{R}} \left(x_r \cdot c_{ar}^{arc} \cdot q_r^{peak}\right) \qquad \forall a \in \boldsymbol{A} \tag{3.37}$$

(3) **Rolling stock acquisition**: Acquire the correct number of trains n_r^{train} for each operating line r. Following equation (3.15), it should equal $\lceil 2 \cdot w_r \cdot t_r \rceil$. This can be linearly defined:

$$n_r^{train} - 1 \leq 2 \cdot w_r \cdot t_r \qquad \forall r \in \boldsymbol{R} \tag{3.38}$$

$$n_r^{train} \geq 2 \cdot w_r \cdot t_r \qquad \forall r \in \boldsymbol{R} \tag{3.39}$$

(4) **Served OD pairs**: An OD pair p is served if the OD flow route r exactly matching that OD pair p is selected (3.41). If not selected, the OD pair is not served (3.40).

$$u_p \leq \sum_{r \in \boldsymbol{R}} \left(x_r \cdot m_{pr}^{ODpair}\right) \qquad \forall p \in \boldsymbol{P} \tag{3.40}$$

$$u_p \geq x_r \cdot m_{pr}^{ODpair} \qquad \forall p \in \boldsymbol{P}, \forall r \in \boldsymbol{R} \tag{3.41}$$

(5) **Served OD pairs, without transfer**: An OD pair p is served directly if any OD flow route r covering that OD pair p is selected (3.43). If none of these routes r are selected, the OD pair is not served directly (3.42). In the second constraint below, M is an arbitrary large constant, it should be defined while satisfying $M \geq \sum_{r \in \boldsymbol{R}} \left(w_r \cdot c_{pr}^{ODpair} \right), \forall p \in \boldsymbol{P}$.

$$v_p \leq \sum_{r \in \boldsymbol{R}} \left(w_r \cdot c_{pr}^{ODpair} \right) \qquad \forall p \in \boldsymbol{P} \tag{3.42}$$

$$M \cdot v_p \geq \sum_{r \in \boldsymbol{R}} \left(w_r \cdot c_{pr}^{ODpair} \right) \qquad \forall p \in \boldsymbol{P} \tag{3.43}$$

(6) **Decision variables:** u_p and v_p are binary, while w_r and n_r^{train} are nonnegative and integer:

$$w_r \in \mathbb{R}_{\geq 0} \qquad \forall r \in \boldsymbol{R} \tag{3.44}$$

$$n_r^{train} \in \mathbb{R}_{\geq 0} \qquad \forall r \in \boldsymbol{R} \tag{3.45}$$

$$u_p \in \{0, 1\} \qquad \forall p \in \boldsymbol{P} \tag{3.46}$$

$$v_p \in \{0, 1\} \qquad \forall p \in \boldsymbol{P} \tag{3.47}$$

Objective function

As stated in section 3.4.4, cash flows considered from the Line Design perspective are rolling stock-related costs. Added to this are transfer penalties, which are not part of HSR profitability, but they are included into the objective function to also account for the user's perspective. This way, the objective function becomes all-encompassing: it contains all factors of profitability, while also considering both sides of perspective. The equations listed below are added to equation (3.34).

Rolling stock acquisition costs

Equation (3.13) as defined in section 3.3.2 can be directly implemented into the objective function. The total rolling stock acquisition costs simply equal a unit cost factor k^X (€ / train) and the total number of trains acquired for operation of the network:

$$k^X \cdot \sum_{r \in \boldsymbol{R}} \left(n_r^{train} \right) \tag{3.48}$$

Rolling stock operation and maintenance costs

Equation (3.17) as defined in section 3.3.2 can be directly implemented into the objective function. The total rolling stock operation and maintenance costs simply equal a unit cost factor k^T (€ / seat-km) and the total number of lifetime seat-km for all selected routes r combined:

$$T^{life} \cdot k^T \cdot s \cdot H \cdot D \cdot \sum_{r \in \boldsymbol{R}} \left(\frac{n_r^{train} \cdot d_r}{t_r} \right) \tag{3.49}$$

Transfer penalties

Earlier in this section, it was defined that transfer penalties are imposed for an OD pair p if $u_p - v_p = 1$. The yearly demand not served directly is the product of the yearly flow over OD pairs for which $u_p - v_p = 1$. This should then be multiplied by the project's lifetime in years T^{life} and the penalty per passenger $k^{transfer}$:

$$T^{life} \cdot k^{transfer} \cdot \sum_{p \in \boldsymbol{P}} \left[\left(u_p - v_p \right) \cdot \sum_{r \in \boldsymbol{R}} \left(c_{pr}^{ODpair} \cdot q_r^{year} \right) \right] \tag{3.50}$$

3.4.7 Model Formulation

This section presents the finalised mathematical programming model formulation for the TNDFSP of this project. It essentially presents an overview of section 3.4.5 and 3.4.6, in the form of a notation and a formulation.

Notation

Below, presents the notation used in the TNDFSP formulation, consisting out of 6 sets, 7 variables and 20 parameters.

Table 3.6 TNDFSP notation

	Sets	
$\boldsymbol{N}$	Set of nodes	$n \in \boldsymbol{N}$
$\boldsymbol{A}$	Set of arcs	$a \in \boldsymbol{A}$
$\boldsymbol{P}$	Set of OD pairs	$p \in \boldsymbol{P}$
$\boldsymbol{R}$	Set of OD flow routes (and potential operating lines)	$r \in \boldsymbol{R}$
$\boldsymbol{N}^{close}$	Set of node pairs breaking the minimum distance boundary	$(i,j) \in \boldsymbol{N}^{close}$
$\boldsymbol{A}^{cross}$	Set of arc pairs crossing each other	$(i,j) \in \boldsymbol{A}^{cross}$
	Decision variables	
x_r	Whether OD flow route $r \in \boldsymbol{R}$ is selected (1) or not (0)	$[-]$
y_a	Whether arc $a \in \boldsymbol{A}$ is selected (1) or not (0)	$[-]$
z_n	Whether node $n \in \boldsymbol{N}$ is selected (1) or not (0)	$[-]$
n_r^{train}	Number of trains acquired to operate line $r \in \boldsymbol{R}$	$[trains]$
w_r	Operating frequency of line $r \in \boldsymbol{R}$	$[trains\ /\ hr]$
u_p	Whether OD pair $p \in \boldsymbol{P}$ flows over the network (1) or not (0)	$[-]$
v_p	Whether OD pair $p \in \boldsymbol{P}$ is served directly by a selected line (1) or not (0)	$[-]$
	Parameters	
f_r^{rev}	Lifetime revenue for OD flow route $r \in \boldsymbol{R}$	$[€]$
f_a^{cost}	Lifetime cost for arc $a \in \boldsymbol{A}$	$[€]$
k^T	Unit rolling stock operating and maintenance cost	$[€\ /\ (pax \cdot km)]$
k^X	Unit rolling stock acquisition cost	$[€\ /\ train]$
$k^{transfer}$	Unit transfer penalty	$[€\ /\ pax]$
T^{life}	Project lifetime	$[years]$
H	Operating hours per day	$[hours\ /\ day]$
D	Operating days per year	$[days\ /\ year]$
s	Seat capacity	$[pax\ /\ train]$
W^{max}	Maximum allowed frequency	$[trains\ /\ hr]$
c_{nr}^{node}	Whether OD flow route $r \in \boldsymbol{R}$ covers node $n \in \boldsymbol{N}$ (1) or not (0)	$[-]$
c_{ar}^{arc}	Whether OD flow route $r \in \boldsymbol{R}$ covers arc $a \in \boldsymbol{A}$ (1) or not (0)	$[-]$
c_{pr}^{ODpair}	Whether OD flow route $r \in \boldsymbol{R}$ covers OD pair $p \in \boldsymbol{P}$ (1) or not (0)	$[-]$
m_{pr}^{ODpair}	Whether OD flow route $r \in \boldsymbol{R}$ exactly matches OD pair $p \in \boldsymbol{P}$ (1) or not (0)	$[-]$
q_r^{year}	Yearly passenger flow for OD flow route $r \in \boldsymbol{R}$ (in both directions)	$[pax\ /\ year]$
q_r^{peak}	Peak hour passenger flow for OD flow route $r \in \boldsymbol{R}$ (per direction)	$[pax\ /\ hr]$
t_r	Travel time along route $r \in \boldsymbol{R}$	$[hr]$
d_r	Length of route $r \in \boldsymbol{R}$	$[km]$
l_r	Whether OD flow route $r \in \boldsymbol{R}$ is a valid operating line (1) or not (0)	$[-]$
M	Arbitrarily large constant	$[-]$

The twenty-three equations below form a Mixed-Integer Linear Program (MILP). Equation (3.51) is the objective function, which maximises profit while also taking the user perspective into account. It consists out of three parts, here listed in the order of how they are presented above:

1. Ticket revenue (see equation (3.9)) and infrastructure costs (see equation (3.10))
2. Rolling stock costs (see equation (3.48) and (3.49))
3. Transfer penalties (see equation (3.50))

Formulation

The formulation of the linear program to solve this project's TNDFSP are presented below:

$$
\begin{aligned}
max \quad & \sum_{r\in \boldsymbol{R}}(x_r \cdot f_r^{rev}) - \sum_{a\in \boldsymbol{A}}(y_a \cdot f_a^{cost}) \\
& - \left[k^X \cdot \sum_{r\in \boldsymbol{R}}(n_r^{train})\right] - \left[T^{life} \cdot k^T \cdot s \cdot H \cdot D \cdot \sum_{r\in \boldsymbol{R}}\left(\frac{n_r^{train} \cdot d_r}{t_r}\right)\right] \\
& - \left[T^{life} \cdot k^{transfer} \cdot \sum_{p\in \boldsymbol{P}}\left[(u_p - v_p) \cdot \sum_{r\in \boldsymbol{R}}(c_{pr}^{ODpair} \cdot q_r^{year})\right]\right] && (3.51)
\end{aligned}
$$

$$
\begin{aligned}
s.t. \quad & z_n \geq x_r \cdot c_{nr}^{node} && \forall n \in \boldsymbol{N}, \forall r \in \boldsymbol{R} && (3.52) \\
& z_n \leq \sum_{r\in \boldsymbol{R}}(x_r \cdot c_{nr}^{node}) && \forall n \in \boldsymbol{N} && (3.53) \\
& y_a \geq x_r \cdot c_{ar}^{arc} && \forall a \in \boldsymbol{A}, \forall r \in \boldsymbol{R} && (3.54) \\
& \sum_{r\in \boldsymbol{R}}(x_r \cdot m_{pr}^{ODpair}) \leq 1 && \forall p \in \boldsymbol{P} && (3.55) \\
& z_i + z_j \leq 1 && \forall (i,j) \in \boldsymbol{N^{close}} && (3.56) \\
& y_i + y_j \leq 1 && \forall (i,j) \in \boldsymbol{A^{cross}} && (3.57) \\
& w_r \leq W^{max} \cdot l_r \cdot x_r && \forall r \in \boldsymbol{R} && (3.58) \\
& s \cdot \sum_{r\in \boldsymbol{R}}(w_r \cdot c_{ar}^{arc}) \geq \sum_{r\in \boldsymbol{R}}(x_r \cdot c_{ar}^{arc} \cdot q_r^{peak}) && \forall a \in \boldsymbol{A} && (3.59) \\
& n_r^{train} - 1 \leq 2 \cdot w_r \cdot t_r && \forall r \in \boldsymbol{R} && (3.60) \\
& n_r^{train} \geq 2 \cdot w_r \cdot t_r && \forall r \in \boldsymbol{R} && (3.61) \\
& u_p \leq \sum_{r\in \boldsymbol{R}}(x_r \cdot m_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (3.62) \\
& u_p \geq x_r \cdot m_{pr}^{ODpair} && \forall p \in \boldsymbol{P}, \forall r \in \boldsymbol{R} && (3.63) \\
& v_p \leq \sum_{r\in \boldsymbol{R}}(w_r \cdot c_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (3.64) \\
& M \cdot v_p \geq \sum_{r\in \boldsymbol{R}}(w_r \cdot c_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (3.65) \\
& u_p \geq v_p && \forall p \in \boldsymbol{P} && (3.66) \\
& x_r \in \{0,1\} && \forall r \in \boldsymbol{R} && (3.67) \\
& y_a \in \{0,1\} && \forall a \in \boldsymbol{A} && (3.68) \\
& z_n \in \{0,1\} && \forall n \in \boldsymbol{N} && (3.69) \\
& w_r \in \mathbb{R}_{\geq 0} && \forall r \in \boldsymbol{R} && (3.70) \\
& n_r^{train} \in \mathbb{R}_{\geq 0} && \forall r \in \boldsymbol{R} && (3.71) \\
& u_p \in \{0,1\} && \forall p \in \boldsymbol{P} && (3.72) \\
& v_p \in \{0,1\} && \forall p \in \boldsymbol{P} && (3.73)
\end{aligned}
$$

All constraints are mentioned and explained in section 3.4.5 and 3.4.6.

3.4.8 Model Input Data

This section will list the various ways that the data regarding sets and parameters (as listed in the model formulation) are acquired. Some adaptations to the approach defined in section 3.2.3 were made:

Set of nodes $\boldsymbol{N}$

As indicated in section 1.1.1, high-speed rail serves to connect urban centres (López-Pita & Robusté, 2004). For this reason, a complete overview of all urban centres within the scope defined in section 1.5 would serve as a starting point. Florczyk et al. (2020), on behalf of the European Commission, used satellite images to identify urban centres based on human settlements. Their method avoids man-made borders and arbitrary population thresholds, focusing instead on the interconnectedness of nearby cities, villages, and towns that function as a unified urban area. The associated data set provides data on 160 metrics for 13,135 urban centres world-wide. The full set of urban centres (nodes) will be obtained as filtering out urban centres based on their coordinates, incompatible with the scope as defined by Figure 1.7 and Appendix C. The coordinates help with defining sets N^{close} and A^{cross}.

Lifetime revenue per OD flow route f_r^{rev}

This data is acquired by the demand forecasting model and its associated optimal fare setting. The data base made by Florczyk et al. (2020) provides information needed regarding the demand forecasting model (equation (3.2)). For this reason, it was deemed easier to make an adaptation in data acquisition when compared to the calibration stage (see section 3.2.3). This was done only for acquiring data on GDP per capita, since the dataset provides both sets of data needed:

- Total Gross Domestic Product GDP_i, which also allows us to calculate the GDP per capita. It is measured in 2011 USD and is adjusted for inflation.
- Population P_i within the defined urban centre boundaries (measured in 2015, which is deemed to be an accurate representation of today's population).

The demand database (see section 3.2.3) is based on air passenger numbers, and thus is reported per airport. As discussed in this section, it has to be translated to being UC-specific. For this reason, a *matching* is needed. To do this, each airport will be coupled the (main) city they serve. The coordinates and Haversine formula as introduced in section 3.2.3 are used here. In the section 3.2, two ways were mentioned in order to represent total trip generation:

- The observed number of air passengers $D_{AIR,ij}$
- The outcome of the gravity model

Due to being a real-life observation, the number of air passengers is perceived as more accurate than the outcome of the gravity model, which is only an approximation. Therefore, this methodology opts to always use observed air demand for forecasting total trip generation, unless (1) air demand is unknown, or (2) the travel distance is shorter than 1000 km, as these OD pairs are subject to competition from other modes (see section 3.2.3). In the latter case, the observed air demand is only used if larger than the outcome of the gravity model. This is done in order to never exclude (the more accurate) observed passengers, while always keeping a realistic estimator in the back of mind. Below, Table 3.7 summarises the methodology used for all possible scenarios.

Table 3.7 Used predictor of total trip generation

Travel distance	Air demand known	
	Yes	No
$<$ **1000 km**	max(gravity, airpax)	gravity
$\geq$ **1000 km**	airpax	gravity

3.4.9 Pre-processing

In this section, the model's pre-processing methodology will be described, which aims to reduce the sizes of sets that are input for the optimisation model. These are the set of nodes N, arcs A, OD pairs P ('network simplification') and OD pair flow routes R ('route generation').

Network simplification

Frei et al. (2010) cites multiple European long-distance travel studies, who all set 100 km as a minimum threshold regarding arc length. For this reason, Deutschel (2022) set the same distance as threshold when researching the potential mode shift to HSR in Europe.

In Europe, the longest distance between two HSR stations is 253 km (European Court of Auditors, 2018). Therefore, considering all potential arcs is unnecessary, as many of them are much longer than 253 km. To have a safe margin, it is decided to set 500 km as the maximum allowed arc length. This also matches the maximum distance over which European high-speed rail generally is competitive with other transport modes, according to last-mentioned authors. The 100 km minimum bound by Frei et al. (2010) will be used for this project as well.

Recall the work of Magnanti & Wong (1984): the TNDFSP is a NP-hard problem, meaning that its running time to solve increases exponentially with the number of nodes, arcs and OD-pairs considered. For this reason it is much more attractive to keep invalid arcs out of the model entirely, during the creation of the set of arcs A. Following up on this, the sets of nodes N and OD pairs P are updated as well.

Route generation

The goal of this algorithm is to find all valid routes for each OD pair's passengers. What makes a route 'valid' or 'invalid' will be described later in this section.

Following the previously-mentioned reasoning of Magnanti & Wong (1984), the potential number of routes is extremely large, increasing exponentially with the size of the network. An efficient algorithm must be used in order to find all valid routes. The Python package '*NetworkX*' will be used, known for efficiently analysing the structure of networks, and is the most commonly used package for this goal (Nvidia, 2024). Our algorithm will iterate over the OD pairs $p \in P$, and follow the steps below for each p.

(1) Potential network definition

This comprises all nodes and arcs the passengers of OD pair p are allowed to travel on. These sets are defined by the maximum detour factor k^{detour}, earlier set at 1.25 in section 3.4.5. Recall the meaning of this factor: passengers will not travel over routes more than 25% longer than the length of the shortest possible route. Below, equation (3.74) will mathematically define its meaning. Let the origin node of OD pair p be defined as n_p^{org} and its destination node as n_p^{dest}. Also, let the distance between two nodes a and b be defined as $d(a,b)$. Then, the set of nodes for OD pair p is defined as N_p:

$$\boldsymbol{N_p} = \{n \in \boldsymbol{N} \mid d(n_p^{org}, n) + d(n, n_p^{dest}) \leq k^{detour} \cdot d(n_p^{org}, n_p^{dest})\} \tag{3.74}$$

The set of arcs for OD pair p can then be defined as $\boldsymbol{A_p}$, which consists only out of arcs for which both end nodes are in $\boldsymbol{N_p}$. Then, for OD pair p, the algorithm only considers graph $G_p = (N_p, A_p)$, which can be considerably smaller than $G = (N, A)$.

(2) Path finding strategy

The algorithm searches for all paths between the origin and destination nodes of the OD pair, stopping when the path exceeds the maximum detour. It uses a breadth-first search (BFS) strategy, which explores nodes progressively by distance, unlike depth-first search (DFS), which goes as deep as possible before backtracking. Since distance is a key constraint due to the maximum detour, BFS is preferred for our algorithm, as it is more efficient for finding shortest paths (Rocha & Ferreira, 2018).

(3) Path constraints

The algorithm will not explore paths any further if the maximum detour for that OD pair p is exceeded. Four other criteria are defined to stop the algorithm from exploring redundant paths any further:

1. If it visits a node for the second time
2. If the added node moves further from the destination (or closer to the origin) of the OD pair
3. Anti-zigzagging rules: the angle of deviation between two consecutive arcs may not be larger than 90 degrees, and cumulative a maximum of 135 degrees is imposed
4. If the route fails to generate a positive level of demand. following equation (3.23), it would then hold $\alpha_p + \beta_p \cdot t_p + \gamma_p \cdot c_p \leq 0$.

The outcome for each OD pair p is a list of routes, defined as sequences of arcs. The algorithm will be ran over all $p \in \boldsymbol{P}$ to produce a complete set of routes $\boldsymbol{R}$.

3.4.10 Optimisation

The focus of this section lies on explaining how the model formulation will be applied in order to solve.

Hardware and software

The model formulation for network and line design will be written in Python 3.7.13 code language, with a loaded in Gurobi Optimizer 9.5.2, which is one of the world's fastest and most popular available commercial solvers for mathematical problems. It is well capable of solving various kinds of complex (non-)linear and other problems, such as transport network related formulations. For example, Air France uses it to design and optimise its flight schedule (Financial Post, 2019).

The mathematical problem related to this project will be solved on a Lenovo laptop with 2.3GHz Intel i7-11800H CPU and 32GB of RAM, and 64-bit Windows 11 as OS.

Strategy

Since it will be unknown a priori how the running time of the optimisation depends on the number of cities taken into account, it is decided to rank them based on their population. At first, the top two cities in terms of population make up the set of nodes. One city will be added at a time, as long as the running time of the model stays within a reasonable duration of six hours. It is assumed that after some point, no new cities will be added to the optimal network as their population has become too small to make an impact on the objective value. As discussed earlier, the Chinese HSR network is considered a leading example in HSR network potential. This network aims to connect all cities with over 500,000 inhabitants (China Daily, 2020). It is aimed for to take all European cities meeting this criterion at the least. Literature review in section 3.2 and 3.3 offered no findings of a 'minimum population'. Therefore, we hope to include more cities, lowering this inhabitant bound by as much as possible, as it increases the result's scientific value.

3.4.11 Model Validation

The workings of the model will be validated by means of a stability analysis and benchmarking. The methodologies for both are explained below.

Model stability analysis

Our model is used to inform important decisions. For this reason, it is important to assess if and how our optimal network design changes if we change the value of input parameters. This way, we can built our confidence in the model's predictions. After inspection of the parameters in section 3.4.7, it was determined that only the fare settings (travel costs) are likely to vary significantly in reality.

In order to assess how our model reacts to realistic fare variations, we multiply the fare of each potential connection with a randomly drawn value from a uniform distribution between 0.9 and 1.1, which simulates a random change deviating within 10% from the original fare. This test is ran six times, and possible changes from the original optimal solution will be assessed. Since our model must be ran a significant number of times, it was determined that the model should be able to find the optimal solution in approximately 30 minutes.

Benchmarking

As stated in section 1.2.3, currently, no model has been developed in order to truly optimise a network design, for medium-to-large sized networks, with reasonable computation times. The only methods currently out there are (meta)heuristics. In order to close the loop, we will see how well our optimisation model measures up against other state-of-the art heuristic TNDFSP solving approaches, it will be *benchmarked* to the network of Mandl (1980), which is the most widely known benchmark problems (Kechagiopoulos & Beligiannis, 2014). Benchmarking comes down to validating the optimisation model.

The network encompasses 15 nodes and 21 arcs, resembling a number of cities in Switzerland. Similar to our problem, demand is exerted bidirectionally, with peak hour demand already provided. In total, this demand equals 15,570 during peak hour. Each arc is attributed a travel time, which has no unit. A graph representation of the network is provided in Figure 3.3, while the demand matrix can be found in Appendix I.

In order to correctly benchmark our model, we must adhere to the nature of Mandl's network and the assumptions that comparable approaches made in order to address it. Only then, we can compare our results to other approaches in literature. The assumptions and adjustments by our model to address the network are listed below:

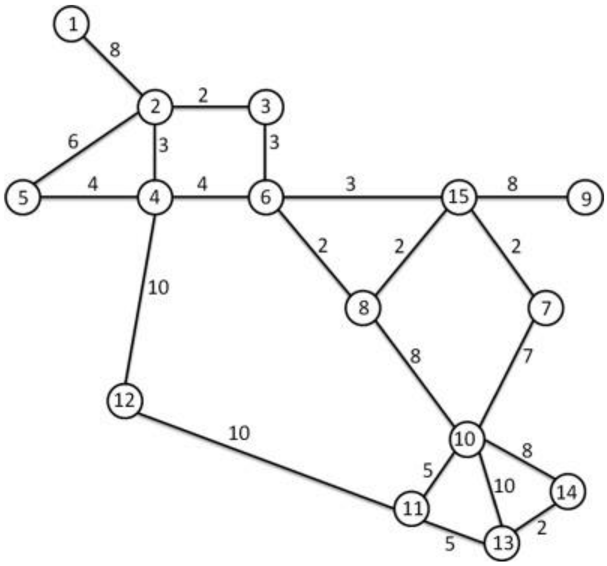


Figure 3.3 Mandl's network (Arbex & da Cunha, 2015)

- Set of nodes now consists out of 15 unnamed nodes
- Set of arcs only considers the arcs in Mandl's graph, not all possible node-pair combinations
- Set of OD pairs does consider all node-pair combinations
- Given the fact that the network is already constructed, sets of close nodes and crossing arcs can be left out.
- We only optimise the design for one peak hour, not for the entire lifetime, as no information is given regarding development of demand over time. Therefore, we can only optimise for operational profitability (equation (3.19)), not lifetime profitability. This means that infrastructure construction costs and rolling stock acquisition costs cannot be taken into account.
- Infrastructure maintenance and operation costs are not taken into account either, since they depend solely on the total length of infrastructure, which is pre-determined for this network.
- The seat capacity of transport vehicles is set to 40, matching the approach of Arbex & da Cunha (2015).
- The demand is kept elastic. For the shortest possible route, the demand is set to the value provided by the demand matrix. In order to exert the same maximum detour factor of 1.25, demand will be linearly interpolated in between, with demand set to zero if this detour factor is reached. Fares are set equal to €0.15 per km, which roughly matches the value found from our database.
- An average operating speed of 250 km/h is assumed, meaning the length of routes in km is equal to 250 times the travel time in h, and $\frac{d_r}{t_r}$ in equation (3.51) is set to 250.

4 Results

AI-generated high-speed train (Alam, n.d.)

4 Results

In this chapter, the methodology defined in the sections of chapter 3 is implemented. The sections of the results follow the same structure as the methodology in the previous chapter. Section 4.1 presents the results related to the demand forecasting model, section 4.2 does so for profitability estimation and section 4.3 for network design.

4.1 Demand Forecasting

As indicated in section 3.2.4, the gravity part of the model defined in equation (3.2) will be calibrated on air passenger data.

4.1.1 Data Collection for Calibration

This section presents the outcomes of the applied methodology described in section 3.2.3: an overview of the data gathered, which serves as input for calibration of the demand forecasting model.

Airports & air demand

The six pre-processing steps were followed. Table 4.1 presents a summary of the found results, with notable findings listed further below:

Table 4.1 Summary of air demand data pre-processing

Pre-processing step	Airport pairs	Unique airports	Passengers ($x10^6$)	Tracking years
1. Data set combining	19,136	716	29,210	1993-2023
2. Time scoping	9,751	680	1,898	2019
3. Area scoping	5,338	294	1,082	2019
4. Airport combining	4,320	268	1,082	2019
5. Directional combining	2,656	268	1,136	2019
6. Competition adaption	514	71	320	2019

- **Data set combining;** The 36 country-specific air passenger data sets (Appendix C) were combined. Yearly passenger data was found over a 31-year tracking period (1993-2023).
- **Time scoping;** In the year of interest (2019), 1.92 billion passengers were transported. Of the raw data, only 9,822 airport pairs (51.3%) contain nonnegative yearly passenger counts, and 71 (0.7%) were reported as 'Unknown'. All countries in the scope are still represented, except for Bosnia and Herzegovina, for which only 2021 data is available (see Appendix C).
- **Area scoping;** More than half (56.8%) of unique airports were deleted, as they were outside of the scope of the project. Their related airport pairs (and thus passengers) were deleted as well. The great majority of deleted airports are 'the other ends' of intercontinental flights, but also airports located on European islands were deleted.
- **Airport combining;** As expected, in some cases multiple airports of the same origin city serve flights to the same destination city. In the database's most extreme case, it contains passenger counts for nine different airport combinations from the city of London to the city of Milan. In fact, by transforming the airport-oriented into city-oriented data, the data set shrinks significantly in the number of airports and associated OD pairs.
- **Directional combining;** For 23% of locations (5% of the passenger numbers), the return passenger count was missing. For these city pairs, the total travel potential was calculated as twice the outward passenger number. The available data shows that this assumption can be made. As seen from the mentioned percentages, the missing data mainly originates from smaller airports. In

total, 268 unique cities, and 2,656 unique city pairs were found when combining outward and return flights. Of these, Table 4.2 shows the five most popular city pairs for air passengers. Note the alphabetical ordering of city names in city pairs, which was chosen deliberately.

Table 4.2 Top 5 most popular air connections (city-pairs) within scope

#	City 1	City 2	Passengers
	Name, Country	Name, Country	(2019)
1	Amsterdam, Netherlands	London, England	9,847,961
2	Barcelona, Spain	London, England	7,209,728
3	Edinburgh, Scotland	London, England	6,746,995
4	Paris, France	Toulouse, France	6,441,863
5	Nice, France	Paris, France	6,364,516

- **Competition adaption;** When adjusted for competition among airports, the data set attained after the previous five pre-processing steps serves as a starting point, as it encompasses all air passenger numbers that could be retrieved. 71 unique airports were found with 514 connections among them, serving 320 million passengers yearly. Even though only 19% of the full data set's city pairs are considered for this sample, they represent 28% of the passengers. The figure below shows the selected airports for calibration. Figure 4.1 presents the eligible set of unique airports.

Figure 4.1 Airports in calibration data

Model-implied impact factors

This section will first look into how missing data was handled. Then, a statistical summary of all impact factor data is presented.

Missing data

Travel cost data missing in Rome2Rio (2024) was found for a number of unforeseeable cases. Here, it is listed how these were solved.

- **Special case 1:** Non-reported travel costs for car-based, train-based or plane-based trips. They will be estimated by using:
  - The average driving cost / hour equals € 22.35 (calculated from data set)
  - The average train travel cost / hour equals € 14.85 (calculated from data set)
  - The average plane travel cost / hour equal € 46.93 (calculated from data set)

- **Special case 2:** Trips making use of the Eurotunnel between the UK and France are classified as train-based trips, even though one may also board a car on the train. To allow car-based trips between the UK and France, their travel times and costs are estimated by adapting the found travel costs for trips through the Eurotunnel, using the following:
  - The Eurotunnel car travel time equals 35 minutes (Eurotunnel, 2024)
  - The average driving cost / hour equals € 22.35 (calculated from data set)
  - The Eurotunnel car fare price equals € 135 (Eurotunnel, 2024)

  The following formula is used to calculate the new car-based Eurotunnel trip's travel cost TC_{new}, based on the reported travel time in hours TT_{old}:

$$TC_{new} = \left(TT_{old} - \frac{35}{60}\right) \cdot 22.35 + 135 \tag{4.1}$$

- **Special case 3:** No car-based travel alternative found, even though all cities lie in continental Europe and connected by road.
  - The average speed of a car-based trip is 105.88 km/h (calculated from data set)
  - The travel distance by car can be found through web scraping (section 3.2.3)

Statistical summary

The data shows great variation among all variables, indicated by the standard deviation being larger than the mean value. This is great news, as the calibrated model should be able to cope with a wide range of population catchments $P_{i,t}$. The same can be said about the other variable that relates to people: the number of air passengers $D_{AIR,ij}$. GDP (city total GDP_i or per capita $GDPCAP_i$) and distances d_{ij} vary much less, which does not impose a problem, since the choice of locations is spread well across the continent. The table below represents the key characteristics of the data used for model calibration.

Table 4.3 Characteristics of calibration data

Variable	Unit	N	MIN	Q1	MEDIAN	Q3	MAX	AVG	STD
$P_{i,15}$	[pax]	71	923	129,989	253,994	636,786	2,342,971	442,383	474,620
$P_{i,30}$	[pax]	71	4,133	356,929	708,944	1,662,124	9,974,841	1,265,734	1,548,360
$P_{i,45}$	[pax]	71	12,388	520,215	1,113,145	2,490,703	13,417,545	1,916,266	2,332,799
$P_{i,60}$	[pax]	71	36,904	797,282	1,515,685	3,082,523	14,474,396	2,504,137	2,835,263
$P_{i,75}$	[pax]	71	43,916	1,060,284	1,952,954	3,722,006	15,108,562	3,121,031	3,270,876
$P_{i,90}$	[pax]	71	59,127	1,309,692	2,580,338	4,941,481	17,432,372	3,820,950	3,768,015
$P_{i,105}$	[pax]	71	68,970	1,628,828	3,298,628	5,920,694	20,158,470	4,618,074	4,424,464
$P_{i,120}$	[pax]	71	72,672	1,996,925	3,646,519	6,874,958	23,084,990	5,473,373	5,280,943
GDP_i	[M€]	71	1,178	15,344	35,447	98,136	757,630	84,055	127,604
$GDPCAP_i$	[K€]	71	5.980	38.293	58.562	103.434	607.256	83.758	84.513
d_{ij}	[km]	514	1,001	1,229	1,517	1,882	3,364	1,623	482
$D_{AIR,ij}$	[pax]	514	34,660	121,820	309,341	736,350	7,209,728	622,587	871,726

N = number of values; MIN = minimum; Q1=first quartile; Q3=third quartile; MAX=maximum; AVG=average; STD=standard deviation
K€ = thousand euros; M€ = million euros

4.1.2 Model Calibration

This section describes the outcomes of the conducted methodology in section 3.2.4, which states the road to attributing values to each of the demand forecasting model's parameters. In the mentioned section, it was determined that only the gravity part is calibrated to observed data.

The gravity model was calibrated for various catchment area sizes, as well as for both total GDP and GDP per capita, to see what GDP indicator explains the data best. The R^2 fit for both models at various catchment area sizes are plotted in Figure 4.2.

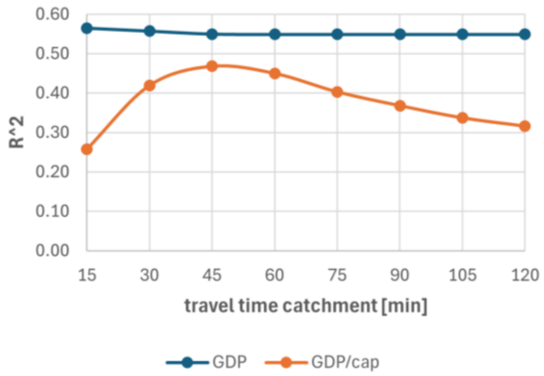


Figure 4.2 Model fit for varying catchments and GDP measures

In terms of R^2, the GDP model outperforms the GDP/cap model for all catchment area sizes. However, this seems due to *overfitting*, as most calibrated parameters are not showing a decent statistical significancy, and some are of the wrong sign. The likely cause is the intercorrelation between two of the GDP model's variables: population and total GDP. The GDP/cap model's fit varies with the choice of catchment area size, with an optimal fit at 45 minutes. This aligns with the findings of Martínez et al. (2016) in section 2.1.6: 80% of HSR users live within a 30 min travel. Even though the GDP/cap model has a poorer fit than the total GDP model, most of its calibrated parameters show strong statistical significance and have the correct sign. Due to the significant parameters, the model is likely to represent true relationships in the data, making it easier to interpret and trust. This also enhances predictive power, as the model likely captures real effects. For these reasons, the GDP/cap model is chosen, as it balances a good fit with great parameter significance. Below, Table 4.4 shows the calibration results.

The GDP/cap model also points out one could better underestimate than overestimate the catchment area size, as for smaller catchment areas the optimal model fit decreases rapidly. It indicates the significant role that large regional airports serve. Typically, 45 minutes is ample time to reach destinations situated reasonably far from the city and its centre. Trying all possibilities of combinations of catchment areas, as described in section 3.2.3, yielded no better results.

Table 4.4 Chosen gravity model's calibration results

Parameter	Coeff.		Std	t-stat	p-value
k	-2.524	**	0.606	-4.16	3.74×10^{-5}
α	0.564	**	0.028	20.31	0
β	0.382	**	0.046	8.22	1.55×10^{-15}
γ	0.139		0.130	1.03	3.02×10^{-1}

Est. = estimated value; Std = standard deviation; Significant at conf. level: 95% (*), 99% (**)

All parameters' estimates of the expected sign, and all but distance parameter γ are showing a great level of significance. This means that in the data, the relationship between distance and demand cannot be picked up as easily as for other model-assumed relationships. The parameter is still significant at a 70% confidence level.

Figure 4.3 shows the relationship between the observed and estimated total passenger flows for the 514 city pairs used for calibration. In red, the line $y = x$ is plotted. For a perfectly accurate demand forecasting model, all points should be on this line. However, it can be seen that popular routes are underestimated and unpopular routes are overestimated. The cause for this bias is already known; literature review in section 2.1.6 states that larger airports generally have larger catchment areas. Even though small airports are excluded, still a wide variety in passenger numbers per year exists among the airports (1-192 million). In this model however, popular airports are assumed to have the same catchment area size as small airports, leading to an underestimation of the potential number of passengers for large airports, and an overestimation for smaller airports. This fact is even more accentuated by the trend line of the data points, indicated in orange in Figure 4.3 below. The model fits the data reasonably well, reaching a R^2 value of 0.468.

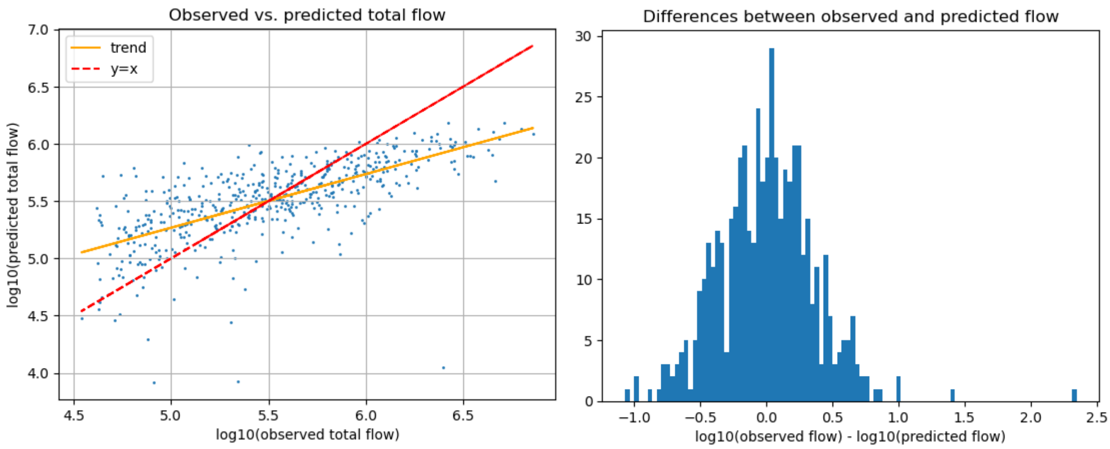


Figure 4.3 Left: observed vs. predicted total flow. Right: accompanying error histogram (N=514, R^2=0.468)

Bias elimination

As described in section 3.2.4, the found bias will be removed by making use of the orange trendline, for which linear parameters $a = 0.4684$ and $b = 2.9257$ were found. This means all predicted flows should be adjusted by equation (3.6). When applied to the data points in Figure 4.3, it can be observed that the bias is indeed eliminated, while increasing the model fit significantly to 0.751:

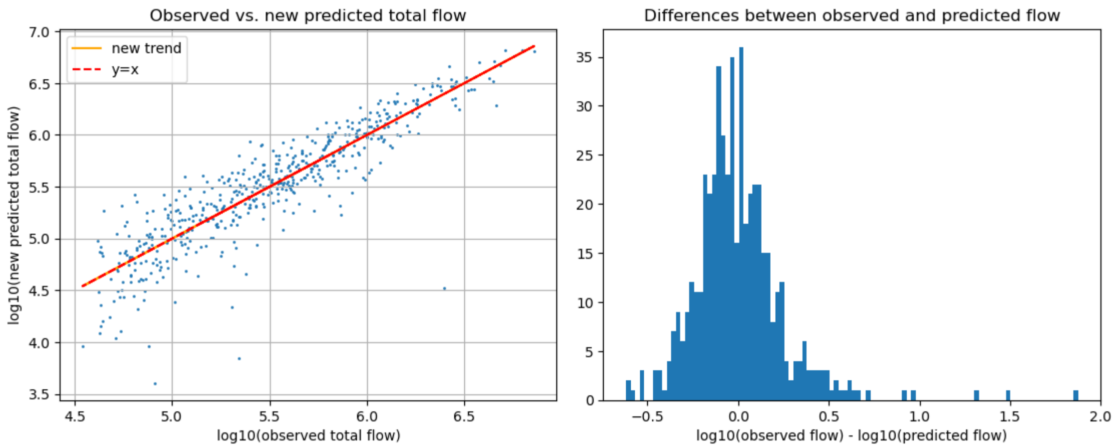


Figure 4.4 Left: observed vs. new predicted total flow. Right: accompanying new error histogram (N=514, R^2=0.751)

4.1.3 Model Validation

Section 3.2.5 described how the calibrated part of the demand forecasting model will be validated. It can be verified using Table 4.1 that taking the data erased by last pre-processing step yields a data set of 816 million passengers, traversing 266 unique cities and 2,142 connections. After elimination of intermodal competition effects by removing connections with length under 1,000 km, a validation data set of 180 million passengers, 187 cities and 890 connections remains. The relatively large share of small Scandinavian airports providing data for regional flights only is responsible for the significant loss of cities in the data set. A statistical summary of this data set is provided in the table below.

Table 4.5 Characteristics of validation data

Variable	Unit	N	MIN	Q1	MEDIAN	Q3	MAX	AVG	STD
$P_{i,45}$	[pax]	187	4,191	1,797,283	4,129,149	7,798,901	29,518,637	6,225,502	6,358,880
$GDPCAP_i$	[€]	187	5,980	33,844	48,534	93,859	607,256	72,364	68,010
d_{ij}	[km]	890	1,000	1,225	1,473	1,741	3,120	1,527	377
$D_{AIR,ij}$	[pax]	890	20,114	67,137	99,356	214,235	3,195,192	202,426	293,041

N = number of values; MIN = minimum; Q1=first quartile; Q3=third quartile; MAX=maximum; AVG=average; STD=standard deviation
K€ = thousand euros

Since the calibration stage determined the catchment area size and the GDP indicator, the table above contains much less rows than Table 4.1. Both the number of cities and connections are larger than they are in the calibration data set.

The validation approach used by Belal et al. (2020) is adopted here. We select five 'interesting' air connections and compare the model's predictions against the observed demand for these routes. To identify 'interesting' connections, we filter the database for routes with an observed yearly demand of over 500 thousand passengers, as the model specifically targets popular connections. To ensure variation in the thousand data, no city is selected more than once. The table below presents the model's accuracy in forecasting demand for these selected connections.

Table 4.6 Validation results

Connection	Observed demand		Predicted demand		Difference
	Ordinary notation	log	Ordinary notation	log	(% in 10-log)
London-Faro	3,195,192	6.50	54,448	4.74	-27.2
Paris-Seville	945,180	5.98	851,616	5.93	-0.8
Milan-Porto	780,703	5.89	843,851	5.93	+0.6
Istanbul-Geneva	680,629	5.83	554,840	5.74	-1.5
Málaga-Düsseldorf	555,346	5.74	509,695	5.71	-0.6

The results show that the model predicts demand reasonably well, with the absolute deviation in terms of 10-log staying within 2%. The only exception here is the London to Faro connection, which is severely underestimated by the model. This indicates that the model is not able to estimate trips that have a heavy touristic character. This comes as no surprise, since the model only takes population, GDP and distance into account. The correlation between observed and predicted demand in the table above, when expressed in base 10 logarithm, equals a R^2 value of 0.773. Without taking connections with a touristic character into account, this value would reach 0.793.

4.1.4 Demand Evolution

This section presents the results after implementation of the related methodology in section 3.2.6.

Economic demand growth

Section 3.2.6 described how the economic growth-related demand factor k_{ij}^{eco} for a connection between city i and city j should be calculated. Regarding data collection, it applies to finding p_{ij}, which is the minimum economic growth rate among the two countries related to city pair ij as a percentage. As explained, for each country this value is found by taking the median yearly growth rate p over the last 40 years. The results of the data provided by World Bank (2023) are presented in the figure below:

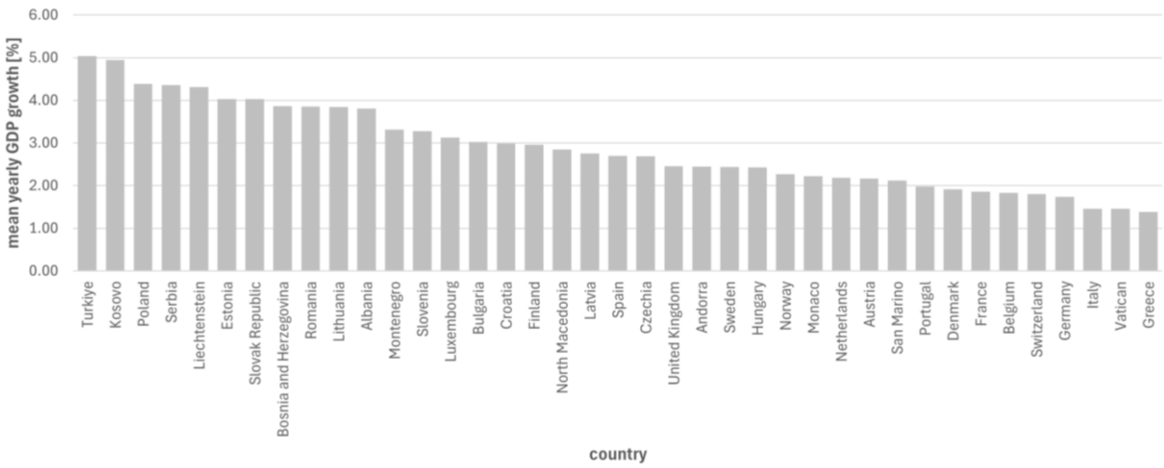


Figure 4.5 Mean yearly GDP growth (1964-2023)

A significant variation exists among the 39 countries within this project's scope, between 5.04% (Turkiye) and 1.38% (Greece). From the graph, one can confirm that the highest p_{ij} value belongs to connections between Turkiye and Kosovo, while the lowest value belongs to any connection made with or within Greece. When inserting the possible values into equation (3.8), it turns out that the factor k_{ij}^{eco} varies between 1.22 and 2.17.

Induced demand

As described in section 3.2.6, the level of induced demand is the same for every connection, unless it already has an upgradeable high-speed rail line. Therefore, induced demand will come into play during the network design stage, and therefore the related results can be found in section 4.3.

4.2 Profitability Estimation

This section provides results related to the profitability model formulated in section 3.3, in which the values of all parameters were already determined. Only the unit construction cost per km (for surface and tunnelling) still has to be calculated. The steps taken to do so, along with the end results for each country are listed in Appendix F.

4.3 Network Design

This section provides results related to the methodology introduced in section 3.4, which has been split into three parts here: data collection (section 4.3.1), pre-processing (section 4.3.2) optimisation (section 4.3.3). The model is validated in section 4.3.4. Lastly, in section 4.3.5, experiments are conducted to identify secondary connections, meaning connections that may become relevant under certain HSR-beneficial conditions.

4.3.1 Data Collection

The process summarises the findings of the implementation of the methodology provided in section 3.4.8, which described how the model input data is gathered. As a starting point, the potential network is defined by the set of nodes N, arcs A and OD pairs P. As highlighted previously in section 3.4.8, the definition of set N fully determines the definition of the two other sets.

Candidate cities

The data set by Florczyk et al. (2020) provides data on 160 metrics for 13,135 urban centres (hereafter referred to as cities) from 184 unique countries. After analysis of Figure 1.7, it was found that the scope is covered by a rectangle bounded by 35 to 72 degrees latitude and -12 to 44 degrees longitude. 1,265 cities lie within this rectangle. The eventual selection of cities is presented in Figure 4.6, which has the exact same boundaries. The following two steps removed cities from the rectangle, for different reasons:

- 391 cities are located in countries outside of the scope defined by Figure 1.7 and Appendix C
- 148 cities are located in countries within the scope, but are not part of continental Europe due to being located on an island, or in the Asian part of Turkiye.

Below, Figure 4.6 shows the remaining 726 cities, originating from 35 countries. Together, they form the set of potential nodes N, sharing 263,175 possible connections among them.

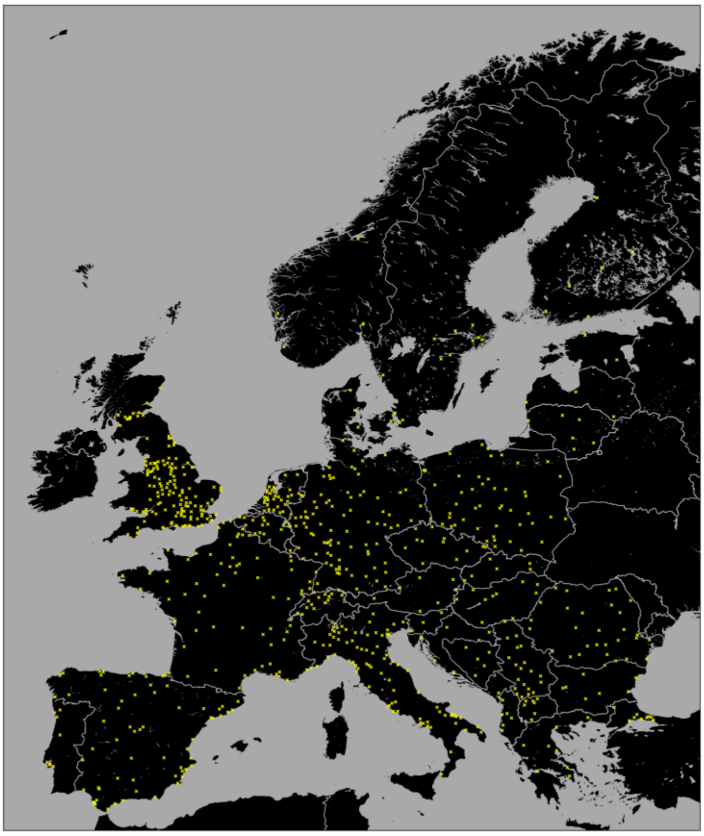


Figure 4.6 Urban centres within scope

Below, Figure 4.7 shows the representation of every country:

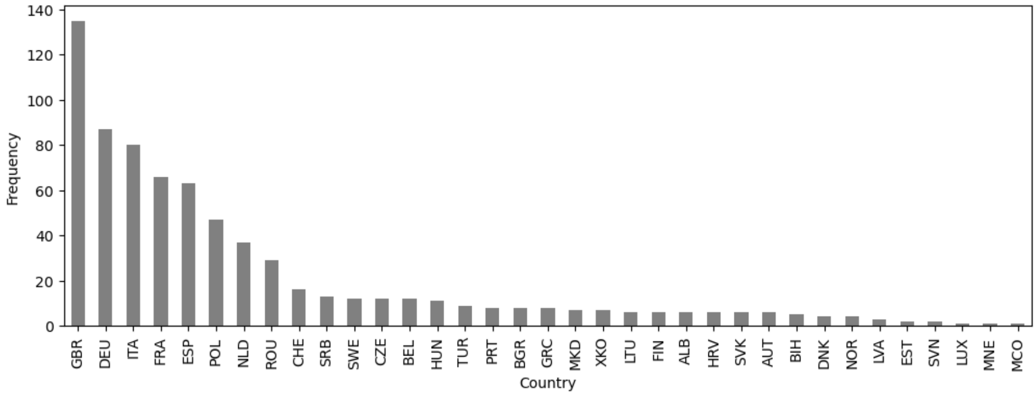


Figure 4.7 Cities per country in potential network

By far, almost a fifth (19%) of all cities are from Great Britain (GBR). A clear top-8 exists, which also includes Germany (DEU), Italy (ITA), France (FRA), Spain (ESP), Poland (POL), the Netherlands (NLD) and Romania (ROU). A full translation of all other ISO codes can be found in Appendix C.

Following the methodology described by section 3.4.8, the needed data was gathered regarding all potential 726 cities. Below, a statistical summary of the city-related model input data is given:

Table 4.7 Statistical summary of potential nodes

Characteristic	Unit	Min	Q1	Med	Q3	Max	Mean	Std
Population	[pax]	50,054	66,977	101,361	199,353	14,111,242	285,165	844,314
Catchment	[pax]	195,736	630,026	1,243,560	2,515,302	13,417,545	1,845,190	1,786,929
GDP / capita	[€]	84	11,919	15,772	20,096	51,478	16,105	6,749
Mean elevation	[m]	-3	32	80	190	909	142	161

It is clear that the 726 cities encompass a wide range of values in all characteristics. With all human settlements of population over 50,000 represented, there is confidence that the set encompasses all potential HSR stations.

Matchings

As stated by the methodology provided in section 3.4.8, airports are matched to cities in set N in order to appropriately use the air demand data. Between the 268 airports from the air passenger demand data set, and the 726 cities, 200 matchings were found. For 198 of these cities, air passenger data is available to at least one other airport. A complete list of these urban centres and their allocated airport can be found in Appendix C. The original air passenger demand data set contains data for 2,656 airport pairs. When translated by use of the matchings with urban centres, it was found that 2,340 of them are connections between urban centres. The size of mentioned data sets are listed in the table below.

Table 4.8 Data set sizes

Data set	Nodes	Connections
Potential network	726	263,175
Demand data set (airports)	268	2,656
Demand data set (urban centres)	198	2,340

While only 74% of the airports could be matched to an urban center, 88% of their connections can. This is due to the fact that the missing airports are not associated with an urban centre. Therefore, these airports are generally small, and thus have a relatively low number of connections.

Potential arcs: web scraping results

Now the information regarding the set of nodes N is gathered, it is used to find all necessary arc- and OD pair-related data by means of web scraping, following the methodology described in section 3.4.8. Web scraping Rome2Rio and Google Maps was firstly performed for all 31,125 city pairs for the 250 most populated cities, as this already took more than three days to complete. The task of performing such large amounts of searches regularly caused errors due to not being able to load web pages in time. The web scraping code had to be adapted to this, to automatically search arc-related data again if the load errors occurred. A complete database encompassing the needed data for all 31,125 city pairs was made successfully. Below, a statistical summary of this data base is given. Note that all characteristics are based on the direct arc between origin and destination.

Table 4.9 Statistical summary of model input data

Characteristic	Unit	Min	Q1	Med	Q3	Max	Mean	Std
Demand forecasting								
Line length	[km]	16	856	1,423	2,064	5,041	1,507	832
Travel time (plane)	[h]	0.0	5.5	6.6	7.9	26.9	6.6	2.3
Travel time (train)	[h]	0.0	8.3	15.4	24.7	100.0	17.9	12.7
Travel time (car)	[h]	0.2	8.3	13.5	19.7	59.6	14.5	8.1
Travel time (HSR)	[h]	0.1	2.8	4.6	6.6	16.1	4.9	2.6
Travel cost (plane)	[€]	0	184	220	268	1,355	228	90
Travel cost (train)	[€]	0	150	268	380	1,341	272	158
Travel cost (car)	[€]	4	205	336	484	1,453	355	196
Travel cost (HSR)	[€]	37	132	170	213	734	177	71
1st year HSR demand	[pax]	14	6,899	24,237	85,628	10,263,681	108,926	313,918
Average train speed	[km / h]	0	68	83	100	282	83	29
Alpha-parameter	[pax]	35	19,010	61,874	195,717	22,732,344	234,875	662,871
Beta-parameter	[pax / h]	-974,438	-8,935	-2,515	-709	-1	-11,355	32,557
Gamma-parameter	[pax / €]	-66,572	-610	-171	-48	0	-772	-2,213
Model linear fit (R^2)	[-]	0.975	0.995	0.997	0.999	1.000	0.997	0.003
Profitability estimation								
Unit construction cost	[M€ / km]	7.0	37.8	47.6	57.7	116.1	48.9	16.7
Lifetime revenue	[M€]	0.1	43.0	139.6	423.7	64,577.4	521.5	1,540.6
Lifetime infrastructural cost	[B€]	0.8	44.4	71.5	100.9	314.4	75.7	42.6

Since all cities are located in continental Europe, travelling between them by car or the potential new HSR line should always be possible. This is reflected in the data, as the minimum travel time of these modes is larger than zero. A relatively low share city pairs are not connected by plane (4.5%) or train (3.9%), indicating the existence of infrastructure related to all modes. Most of the city pairs can be traversed by plane within a day, while the median train travel time already exceeds 15 hours. A comparison to the median 4.6 hours of potential HSR infrastructure is indicative to the opportunity of HSR development in continental Europe. This is even further accentuated by comparing travel costs: when instructed to maximise revenue, high-speed rail can be significantly cheaper than competing modes.

This does not mean that HSR infrastructure should be developed everywhere across the continent. As the data shows, in many cases the demand is low, also indicated by a median of only 24,000 passengers per year. Only 1.6% of all city pairs would be able to generate a first-year demand of more than 1 million. Considered solely, only eleven (0.03%) of them would be profitable. One should keep in mind that this only considers direct connections – in networks, most city pairs are not directly connected. It is already indicative to how the optimal network would look like: an optimal configuration centred around the biggest few European cities. The 'average train speed' was calculated in order to find upgradeable infrastructure, for which the value must be at least 200 km/h. A total of 29 arcs met this criterion. An overview is provided in Appendix G. Linearisation of the demand forecasting model was needed in order to be suitable input for the optimisation model. For all arcs, the parameters are of the expected sign, showing a great variety similarly to that of the estimated demand itself. It can be verified that the linearised demand model closely approaches the original model, as R^2 values are close to one for all arcs.

4.3.2 Pre-Processing

This section presents the results of pre-processing, which aims to reduce the sizes of the set of nodes N, arcs A, OD pairs P ('network simplification') and OD pair flow routes R ('route generation') by removing their unrealistic elements, in hopes to significantly reduce the model complexity and solving times, while not affecting the optimal solution. The methodology for this section was described in section 3.4.9. The largest potential network size our laptop could solve for within six hours considers the 111 most populated cities within the scope, which are all cities with a population exceeding 315,000.

Network simplification

Without pre-processing, our optimisation model would consider 111 nodes, connected by 6,105 arcs and the same number of OD pairs. Following the steps described in section 3.4.9, the size of the network and mainly its related set A was reduced significantly: only 109 nodes, 589 arcs and 5,886 OD pairs. This is due to the fact that our network simplification process primarily imposes constraints to arc lengths.

Route generation

The *NetworkX* route generating algorithm proved to be significantly faster than the optimisation process. Considering the final potential network size of 111 cities, it found 77,062 valid potential routes in approximately eight minutes. Even though the number of routes averages to 13 for each OD pair, the distribution of routes among the OD pairs is extremely uneven. Only 2,269 out of 5,886 OD pairs have at least one route, indicating that many city pairs cannot by connected by a valid demand-generating route. In fact, the top 10% OD pairs with the most routes account for over 90% of all routes found. This can be attributed to the fact that a relatively low number of city pairs are far apart and are able to generate demand, exponentially increasing their number of potential routes. For illustration, Cardiff-Munich has the highest number of routes: 1,722, which is 132 times higher than the average. On the other hand, many city pairs fail to generate demand, often caused by at least one of them having a low population while having a medium-to-long distance in between. Since the set of routes together make use of all 109 nodes and 589 arcs in the potential network, the relatively low representation of OD pairs cannot be attributed to any disconnections in the network.

4.3.3 Optimisation

The methodology for this section was described in section 3.4.10. Optimising for a network of 111 cities (effectively: 109, as explained earlier), 589 arcs, 2,269 OD pairs and 77,067 routes resulted in construction of a model with 243,671 integer decision variables (of which 89,537 are binary) and 1,035,732 constraints. The optimal solution was found after a little under six hours, an optimal lifetime profitability of €222.8 bn was reported, which can be broken down into the cash-flows present in the objective function:

Table 4.10 Profitability breakdown in optimal network (B€)

Revenue		Costs	
Ticket revenue	€ 655.413	Infrastructure construction	€ 193.809
		Infrastructure maintenance & operation	€ 15.876
		Rolling stock acquisition	€ 5.062
		Rolling stock operation & maintenance	€ 194.134
		Transfer penalty	€ 12.739
Total revenue	**€ 644.413**	**Total cost**	**€ 421.620**
			Total profit: € 222.793

Topology

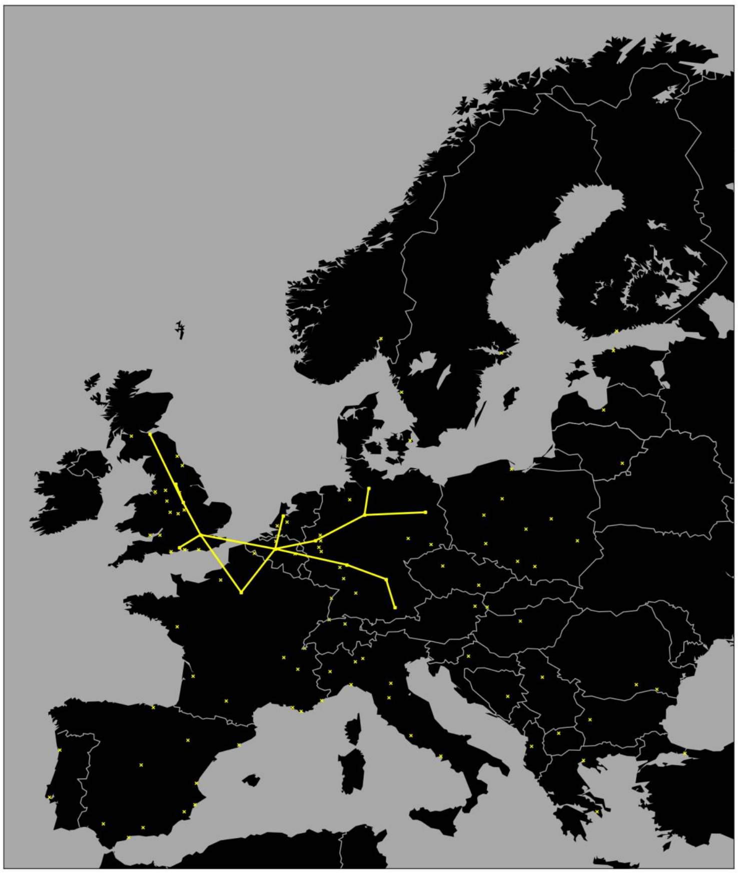


Figure 4.8 Optimal network topology

The optimal topology is presented by Figure 4.8, consisting out of 15 cities, connected by 15 arcs. The yellow dots not connected by lines, are cities that the model considered, but did not add to the network. Most of the arcs will be newly built, as only two are currently in high-speed operation: Brussels-Paris (average speed: 229 km/h) and London-Paris (200 km/h).

It heavily focuses on north-western Europe. Since our model accounts for already existent HSR infrastructure, it becomes evident that it's not worth the investment of upgrading domestic lines in the networks of France, Spain and Italy, as their current quality and coverage is sufficient. Simultaneously, it shows that mainly Germany and Great Britain are in dire need of more border-crossing HSR infrastructure. The network avoids mountainous terrain, indicating that this might affect the viability of HSR operation. Some connections are drawn as if they would cross water, but HSR travel times are determined by the shortest distance over land.

The network covers the largest cities in the remaining western countries: Great Britain, Germany, Belgium and the Netherlands. One can notice a clear triangle-structure, with extending arms in multiple directions to large cities such as Berlin, Amsterdam, Hamburg, Munich and Edinburgh, which all show great air demand with at least one of the cities in the triangle. Having one of the largest airports in the world, the city of Frankfurt is also part of the network. Some medium-to-large cities function as intermediate stops between larger cities, as they benefit from their location, while still adding a sufficient number of passengers to the network. Examples of these cities are Dusseldorf, Hannover, Stuttgart, Leeds, Nottingham and Nuremberg. Perhaps the most surprising addition to the network is the city of Southampton, which is located closely to London, and shows sufficient demand levels with the latter city and Paris.

All in all, the found optimal topology underlines our expectations of being exclusively centred around Europe's largest cities, while showing the currently unserved HSR demand potential in countries such as Great Britain and Germany. Below, an overview is given of the network's nodes:

Table 4.11 Stations in optimal network

Country	Selected node(s)
Germany	Berlin, Munich, Frankfurt, Nuremberg, Dusseldorf, Hamburg, Hannover
Great Britain	London, Leeds, Nottingham, Edinburgh, Southampton
France	Paris
Netherlands	Amsterdam
Belgium	Brussels

It should be noted that several markets with high potential for HSR are neglected by the network, with Spain, Italy, and much of France being the most prominent examples. As previously discussed, these countries already possess extensive, well-functioning national HSR networks, reducing the incentive to develop new infrastructure or upgrade existing lines. In our model, international connections between these countries and our network (Figure 4.8) are significantly hindered by the so-called 'Empty Diagonal'—a vast rural area south of Paris that occupies roughly a third of France's land area (Bopp & Douvinet, 2020), creating a barrier for HSR development. In Switzerland and Austria, despite the presence of several major cities, connections seem to be excluded by higher construction costs, largely due to the need for extensive tunnelling through challenging terrain. Though construction costs are more favourable in eastern European countries, the significantly lower GDP reduces travel potential, which likely explains why this region is excluded from our final network.

The combined length of all arcs equals 3,969 km, with lengths of individual arcs varying between 120 and 464 km. The network would serve close to 600 thousand passengers per day, on average over its lifetime. Focussing purely on ticket revenue and infrastructural costs, it appears that almost two-thirds (63%) of all profit originates from the London-Paris and London-Brussels arcs. Another remarkable finding is that for a maximally profitable network, not all individual arcs have to be profitable on their own: 5 out of 15 arcs are not. An example of this is the arc Hamburg-Hannover, which is expected to make a loss of €4.4 bn over its lifetime, when considered individually. However, many passengers originating from Hamburg have destinations reaching much further than Hannover, therefore increasing the profitability of other arcs, compensating for the loss on their 'home' arc. The same story can be told for the other four unprofitable arcs, as they all are situated at an end point of the network. The table below shows the profitability for each selected arc, only considering ticket revenue and infrastructural costs, as rolling stock-related costs depend on the design of lines, which will be addressed in the next section.

Table 4.12 Profitability and demand data of individual arcs

Connection name	Length [km]	Unit cost [M€ / km]	Flow [pax / day]	Min. freq [h^{-1}]	Revenue [B€]	Costs [B€]	Profitability [B€ / lifetime]
Brussels-London	364	44.1	131,680	6.9	199.943	17.513	**182.430**
London-Paris	464	52.1	42,105	2.2	106.965	13.023	**93.941**
Brussels-Dusseldorf	201	45.2	67,474	3.5	54.182	9.890	**44.292**
Amsterdam-Brussels	211	41.3	56,345	2.9	45.245	9.563	**35.683**
Brussels-Paris	317	39.7	45,726	2.4	40.057	6.934	**33.122**
Brussels-Frankfurt am Main	397	50.0	46,779	2.4	46.443	21.436	**25.007**
London-Nottingham	206	55.4	42,771	2.2	33.721	12.233	**21.488**
Dusseldorf-Hanover	280	58.0	33,654	1.7	26.042	17.371	**8.671**
London-Southampton	123	55.5	13,021	0.6	9.125	7.323	**1.803**
Edinburgh-Nottingham	449	53.9	17,155	0.9	26.800	26.017	**0.783**
Leeds-Nottingham	120	59.7	11,931	0.6	7.490	7.646	**-0.155**
Frankfurt am Main-Nuremberg	223	78.8	22,566	1.1	16.473	18.474	**-2.000**
Munich-Nuremberg	172	80.6	16,618	0.8	12.131	14.557	**-2.426**
Berlin-Hanover	290	57.6	20,093	1.0	14.375	17.865	**-3.490**
Hamburg-Hanover	152	60.7	8,249	0.4	5.420	9.840	**-4.420**
TOTAL	**3,969**		**576,167**		**644.413**	**209.685**	**434.728**

One can verify that the total reported revenue and costs match the data of Table 4.10.

OD pairs

Of the 105 potential OD pairs in the network, 60 (57%) are currently served. OD pairs remain unserved when no valid route is selected between them, typically due to the shape and orientation of the chosen arcs or the OD pair's demand level, leading to a too high detour factor. Passengers on these unserved OD pairs can still travel across the network by making one transfer, but their detour exceeds the maximum allowable threshold. To achieve full coverage (100%), additional arcs would need to be selected to provide more direct connections for all OD pairs. While this would improve coverage, it would likely reduce the network's profitability. Therefore, a balance must be found between OD pair coverage and profitability, with optimal profitability reached at 57% coverage, in this case.

Travel times, costs and distances of selected OD pairs vary between 0.46 - 4.74 hours, €43 – 326 and 120 – 1,362 km, respectively. First-year passenger demand (accounted for induced demand) varies from 36,000 (Edinburgh-Southampton) to 20.2 million (Amsterdam-London). A statistical summary of the selected OD pairs is given below. HSR market share per served OD pair varies between 30 and 92%, with a mean of 75% per OD pair. The market share of the entire network equals 79% (with a roughly equal share for the other modes). A complete overview of all served OD pairs is provided in Appendix H.

Table 4.13 Statistical summary of selected OD pairs

Characteristic	Unit	Min	Q1	Med	Q3	Max	Mean	Std
Length	[km]	120	445	663	903	1,362	673	321
Travel time	[h]	0.46	1.53	2.28	3.18	4.74	2.34	1.11
Travel cost	[€]	43	104	165	223	326	168	81
First year demand	[Mpax]	0.037	0.347	0.971	1.755	20.225	1.990	3.326
Lifetime revenue	[B€]	0.235	1.829	4.556	9.111	107.190	10.740	18.684
HSR market share	[-]	0.301	0.687	0.832	0.869	0.918	0.751	0.168

Line design

The topology displayed in Figure 4.8 will be served by eleven lines and a fleet of 81 trains. A complete overview is provided below.

Table 4.14 Optimal lines

#	Stops	Length [km]	Travel time [h]	Served freq [h^{-1}]	Fleet [-]
1	(2): **London**, **Paris**	464	1.55	3	10
2	(5): **Berlin**, Hannover, Dusseldorf, Brussels, **Paris**	1,088	3.78	1	8
3	(3): **Amsterdam**, Brussels, **Paris**	528	1.84	1	4
4	(3): **Amsterdam**, Brussels, **London**	575	1.99	2	8
5	(5): **Frankfurt am Main**, Brussels, London, Nottingham, **Leeds**	1,087	3.78	1	8
6	(3): **Dusseldorf**, Brussels, **London**	565	1.96	1	4
7	(6): **Hamburg**, Hannover, Dusseldorf, Brussels, London, **Nottingham**	1,203	4.24	1	9
8	(4): **Nuremberg**, Frankfurt am Main, Brussels, **Paris**	937	3.22	1	7
9	(3): **Edinburgh**, Nottingham, **London**	655	2.25	1	5
10	(6): **Munich**, Nuremberg, Frankfurt am Main, Brussels, London, **Southampton**	1,279	4.47	1	9
11	(6): **Berlin**, Hannover, Dusseldorf, Brussels, London, **Southampton**	1,258	4.40	1	9
TOTAL		**9,639**	-	-	**81**

Frequencies of lines are most often set to one, with the exception of two cases: London-Paris (3) and Amsterdam-London (2), which are two of the most busy and profitable corridors. The line design ensures direct connections for 52 out of 60 (87%) of OD pairs and 95% of passengers. All OD pairs are served with at most one transfer. Brussels can be considered a main hub, being associated with nine out of eleven lines, while having a direct connection with all but one of the other cities.

Serving 95% of passengers directly, the inclusion of transfer penalties resulted in a well-balanced line design. This design considers the number of transfers passengers make but avoids the costly approach of serving each OD pair with a separate line. Together, the lines serve all arcs, most often with the minimum required frequency (see Table 4.12).

Below, Figure 4.9 shows the optimal line design. The numbers denote the joint frequency per arc. A legend of lines is provided by Table 4.14 above.

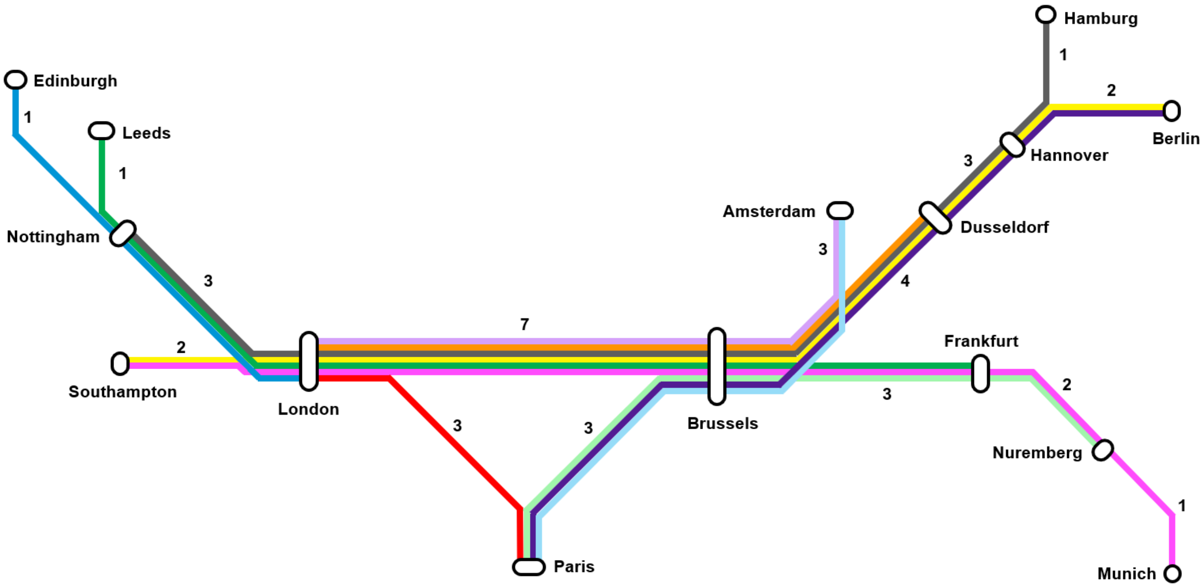


Figure 4.9 Optimal network; the numbers denote joint frequency per arc

Fares

The table below illustrates the competitiveness of the HSR fare setting among available travel mode alternatives. In the table below, the cheapest travel cost for each connection is indicated in bold.

Table 4.15 Fares of HSR and competing modes

Connection name	Length [km]	Travel cost or fare [€]			
		Plane	Train	Car	HSR
London-Paris	464	201	-	226	**174**
Brussels-Paris	317	221	**37**	62	60
London-Southampton	123	-	62	**30**	48
Leeds-Nottingham	120	-	**17**	27	43
Brussels-Düsseldorf	201	-	76	**41**	55
Amsterdam-Brussels	211	205	62	**47**	55
Brussels-London	364	206	127	212	**104**
Hamburg-Hanover	152	-	40	**29**	45
London-Nottingham	206	-	60	**47**	54
Munich-Nuremberg	172	-	93	**31**	50
Düsseldorf-Hanover	280	222	**40**	52	53
Berlin-Hanover	290	258	**37**	53	49
Frankfurt am Main-Nuremberg	223	168	47	**41**	50
Brussels-Frankfurt am Main	397	157	70	81	**68**
Edinburgh-Nottingham	449	184	308	**95**	107

Even though the car most often is the cheapest alternative, high-speed rail outprices air travel on every arc by a significant margin.

4.3.4 Model Validation

As described in section 3.4.11, our model will be validated through stability analysis and benchmarking.

Model Stability Analysis

Since our model must be ran a significant number of times, it was determined that the model should be able to find the optimal solution in approximately 30 minutes. Therefore, the variations are applied to an optimisation model version considering the 77 most populous cities in Europe.

In total, our model was run six times, which generated the optimal network designs in running times varying between 1200 and 2300 seconds. This result already indicates that our fare variations created problems with a large variation in complexity. However, the six optimal networks look much alike:

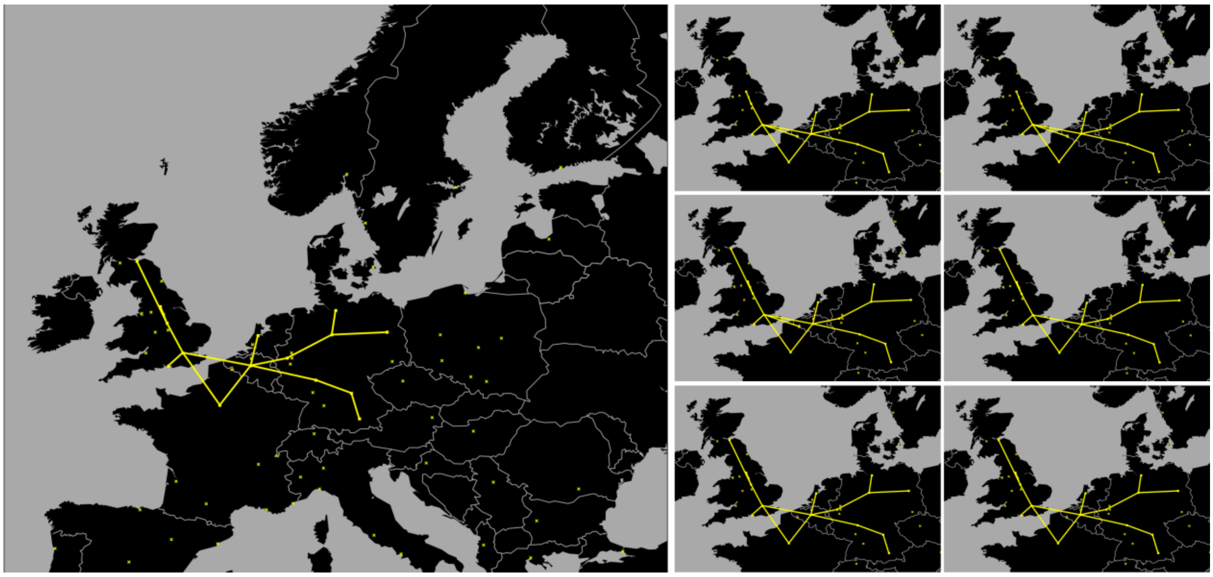


Figure 4.10 Large image on the left: optimal network without fare variations. Smaller images on the right: generated optimal networks for random variations in fare setting

As can be seen from the figure above, the optimal network stayed exactly the same in three out of six simulations. The other half showed minor changes compared to the original network: in three cases, a connection London-Lille was added, and in one case Nottingham-Edinburgh was removed.

Both these connections are considered non-stable, as their addition to the network for the original fare setting makes it slightly more (in the Edinburgh case) or slightly less profitable (in the Lille case), at a maximum deviation of 2.5% from the original value. The remainder of the network's fifteen arcs remained in all six simulations. This shows that our model produces stable results, while the inclusion of only a very few number of arcs varies. Therefore, the results show that the inclusion of arcs often is insensitive to relatively small changes in fare setting, which builds confidence in our model's predictions, which is important in order to be able to make trustworthy recommendations. The line design deviates from the original solution for every test run, indicating a much higher sensitivity to small changes in fare setting when compared to the network's topology. The table below accentuates this statement:

Table 4.16 Characteristics of model stability analysis solutions

Characteristic	Unit	Original	Lower	Upper
Revenue	€bn	655,477	598,244	657,631
Cost	€bn	433,234	377,602	440,970
Profit	€bn	222,243	216,661	224,988
Number of nodes / arcs	-	15 / 15	15 / 15	15 / 15
Network length	km	3,966	3,801	4,250
Fleet	-	87	70	87
Number of lines	-	11	10	12
Maximum frequency of line / arc	h^{-1}	3 / 7	2 / 7	3 / 8

Benchmarking

The methodology for this section was described in section 3.4.11. The route generating algorithm finds 446 valid routes, representing all 105 node pairs, while together covering all 15 nodes and 21 arcs. Altogether, it results in an optimisation model consisting of 1,443 decision variables (of which 551 are binary), and 782 constraints.

Our model finds an optimal solution consisting of 31 lines and a fleet of 134 vehicles, with the length of lines varying between two and eight nodes. The solving time of our model equals 20 seconds on average, which is considerably lower than algorithms proposed by Kechagiopoulos & Beligiannis (2014).

We can verify that each OD pair is served (which is a constraint) with the correct fleet size and frequency. The line configuration is able to serve all arcs with just enough capacity (the surplus varies between 0 and 55), which is needed in order to maximise profitability. Even though demand is treated elastic here, all routes are the shortest path between their terminal nodes. We can now see how our results measure up against other studies who have benchmarked their model to same network, with help of the table below:

Table 4.17 Mandl benchmark comparison with other studies. Table adapted from Asadi Bagloee & Ceder (2011)

Study	Fleet size	No. of lines	% of demand			Time elements (min)	
			No transfer	One transfer	Two transfers	Total travel time	Transfer penalty
Mandl (1980)	99	4	69.6	29.9	0.1	219,094	23,500
Baaj & Mahmassani (1991)	82	7	81.0	17.4	0.0	217,954	14,800
Shih et al. (1998)	87	6	82.6	17.4	0.0	225,102	13,550
Asadi Bagloee & Ceder (2011)	87	12	83.7	29.9	1.0	202,255	10,465
Kechagiopoulos & Beligiannis (2014)	N/A	8	97.5	2.5	0.0	158,357	1,946
Jha et al. (2019)	80	12	99.7	0.3	0.0	156,945	2,336
This work	134	31	100.0	0.0	0.0	77,895	0

Various types of algorithms have been developed and applied to the Mandl network over the past decades. Among the studies listed in the table, the first three are considered important milestones regarding Mandl network benchmarking and were considered state-of-the-art at the time of publication (Kechagiopoulos & Beligiannis, 2014). Both the works of Asadi Bagloee & Ceder (2011) and Kechagiopoulos & Beligiannis (2014) are considered state-of-the art by Jha et al. (2019). One can clearly see the improvement of results over time, with decreasing total travel times and transfer penalties, and an increased level of service regarding the number of transfers. It also becomes evident that increasing the fleet size and the number of lines results in an increased level of service, with diminishing returns.

Our solution clearly favours the user's perspective, serving all OD pairs without transfers, which results in the passengers' lower total travel time and zero minutes in transfer penalty. Compared to other studies mentioned in the table, and nine other studies compared by Kechagiopoulos & Beligiannis (2014), our approach is the only resulting in all OD pairs being served directly, and significantly faster to generate results.

This however comes at the expense of operator costs, with a significantly larger fleet size and number of lines. The balance between operator and user costs is determined by the choice of parameters such as transfer penalty per passenger and unit maintenance & operating costs. Given the experimental character, addressing these parameters and their interaction regarding this network is considered outside of this project's scope and saved for further research. However, we can conclude that our solution is much more realistic than the other heuristic approaches here, given the fact that it takes elastic demand into account.

4.3.5 Post-Optimisation Experiments

Two experiments were conducted in order to map potentially interesting extensions to the optimal network as presented in Figure 4.8: a mapping of new connections under HSR-beneficial scenarios.

HSR-beneficial scenarios

Since the profitability of HSR connections and networks varies significantly based on the chosen principles, it is important to identify 'secondary connections': these are additions to the basic network (Figure 4.8), which are added to the optimal network in HSR-favourable scenarios only. The following three scenarios are universally applied to all considered connections:

1. **Lower costs**. Since construction costs are the largest cash flow and the hardest to pin-point by far, it is likely that they could be lower in reality.
2. **Higher demand (growth)**. Currently, economic growth and induced demand are set to (country-specific) percentages. As mentioned before, these percentages are hard to pin-point and in some cases can exceed expectations.
3. **Aviation fuel tax**. In the EU, commercial aircraft fuel is tax exempt (EU, 2024). If a tax would be imposed, air travel costs rise, which impacts HSR demand positively, following our demand forecasting model. It is assumed that air travel costs increase proportionally with the tax.

In all cases, the variables are adjusted by 0% to 50% (with increments of 5%) in the direction favourable to HSR. As mentioned previously, optimising the model for 111 cities takes approximately 6 hours. Given that the experiments described above require multiple re-runs of the model, we decided to reduce the number of cities. The number was set to 77, the smallest at which the optimal solution differs by only one connection compared to Figure 4.8. Optimising this reduced network takes roughly 1.5 hours.

The results show identical 'secondary connections' in all three scenarios. The figure below shows these connections in white.

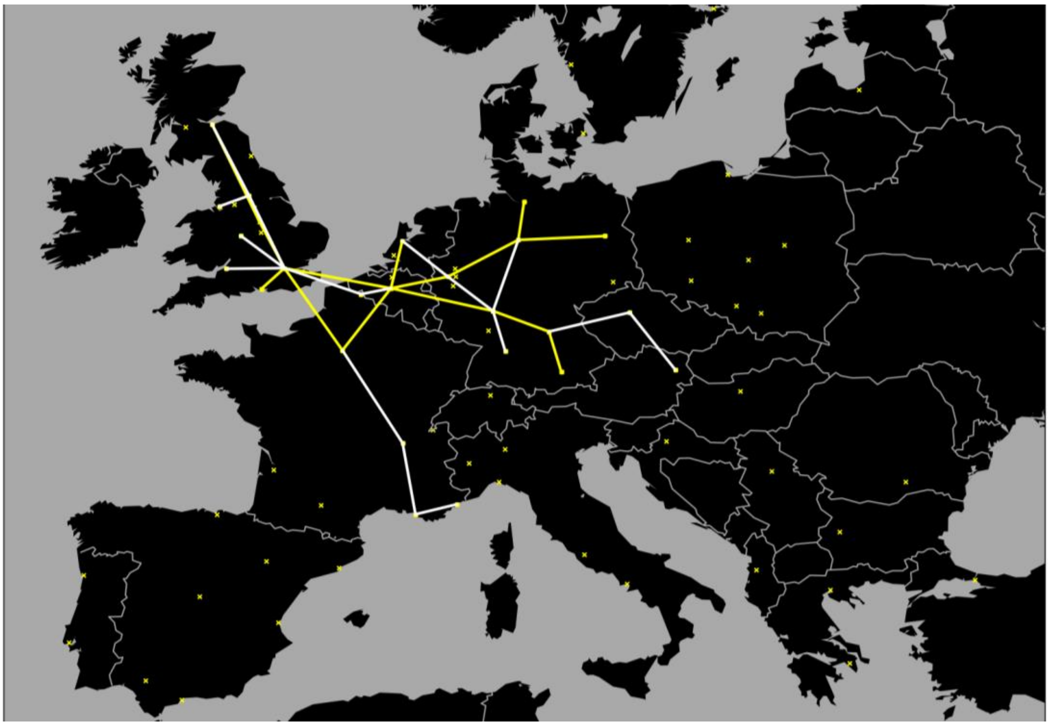


Figure 4.11 Secondary connections

The most remarkable to note about the found ‘secondary connections’, is that they most often densify the existing network, instead of expanding it, indicating that increasing the profitability can be mainly found in reinforcement of the existing structure. Only for extremely HSR-beneficial scenarios (≥35%), our network would expand into new territories. This is further accentuated by Table 4.18 below. One may note that the Lille-London connection, which was the only new connection found in model stability analysis (section 4.3.4), is the first new connection to arise here.

Table 4.18 List of added secondary connections per HSR-beneficial scenario

HSR-beneficial scenario	Added secondary connections
10%	Lille-London
25%	Amsterdam-Dusseldorf • Dusseldorf-Frankfurt am Main
35%	Brussels-Lille • Prague-Vienna • Nuremberg-Prague
40%	Frankfurt am Main-Hanover • Edinburgh-Leeds • Bristol-London • Frankfurt am Main-Stuttgart • Leeds-London • Birmingham-London
50%	Leeds-Liverpool • Lyon-Paris • Lyon-Marseille • Marseille-Nice

5

Policy Implications

AI-generated high-speed train (Balbatunou, n.d.)

5 Policy Implications

This chapter aims to bring the results presented by chapter 1 into perspective: what policy implications can we derive from it? Therefore, it closes the loop and provides answers to our last research sub-question as formulated in section 1.3.2, in order to advice and inform policy makers in a short note about the policy implications of our final and optimal network design.

5.1 Introduction

The subject of policy implications can be approached from several perspectives. The few most prominent are addressed in their respective sections below: international cooperation (section 5.2), funding and financing (section 5.3) and environmental sustainability (section 5.4).

5.2 International Cooperation

Despite the fact that our network design has multiple stations in the same countries – in particular the UK and Germany – 84.8% of all passengers travel internationally. Therefore, cross-border operations are vital for the vitality of our network. For this reason, it's of crucial importance that national and regional governmental bodies are coordinated. The heap of national and regional laws and regulations, as discussed in section 1.2.2, complexify the development of cross-border connections significantly. These should be lifted, and new regulation should be made in order to make it easier for countries to cooperate. A specific task group should be founded, which should take on this task. International cooperation and communication must be made possible and functions as the crucial starting point in truly being able to develop a viable HSR network. The importance of this is no new information, but our results endorse it even more. The fact that all involved countries in our network are part of the EU, should only be beneficial in order to establish sufficient international cooperation and communication.

5.3 Funding and Financing

As stated before, high-speed rail infrastructure requires substantial investments. Our optimal network is able to underline this statement with numbers: for a total revenue of €644.4 bn and lifetime profit of €222.8 bn, approximately €421.6 bn must be invested in the coming 40 years. A substantial part of €193.8 bn is invested initially for construction and acquisition, which is 46% of all expenses during the project's lifetime. The rest will be spent for maintenance and operation for both the infrastructure and rolling stock, and amounts to a yearly expense of €5.69 bn. Most of the expenses are made for HSR development and exploitation in the UK and Germany, given the design of the network and the fact that HSR construction is relatively expensive here. We can split the remainder of section into two parts: general and specific subsidy recommendations, where the former looks at a general strategy to realise the optimal network design, and the latter looks at promising individual projects that would be viable with a relatively minor subsidy.

General subsidy recommendations

The required investments are significantly larger than subsidiary budgets allocated for rail development in the past, such as the €25 bn as part of the European Green Deal (European Council, 2024) and much more than the €23.7 bn already invested into high-speed rail infrastructure since 2000 (European Court of Auditors, 2018). They are a fraction of the total amount needed, which is also supported by last-mentioned source. It is indicative to the fact that subsidiary help is needed from not only the European Union, but also from national and/or regional governmental bodies, a message Deutsche Bahn (2023) subscribes to. The willingness of these organisations are crucial to establishment of a vital HSR network.

Our network is designed with the basic premise of being able to make up for its costs, but following the mentioned cash-flows, this will take 19 years. As mentioned in section 1.2.1, the European Commission and national governmental bodies are hesitant with guaranteeing high initial investments out of fear for disappointing returns. In these first 19 years, they must be willing to spent this amount of money initially. As the vitality of the network depends on it, the parties must come together, and discuss how to set aside their fears and trust science. The massive initial investments underline the importance of subsidiary help in the initial stages. Given the fact that our network is able to be operationally profitable without subsidiary help, it can be confirmed that it is only needed during the initial stage. Of the €193.8 bn in total initial investments, 42% should be invested in cross-border connections, which encompass half of the total network length.

Specific subsidy recommendations

Currently, no long-term high-speed rail investment plan exists (European Court of Auditors, 2018). This section will therefore formulate one. Our analysis of all 31,125 OD pairs among the 250 most populous cities delivers only three connections that would be profitable on their own: London-Paris, Amsterdam-London and Lille-London. However, many connections would become profitable with help of investments.

Altogether, it can be concluded that subsidiary help from both the European Commission and national and/or regional governmental bodies is crucial in order to establish a viable European HSR network. In total, €193.8 bn is needed in the initial stages, which must be spent on infrastructure construction and rolling stock acquisition and will pay itself back in 19 years. For this reason, the subsidy might as well be a loan. Subsidising cross-border connections in the initial stages is the cornerstone of creating a viable European HSR network. Based on the subsidies granted by the European Commission since 2000 (European Court of Auditors, 2018), the table below displays all thirteen projects -not part of the network design- that would become profitable with a subsidy of no more than €3 bn.

Table 5.1 Most promising HSR projects with minor subsidy

Connection		Length	Subsidy required (€bn)
City A	City B		
Kaunas, Lithuania	Vilnius, Lithuania	100	1.27
Katowice, Poland	Krakow, Poland	100	2.04
Radom, Poland	Warsaw, Poland	317	2.20
Lublin, Poland	Radom, Poland	105	2.26
Riga, Latvia	Tallinn, Estonia	114	2.37
Częstochowa, Poland	Kielce, Poland	306	2.51
Bergen, Norway	Oslo, Norway	128	2.61
Kielce, Poland	Krakow, Poland	468	2.63
Częstochowa, Poland	Lodz, Poland	125	2.67
Oslo, Norway	Stavanger, Norway	129	2.83
Lodz, Poland	Warsaw, Poland	440	2.88
Bydgoszcz, Poland	Poznan, Poland	129	2.92
Blackwater, United Kingdom	Chatham, United Kingdom	148	2.92

Despite our optimal network being focused solely on western Europe, the connections that come the closest to being profitably are located in the eastern part of the continent. In particular, the country of Poland is heavily represented, along with connections in the Baltic states. The data thus shows that, with help of subsidies, the Baltic states could be connected in a profitable network, significantly increasing the span of our network.

5.4 Environmental Sustainability

As mentioned in section 1.1.2 and briefly in the previous section, the motivation of European governmental bodies to invest into high-speed rail comes from its greener characteristics. Moving people from planes or cars to trains is needed in order to reduce carbon footprint and reach climate goals, which EU member states agreed to. The most recent and prominent agreement concerns the European Green Deal, which aims to reduce transport-related emissions by 90% and triple HSR's traffic volume between 1990 and 2050 (European Council, 2024). As the total transport-related greenhouse gas emissions in fact has increased since 1990 (EEA, 2023), it can be concluded that the EU is not on track to meet its climate goals.

In order to estimate the contribution of our network in the light of the Green Deal goals, we use the following facts and assumptions:

Table 5.2 Key variables for sustainability calculations

Key variable	Unit	Air travel	Car travel	HSR travel	Source
CO_2 emissions	[g CO_2 / paxkm]	200	170	4	(UK Government, 2022)
Average travel distance	[km]	1,000	500	1,000	Estimate from results
Passengers shifted to HSR	[10^6 pax]	42	21.3	63.3	Optimisation results

Our network would serve 94.3 million passengers on average per year during its lifetime, which would increase the total HSR traffic volume in the EU by 72%, from 131 (Statistica, 2018) to 225 billion km travelled on a yearly basis. Thus, even with a partial OD pair coverage of (57%, see section 4.3.3), our network would significantly enhance service, improving connectivity on key high-demand routes. Among these passengers, 42 million used to travel by plane and 21.3 million by car. The mentioned shift alone would significantly contribute to that, reducing emissions of the entire European transport sector by 33% by 2050. This percentage is based on an average of 200g CO2/paxkm for air travel, 170g for car travel and 4g for HSR travel (UK Government, 2022), an average travel distance of 1,000 km for both HSR and air and 500 km for car travel, and 714 Mt CO2 total transport-related emissions in 1990 (EEA, 2023). It is assumed that HSR operations would start in 2030. These results illustrate the fact that a HSR network would not only be profitable on its own, but also provide significant benefits in non-monetary terms. As our model is purposely designed to only look at profitability in monetary terms, it makes one wonder how much more HSR could be developed if these additional benefits were monetarised.

Altogether, it can be concluded that development of an European HSR network would significantly play a role in reducing emissions and reaching the goals aimed for by the Green Deal agreement. Given the large impact, its development should start rather sooner than later.

6 Conclusions & Discussion

AI-generated high-speed train (Balbatunou, n.d.)

6 Conclusions & Discussion

This chapter finalises the project by reviewing the results and providing answers to the previously stated research questions. Section 6.1 provides answers to the research questions as formulated in section 1.3 initially, while section 6.2 discusses the previous chapters.

6.1 Answers to Research Questions

This first section will answer the research questions related to each of the four sub-problems introduced in section 1.3, in chronological order: demand forecasting in section 6.1.1, profitability estimation in section 6.1.2 and network optimisation in section 6.1.3.

6.1.1 Demand Forecasting

All three research questions regarding this sub-problem can now be answered:

Question 1: What can be learned from completed high-speed rail projects, regarding their demand?

The problem definition in section 1.2.1 illustrates the uncertainty of demand, revenue, construction time and the values of various cash-flows, world-wide, across all times. Many HSR projects world-wide and more than half of European projects fail to meet expected demands, therefore becoming unprofitable and relying on subsidies. Thus, accurate demand forecasting is crucial for the success of high-speed rail. Later, section 3.2.1 dived more deeply into this problem and found out forecasting knowledge has grown, but not the accuracy of forecasting models: financial performances often disappoint as they generally overestimate demand figures. This can be attributed to political causes, which have a substantial influence on rail projects: decision-makers generally ignore or downplay financial risks under the guise of social welfare or other variables that are impossible to measure accurately (Flyvbjerg et al., 2005). In conclusion, passenger forecasts for rail projects can be accurate if done in a scientific and independent manner.

Question 2: What models and impact factors can be used to forecast high-speed rail demand?

In literature, high-speed rail demand forecasts focus on either trip distribution (market shares) or trip generation (total available demand across all modes).

Trip distribution

Particularly logistic regression (logit) models are used to forecast high-speed rail market shares. Multiple types of logit models are applied in practice, of which overviews are presented Table 2.2 in and Table 2.5. Binomial (BNL) an multinomial logit (MNL) models are the simplest and most popular logit types used in literature. However, their simplicity results in a number of drawbacks, particularly regarding the IIA property as explained in section 2.2.1. Both nested logit (NL) and mixed logit (ML) are able to mitigate this, leading to more accurate forecasting results. Their increased complexity however leads to new drawbacks regarding computation difficulty. This project uses the MNL model.

A wide variety of demand-impacting factors are used in models. A full overview is shown in Table 2.1. Their usage popularity depends of the model implemented, as some factor-model combinations work better than others. Section 2.2 provides an overview of this, and shows that a logit model almost always includes travel time and travel costs as independent variables. Service frequency is the only other factor implemented by at least half of demand forecasting studies. Population-related data is regularly used, but only from the economic perspective (focussing on value instead of quantity). Population numbers and destination attractiveness are almost never used in logit models.

Trip generation

To predict the total travel demand, multiple models are in existence, most prominently linear regression and gravity models. Of these two, linear regression is by far the most used in literature. Table 2.3 presents an overview of different linear regression models used in literature. Numerous adaptations of the standard simple regression model are in use, with different levels of complexity and computation difficulty. However, as shown in section 2.1, realistic effects often cannot be linearised, which remains a crucial critique against these models. On top of that, literature review in section 2.2.3 pointed out that gravity models are the preferred method when no direct connections are available currently.

Figure 2.7 shows that gravity models are suitable for different variables: most prominently travel time and population. Travel cost and economic factors also are used regularly. Gravity models are also able to include the attractiveness of destinations as independent variables.

From the answers to the previous question, it can be seen that gravity models and logit models complement each other in terms of demand-impacting factors considered, allowing for an all-encompassing forecasting method when these two models are combined. Most of the impact factors are based on one, practically undebatable number, with the exception of catchment area population, which depends on the choice of trip duration. Calibration showed that 45-minute catchment areas yield the most accurate forecasting result, while creating a weighted average of different 'catchment rings' do not. It was also found that usage of 'GDP/capita' as impact factor results in much better calibrated parameters. Simply using 'GDP' does not, despite yielding a better model fit. The demand for high-speed rail depends on more than just travel time and travel costs, and the size of the catchment area also depends on the city's attractiveness. The conclusion of question 2 can be recalled here as well - passenger forecasts for rail projects can be accurate if done in a scientific and independent manner. The final forecasting model is presented by equation (6.1), which was calibrated having a model fit R^2 of 0.75, and all but one parameter being over 99% statistically significant. The model is proof from any bias (with respect to airport size) and is, next to diverted demand, further extended to take induced demand and economic demand growth into account.

6.1.2 Profitability Estimation

All five research questions regarding this sub-problem can now be answered:

Question 3: What can be learned from completed high-speed rail projects, regarding their profitability?

Many high-speed rail projects have turned out to not be profitable. Analysis of each individual HSR line in China's network shows a clear correlation between operating speed and profitability, indicating the importance of the former term. For this reason, this project investigates whether the maximum operating speed allows for profitable connections, to find the true potential of HSR. In Europe, a large share of connections is unprofitable as well. There are multiple causes for this. Pre-project cost estimates often are unrealistic and budgets are easily exceeded with the slightest amount of setbacks. The same holds for demand projections, which in 90% of cases are too optimistic, while having an average overestimation of 106%. These two reasons combined result in insufficient research and inaccurate projections on both sides of the balance sheet, which is exaggerated since the difference in both cases is present on the pessimistic side. Truly independent research and forecasting is needed to produce accurate predictions. Also, the natural variation of costs and revenue should be taken into account since even the slightest setback can increase costs or decrease revenue substantially.

Question 4: Into what cash-flows does high-speed rail operational profitability break down?

High-speed rail projects can be broken down into two parts: the infrastructure and the rolling stock. Each of these two have their own related cash-flows. Infrastructural costs can be broken down into construction, operating and maintenance costs. The latter two terms often are combined in literature under the heading of operating costs. It's important to consider these two separately, to avoid double counting. Construction terms can be further broken down into planning & land costs, infrastructure building costs and superstructure costs. But in literature, they are often all combined and are considered as simply 'construction costs'. The rolling stock must be acquired, in literature often referred to as 'acquisition costs'. Furthermore, just like the infrastructure, it needs to be operated and maintained, each adding extra costs.

All cash-flows mentioned can be divided into two types: initial investments (construction / acquisition costs) and recurring payments (operating / maintenance costs).

Question 5: What factors influence these cash-flows and how can this relationship be captured?
The acquisition costs of trainsets only depend on the choice of train and the amount needed, and therefore arguably is the most easy to predict. The operation and maintenance costs of rolling stock is much less set in stone and depend on the characteristics of the chosen rolling stock and the degree of use. However, it shows relatively low variability when expressed in terms of [€ / km] – indicating a relationship with only the line length. In literature, terms of [€ / seat-km] are used more often, with their popularity linked to the inclusion of the degree of use.

Infrastructure-related costs show much more variation in practice. Their costs vary much in relative sense between different countries, but less when comparing projects within the same country. Literature describes a certain 'base' cost per km (indicating that line length is the most important factor of influence), but highly sensitive to mainly environmental factors: rough terrain yields more tunnelling and related costs, while also more densely populated areas generally induce more construction costs. Also, the economical price level might further influence costs, which is a country- or region-dependent factor. It's impossible to estimate these values directly and accurately by formulas. Therefore, we must rely on literature to estimate them through thorough investigation of the region's characteristics. Operation and maintenance costs for infrastructure vary much less, relatively. In this sense, they are comparable to the same type of costs for rolling stock, and similarly are often expressed in terms of [€ / seat-km], pointing to the degree of use as its biggest impact factor.

Question 6: What model can be developed to forecast high-speed rail operational profitability?
Since this term consists out of a summation and subtraction of a selection of the HSR cash-flows, a forecasting model should follow the same approach. A linear addition model thus matches this philosophy. This model is developed in section 3.3, following the cash-flows and their influence factors that arose in literature in section 2.5: infrastructure construction, rolling stock acquisition, and the operation and maintenance of infrastructure and rolling stock. It is section 3.4.7 where all elements of are put together and made ready to serve as input for the optimisation model as the objective function.

Question 7: How do the cash-flows of 'profitability' translate into 'operational profitability' and 'justifiability'?
Section 3.3.3 specifically focuses on this. Operational profitability consists out of all cash-flows occurring during the operational stage: maintenance and operation costs of infrastructure and rolling stock, as well as ticket revenue. The extent in which this operational profitability is able to pay back the initial investments (infrastructure construction and rolling stock acquisition) determines the justifiability.

6.1.3 Network Optimisation
All four research questions regarding this sub-problem can now be answered:

Question 8: What can be learned from completed high-speed rail projects regarding the requirements cities and connections need to fulfil?
Given the earlier addressed uncertainty in demand and profitability forecasting, it's hard to find conclusive answers to this. One conclusion that may be drawn can be derived from this statement: it is important to

keep this uncertainty in mind. Connected cities must be large, and their respective stations must be reachable to a large number of passengers within a relatively low amount of time. Stronger conclusions can be made regarding operating speeds of HSR services. Section 2.5.2 indicated higher design speeds result in more profitable HSR connection, with diminishing effects. This has been shown in theory by Barrón et al. (2012) and Belal et al. (2020) as well as in real life data (Zhang, 2024).

Question 9: How can linear programs be formulated in order to solve a realistic TNDFSP formulation to optimality within reasonable computation times?

Current methods lack the capability to solve to optimality for medium-to-large sized networks, while also taking demand elasticity into account. By analysis of literature, it was found that the Multi-Commodity Flow Problem (MCFP) formulation is needed to do so, and was adapted to incorporate a set of pre-defined valid OD flow routes. Compared to the original MCFP formulation, this yielded a TNDP model with the same intrinsic logic and outcomes, while requiring much less computation power, increasing the network size that could be considered. The model was further enhanced to also incorporate line design elements, so that it functions as a fully integrated TNDFSP model. The objective function considers all cash-flows found previously, accompanied by transfer penalties in order to capture the user's perspective. The constraints ensure a valid network and line design. Within 6 hours, the model can be solved to optimality for up to 111 nodes (all cities with a population exceeding 315,000) and all their potential OD pairs.

Question 10: What would an optimal high-speed rail network look like, with respect to its topology, operating lines and associated frequencies?

Connections should be made to serve and bundle multiple OD-markets, as purely connecting two cities will almost certainly lead to a unprofitable solution. Thus, in order to make HSR profitable in Europe, a network should be designed rather than just one single connection. New development potential can be found in North-Western Europe (mainly the UK and Germany), centred around large cities such as London, Paris, Amsterdam, and Berlin. Brussels functions as the hub of the network, with lines radiating outwards towards the mentioned large cities. Medium-sized cities such as Dusseldorf, Hannover, Stuttgart, Frankfurt and Leeds are included, primarily due to their beneficial location directly between two much larger cities, while adding enough passengers to the network to make up for the time loss of stopping at the city itself. Population-dense regions are considered by using a 45-minute drive population catchment for predicting demand, rather than relying solely on individual city populations. This approach likely explains the selection of Düsseldorf over the larger nearby city of Cologne. Despite Cologne's size, Düsseldorf benefits from a central location within the densely and heavily populated Ruhr Area, giving it access to a significantly higher population catchment, making it the more attractive choice. The network would consist of 15 cities, 15 arcs and would serve 60 OD-pairs. Only two arcs are upgraded lines already existing as of today. Eleven lines would serve the network, most with operating frequencies of one train per hour, increasing to a maximum of three for the busiest connections. The addition of transfer penalties makes up for a seemingly well-thought line design, which serves most connections with the minimum frequency needed.

Question 11: What are the policy implications from this optimal design?

The policy implications are listed in chapter 5 specifically, and come down to advice in three areas:

- International cooperation is vital, as our design involves multiple countries and requires significant initial investments.
- Funding and financial is critically needed, particularly at cross-border connections in the initial stages of infrastructure construction and rolling stock acquisition.
- Our network design would play a major role in meeting set climate goals by the Green Deal agreement.

6.2 Discussion

This section aims to summarise all points of discussion that arose during the drafting of the methodology and the results of this project. Alike the chapters mentioned in the previous sentence, the discussion will be split into three parts: demand forecasting, profitability estimation and network design.

6.2.1 Demand Forecasting

This section lists all points of potential discussion regarding the demand forecasting parts (section 3.2 and 4.1).

Model choice

In section 3.2.2, a forecasting model was developed, which only took diverted demand into account, for the first year. The model estimates modal split per mode and trip generation by means of a logit and gravity model, respectively. As mentioned in the section, this combination of models has been used before in multiple studies. However, the model types chosen have their drawbacks and limitations.

- **Modal split:** Demand forecasting has already been widely researched in literature. Therefore, it is not the main focus as to what this project hopes to contribute to science. This study has opted for MNL, also due to its simplicity and popularity among transport studies. However, our literature review in section 2.2, and Table 2.2 in particular shows that ideally a mixed logit (ML) should be used, given the fact that it should be able to mitigate the IIA property and account for non-constant taste parameter values among the population. If the variation of the taste parameters is deemed significantly small, the more simpler nested logit (NL) can be used safely. These effects are not taken into account by the MNL model, which results in less accurate demand forecasting results. Both 'train' and 'HSR' are seen as two separate travel modes, despite being very alike. Since our model does not capture the IIA property, it makes the outcomes of our model less reliable. Since forecasting demand is not considered the main focus as to what this project hopes to contribute to science, the implementation of ML and/or NL models is saved for further research.
- **Trip generation:** as described by section 2.2.3, there is little to argue against the choice for using a gravity model, especially when it comes to forecasting demand for new connections, a primary interest for this project. Potentially better versions of this model are available, for example the dynamic gravity model as developed by Yu et al. (2021) could be used as well, given the fact that it is able to incorporate changing parameter values over time. Further studies could experiment with this.

Considered travel modes

The bus is not taken into account as a travel mode, given its marginal market share along the complete spectrum of travel distances (Figure 1.1). However, for a truly accurate demand forecasting model, the decision to exclude can be argued against. Further studies could look into accounting for the bus mode and whether it affects HSR network development.

Impact factors

Together, the logit and gravity models incorporate most of the suitable demand impact factors as listed in section 2.1. However, for simplicity reasons, some were left out. These factors are unpopular among demand forecasters, but their effects can be nonnegligible. During model validation, it showed that our model is unable to accurately forecast demand for connections with a touristic character, which can be attributed to the fact that related impact factors such as 'destination attraction' are left out from our forecasting model. Our research shows that these impact factors cannot be neglected, and future research should incorporate this factor to develop more accurate forecasting models. The same holds for other factors left out of our model.

For simplicity, the values of population, GDP, distance as well as the travel time and travel costs of all travel modes are assumed to remain constant along the lifetime of a high-speed rail project. In fact, the values regarding all impact factors are assumed constant. Of course, this is highly unlikely. Furthermore, it can be expected that once a high-speed rail network is developed, competing modes change their fares (and perhaps shorten their travel times) in order to regain an optimal position in the travel market. This will affect the demand and viability of HSR negatively. Due to the extreme complexity and unpredictability of this secondary effect, it is left out of consideration for this research. Already during the web scraping stage, which was re-run many times during our project due to code enhancements, it was discovered that among the thousands of travel options and their characteristics, a handful change every day, which could affect the optimal solution. For this reason, the final web scraped data set was collected *in one go*. Also, within the set lifetime of forty years, many other events could occur that could crucially affect HSR's viability: the rise and fall of new transport modes, changes in travel habits and preferences, a new pandemic, etc. Events like these often cannot be predicted, as well as their influence on HSR, and we have to assume that they do not happen in order to design our network.

The values of beta parameters are currently taken as a median from 57 studies. However, it should be noticed that the variation in these studies is large, given that each of these studies has a different approach when it comes to the determination of these values. We can therefore not be sure whether the beta values are correctly chosen. Also, as mentioned before, it could be the case that these values change from country to country or even from person to person. Some studies have estimated parameter values for each travel purpose (e.g. commuting or leisure), as they weigh travel times and costs differently. For simplicity, we assume that all passengers in Europe weigh off travel times and costs in the same way, along the entire 40-year lifespan of HSR projects. Also, these values are prone to change for each travel mode. Furthermore, base preferences for certain travel modes are not taken into account. Further studies can look into the effect of varying values of taste parameters, and their effect on the optimal design.

Despite that fact that the size of catchment areas varies from city to city (Boelrijk, 2019), our model assumes a constant size for all cities. This choice was made as this variation requires new research, as studies on this subject are scarce. The issue is attempted to be addressed through our 'bias elimination' process, but more research is needed to assess the accuracy and shortcomings of our model for various inputs. This subject can be recommended for further study.

Our model is calibrated to air passenger data. However, as discovered during the data collection stage, many data points are missing from the Eurostat data sets. This applies to data regarding return flights (if the outward flight is represented), but also to some connections in its entirety. For some countries and airports, a large share of data is not reported. The influence of the missing data on the quality of our final calibrated model is unknown.

Calibration

To mitigate effects of competition among modes, and our intention to create an accurate forecasting model, our only choice was to calibrate the gravity model only for distances larger than 1,000 km. However, the model is used to forecast demand over distances shorter than 1,000 km. This approach results in concerns regarding the accuracy of our model for shorter travel distances.

The exclusion of some impact factors resulted in a forecasting bias. The bias is eliminated by scaling the model outcomes by a factor depending on the outcome's value itself, which results in a much more accurate forecasting model, while questions about the scientific integrity of this approach should be asked. Further research should look into adding more impact factors, and whether they can eliminate the bias, making a 'bias elimination' step unnecessary.

Demand Evolution

The forecasting model was further refined in section 3.2.6, to also incorporate the effects of induced demand and economic demand growth. Both effects have been added. As future demand is very complex to accurately estimate, and prone to many unpredictable factors that could heavily influence its value, this project has opted for a relatively simple approach to include them. For induced demand, a simple factor is added, uniformly for all potential connections in Europe, even though it was mentioned that this value can significantly vary from location to location. Economic demand growth is implemented with more detail, and in this extent, mainly country-based GDP growth is estimated with a significant amount background research and data analysis. The flaws with our approach arise during translating GDP growth into demand growth, which was performed by a simple calculation of elasticity, which was assumed to be constant across the entirety of Europe, and for the decades to come. Also, section 2.1.3 noted that elasticities are only confirmed to be accurate for the small frame of values (in this case, GDP growth) they were calculated for. Our study uses the same elasticity for a wide frame of GDP growth values, and therefore we cannot be certain whether the projected demand growth is accurate for all connections. As our study does not take demand forecasting into account as its prime focus, we leave further research regarding this part for future studies.

Altogether, it can be concluded that the demand forecasting model complies with the level of detail required for this project. However, many enhancements could still be made in order to make it significantly more accurate. This requires a study having it as primary focus.

6.2.2 Profitability Estimation

This section lists all points of potential discussion regarding the profitability estimation parts (section 3.3 and 4.2).

For our project, a reasonably simple profitability estimation model was developed. The literature review section successfully mapped all cash-flows that significantly impact high-speed rail profitability. Mapping the impact factors of each cash-flow was done, but did not go much in depth deliberately, as profitability estimation is not seen as this project's primary focus. Our analysis sticks with the few most obvious impact factors, such as infrastructure length.

Our HSR fare setting is based on travel times and fares for competing modes of transportation, and optimally chosen to maximise the expected revenue. It was shown that this is a justifiable approach and matches common real-life practice. However, our model fails to take into account changing competition caused by our fare setting. Given the fact that we chose a HSR fare optimising our revenue, it is very likely that other travel modes change their fares in response, which benefits them optimally but will definitely not benefit HSR. Given the complexity and unpredictability of these interactions, our project does not take them into account. More research regarding this matter is needed.

Instead of setting HSR fares in order to maximise revenue, a much greener policy could be applied: optimise the total HSR demand served by the network. Since our project aims to optimise for profitability to demonstrate the limits of its economic viability, other policies were disregarded. However, other optimisation approaches could be taken, depending on the wishes of both operators and governmental bodies. More research is needed in order to further investigate this.

In order to reduce complexity, our model does not take unpredictable variabilities into account. For example, the construction costs are determined by the origin's and destination's elevation and country. This gives our model a reasonable level of realism, but does not take factors into account that are unpredictable, such as the ones mentioned by section 1.2.1, causing construction delays and heavily influencing the project's economic viability. The values produced by our model do not specifically take these factors into account, but it was based on unit costs of completed projects.

Also, even though our model projects codes decades into the future, it does not take possible inflation (or deflation) into account. This choice was made deliberately, as it is extremely hard to predict 40 years into the future.

Altogether, our profitability model is able to produce reasonable estimates of HSR profitability, complying with the level of detail required for this project, but scrapes the surface of the science behind it. A study should investigate this to produce significantly more accurate models.

6.2.3 Network Design

This section lists all points of potential discussion regarding the network design parts (section 3.4 and 4.3).

Being the part considered the core of what this project hopes to offer scientifically, our network design model was developed to have as few flaws as possible. However, given the complexity of transportation problems, a number of assumptions had to be made in order to find the right balance between the applicable scale of this model (larger being better) and the scientific value of its outcome. To increase the scale of our model, assumptions and simplifications were made in section 3.4.9, which were carefully chosen and funded on literature in order to stay confident that the model's output is optimal. We call these 'smart restrictions'. Even though our 'smart restrictions' were chosen very cautiously and funded in literature, their existence automatically results in that it cannot be proven that our solution would be the optimal in real life. However, our solution is optimal with respect to the data fed to our optimisation model. More computation power is needed in order to run our model without 'smart restrictions', or to run it for larger potential networks. Supercomputers might be able to provide the necessary computation power, but test runs on the DelftBlue Supercomputer showed that our code should then first be rewritten in order to be solved in parallel. These experiments are left up for future study.

During model stability analysis, it was proven that unlike the network's topology, its line design is sensitive to small changes in fare setting. Luckily, the line design is easier to change in real life, since being a tactical decision. However, more research is necessary to look deeper into making robust line designs, that are able to adapt to multiple demand or fare setting scenarios.

During benchmarking, it was shown that our model outperforms comparable studies when viewed from the user's perspective. However, it did come at the cost of operators, requiring a significantly larger fleet and much more lines. It should be noted that our model's nature is much different from the models we compare ourselves to, which makes it harder to make conclusions about the quality of each model.

The balance between operator and user costs is determined by the choice of parameters such as transfer penalty per passenger and unit maintenance & operating costs. Given the experimental character, addressing these parameters and their interaction regarding this network is considered outside of this project's scope and saved for further research. However, we can conclude that our solution is much more realistic than the other heuristic approaches here, given the fact that it takes elastic demand into account.

The sustainability calculations illustrate the HSR network would not only be profitable on its own, but also provide significant benefits in non-monetary terms. As our model is purposely designed to only look at profitability in monetary terms, it makes one wonder how much more HSR could be developed if these additional benefits were monetarised. This is left up to further research.

Fluctuations in travel demand, costs, and revenue over a 40-year horizon are inherently uncertain. However, a strictly robust or stochastic approach may not be needed, as our current model is already complex and computationally demanding. The additional value from using robust or stochastic optimisation is unlikely to justify the effort in large-scale HSR network design. Furthermore, once the infrastructure is built, the line design can be adapted relatively flexibly to accommodate unforeseen changes in demand, by changing timetables, operating lines or frequencies.

Altogether, our optimisation model is able to produce very realistic and optimal high-speed rail networks for larger problem sizes than currently available in literature, especially if elastic demand should be taken into account. However, research on the subject of line design robustness could potentially further enhance the quality of our model's solution when it comes to sensitivity to small changes is fare setting.

Bibliography

Agramanisti Azdy, R., & Darnis, F. (2020). Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites. *Journal of Physics: Conference Series*, *1500*(1), 012104. https://doi.org/10.1088/1742-6596/1500/1/012104

Agrawal, J., & Mathew, T. V. (2004). Transit Route Network Design Using Parallel Genetic Algorithm. *Journal of Computing in Civil Engineering*, *18*(3), 248–256. https://doi.org/10.1061/(asce)0887-3801(2004)18:3(248)

Ahmed, L., Mumford, C., & Kheiri, A. (2019). Solving urban transit route design problem using selection hyper-heuristics. *European Journal of Operational Research*, *274*(2), 545–559. https://doi.org/10.1016/j.ejor.2018.10.022

Alam, I. (n.d.-a). *AI generated High-speed train on the road in the city. 3D rendering, High-speed train on the road to the modern city, presented in 3D rendering, AI Generated Free Photo*. Vecteezy. Retrieved September 1, 2024, from https://www.vecteezy.com/photo/35973250-ai-generated-high-speed-train-on-the-road-in-the-city-3d-rendering-high-speed-train-on-the-road-to-the-modern-city-presented-in-3d-rendering-ai-generated

Alam, I. (n.d.-b). *High-speed train moves through the city. 3D rendering. High Speed train on the road to the modern city, AI Generated Free Photo*. Retrieved September 1, 2024, from https://www.vecteezy.com/photo/32478046-high-speed-train-moves-through-the-city-3d-rendering-high-speed-train-on-the-road-to-the-modern-city-ai-generated

Albalate, D., Bel, G., & Fageda, X. (2015). Competition and cooperation between high-speed rail and air transportation services in Europe. *Journal of Transport Geography*, *42*, 166–174. https://doi.org/10.1016/j.jtrangeo.2014.07.003

Almujibah, H., & Preston, J. (2019). The Total Social Costs of Constructing and Operating a High-Speed Rail Line Using a Case Study of the Riyadh-Dammam Corridor, Saudi Arabia. *Frontiers in Built Environment*, *5*. https://doi.org/10.3389/fbuil.2019.00079

Arbex, R. O., & da Cunha, C. B. (2015). Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, *81*, 355–376. https://doi.org/10.1016/j.trb.2015.06.014

Arellano, M., & Honoré, B. (2001). Panel Data Models: Some Recent Developments. *Handbook of Econometrics*, *5*, 3229–3296. https://doi.org/10.1016/s1573-4412(01)05006-1

Arentze, T. A., & Molin, E. J. E. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. *Transportation Research Part A: Policy and Practice*, *58*, 15–28. https://doi.org/10.1016/j.tra.2013.10.005

ARUP, & OXERA. (2010). *What are the findings from the econometric analysis? Findings report Prepared for the Department for Transport, Transport Scotland, and the Passenger Demand Forecasting Council*. https://www.oxera.com/wp-content/uploads/2018/03/Report-2-Findings-report-1.pdf

Asadi Bagloee, S., & Ceder, A. (Avi). (2011). Transit-network design methodology for actual-size road networks. *Transportation Research Part B: Methodological*, *45*(10), 1787–1804. https://doi.org/10.1016/j.trb.2011.07.005

Ashiabor, S., Baik, H., & Trani, A. (2007). Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States. *Transportation Research Record: Journal of the Transportation Research Board*, *2007*(1), 1–12. https://doi.org/10.3141/2007-01

Baaj, M. H., & Mahmassani, H. S. (1991). An AI-based approach for transit route system planning and design. *Journal of Advanced Transportation*, *25*(2), 187–209. https://doi.org/10.1002/atr.5670250205

Balbatunou, K. (n.d.). *High-speed train on the background of a futuristic city Free Photo*. Vecteezy. Retrieved September 1, 2024, from https://www.vecteezy.com/photo/28845484-high-speed-train-on-the-background-of-a-futuristic-city

Barra, A., Balassiano, R., Carvalho, L., Teypaz, N., & Cung, V. D. (2007). Solving the Transit Network Design problem with Constraint Programming. *HAL (Le Centre Pour La Communication Scientifique Directe)*, 24–28.

Barrón, I., Campos, J., Gagnepain, P., Nash, C., Ulied, A., & Vickerman, R. (2012). *Economic Analysis of High Speed Rail in Europe* (G. de Rus, Ed.). Fundacion BBVA. https://www.researchgate.net/publication/326159879_Economic_Analysis_of_High_Speed_Rail_in_Europe

BBC. (2019, August 23). Climate change: Should you fly, drive or take the train? *BBC News*. https://www.bbc.com/news/science-environment-49349566

BBD. (2015). *2015 FINANCIAL REPORT*. https://www.annualreports.com/HostedData/AnnualReportArchive/B/TSX_BBD_2015.pdf

Behrens, C., & Pels, E. (2012). Intermodal competition in the London–Paris passenger market: High-Speed Rail and air transport. *Journal of Urban Economics*, *71*(3), 278–288. https://doi.org/10.1016/j.jue.2011.12.005

Belal, E. M., Khalil, A. A., & El-Dash, K. M. (2020). Economic investigation for building a high-speed rail in developing countries: The case of Egypt. *Ain Shams Engineering Journal*, *11*(4), 1001–1011. https://doi.org/10.1016/j.asej.2020.02.003

Ben-Akiva, M., Cascetta, E., Coppola, P., Papola, A., & Velardi, V. (2010). High speed rail demand forecasting in a competitive market: the Italian case study. *Proceedings of the World Conference of Transportation Research (WCTR)*. https://www.academia.edu/download/45596565/High_Speed_Rail_Demand_Forecasting_in_a_20160513-12210-9khd4y.pdf

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*.

Bergantino, A. S., & Capozza, C. (2015). One price for all? Price discrimination and market captivity: Evidence from the Italian city-pair markets. *Transportation Research Part A: Policy and Practice*, *75*, 231–244. https://doi.org/10.1016/j.tra.2015.03.005

Bergantino, A. S., & Madio, L. (2020). Intermodal competition and substitution. HSR versus air transport: Understanding the socio-economic determinants of modal choice. *Research in Transportation Economics*, *79*(100823). https://doi.org/10.1016/j.retrec.2020.100823

Bielli, M., Caramia, M., & Carotenuto, P. (2002). Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*, *10*(1), 19–34. https://doi.org/10.1016/s0968-090x(00)00048-6

Bleijenberg, A. (2020). *Air2Rail: Reducing CO2 from intra-European aviation by a modal shift from air to rail* (p. 21). https://www.transportenvironment.org/wp-content/uploads/2021/07/2020_03_Air2Rail_Koios_strategy_rev.pdf

Boelrijk, W. (2019). *Gravity model for air passenger demand estimation: the addition of big data* [Master Thesis]. https://repository.tudelft.nl/islandora/object/uuid%3A2f2ee8aa-3ed3-4e5d-802d-b64e85853d7a

Bopp, E., & Douvinet, J. (2020). Spatial performance of location-based alerts in France. *International Journal of Disaster Risk Reduction*, *50*, 101909. https://doi.org/10.1016/j.ijdrr.2020.101909

Borgogno, F. (2023). *Roadmap Towards a Unified European High-Speed Rail Infrastructure*. https://repository.tudelft.nl/record/uuid:2c95aa94-039e-4539-8f61-6ae95184e1f7

Börjesson, M. (2014). Forecasting demand for high speed rail. *Transportation Research Part A: Policy and Practice*, *70*, 81–92. https://doi.org/10.1016/j.tra.2014.10.010

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2005). *A path based model for line planning in public transport*. Zuse Institute Berlin. Report ZR-05-18.

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2007). A Column-Generation Approach to Line Planning in Public Transport. *Transportation Science*, *41*(1), 123–132. https://doi.org/10.1287/trsc.1060.0161

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2008). Models for Line Planning in Public Transport. In *Computer-aided Systems in Public Transport* (Vol. 600, pp. 363–378). Springer. https://doi.org/10.1007/978-3-540-73312-6_18

Bourbonnais, P.-L., Morency, C., Trépanier, M., & Martel-Poliquin, É. (2019). Transit network design using a genetic algorithm with integrated road network and disaggregated O–D demand data. *Transportation*, *48*(1), 95–130. https://doi.org/10.1007/s11116-019-10047-1

Bradsher, K. (2013, September 23). Speedy Trains Transform China . *The New York Times*. https://www.nytimes.com/2013/09/24/business/global/high-speed-train-system-is-huge-success-for-china.html

Brand, D., Parody, T. E., Hsu, P. S., & Tierney, K. F. (1992). Forecasting High-Speed Rail Ridership. *Transportation Research Record*, *1341*, 12–18. https://onlinepubs.trb.org/Onlinepubs/trr/1992/1341/1341-002.pdf

Brons, M., Dijkstra, L., & Poelman, H. (2023). How fast are rail trips between EU cities and is rail faster than air? In *European Commission* (p. 15). https://ec.europa.eu/regional_policy/sources/work/2023-rail-vs-air_en.pdf

Burge, P., Rohr, C., & Kim, C. W. (2010). Modelling choices for long-distance travellers in the UK: an SP analysis of mode choice. *Proceeding from the European Transport Conference, Glasgow*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2873d3ac9de2b1f5c3e80f587401801738c8e055

Bussieck, M. R. (1998). *Optimal Lines in Public Rail Transport*. Braunschweig University of Technology. https://d-nb.info/955665965/34

Cabanne, I. (2003). *A Long Term Model For Long Distance Travel In France*. Association for European Transport.

Cadarso, L., Codina, E., Escudero, L. F., & Marín, A. (2017). Rapid transit network design: considering recovery robustness and risk aversion measures. *Transportation Research Procedia*, *22*, 255–264. https://doi.org/10.1016/j.trpro.2017.03.032

Cadarso, L., & Marín, Á. (2016). Rapid transit network design considering risk aversion. *Electronic Notes in Discrete Mathematics*, *52*, 29–36. https://doi.org/10.1016/j.endm.2016.03.005

Canca, D., José Luis Andrade-Pineda, De, A., & Calle, M. (2018). The Railway Rapid Transit frequency setting problem with speed-dependent operation costs. *Transportation Research. Part B: Methodological/Transportation Research. Part B, Methodological*, *117*(A), 494–519. https://doi.org/10.1016/j.trb.2018.09.013

Cancela, H., Mauttone, A., & Urquhart, M. E. (2015). Mathematical programming formulations for transit network design. *Transportation Research Part B: Methodological*, *77*, 17–37. https://doi.org/10.1016/j.trb.2015.03.006

Capali, B., & Ceylan, H. (2020). A multi-objective meta-heuristic approach for the transit network design and frequency setting problem. *Transportation Planning and Technology*, *43*(8), 851–867. https://doi.org/10.1080/03081060.2020.1829093

Capozza, C. (2016). The effect of rail travel time on airline fares: First evidence from the Italian passenger market. *Economics of Transportation*, *6*, 18–24. https://doi.org/10.1016/j.ecotra.2016.06.002

Carrese, S., & Gori, S. (2002). An Urban Bus Network Design Procedure. In *Applied Optimization* (Vol. 64, pp. 177–195). https://doi.org/10.1007/0-306-48220-7_11

Cartenì, A., Pariota, L., & Henke, I. (2017). Hedonic value of high-speed rail services: Quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities. *Transportation Research Part A: Policy and Practice*, *100*, 348–365. https://doi.org/10.1016/j.tra.2017.04.018

Cascetta, E., & Cartenì, A. (2014). The hedonic value of railways terminals. A quantitative analysis of the impact of stations quality on travellers behaviour. *Transportation Research Part A: Policy and Practice*, *61*, 41–52. https://doi.org/10.1016/j.tra.2013.12.008

Cascetta, E., & Coppola, P. (2011). High speed rail demand: empirical and modelling evidences from Italy. *Proceedings of European Transport Conference 2011*. https://www.researchgate.net/profile/Ennio-Cascetta/publication/258450854_High_Speed_Rail_demand_empirical_and_modeling_evidences_from_Italy/links/00463528b57ae48995000000/High-Speed-Rail-demand-empirical-and-modeling-evidences-from-Italy.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19. Association for European Transport.

Cascetta, E., & Coppola, P. (2012). An elastic demand schedule-based multimodal assignment model for the simulation of high speed rail (HSR) systems. *EURO Journal on Transportation and Logistics*, *1*(1-2), 3–27. https://doi.org/10.1007/s13676-012-0002-0

Castillo-Manzano, J. I., Pozo-Barajas, R., & Trapero, J. R. (2015). Measuring the substitution effects between High Speed Rail and air transport in Spain. *Journal of Transport Geography*, *43*, 59–65. https://doi.org/10.1016/j.jtrangeo.2015.01.008

CD. (2009, March 16). *China to purchase 350km/h high-speed trains for 39.2b yuan*. https://www.chinadaily.com.cn/china/2009-03/16/content_7584281.htm

CD. (2017, June 30). *China to roll out new class of bullet trains*. https://www.chinadaily.com.cn/business/2017-06/30/content_29941631.htm

Ceder, A. (2001). Operational objective functions in designing public transport routes. *Journal of Advanced Transportation*, *35*(2), 125–144. https://onlinelibrary.wiley.com/doi/abs/10.1002/atr.5670350205

Ceder, A. (2003). Chapter 3: designing public transport network and routes. In *Advanced Modeling for Transit Operations and Service Planning* (pp. 59–91). Elsevier Science Ltd. Pub., Pergamon Imprint.

Ceder, A., & Wilson, N. H. M. (1986). Bus network design. *Transportation Research Part B: Methodological*, *20*(4), 331–344. https://doi.org/10.1016/0191-2615(86)90047-0

Chai, J., Zhou, Y., Zhou, X., Wang, S., Zhang, Z. G., & Liu, Z. (2018). Analysis on shock effect of China's high-speed railway on aviation transport. *Transportation Research Part A: Policy and Practice*, *108*, 35–44. https://doi.org/10.1016/j.tra.2017.12.001

Chakroborty, P., & Dwivedi, T. (2002). Optimal Route Network Design for Transit Systems Using Genetic Algorithms. *Engineering Optimization*, *34*(1), 83–100. https://www.researchgate.net/publication/228551307_Optimal_Route_Network_Design_for_Transit_Systems_Using_Genetic_Algorithms

Chen, A., & Subprasom, K. (2007). Analysis of regulation and policy of private toll roads in a build-operate-transfer scheme under demand uncertainty. *Transportation Research Part A: Policy and Practice*, *41*(6), 537–558. https://doi.org/10.1016/j.tra.2006.11.009

Chen, A., Subprasom, K., & Ji, Z. (2003). Mean-variance model for the build-operate-transfer scheme under demand uncertainty. *Transportation Research Record*, *1857*(1), 93–101. https://doi.org/10.3141/1857-11

Chen, A., Subprasom, K., & Ji, Z. (2006). A simulation-based multi-objective genetic algorithm (SMOGA) for build-operate-transfer network design problem. *Optimization and Engineering Journal*, *7*(3), 225–247. https://doi.org/10.1007/s11081-006-9970-y

Chen, A., Zhou, Z., Chootinan, P., Ryu, S., Yang, C., & Wong, S. C. (2011). Transport Network Design Problem under Uncertainty: A Review and New Developments. *Transport Reviews*, *31*(6), 743–768. https://doi.org/10.1080/01441647.2011.589539

Chen, F. (2020, August 24). *China sets railway building spree in high-speed motion*. Asia Times. https://asiatimes.com/2020/08/china-sets-railway-building-spree-in-high-speed-motion/

Chen, X. (2010). *Modeling demand for High Speed Rail in Sweden: Private Trips*. http://www.diva-portal.se/smash/get/diva2:456640/FULLTEXT01.pdf

Chen, Z. (2017). Impacts of high-speed rail on domestic air transportation in China. *Journal of Transport Geography*, *62*, 184–196. https://doi.org/10.1016/j.jtrangeo.2017.04.002

Chen, Z., Wang, Z., & Jiang, H. (2019). Analyzing the heterogeneous impacts of high-speed rail entry on air travel in China: A hierarchical panel regression approach. *Transportation Research Part A: Policy and Practice*, *127*, 86–98. https://doi.org/10.1016/j.tra.2019.07.004

China Daily. (2020). *Chinese cities with a population of 500,000 to have high-speed rail in 15 years*. Chinadaily.com.cn. https://www.chinadaily.com.cn/a/202008/13/WS5f34ddfaa3108348172601d0.html

Cipriani, E., Gori, S., & Petrelli, M. (2012). Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, *20*(1), 3–14. https://doi.org/10.1016/j.trc.2010.09.003

Clever, R., & Hansen, M. M. (2008). Interaction of Air and High-Speed Rail in Japan. *Transportation Research Record: Journal of the Transportation Research Board*, *2043*(1), 1–12. https://doi.org/10.3141/2043-01

Clewlow, R. R., Sussman, J. M., & Balakrishnan, H. (2014). The impact of high-speed rail and low-cost carriers on European air passenger traffic. *Transport Policy*, *33*, 136–143. https://doi.org/10.1016/j.tranpol.2014.01.015

Coker, A. K. (1995). Numerical Computation. *Elsevier EBooks*, 1–102. https://doi.org/10.1016/b978-088415280-4/50002-9

Costa, A. M. (2005). A survey on benders decomposition applied to fixed-charge network design problems. *Computers & Operations Research*, *32*(6), 1429–1450. https://doi.org/10.1016/j.cor.2003.11.012

Couto, A., & Graham, D. J. (2007). The impact of high-speed technology on railway demand. *Transportation*, *35*(1), 111–128. https://doi.org/10.1007/s11116-007-9138-4

Curtin, K. M., & Biba, S. (2011). The Transit Route Arc-Node Service Maximization problem. *European Journal of Operational Research*, *208*(1), 46–56. https://doi.org/10.1016/j.ejor.2010.07.026

Daly, A. (2010). *Cost damping in travel demand models: Report of a study for the department for transport*. https://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR717.pdf

Danapour, M., Nickkar, A., Jeihani, M., & Khaksar, H. (2018). Competition between high-speed rail and air transport in Iran: The case of Tehran–Isfahan. *Case Studies on Transport Policy*, *6*(4), 456–461. https://doi.org/10.1016/j.cstp.2018.05.006

Dargay, J. M., & Clark, S. (2012). The determinants of long distance travel in Great Britain. *Transportation Research Part A: Policy and Practice*, *46*(3), 576–587. https://doi.org/10.1016/j.tra.2011.11.016

de Bok, M., Costa, Á., Melo, S., Palma, V., & Frias, R. D. (2010, July). Estimation of a mode choice model for long distance travel in Portugal. *Proceedings of the World Conference of Transport Research*.

de Keizer, B., Kouwenhoven, M., & Hofker, F. (2015). New Insights in Resistance to Interchange. *Transportation Research Procedia*, *8*, 72–79. https://doi.org/10.1016/j.trpro.2015.06.043

de Palma, A., Fosgerau, M., Melo, E., & Shum, M. (2019). *Discrete choice and rational inattention: A general equivalence result*. Mimeo.

de Rus, G., Campos, J., Socorro, M., Valido, J., & Ortuño, A. (2020). Cost-benefit analysis of railway projects: high-speed lines and suburban lines. *Independent Authority for Fiscal Responsibility(AIReF)*. https://www.researchgate.net/publication/344046638_Cost-benefit_analysis_of_railway_projects_high-speed_lines_and_suburban_lines

Desmos. (2011). *Desmos Graphing Calculator*. Desmos Graphing Calculator; Desmos. https://www.desmos.com/calculator

Deutsche Bahn. (2023). *Metropolitan Network: A strong European railway for an ever closer union*. https://www.deutschebahn.com/resource/blob/10910200/bb0f84cddf0b03ca924325579ebdaa55/Study-EU-HGV_MK-data.pdf

Deutschel, J. L. (2022). *High-speed rail and its competitors on long-distance travel in Western Europe* [Bachelor's thesis]. https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/581664/1/sa759.pdf

Díez-Prisonero, R. (2012). High speed rail (HSR) vs air transportation: Trendy competition in the transport geography of Spain. *Inf. Technol. Res. J*, *2*, 8–19. https://www.researchgate.net/publication/260385247_High_speed_rail_HSR_vs_air_transportation_Trendy_competition_in_the_transport_geography_of_Spain

Dixit, V., & Niu, C. (2023). Quantum computing for transport network design problems. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-38787-2

Dobruszkes, F. (2011). High-speed rail and air transport competition in Western Europe: A supply-oriented perspective. *Transport Policy*, *6*, 870–879. https://doi.org/10.1016/j.tranpol.2011.06.002

Dobruszkes, F., Dehon, C., & Givoni, M. (2014). Does European high-speed rail affect the current level of air services? An EU-wide analysis. *Transportation Research Part A: Policy and Practice*, *69*, 461–475. https://doi.org/10.1016/j.tra.2014.09.004

Doomernik, J. E. M. J. (2017). *Strategieën voor internationale hogesnelheidsspoorvervoerders in de Europese reizigersvervoersmarkt* (p. 231). https://medialibrary.uantwerpen.be/oldcontent/container2629/files/ThesisDoomernik.pdf

Durán-Micco, J., & Vansteenwegen, P. (2021). A survey on the transit network design and frequency setting problem. *Public Transport*, *14*, 155–190. https://doi.org/10.1007/s12469-021-00284-y

E.H. Michell. (2024). Rail market share against air. In *Hot Rails*. https://hotrails.net/2024/04/rails-market-share-against-air-a-global-review/

EC. (2018). *Flash Eurobarometer 463 — Europeans' satisfaction with passenger rail services*. European Commission.

EEA. (2020). *Transport and Environment Report 2020: Train or plane?* https://www.eea.europa.eu/publications/transport-and-environment-report-2020/download

EEA. (2023, October 24). *Greenhouse gas emissions from transport in Europe*. Www.eea.europa.eu. https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emissions-from-transport?activeAccordion=546a7c35-9188-4d23-94ee-005d97c26f2b

EEA. (2024, January 19). *Transport and mobility*. Www.eea.europa.eu. https://www.eea.europa.eu/en/topics/in-depth/transport-and-mobility#:~:text=Travel%20by%20a%20petrol%20or

Erkut, E., & Gzara, F. (2008). Solving the hazmat transport network design problem. *Computers & Operations Research*, *35*(7), 2234–2247. https://doi.org/10.1016/j.cor.2006.10.022

EU. (2024). *https://taxation-customs.ec.europa.eu/excise-duties-other-energy-tax-legislation_en*. Europa.eu. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32003L0096

Eurocities. (2020, August 6). *Essen - Eurocities*. Eurocities.eu. https://eurocities.eu/cities/essen/

European Commission. (2013). *Trans-European Transport Network (TEN-T)*. Transport.ec.europa.eu. https://transport.ec.europa.eu/transport-themes/infrastructure-and-investment/trans-european-transport-network-ten-t_en

European Commission. (2014). *Communication from the Commission — Guidelines on State aid to airports and airlines*. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52014XC0404%2801%29

European Commission. (2021). *Reducing emissions from aviation*. Climate.ec.europa.eu. https://climate.ec.europa.eu/eu-action/transport/reducing-emissions-aviation_en#aviation-emissions

European Council. (2024, January 3). *Rail transport policy*. European Council | Council of the European Union. https://www.consilium.europa.eu/en/policies/rail-transport-policy/#:~:text=Sustainable%20and%20smart%20mobility&text=The%20strategy%20outlines%20the%20EU's,adopted%20conclusions%20on%20this%20topic.

European Court of Auditors. (2018). *Special report: A European high-speed rail network*. Op.europa.eu. https://op.europa.eu/webpub/eca/special-reports/high-speed-rail-19-2018/en/

European Parliament. (2019, March 22). *CO2 emissions from cars: facts and figures (infographics)*. Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20190313STO31218/co2-emissions-from-cars-facts-and-figures-infographics#:~:text=Passenger%20cars%20are%20a%20major

Eurostat. (2019). *Eurostat Air Transport Statistics* (pp. 1–2). https://www.sipotra.it/wp-content/uploads/2019/12/Air-transport-statistics.pdf

Eurostat. (2024a). *Data browser - avia_pa*. Europa.eu. https://ec.europa.eu/eurostat/databrowser/explore/all/transp?lang=en&subtheme=avia.avia_pa&display=list&sort=category

Eurostat. (2024b). Gross domestic product (GDP) at current market prices by metropolitan regions. *Europa.eu*. met_10r_3gdp. https://ec.europa.eu/eurostat/databrowser/view/met_10r_3gdp/default/table?lang=en

Eurostat. (2024c). *Population on 1 January by age, sex and NUTS 2 region*. demo_r_d2jan. https://ec.europa.eu/eurostat/databrowser/view/DEMO_R_D2JAN/default/table?lang=en

Eurostat. (2024d). Regional gross domestic product by NUTS 2 regions - million EUR. *Europa.eu*. tgs00003. https://ec.europa.eu/eurostat/databrowser/view/tgs00003/default/table?lang=en

Eurostat. (2024e). *Who we are - Eurostat*. Ec.europa.eu. https://ec.europa.eu/eurostat/web/main/about-us/who-we-are

Eurotunnel. (2024). *About My Journey*. LeShuttle. https://www.eurotunnel.com/uk/travelling-with-us/about-my-journey/#:~:text=Depending%20on%20the%20height%20of

Fan, L., Chen, H., & Gao, Y. (2019). An improved flower pollination algorithm to the urban transit routing problem. *Soft Computing*, *24*(7), 5043–5052. https://doi.org/10.1007/s00500-019-04253-3

Fan, L., & Mumford, C. L. (2010). A metaheuristic approach to the urban transit routing problem. *Journal of Heuristics*, *16*(3), 353–372. https://doi.org/10.1007/s10732-008-9089-8

Fan, W., & Machemehl, R. (2004). *Optimal transit route network design problem: algorithms, implementations, and numerical results*. Center for Transportation Research, University of Texas. Tech. Rep. SWUTC/04/ 167244-1.

Fan, W., & Machemehl, R. B. (2006a). Optimal Transit Route Network Design Problem with Variable Transit Demand: Genetic Algorithm Approach. *Journal of Transportation Engineering*, *132*(1), 40–51. https://doi.org/10.1061/(asce)0733-947x(2006)132:1(40)

Fan, W., & Machemehl, R. B. (2006b). Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering*, *132*(2), 122–132. https://doi.org/10.1061/(asce)0733-947x(2006)132:2(122)

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., & Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, *229*(2), 281–302. https://doi.org/10.1016/j.ejor.2013.01.001

Feng, X., Zhu, X., Qian, X., Jie, Y., Ma, F., & Niu, X. (2019). A new transit network design study in consideration of transfer time composition. *Transportation Research Part D: Transport and Environment*, *66*, 85–94. https://doi.org/10.1016/j.trd.2018.03.019

Feremans, C., Labbé, M., & Laporte, G. (2003). Generalized network design problems. *European Journal of Operational Research*, *148*(1), 1–13. https://doi.org/10.1016/s0377-2217(02)00404-6

Fiig, T., Härdling, R., Pölt, S., & Hopperstad, C. (2014). Demand forecasting and measuring forecast accuracy in general fare structures. *Journal of Revenue and Pricing Management*, *13*(6), 413–439. https://doi.org/10.1057/rpm.2014.29

Financial Post. (2019). Gurobi Optimization Signs Partnership with Air France for Its Operations Research Projects. *Financial Post*. https://financialpost.com/pmn/press-releases-pmn/business-wire-news-releases-pmn/gurobi-optimization-signs-partnership-with-air-france-for-its-operations-research-projects

FlightsFrom. (2024). *Worldwide routes and flights from all airports*. https://www.flightsfrom.com/

Florczyk, A., Corbane, C., Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2020). *GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A*. https://doi.org/10.2905/53473144-b88c-44bc-b4a3-4583ed1f547e. European Commission, Joint Research Centre (JRC).

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2005). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association*, *71*(2), 131–146. https://doi.org/10.1080/01944360508976688

Frei, A., Kuhnimhof, T., & Axhausen, K. W. (2010). Long distance travel in Europe today. *89th Annual Meeting of the Transportation Research Board 2010*, *611*. https://doi.org/10.3929/ethz-a-005976787

Fröidh, O. (2005). Market effects of regional high-speed trains on the Svealand line. *Journal of Transport Geography*, *13*(4), 352–361. https://doi.org/10.1016/j.jtrangeo.2004.12.006

Fröidh, O. (2006). *Modelling operational costs of a future high-speed train*. Royal Institute of Technology (KTH). http://www.gronataget.se/upload/PublikaDokument/Modelling%20operational.pdf

Fu, X., Oum, T., & Yan, J. (2014). An Analysis of Travel Demand in Japan's Intercity Market: Empirical Estimation and Policy Simulation. *Journal of Transport Economics and Policy (JTEP)*, *48*(17), 97–113. https://www.ingentaconnect.com/content/lse/jtep/2014/00000048/00000001/art00007

Fusco, G., Gori, S., & Petrelli, M. (2002). A heuristic transit network design algorithm for medium size towns. *Proceedings of the 13th Mini-EURO Conference*. https://www.researchgate.net/profile/Gaetano-Fusco/publication/228598844_A_heuristic_transit_network_design_algorithm_for_medium_size_towns/links/09e4151472279c118c000000/A-heuristic-transit-network-design-algorithm-for-medium-size-towns.pdf

Gallo, M., Montella, B., & D'Acierno, L. (2011). The transit network design problem with elastic demand and internalisation of external costs: An application to rail frequency optimisation. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1276–1305. https://doi.org/10.1016/j.trc.2011.02.008

Gao, Z., Wu, J., & Sun, H. (2005). Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological*, *39*(6), 479–495. https://doi.org/10.1016/j.trb.2004.06.004

Gaudrey, M. (2008). Non linear logit modelling developments and high speed rail profitability. *Agora Jules Dupuit, Publication AJD-127*. http://www.e-ajd.net/source-pdf/AJD%20127%20DEST%2012%20Gaudry_HSR_%20Profitability_23_July_2008.pdf

Gerdtham, U. G., & Jönsson, B. (2000). Chapter 1 - International comparisons of health expenditure: Theory, data and econometric analysis. *Handbook of Health Economics*, *1*(A), 11–53. https://doi.org/10.1016/s1574-0064(00)80160-2

Geurts, M. D., & Whitlark, D. (1992). Forecasting market share. *The Journal of Business Forecasting Methods & Systems*, *17*. https://www.proquest.com/openview/099522dd166734216ac6076b5c0ada67/1?cbl=28144&pq-origsite=gscholar&parentSessionId=T1wopyNQaNnNBP3xO1Jfi5YM4wYZkcVGX%2FZjdjF%2BUk0%3D. Winter 1992-93.

Gössling, S., & Humpe, A. (2020). The global scale, distribution and growth of aviation: Implications for climate change. *Global Environmental Change*, *65*(102194), 102194. https://doi.org/10.1016/j.gloenvcha.2020.102194

Great Brittain Navy Dept. (1987). *Admiralty Manual of Navigation* (45th ed., Vol. 1). The Stationery Office. p. 10.

Grolle, J., Donners, B., Annema, J. A., Duinkerken, M., & Cats, O. (2024). Service design and frequency setting for the European high-speed rail network. *Transportation Research Part A: Policy and Practice*, *179*(103906). https://doi.org/10.1016/j.tra.2023.103906

Grosche, T., Rothlauf, F., & Heinzl, A. (2007). Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, *13*(4), 175–183. https://doi.org/10.1016/j.jairtraman.2007.02.001

GRR. (2021). *Bombardier's Chinese JV to supply 16 new high-speed train CR400AF cars*. Global Railway Review. https://www.globalrailwayreview.com/news/116638/bombardier-chinese-jv-high-speed-train-cars/

Gu, H., & Wan, Y. (2020). Can entry of high-speed rail increase air traffic? Price competition, travel time difference and catchment expansion. *Transport Policy*, *97*, 55–72. https://doi.org/10.1016/j.tranpol.2020.07.011

Guan, J. F., Yang, H., & Wirasinghe, S. C. (2003). Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B*, *40*(10), 885–902. https://ideas.repec.org/a/eee/transb/v40y2006i10p885-902.html

Guan, J. F., Yang, H., & Wirasinghe, S. C. (2006). Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological*, *40*(10), 885–902. https://doi.org/10.1016/j.trb.2005.12.003

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273. https://doi.org/10.1016/j.tra.2008.03.011

Gundelfinger-Casar, J., & Coto-Millán, P. (2017). Intermodal competition between high-speed rail and air transport in Spain. *Utilities Policy*, *47*, 12–17. https://doi.org/10.1016/j.jup.2017.06.001

Gutiérrez-Jarpa, G., Laporte, G., Marianov, V., & Moccia, L. (2017). Multi-objective rapid transit network design with modal competition: The case of Concepción, Chile. *Computers & Operations Research*, *78*, 27–43. https://doi.org/10.1016/j.cor.2016.08.009

Hansen, M., & Liu, Y. (2015). Airline competition and market frequency: A comparison of the s-curve and schedule delay models. *Transportation Research Part B: Methodological*, *78*, 301–317. https://doi.org/10.1016/j.trb.2015.04.012

Harnish, R. (2023, January 22). *What makes high-speed rail successful?* High Speed Rail Alliance. https://www.hsrail.org/blog/success-factors/

Hasselström, D. (1979). *A Method for Optimization of Urban Bus Route Networks*. Volvo Bus Corporation.

Hasselström, D. (1981). *Public Transportation Planning: A Mathematical Programming Approach*. BAS.

Hensher, D. A. (1997). A practical approach to identifying the market potential for high speed rail: A case study in the Sydney-Canberra corridor. *Transportation Research Part A: Policy and Practice*, *31*(6), 431–446. https://doi.org/10.1016/s0965-8564(97)00001-3

Hepher, T., & Frost, L. (2021, June 10). Airbus tells EU hydrogen won't be widely used in planes before 2050. *Reuters*. https://www.reuters.com/business/aerospace-defense/airbus-tells-eu-hydrogen-wont-be-widely-used-planes-before-2050-2021-06-10/

Heyken Soares, P., Ahmed, L., Mao, Y., & Mumford, C. L. (2020). Public transport network optimisation in PTV Visum using selection hyper-heuristics. *Public Transport*, *13*(1), 163–196. https://doi.org/10.1007/s12469-020-00249-7

Heyken Soares, P., Mumford, C. L., Amponsah, K., & Mao, Y. (2019). An adaptive scaled network for public transport route optimisation. *Public Transport*, *11*(2), 379–412. https://doi.org/10.1007/s12469-019-00208-x

Hiscock, R., Macintyre, S., Kearns, A., & Ellaway, A. (2002). Means of transport and ontological security: Do cars provide psycho-social benefits to their users? *Transportation Research Part D: Transport and Environment*, *7*(2), 119–135. https://doi.org/10.1016/s1361-9209(01)00015-3

Hong, S.-J., & Najmi, H. (2022). Impact of High-speed rail on air travel demand between Dallas and Houston applying Monte Carlo simulation. *Journal of Air Transport Management*, *102*(102222). https://doi.org/10.1016/j.jairtraman.2022.102222

Hood, C. P. (2006). *Shinkansen: From Bullet Train to Symbol of Modern Japan*. Routledge.

Hsu, C.-I., & Chung, W.-M. (1997). A model for market share distribution between high-speed and conventional rail services in a transportation corridor. *The Annals of Regional Science*, *31*(2), 121–153. https://doi.org/10.1007/s001680050042

Hu, J., Shi, X., Song, J., & Xu, Y. (2005). Optimal Design for Urban Mass Transit Network Based on Evolutionary Algorithms. *Lecture Notes in Computer Science*, *3611*, 429–429. https://doi.org/10.1007/11539117_148

IATA. (2023). *Global Outlook for Air Transport: Highly Resilient, Less Robust* (p. 15). https://www.iata.org/en/iata-repository/publications/economic-reports/global-outlook-for-air-transport----june-2023/

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, *77*, 38–75. https://doi.org/10.1016/j.trb.2015.03.002

ICAO. (2019). *Trends in Emissions that affect Climate Change*. Icao.int. https://www.icao.int/environmental-protection/Pages/ClimateChange_Trends.aspx

ICAO. (2023). *ICAO Document 7910: Location Indicators* (190th ed.). https://store.icao.int/en/location-indicators-doc-7910?_gl=1*rmajyq*_ga*MTQ4NTk2MDUxNS4xNzA4NDY5MzM1*_ga_992N3YDLBQ*MTcwOTgyMDg1NC4zLjEuMTcwOTgyMTg0Ny4wLjAuMA..

Iliopoulou, C., Kepaptsoglou, K., & Vlahogianni, E. (2019). Metaheuristics for the transit route network design problem: a review and comparative analysis. *Public Transport*, *11*(3), 487–521. https://doi.org/10.1007/s12469-019-00211-2

Iliopoulou, C., Tassopoulos, I., Kepaptsoglou, K., & Beligiannis, G. (2019). Electric Transit Route Network Design Problem: Model and Application. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(8), 264–274. https://doi.org/10.1177/0361198119838513

InflationTool. (2024). *Inflation Calculator*. Www.inflationtool.com. https://www.inflationtool.com/euro/2005-to-present-value?amount=1000&year2=2024&frequency=yearly

Inoue, G., Ono, M., Uehara, K., & Isono, F. (2015). Stated-preference analysis to estimate the domestic transport demand following the future entry of LCCs and the inauguration of the Linear Chuo Shinkansen in Japan. *Journal of Air Transport Management*, *47*, 199–217. https://doi.org/10.1016/j.jairtraman.2015.06.004

IRJ. (2010, September 1). *First Chinese--designed HS train breaks cover: CSR is building a fleet of very-high-speed trains will be the first such trains built in China to use only Chinese technology. They will be launched on the new Beijing - Shanghai high-speed line, due to open next year, as well as the Beijing - Guangzhou line.(Innovative trains) - International Railway Journal | HighBeam Research*. Web.archive.org. https://web.archive.org/web/20181115185155/https://www.highbeam.com/doc/1G1-238084330.html

Islam, K. A., Moosa, I. M., Mobin, J., Nayeem, M. A., & Rahman, M. S. (2019). A heuristic aided Stochastic Beam Search algorithm for solving the transit network design problem. *Swarm and Evolutionary Computation*, *46*, 154–170. https://doi.org/10.1016/j.swevo.2019.02.007

Janić, M. (1999). Aviation and externalities: the accomplishments and problems. *Transportation Research Part D: Transport and Environment*, *4*(3), 159–180. https://doi.org/10.1016/s1361-9209(99)00003-6

Janić, M. (2017). Transport Systems. In *CRC Press eBooks*. Informa. https://doi.org/10.1201/9781315371023

JawedGfx. (n.d.). *Swift train blurs station backdrop High speed elegance in rapid transit spectacle For Social Media Post Size AI Generated Free Photo*. Vecteezy. Retrieved September 1, 2024, from https://nl.vecteezy.com/foto/33118048-snel-trein-vervaagt-station-backdrop-hoog-snelheid-elegantie-in-snel-doorvoer-schouwspel-voor-sociaal-media-post-grootte-ai-gegenereerd

Jha, S. B., Jha, J. K., & Tiwari, M. K. (2019). A multi-objective meta-heuristic approach for transit network design and frequency setting problem in a bus transit system. *Computers & Industrial Engineering*, *130*, 166–186. https://doi.org/10.1016/j.cie.2019.02.025

Jiang, X., Zhang, L., & Michael) Chen, X. (2014). Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C: Emerging Technologies*, *44*, 110–127. https://doi.org/10.1016/j.trc.2014.03.016

Jiménez, J. L., & Betancor, O. (2012). When trains go faster than planes: The strategic reaction of airlines in Spain. *Transport Policy*, *23*, 34–41. https://doi.org/10.1016/j.tranpol.2012.06.003

Jorritsma, P. (2009). Substitution Opportunities of High Speed Train for Air Transport. In *Aerlines Magazine*. https://aerlinesmagazine.wordpress.com/2009/05/01/substitution-opportunities-of-high-speed-train-for-air-transport/

Jung, S.-Y., & Yoo, K.-E. (2014). Passenger airline choice behavior for domestic short-haul travel in South Korea. *Journal of Air Transport Management*, *38*, 43–47. https://doi.org/10.1016/j.jairtraman.2013.12.017

Kanafani, A., Wang, R., & Griffin, A. (2012). The Economics of Speed – Assessing the performance of High Speed Rail in Intermodal Transportation. *Procedia - Social and Behavioral Sciences*, *43*, 692–708. https://doi.org/10.1016/j.sbspro.2012.04.143

Kechagiopoulos, P. N., & Beligiannis, G. N. (2014). Solving the Urban Transit Routing Problem using a particle swarm optimization based algorithm. *Applied Soft Computing*, *21*, 654–676. https://doi.org/10.1016/j.asoc.2014.04.005

Kepaptsoglou, K., & Karlaftis, M. (2009). Transit Route Network Design Problem: Review. *Journal of Transportation Engineering*, *135*(8), 491–505. https://doi.org/10.1061/(asce)0733-947x(2009)135:8(491)

King, J. (1996). *Economic significance of high speed rail*. https://ses.library.usyd.edu.au/bitstream/handle/2123/19404/ITS-WP-96-10.pdf?sequence=1&isAllowed=y. ITS Working Paper Series 1991-2000, CD-ROM, the Institute of Transport Studies, Faculty of Economics & Business, The University of Sydney.

Kroes, E., & Savelberg, F. (2019). Substitution from Air to High-Speed Rail: The Case of Amsterdam Airport. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(5), 166–174. https://doi.org/10.1177/0361198119839952

Lampkin, W., & Saalmans, P. D. (1967). The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study. *Journal of the Operational Research Society*, *18*(4), 375–397. https://doi.org/10.1057/jors.1967.70

Lee, J.-K., Yoo, K.-E., & Song, K.-H. (2016). A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the Seoul-Jeju route. *Journal of Air Transport Management*, *56*, 131–137. https://doi.org/10.1016/j.jairtraman.2016.04.020

Lee, Y.-J., & Vuchic, V. R. (2005). Transit Network Design with Variable Demand. *Journal of Transportation Engineering*, *131*(1), 1–10. https://doi.org/10.1061/(asce)0733-947x(2005)131:1(1)

Leng, N., Nie, L., Guo, G., & Wu, X. (2015). Passenger Flow Forecasting for Chinese High Speed Rail Network. *Proceedings of the 2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering*. https://doi.org/10.2991/meic-15.2015.153

Levinson, D., Mathieu, J. M., Gillen, D., & Kanafani, A. (1997). The full cost of high-speed rail: an engineering approach. *The Annals of Regional Science*, *31*(2), 189–215. https://doi.org/10.1007/s001680050045

Li, H., Strauss, J., & Lu, L. (2019). The impact of high-speed rail on civil aviation in China. *Transport Policy*, *74*, 187–200. https://doi.org/10.1016/j.tranpol.2018.11.015

Li, X., Ma, R., Guo, Y., Wang, W., Yan, B., & Chen, J. (2021). Investigation of factors and their dynamic effects on intercity travel modes competition. *Travel Behaviour and Society*, *23*, 166–176. https://doi.org/10.1016/j.tbs.2021.01.003

Li, Y. T., & Schmöcker, J. D. (2014). Demand impact of access links to Taiwan high speed rail. *Proceeding of the 19th International Conference of Hong Kong Society for Transportation Studies*, 103–110. https://www.researchgate.net/profile/Yeun-Touh-Li/publication/271345984_DEMAND_IMPACT_OF_ACCESS_LINKS_TO_TAIWAN_HIGH_SPEED_RAIL/links/54c5fa2d0cf256ed5a9c41bd/DEMAND-IMPACT-OF-ACCESS-LINKS-TO-TAIWAN-HIGH-SPEED-RAIL.pdf

Li, Z.-C., & Sheng, D. (2016). Forecasting passenger travel demand for air and high-speed rail integration service: A case study of Beijing-Guangzhou corridor, China. *Transportation Research Part A: Policy and Practice*, *94*, 397–410. https://doi.org/10.1016/j.tra.2016.10.002

Liang, J., Wu, J., Gao, Z., Sun, H., Yang, X., & Hong Kam Lo. (2019). Bus transit network design with uncertainties on the basis of a metro network: A two-step model framework. *Transportation Research Part B: Methodological*, *126*, 115–138. https://doi.org/10.1016/j.trb.2019.05.011

Lieshout, R. (2012). Measuring the size of an airport's catchment area. *Journal of Transport Geography*, *25*, 27–34. https://doi.org/10.1016/j.jtrangeo.2012.07.004

Liu, S., Wan, Y., Ha, H.-K., Yoshida, Y., & Zhang, A. (2019). Impact of high-speed rail network development on airport traffic and traffic distribution: Evidence from China and Japan. *Transportation Research Part A: Policy and Practice*, *127*, 115–135. https://doi.org/10.1016/j.tra.2019.07.015

Liu, Y., Feng, X., Ding, C., Hua, W., & Ruan, Z. (2020). Electric Transit Network Design by an Improved Artificial Fish-Swarm Algorithm. *Journal of Transportation Engineering, Part A: Systems*, *146*(8), 04020071. https://doi.org/10.1061/jtepbs.0000393

López-Pita, A., & Robusté, F. (2004). *High-Speed Line Airport Connections in Europe: State-of-the-Art Study*. https://doi.org/10.3141/1863-02

Lu, Y., Kimpton, A., Prato, C. G., Sipe, N., & Corcoran, J. (2023). First and last mile travel mode choice: A systematic review of the empirical literature. *International Journal of Sustainable Transportation*, *18*(1), 1–14. https://doi.org/10.1080/15568318.2023.2218285

Lubis, H. A., Pantas, V. B., & Farda, M. (2019). Demand forecast of Jakarta-Surabaya high speed rail based on stated preference method. *Int. J. Technol*, *10*(2). https://www.academia.edu/download/70809448/2442.pdf

Ma, W., Wang, Q., Yang, H., Zhang, A., & Zhang, Y. (2019). Effects of Beijing-Shanghai high-speed rail on air travel: Passenger types, airline groups and tacit collusion. *Research in Transportation Economics*, *74*, 64–76. https://doi.org/10.1016/j.retrec.2018.12.002

Magnanti, T. L., & Wong, R. T. (1984). Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, *18*(1), 1–55. https://doi.org/10.1287/trsc.18.1.1

Mahardika, M. D., Irawan, M. Z., & Bastarianto, F. F. (2022). Exploring the potential demand for Jakarta–Bandung high-speed rail. *Transportation Research Interdisciplinary Perspectives*, *15*(100658). https://doi.org/10.1016/j.trip.2022.100658

Maix. (2007, February 4). *File:Blank map of Europe (with disputed regions).svg*. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Blank_map_of_Europe_(with_disputed_regions).svg

Malonda, A. G., & Carles, A. G. (2003). RADIOACTIVITY COUNTING STATISTICS. *Handbook of Radioactivity Analysis*, 609–654. https://doi.org/10.1016/b978-012436603-9/50012-0

Mandel, B., Gaudry, M., & Rothengatter, W. (1994). Linear or nonlinear utility functions in logit models? The impact on German high-speed rail demand forecasts. *Transportation Research Part B: Methodological*, *28*(2), 91–101. https://doi.org/10.1016/0191-2615(94)90019-1

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, *5*(6), 396–404. https://doi.org/10.1016/0377-2217(80)90126-5

MarcoPolo. (2024). *Is The HSR Worth It?* MacroPolo. https://macropolo.org/digital-projects/high-speed-rail/introduction/

Marcucci, E., & Gatta, V. (2011). Regional airport choice: Consumer behaviour and policy implications. *Journal of Transport Geography*, *19*(1), 70–84. https://doi.org/10.1016/j.jtrangeo.2009.10.001

Marín, Á., & García-Ródenas, R. (2009). Location of infrastructure in urban railway networks. *Computers & Operations Research*, *36*(5), 1461–1477. https://doi.org/10.1016/j.cor.2008.02.008

Martín, J. C., & Nombela, G. (2007). Microeconomic impacts of investments in high speed trains in Spain. *The Annals of Regional Science*, *41*(3), 715–733. https://doi.org/10.1007/s00168-007-0116-8

Martín, J. C., Román, C., García-Palomares, J. C., & Gutiérrez, J. (2014). Spatial analysis of the competitiveness of the high-speed train and air transport: The role of access to terminals in the Madrid–Barcelona corridor. *Transportation Research Part A: Policy and Practice*, *69*, 392–408. https://doi.org/10.1016/j.tra.2014.09.010

Martínez, H. S., Moyano, A., Coronado, J. M., & Garmendia, M. (2016). Catchment areas of high-speed rail stations: a model based on spatial analysis using ridership surveys. *European Journal of Transport and Infrastructure Research*, *16*(2). https://journals.open.tudelft.nl/ejtir/article/view/3143/3329

Martínez, H., Mauttone, A., & Urquhart, M. E. (2014). Frequency optimization in public transportation systems: Formulation and metaheuristic approach. *European Journal of Operational Research*, *236*(1), 27–36. https://doi.org/10.1016/j.ejor.2013.11.007

Mauttone, A., & Urquhart, M. E. (2009). A route set construction algorithm for the transit network design problem. *Computers & Operations Research*, *36*(8), 2440–2449. https://doi.org/10.1016/j.cor.2008.09.014

May, M., J., S., & Bart, A. (2022). *2.2. Modeling Revenue, Costs, and Profit*. Slu.edu. https://mathstat.slu.edu/~may/ExcelCalculus/sec-2-2-ModelingRevenueCostProfit.html

Miyoshi, C., & Givoni, M. (2013). The Environmental Case for the High-Speed Train in the UK: Examining the London–Manchester Route. *International Journal of Sustainable Transportation*, *8*(2), 107–126. https://doi.org/10.1080/15568318.2011.645124

Mizutani, J., & Sakai, H. (2021). Which is a stronger competitor, High Speed Rail, or Low Cost Carrier, to Full Service Carrier? – Effects of HSR network extension and LCC entry on FSC's airfare in Japan. *Journal of Air Transport Management*, *90*(101965). https://doi.org/10.1016/j.jairtraman.2020.101965

Mokhtarian, P. L., & Salomon, I. (2001). How derived is the demand for travel? Some conceptual and measurement considerations. *Transportation Research Part A: Policy and Practice*, *35*(8), 695–719. https://doi.org/10.1016/S0965-8564(00)00013-6

Morgan, J., Daugherty, R., Hilchie, A., & Carey, B. (2003). Sample Size and Modeling Accuracy of Decision Tree based Data Mining Tools. *Academy of Information and Management Sciences Journal*, *6*(2), 77–92. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=52af82237a3da053481d1afc17c4c853b429e9bf

Moyano, A., Moya-Gómez, B., & Gutiérrez, J. (2018). Access and egress times to high-speed rail stations: a spatiotemporal accessibility analysis. *Journal of Transport Geography*, *73*, 84–93. https://doi.org/10.1016/j.jtrangeo.2018.10.010

Murray, A. T. (2003). A Coverage Model for Improving Public Transit System Accessibility and Expanding Access. *Annals of Operations Research*, *123*(1/4), 143–156. https://doi.org/10.1023/a:1026123329433

Nash, C. (2010). *nvironmental and Other Co-benefits of Developing a High Speed Rail System in California: A Prospective Vision 2010–2050*. Center for Environmental Public Policy.

Nelldal, B. L., & Jansson, K. (2010). High-speed trains in Sweden–a good idea? *12th WCTR Congress in Lisbon, 11-15th July 2010*.

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*.

Ng, M. T. M., Mahmassani, H. S., Verbas, Ö., Cokyasar, T., & Engelhardt, R. (2024). *Redesigning large-scale multimodal transit networks with shared autonomous mobility services*. https://arxiv.org/ftp/arxiv/papers/2307/2307.16075.pdf

Nikolić, M., & Teodorović, D. (2013). Transit network design by Bee Colony Optimization. *Expert Systems with Applications*, *40*(15), 5945–5955. https://doi.org/10.1016/j.eswa.2013.05.002

Nikolić, M., & Teodorović, D. (2014). A simultaneous transit network design and frequency setting: Computing with bees. *Expert Systems with Applications*, *41*(16), 7200–7209. https://doi.org/10.1016/j.eswa.2014.05.034

Noland, R. B. (2001). Relationships between highway capacity and induced vehicle travel. *Transportation Research Part A: Policy and Practice*, *35*(1), 47–72. https://doi.org/10.1016/s0965-8564(99)00047-6

Nurhidayat, A. Y., Widyastuti, H., Sutikno, & Upahita, D. P. (2023). Research on Passengers' Preferences and Impact of High-Speed Rail on Air Transport Demand. *Sustainability*, *15*(4), 3060. https://doi.org/10.3390/su15043060

Nurhidayat, A. Y., Widyastuti, H., & Utomo, D. P. (2018). Model of transportation mode choice between aircraft and high speed train of Jakarta-Surabaya route. *IOP Conference Series: Earth and Environmental Science*, *202*(1). https://iopscience.iop.org/article/10.1088/1755-1315/202/1/012002/pdf

Nurhidayat, A., Utomo, D., Fajar, R., Sucipto, & Widyastuti, H. (2019). Aircraft and High Speed Train Using the Logit Model A Case Study of The Jakarta-Surabaya Route. *Advances in Engineering Research*. https://doi.org/10.2991/apte-18.2019.16

Nvidia. (2024). *What is NetworkX?* NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/networkx/

ODS. (2024). *Airports code*. Opendatasoft.com. https://data.opendatasoft.com/explore/dataset/airports-code%40public/export/

OECD. (2022). *Gross domestic product (GDP)*. OECD. https://data.oecd.org/gdp/gross-domestic-product-gdp.htm

OECD. (2023). *Real GDP forecast*. OECD. https://data.oecd.org/gdp/real-gdp-forecast.htm

Oliker, N., & Bekhor, S. (2020). An infeasible start heuristic for the transit route network design problem. *Transportmetrica A: Transport Science*, *16*(3), 388–408. https://doi.org/10.1080/23249935.2020.1719551

Ollivier, G., Bullock, R., Jin, Y., & Zhou, N. (2014). High-Speed Railways in China: A Look at Traffic. In *World Bank*. https://documents1.worldbank.org/curated/en/451551468241176543/pdf/932270BRI0Box30ffic020140final000EN.pdf

Ortúzar, J. de D., & Simonetti, C. (2008). Modelling the demand for medium distance air travel with the mixed data estimation method. *Journal of Air Transport Management*, *14*(6), 297–303. https://doi.org/10.1016/j.jairtraman.2008.08.002

Outwater, M., Tierney, K., Bradley, M., Sall, E., Kuppam, A., & Modugula, V. (2010). California Statewide Model for High-Speed Rail. *Journal of Choice Modelling*, *3*(1), 58–83. https://doi.org/10.1016/s1755-5345(13)70029-0

Pagliara, F., La Pietra, A., Gomez, J., & Manuel Vassallo, J. (2015). High Speed Rail and the tourism market: Evidence from the Madrid case study. *Transport Policy*, *37*, 187–194. https://doi.org/10.1016/j.tranpol.2014.10.015

Pagliara, F., Vassallo, J. M., & Román, C. (2012). High-Speed Rail versus Air Transportation. *Transportation Research Record: Journal of the Transportation Research Board*, *2289*(1), 10–17. https://doi.org/10.3141/2289-02

Pan, & Truong. (2020). Low-Cost Carriers versus High-Speed Rail: Understanding Key Drivers of Passengers' Choice in China. *Transportation Journal*, *59*(1), 1. https://doi.org/10.5325/transportationj.59.1.0001

Park, Y., & Ha, H.-K. (2006). Analysis of the impact of high-speed railroad service on air transport demand. *Transportation Research Part E: Logistics and Transportation Review*, *42*(2), 95–104. https://doi.org/10.1016/j.tre.2005.09.003

Pattnaik, S. B., Mohan, S., & Tom, V. M. (1998). Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering*, *124*(4), 186–195. https://doi.org/10.1061/(asce)0733-947x(2003)129:2(186)

Patz, A. (1925). Die richtige Auswahl von Verkehrslinien bei großen Straßenbahnnetzen. *Verkehrstechnik*, *50*(51), 977–983.

Pazour, J. A., Meller, R. D., & Pohl, L. M. (2010). A model to design a national high-speed rail network for freight distribution. *Transportation Research Part A: Policy and Practice*, *44*(3), 119–135. https://doi.org/10.1016/j.tra.2009.11.006

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., & Ellison, J. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, *38*(3). sciencedirect. https://doi.org/10.1016/j.ijforecast.2021.11.001

Preston, J. (2013). *The economics of investment in high speed rail: Summary and conclusions*. Www.econstor.eu. https://hdl.handle.net/10419/97087

Preston, R. (2023, March 15). Europe Needs a High-Speed Masterplan. *International Railway Journal*. https://www.railjournal.com/in_depth/europe-needs-a-high-speed-masterplan/

Profillidis, V. A., & Botzoris, G. N. (2019). Trend Projection and Time Series Methods. *Modeling of Transport Demand*, 225–270. https://doi.org/10.1016/b978-0-12-811513-8.00006-6

Prov. N-Holland, & Hardt. (2020). *Role of hyperloop on the Development perspective "Compacte Metropool."* https://hardt.global/files/7815/8641/9870/ENGLISH_-_NH_results_one-pager.pdf

Qian, C., Siyuan, Q., & Qizhou, H. (2023). Study on Induced Passenger Flow Forecast for Intercity High-Speed Rail. *Transportation Research Record*. https://doi.org/10.1177/0361198123120928

Qin, J., Qu, W., Wu, X., & Zeng, Y. (2019). Differential Pricing Strategies of High Speed Railway Based on Prospect Theory: An Empirical Study from China. *Sustainability*, *11*(14), 3804. https://doi.org/10.3390/su11143804

Railtech. (2024, January 8). *Spain has Europe's most efficient high-speed rail network, says report*. RailTech.com. https://www.railtech.com/all/2024/01/08/spain-has-europes-most-efficient-high-speed-rail-network-says-report/

Rajendran, S., & Popfinger, M. (2022). Evaluating the substitutability of short-haul air transport by high-speed rail using a simulation-based approach. *Transportation Research Interdisciplinary Perspectives*, *15*(102194), 100632. https://doi.org/10.1016/j.trip.2022.100632

Ramesh Chirania, S. (2012). *Forecasting Model for High-Speed Rail in the United States* [Doctoral dissertation]. https://vtechworks.lib.vt.edu/server/api/core/bitstreams/0f5efebf-9a14-486c-85ae-b78ac2a2f18e/content

Ranjbari, A., Hickman, M., & Chiu, Y.-C. (2020). A network design problem formulation and solution procedure for intercity transit services. *Transportmetrica A: Transport Science*, *16*(3), 1156–1175. https://doi.org/10.1080/23249935.2020.1719547

RAVE. (2003). *Estudo da procura nos corredores da rede ferroviária de alta velocidade*. AT Kearney. [in Portugese].

Reeves, C. R. (1995). *Modern heuristic techniques for combinatorial problems*. Mcgraw-Hill.

Ren, X., Chen, Z., Wang, F., Dan, T., Wang, W., Guo, X., & Liu, C. (2020). Impact of high-speed rail on social equity in China: Evidence from a mode choice survey. *Transportation Research Part A: Policy and Practice*, *138*, 422–441. https://doi.org/10.1016/j.tra.2020.05.018

Rich, J., & Mabit, S. L. (2012). A Long-Distance Travel Demand Model for Europe. *European Journal of Transport and Infrastructure Research*, *12*(1). https://doi.org/10.18757/ejtir.2012.12.1.2946

Rocha, M., & Ferreira, P. G. (2018). Graphs: Concepts and Algorithms. *Elsevier EBooks*, 275–288. https://doi.org/10.1016/b978-0-12-812520-5.00013-4

Rodrigue, J.-P. (2017, December 28). *B.4 – High Speed Rail Systems | The Geography of Transport Systems*. The Geography of Transport Systems. https://transportgeography.org/contents/applications/high-speed-rail-systems/#:~:text=A%20distance%20of%2050%20km

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., & Tsang, F. (2010). Modelling long-distance travel in the UK. *Proceedings of the 38th European Transport Conference, Glasgow, Scotland. He Danish National Passenger Model–Model Specification and Results.* https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e35e1436db6733d909466dd514b26ea336dabddd

Román , C., & Martín, J. C. (2010). Potential demand for new high speed rail services in high dense air transport corridors. *International Journal of Substainable Development and Planning*. https://books.google.nl/books?hl=en&lr=&id=wowQePT4IHAC&oi=fnd&pg=PA35&ots=cheVc4VbiN&sig=IHW_-w83eqrfvhQz-D1Lk3mm334&redir_esc=y#v=onepage&q&f=false

Román, C., Espino, R., & Martín, J. C. (2007). Competition of high-speed train with air transport: The case of Madrid–Barcelona. *Journal of Air Transport Management*, *13*(5), 277–284. https://doi.org/10.1016/j.jairtraman.2007.04.009

Román, C., Espino, R., & Martín, J. C. (2010). Analyzing Competition between the High Speed Train and Alternative Modes. The Case of the Madrid-Zaragoza-Barcelona Corridor. *Journal of Choice Modelling*, *3*(1), 84–108. https://doi.org/10.1016/s1755-5345(13)70030-7

Rome2Rio. (2024, March 26). *Rome2rio: discover how to get anywhere*. https://www.rome2rio.com/

RSW. (2022, August 5). *CRRC is preparing for the first export delivery of high-speed trains to Indonesia*. ROLLINGSTOCK. https://rollingstockworld.com/passenger-cars/crrc-is-preparing-for-the-first-export-delivery-of-high-speed-trains-to-indonesia/

Russo, F., Sgro, D., & Musolino, G. (2023). Sustainable Development of Railway Corridors: Methods and Models for High Speed Rail (HSR) Demand Analysis. *Lecture Notes in Computer Science*, *14110*, 527–538. https://doi.org/10.1007/978-3-031-37123-3_36

Sadeghi, M., Aghayan, I., & Ghaznavi, M. (2020). A cuckoo search based approach to design sustainable transit network. *Transportation Letters*, 1–14. https://doi.org/10.1080/19427867.2020.1750767

Sánchez-Borràs, M., Nash, C., Abrantes, P., & López-Pita, A. (2010). Rail access charges and the competitiveness of high speed trains. *Transport Policy*, *17*(2), 102–109. https://doi.org/10.1016/j.tranpol.2009.12.001

Sane, Y. (2020). *Multi-Criteria Analysis of the proposed Hyperloop transport project in Northern Holland* [Thesis]. https://www.diva-portal.org/smash/get/diva2:1514003/FULLTEXT01.pdf

Shih, M.-C., Mahmassani, H. S., & Baaj, M. H. (1998). Planning and Design Model for Transit Route Networks with Coordinated Operations. *Transportation Research Record: Journal of the Transportation Research Board*, *1623*(1), 16–23. https://doi.org/10.3141/1623-03

Shilton, D. (1982). Modelling the Demand for High Speed Train Services. *The Journal of the Operational Research Society*, *33*(8), 713. https://doi.org/10.2307/2634319

Sinharay, S. (2010). An Overview of Statistics in Education. *International Encyclopedia of Education*, 1–11. https://doi.org/10.1016/b978-0-08-044894-7.01719-x

Smappen. (2024). *Map*. Smappen.com. https://www.smappen.com/app/map/D1xHgpKf6ylUN8bY

Sonntag, H. (1978). Linienplanung im öffentlichen Personennahverkehr. *Vorträge Der Jahrestagung*, *1977*, 430–439. https://doi.org/10.1007/978-3-662-00409-8_62

Statistica. (2018). *EU-28: high-speed rail transport 2000-2018 | Statista*. Statista; Statista. https://www.statista.com/statistics/279576/high-speed-rail-transport-in-eu-27/

Strauss, J., Li, H., & Cui, J. (2021). High-speed Rail's Impact on Airline Demand and Air Carbon Emissions in China. *Transport Policy*, *109*, 85–87. https://doi.org/10.1016/j.tranpol.2021.05.019

Su, M., Luan, W., Fu, X., Yang, Z., & Zhang, R. (2020). The competition effects of low-cost carriers and high-speed rail on the Chinese aviation market. *Transport Policy*, *95*, 37–46. https://doi.org/10.1016/j.tranpol.2020.05.025

Suman, H. K., & Bolia, N. B. (2019). Improvement in direct bus services through route planning. *Transport Policy*, *81*, 263–274. https://doi.org/10.1016/j.tranpol.2019.07.001

Szeto, W. Y., & Wu, Y. (2011). A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research*, *209*(2), 141–155. https://doi.org/10.1016/j.ejor.2010.08.020

Tetlow, G., & Pattison, J. (2023, October 5). *HS2: costs and controversies*. Institute for Government. https://www.instituteforgovernment.org.uk/explainer/hs2-costs

Tian, C., & Zhang, S. (2024). Study on the pattern of train arrival headway time in high-speed railway. *Railway Sciences*, *3*(3), 344–366. https://doi.org/10.1108/rs-03-2024-0007

Timperley, J. (2020, February 19). *Should we give up flying for the sake of the climate?* BBC. https://www.bbc.com/future/article/20200218-climate-change-how-to-cut-your-carbon-emissions-when-flying

Tirana Times. (2018, October 29). *Albania selects audit firm to probe airport's alleged abuse of monopoly position*. https://www.tiranatimes.com/?p=139054

Tom, V. M., & Mohan, S. (2003). Transit Route Network Design Using Frequency Coded Genetic Algorithm. *Journal of Transportation Engineering*, *129*(2), 186–195. https://doi.org/10.1061/(asce)0733-947x(2003)129:2(186)

Trabo, I., Landex, A., Nielsen, O. A., & Schneider-Tili, J. E. (2013). Cost benchmarking of railway projects in Europe – can it help to reduce costs? *5th International Seminar on Railway Operations Modelling and Analysis-RailCopenhagen*. https://backend.orbit.dtu.dk/ws/portalfiles/portal/106565055/RailCPH_010213.pdf

Trafikverket. (2021). *Travel Demand Review: A study performed within the government assignment regarding New Main Lines for high-speed trains 2020/2021*. https://www.almega.se/app/uploads/sites/9/2021/06/travel-demand-review-report-public-report.pdf

Transit Costs Project. (2024). *High Speed Rail Preliminary Data Analysis*. Transit Costs Project. https://transitcosts.com/high-speed-rail-preliminary-data-analysis/

UIC. (2002). *Infracost the cost of railway infrastructure*. Final report.

UIC. (2005). *Estimation des resources et des activités économiques liées a la grande vitesse*. CENIT (Center for Innovation in Transport, Universitat Politecnica de Catalunya).

UIC. (2018). *High Speed Rail: Fast Track To Sustainable Mobility*. https://www.uic.org/com/IMG/pdf/uic_high_speed_brochure.pdf

UIC High-Speed. (2018). *World High Speed Rolling Stock*. https://uic.org/IMG/pdf/20180910_highspeed_rolling_stock.pdf

UK Government. (2022). *Greenhouse gas reporting: conversion factors 2022*. GOV.UK. https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2022

Utomo, D. P., Lestari, P. D. R., Sucipto, S., & Nurhidayat, A. Y. (2020). IMPACT OF THE JAKARTA - SURABAYA HIGH - SPEED TRAIN ON PASSENGERS OF THE JAKARTA - SURABAYA EXECUTIVE TRAIN CORRIDOR. *Majalah Ilmiah Pengkajian Industri*, *14*(1), 1–10. https://doi.org/10.29122/mipi.v14i1.3897

van Nes, R., Hamerslag, R., & Immers, B. (1988). Design of Public Transport Networks. *Transportation Research Record*, *1202*, 74–83. https://onlinepubs.trb.org/Onlinepubs/trr/1988/1202/1202-010.pdf

Vision of Humanity. (2024). *A Closer Look at Russia, Ukraine and the Eurasia Region*. Vision of Humanity. https://www.visionofhumanity.org/russia-ukraine-the-eurasia-region/

Wan, Q. K., & Lo, H. K. (2003). A Mixed Integer Formulation for Multiple-Route Transit Network Design. *Journal of Mathematical Modelling and Algorithms*, *2*(4), 299–308. https://doi.org/10.1023/b:jmma.0000020425.99217.cd

Wan, Y., Ha, H.-K., Yoshida, Y., & Zhang, A. (2016). Airlines' reaction to high-speed rail entries: Empirical study of the Northeast Asian market. *Transportation Research Part A: Policy and Practice*, *94*, 532–557. https://doi.org/10.1016/j.tra.2016.10.014

Wang, C., Jiang, C., & Zhang, A. (2021). Effects of Airline Entry on High-Speed Rail. *Transportation Research Part B: Methodological*, *154*, 242–265. https://doi.org/10.1016/j.trb.2021.10.004

Wang, J., Jiao, J., Du, C., & Hu, H. (2015). Competition of spatial service hinterlands between high-speed rail and air transport in China: Present and future trends. *Journal of Geographical Sciences*, *25*(9), 1137–1152. https://doi.org/10.1007/s11442-015-1224-5

Wang, K., Jiang, C., Ng, A. K. Y., & Zhu, Z. (2020). Air and rail connectivity patterns of major city clusters in China. *Transportation Research Part A: Policy and Practice*, *139*, 35–53. https://doi.org/10.1016/j.tra.2020.07.002

Wang, K., Xia, W., Zhang, A., & Zhang, Q. (2018). Effects of train speed on airline demand and price: Theory and empirical evidence from a natural experiment. *Transportation Research Part B: Methodological*, *114*, 99–130. https://doi.org/10.1016/j.trb.2018.05.017

Wang, Q. (2011). *Travel Demand Forecasting with Stated Choice Data - Swedish Domestic Long-distance Trip* . https://www.diva-portal.org/smash/get/diva2:456644/FULLTEXT01.pdf

Wang, Y., Sun, L., Teunter, R. H., Wu, J., & Hua, G. (2020). Effects of introducing low-cost high-speed rail on air-rail competition: Modelling and numerical analysis for Paris-Marseille. *Transport Policy*, *99*, 145–162. https://doi.org/10.1016/j.tranpol.2020.08.006

Want China Times. (2013, November 10). *China's high-speed rail network reports staggering losses ｜Markets ｜Business ｜ WantChinaTimes.com*. Web.archive.org. https://web.archive.org/web/20131110133055/http://www.wantchinatimes.com/news-subclass-cnt.aspx?id=20130203000075&cid=1202

Wardman, M. (2006). Demand for rail travel and the effects of external factors. *Transportation Research Part E: Logistics and Transportation Review*, *42*(3), 129–148. https://doi.org/10.1016/j.tre.2004.07.003

Wardman, M., Chintakayala, P., de Jong, G., & Ferrer, D. (2012). *EUROPEAN WIDE META-ANALYSIS OF VALUES OF TRAVEL TIME: FINAL REPORT TO THE EUROPEAN INVESTMENT BANK* . https://significance.nl/wp-content/uploads/2019/03/2012-GDJ-European-wide-meta-analysis-of-values-of-travel-time.pdf

Wheeler, J. O. (2005). Geography. *Encyclopedia of Social Measurement*, 115–123. https://doi.org/10.1016/b0-12-369398-5/00277-2

World Bank. (2023). *GDP growth (annual %)*. World Bank Open Data. https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

WPR. (2024). *Population of Cities in Germany (2024)*. Worldpopulationreview.com. https://worldpopulationreview.com/countries/cities/germany

Xia, W., & Zhang, A. (2017). Air and high-speed rail transport integration on profits and welfare: Effects of air-rail connecting time. *Journal of Air Transport Management*, *65*, 181–190. https://doi.org/10.1016/j.jairtraman.2017.06.008

Xiong, Y., & Schneider, J. B. (1992). Transportation network design using a cumulative genetic algorithm and neural network. *Transportation Research Record*, *1364*.

Yan, S., & Chen, H.-L. (2002). A scheduling model and a solution algorithm for inter-city bus carriers. *Transportation Research Part A: Policy and Practice*, *36*(9), 805–825. https://doi.org/10.1016/s0965-8564(01)00041-6

Yang, H., Burghouwt, G., Wang, J., Boonekamp, T., & Dijst, M. (2018). The implications of high-speed railways on air passenger flows in China. *Applied Geography*, *97*, 1–9. https://doi.org/10.1016/j.apgeog.2018.05.006

Yang, H., Dobruszkes, F., Wang, J., Dijst, M., & Witte, P. (2018). Comparing China's urban systems in high-speed railway and airline networks. *Journal of Transport Geography*, *68*, 233–244. https://doi.org/10.1016/j.jtrangeo.2018.03.015

Yang, H., & Zhang, A. (2012). Effects of high-speed rail and air transport competition on prices, profits and welfare. *Transportation Research Part B: Methodological*, *46*(10), 1322–1333. https://doi.org/10.1016/j.trb.2012.09.001

Yang, J., & Jiang, Y. (2020). Application of Modified NSGA-II to the Transit Network Design Problem. *Journal of Advanced Transportation*, *2020*, 1–24. https://doi.org/10.1155/2020/3753601

Yang, W., Chen, Q., & Yang, J. (2022). Factors Affecting Travel Mode Choice between High-Speed Railway and Road Passenger Transport—Evidence from China. *Sustainability*, *14*(23). https://doi.org/10.3390/su142315745

Yang, Z., Yu, B., & Cheng, C. (2007). A Parallel Ant Colony Algorithm for Bus Network Optimization. *Computer-Aided Civil and Infrastructure Engineering*, *22*(1), 44–55. https://doi.org/10.1111/j.1467-8667.2006.00469.x

Yao, B., Hu, P., Lu, X., Gao, J., & Zhang, M. (2014). Transit network design based on travel time reliability. *Transportation Research Part C: Emerging Technologies*, *43*, 233–248. https://doi.org/10.1016/j.trc.2013.12.005

Yao, E., & Morikawa, T. (2005). A study of on integrated intercity travel demand model. *Transportation Research Part A: Policy and Practice*, *39*(4), 367–381. https://doi.org/10.1016/j.tra.2004.12.003

Yoon, G., & Chow, J. Y. J. (2020). Contextual Bandit-Based Sequential Transit Route Design under Demand Uncertainty. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(5), 613–625. https://doi.org/10.1177/0361198120917388

Yu, B., Yang, Z., Cheng, C., & Liu, C. (2005). Optimizing bus transit network with parallel ant colony algorithm. *Roceedings of the Eastern Asia Society for Transportation Studies*, *5*, 374–389.

Yu, B., Yang, Z.-Z., Jin, P.-H., Wu, S.-H., & Yao, B.-Z. (2012). Transit route network design-maximizing direct and transfer demand density. *Transportation Research Part C: Emerging Technologies*, *22*, 58–75. https://doi.org/10.1016/j.trc.2011.12.003

Yu, D., Zhang, Y., Wu, X., Li, D., & Li, G. (2021). The varying effects of accessing high-speed rail system on China's county development: A geographically weighted panel regression analysis. *Land Use Policy*, *100*, 104935. https://doi.org/10.1016/j.landusepol.2020.104935

Yu, K., Strauss, J., Liu, S., Li, H., Kuang, X., & Wu, J. (2021). Effects of railway speed on aviation demand and CO2 emissions in China. *Transportation Research Part D: Transport and Environment*, *94*, 102772. https://doi.org/10.1016/j.trd.2021.102772

Yue, Y., Chen, J., Feng, T., Ma, X., Wang, W., & Bai, H. (2023). Classification and determinants of high-speed rail stations using multi-source data: A case study in Jiangsu Province, China. *Sustainable Cities and Society*, *96*(104640), 104640–104640. https://doi.org/10.1016/j.scs.2023.104640

Zeng, N., & Wang, Z. (2010). Analysis of Beijing-guangzhou high-speed railway competitiveness based on generalized travel cost model. *IOP Conference Series: Earth and Environmental Science*, *587*(1). https://iopscience.iop.org/article/10.1088/1755-1315/587/1/012099/pdf

Zhang, F., Graham, D. J., & Wong, M. S. C. (2018). Quantifying the substitutability and complementarity between high-speed rail and air transport. *Transportation Research Part A: Policy and Practice*, *118*, 191–215. https://doi.org/10.1016/j.tra.2018.08.004

Zhang, J., & Zhang, J. (2021). Comprehensive Evaluation of Operating Speeds for High-Speed Railway: A Case Study of China High-Speed Railway. *Mathematical Problems in Engineering*, *2021*, 1–16. https://doi.org/10.1155/2021/8826193

Zhang, L., & Lu, Y. (2015). *US National and Inter-Regional Travel Demand Analysis: Person-Level Microsimulation Model and Application to High-Speed Rail Demand Forecasting*. https://mti.umd.edu/sites/mti.umd.edu/files/documents/NTC2014-SU-R-21%20Lei%20Zhang.pdf. No. Project ID: NTC2014-SU.

Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A. (2012). Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review. *Transport Reviews*, *32*(4), 399–433. https://doi.org/10.1080/01441647.2012.688174

Zhang, L., Yang, H., Wu, D., & Wang, D. (2014). Solving a discrete multimodal transportation network design problem. *Transportation Research Part C: Emerging Technologies*, *49*, 73–86. https://doi.org/10.1016/j.trc.2014.10.008

Zhang, Q., Yang, H., & Wang, Q. (2017). Impact of high-speed rail on China's Big Three airlines. *Transportation Research Part A: Policy and Practice*, *98*, 77–85. https://doi.org/10.1016/j.tra.2017.02.005

Zhang, Q., Yang, H., Wang, Q., & Zhang, A. (2014). Market power and its determinants in the Chinese airline industry. *Transportation Research Part A: Policy and Practice*, *64*, 1–13. https://doi.org/10.1016/j.tra.2014.03.003

Zhang, Q., Yang, H., Wang, Q., Zhang, A., & Zhang, Y. (2020). Impact of high-speed rail on market concentration and Lerner index in China's airline market. *Journal of Air Transport Management*, *83*, 101755. https://doi.org/10.1016/j.jairtraman.2019.101755

Zhang, W. (2024, January 3). *China: passenger transport volume of highspeed rail*. Statista. https://www.statista.com/statistics/1120071/china-passenger-transport-volume-of-highspeed-rail/

Zhao, F. (2006). Large-Scale Transit Network Optimization by Minimizing User Cost and Transfers. *Journal of Public Transportation*, *9*(2), 107–129. https://doi.org/10.5038/2375-0901.9.2.6

Zhao, F., & Gan, A. (2003). *Optimization of transit network to minimize transfers*. https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/research/reports/fdot-bd015-02rpt.pdf

Zhao, F., & Ubaka, I. (2004). Transit Network Optimization - Minimizing Transfers and Optimizing Route Directness. *Journal of Public Transportation*, *7*(1), 63–82. https://doi.org/10.5038/2375-0901.7.1.4

Zhao, F., & Zeng, X. (2006). Simulated Annealing–Genetic Algorithm for Transit Network Optimization. *Journal of Computing in Civil Engineering*, *20*(1), 57–68. https://doi.org/10.1061/(asce)0887-3801(2006)20:1(57)

Zhao, F., & Zeng, X. (2007). Optimization of User and Operator Cost for Large-Scale Transit Network. *Journal of Transportation Engineering*, *133*(4), 240–251. https://doi.org/10.1061/(asce)0733-947x(2007)133:4(240)

Zhong, C., Bel, G., & Warner, M. E. (2014). High-speed rail accessibility: a comparative analysis of urban access in Los Angeles, San Francisco, Madrid, and Barcelona. *European Journal of Transport and Infrastructure Research*, *14*(4), 473–495. https://journals.open.tudelft.nl/ejtir/article/download/3048/3239/8431

Zhou, Y., Yang, H., Wang, Y., & Yan, X. (2021). Integrated line configuration and frequency determination with passenger path assignment in urban rail transit networks. *Transportation Research Part B: Methodological*, *145*, 134–151. https://doi.org/10.1016/j.trb.2021.01.002

Appendices

AI-generated high-speed train (Alam, n.d.)

Appendices

A. TNDFSP Literature

Overviews of studies using different network design problems are provided by Guihaire & Hao (2008), Ibarra-Rojas et al. (2015), Chen et al. (2011) and Farahani et al. (2013). In practice, these problems are applied to urban public transport only. In total, 21 TNDP and 23 TNDFSP-related works were found with at least some relation to this work, highlighted in green in Table 0.1 below. It serves as a preliminary step in determining closely-related works, as presented in Table 1.1 in section 1.2.3.

Table 0.1 Complete overview of works using related network design problems. Related parts of the work are highlighted in green.

Reference	Constraints	Objective(s)	Solution method				Real case
			E	H	N	A	
TNDP studies							
Patz (1925)	capacity, demand	min. number of empty seats		✓			
Sonntag (1978)	restricted set of possible lines	- min. average travel time - min. number of transfers		✓			
Mandl (1980)	- constant frequency - area coverage	- min. travel time - max. route directness		✓			
Xiong & Schneider (1993)		- min. total travel time - min. construction cost				✓	
Chakroborty & Dwivedi (2002)	route feasibility	- min. total travel time - min. unsatisfied demand - max. 2-transfer passengers				✓	
Chen et al. (2003) Chen et al. (2006)		max. expected profit, social welfare	✓				
Guan (2003)	- transit capacity - number of transfers per OD pair - line length	- min. total line length - min. total number of lines taken - min. total travel length	✓				✓
Murray (2003)	- service coverage - access	- max. number of stops	✓				✓
Zhao & Gan (2003) Zhao & Ubaka (2004) Zhao (2006)	- predefined routes and areas - number of lines and stops - route length and network directness - deviation from main routes	- min. number of transfers - max. route directness - max. area coverage		✓			
Yu et al. (2005)	- line length - route directness	- min. number of transfers - max. passenger flow / route				✓	✓
Guan et al. (2006)	- link capacity - line length - number of transfers	- min. total route length - min. number of passenger routes - min. total travelled distance	✓				✓
Zhao & Zeng (2006)	- route directness - route feasibility - number of routes - route length - budget	- min. average number of transfers - max. service coverage		✓			✓
Chen & Subprasom (2007)		max. expected profit, social welfare, equity	✓				
Barra et al. (2007)	- travel demand satisfaction - budget - service level	min. total route length	✓				
Yang et al. (2007)	route length	max. number of direct travellers / length					
Mauttone & Urquhart (2009)	demand	- min. number of routes - min. total travel time		✓			
Fan & Mumford (2010)	number of lines	- min. total travel time - min. number of transfers		✓			
Curtin & Biba (2011)	route length	max. sum of arc and node service value		✓			✓

***Table 0.1* continued**

TNDFSP studies							
Hasselström (1979) Hasselström (1981)	budget	- min. number of transfers - max. number of passengers	✓				
Ceder & Wilson (1986)	- frequency - fleet size - route length	- min. excess travel, transfer & waiting time - min. vehicle costs		✓			
Van Nes et al. (1988)	fleet size	- max. demand satisfaction - max. number of direct trips	✓				
Bussieck (1998)	- number of transit vehicles - frequency - line and vehicle capacity	- max. number of direct passengers - min. operator costs	✓				
Pattnaik et al. (1998)	- headway - load factor	- min. operator costs - min. travel time				✓	✓
Bielli et al. (2002)	pre-defined lines	max. 24 network performance criteria				✓	✓
Carrese & Gori (2002)	- demand - route length - number of transfers - total travel time - fleet size	- min. user waiting & excess time (compared to minimum path) - min. operator costs		✓			✓
Fusco et al. (2002)	- level of service - demand - lines configuration - frequency - route length	min. overall cost		✓			
Ceder (2003)	- route length - deviation from shortest path	- min. operator and user costs		✓			
Tom & Mohan (2003)		- min. operator costs - min. total travel time				✓	
Wan & Lo (2003)	- line frequency - capacity	min. operator costs	✓				
Agrawal & Matthew (2004)	- line frequency - load factor	- min. operator and user costs	✓				✓
Fan and Machemehl (2004) Fan and Machemehl (2006a) Fan and Machemehl (2006b)	route length	- min. waiting, traveling, walking time - min. fleet size - min. cost of unsatisfied demand			✓	✓	
Hu et al. (2005)	- route length - average transfer, stop, headway times	- max. nonstop passenger flow - min. operator and user costs				✓	✓
Zhao (2006)	route directness	- min. number of transfers - max. demand coverage		✓			
Zhao & Zeng (2007)	- headway - fleet size - route length - load factor	- min. weighted sum of operator and user costs		✓			
Borndörfer et al. (2005) Borndörfer et al. (2008)	demand	- min. operator costs - min. total travel time				✓	
Szeto & Wu (2011)	- fleet size - number of stops - frequency - route length	- min. weighted sum of transfers & travel time				✓	✓
Cipriani et al. (2012)	- capacity - frequency - route length	- min. sum of operator and user costs		✓		✓	✓
This work	- demand - travel time - constant frequency - level of service - route/line feasibility & length - frequency - pre-defined lines	- max. operating profit (or min. operator costs, min. overall cost, max. profit) - min. number of transfers	✓				✓

Problem: ND (Transit Network Design), FS (Transit Network Design & Frequency Setting)
Methods: E (Exact), H (Heuristic), N (Neighbourhood Search), A ((Evolutionary) Algorithm)

B. Demand Forecasting

For this project, the web was searched for papers attempting to forecast high-speed rail demand, or attempting to forecast demand for other modes while including HSR. In total, exactly 100 studies were found to meet the mentioned requirements. These papers form the basis of the literature reviews in section 2.1 and section 2.2. An overview of these papers can be found here.

Table 0.2 Overview of 100 papers dedicated to high-speed rail demand forecasting, listed with the demand-impacting factors, models used and location of the case study.

Reference	Model	Case Study	Factors
(Albalate et al., 2015)	regression (GLS-random effects)	Europe	fare, distance, seat cost, population, GDP
(Ashiabor et al., 2007)	logit (nested, mixed)	USA	travel time, travel cost, household income
(Behrens & Pels, 2012)	logit (multinomial, mixed)	UK, France	fare, distance, frequency, travel time
(Ben-Akiva et al., 2010)	logit (nested), regression (linear), formula calibration	Italy	travel time, travel cost, access/egress time, frequency
(Bergantino & Capozza, 2015)	regression (GLS-random effects)	Italy	fare, income
(Bergantino & Madio, 2020)	logit (multinomial)	Italy	in-vehicle time, access / exit time, reliability, price, frequency
(Börjesson, 2014)	logit (nested)	Sweden	fare, travel time, income
(Brand et al., 1992)	logit (multinomial)	USA	travel time (in-vehicle, wait), transfers, travel cost, income
(Burge et al., 2010)	choice model (stated preference)	UK	population, income, travel time, travel cost, frequency
(Cabanne, 2003)	time series (direct demand, generation/modal split)	France	(access, waiting, in-vehicle) time, punctuality, frequency, interchanges, travel cost
(Capozza, 2016)	regression (linear)	Italy	GDP, travel cost, travel time
(Cartenì et al., 2017)	logit (binomial)	Italy	city attractiveness, distance, time, frequency, ticket price
(Cascetta & Cartenì, 2014)	logit (binomial)	Italy	(access, egress, transfer, waiting, in-vehicle, total) time, ticket fare
(Cascetta & Coppola, 2011)	logit (nested), regression (linear)	Italy	travel time, travel cost, access/egress time, punctuality, frequency
(Cascetta et al., 2011)	logit (nested)	Italy	travel cost, travel time, access + egress time
(Castillo-Manzano et al., 2015)	regression (linear, dynamic)	Spain	population, number of air operations, air passengers, unemployment rate
(Chai et al., 2018)	regression (panel threshold)	China	frequency, GDP, population, pollution, speed, travel time
(Chen et al., 2019)	regression (panel, hierarchical)	China	distance, population, GDP
(Chen, 2010)	logit (multinomial, nested)	Sweden	travel cost, travel time, access, egress, in vehicle-time
(Chen, 2017)	regression (panel)	China	distance, population, GDP, hub status
(Chirania, 2012)	logit (Box-Cox)	USA	travel time, travel cost
(Clever & Hansen, 2008)	logit (nested)	Japan	frequency, transfers, fare, distance
(Clewlow et al., 2014)	regression (linear)	Europe	travel time, population, GDP, density, fuel price
(Couto & Graham, 2007)	regression (multiplicative heteroscedastic)	Worldwide	fare, GDP/capita, population
(Daly, 2010)	logit (nested)	UK	distance, travel cost
(Danapour et al., 2018)	logit (binomial)	Iran	ticket price, travel time, hospitality, convenience
(Dargay & Clark, 2012)	regression (linear, dynamic)	UK	income
(De Bok et al., 2010)	logit (multinomial)	Portugal	travel cost, travel time, frequency, accessibility, wait time
(Diez-Pisonero, 2012)	none (case studies)	Spain	transit time, comfort, fixed cost
(Dobruszkes et al., 2014)	regression (Tobit)	Europe	frequency, distance, travel time, population, GDP, number of air transit, seat availability
(Dobruszkes, 2011)	none (case studies)	Europe	fare, frequency, travel time
(Fröidh, 2005)	none (case studies)	Sweden	travel time, fare, frequency
(Fu et al., 2014)	logit (three-level, nested)	Japan	fare, travel time, distance, frequency, capacity
(Gaudry, 2008)	logit (Box-Cox)	Canada	fare, travel time, access time, frequency of service, income
(Gu & Wan, 2020)	regression (linear)	China	frequency, travel time, population, GDP, seat availability, welfare
(Gundelfinger-Casar & Coto-Millán, 2017)	gravity	Spain	fare, travel time

(Hensher, 1997)	logit (heteroskedastic extreme value)	Australia	travel time, frequency, fares
(Hong & Najmi, 2022)	logit (conditional choice)	USA	ticket price, travel time, frequency
(Hsu & Chung, 1997)	new (discrete choice model)	China	value of time, speed, distance, fares
(Inoue et al., 2015)	logit (nested)	Japan	fare, frequency, access time, egress time, travel time
(Jiménez & Betancor, 2012)	regression (linear)	Spain	number of air passengers, tourism, GDP, distance, time
(Jung & Yoo, 2014)	logit (multinomial, nested)	South Korea	fare, access time, travel time, frequency
(Kroes & Savelberg, 2019)	new (substitution model)	The Netherlands	travel time, fares, transfers
(Lee et al., 2016)	logit (mixed)	South Korea	frequency, travel cost, travel time, safety
(Leng et al., 2015)	gravity, logit (multinomial)	China	GDP, population, distance
(Li & Schmöcker, 2014)	time series (log-linear first-order moving average)	Taiwan	population, GDP
(Li & Sheng, 2016)	logit (multinomial)	China	access time, egress time, travel time, income, trip purpose
(Li et al., 2019)	regression (panel)	China	fare, frequency, distance, access time, population, GDP, internet usage
(Li et al., 2021)	regression (Bayesian binary logistic)	China	distance, travel cost, travel time, safety, comfort, punctuality, access time
(Liu et al., 2019)	regression (panel)	China, Japan	fare, population, GDP
(Lubis et al., 2019)	logit (multinomial)	Indonesia	travel time, price, frequency
(Ma et al., 2019)	regression (reduced-form)	China	fare, travel time, population, GDP
(Mahardika et al. 2021)	logit (mixed)	Indonesia	travel time, travel cost
(Mandel et al. 1994)	logit (linear, Box-Cox)	Germany	travel cost, travel time, frequency, trip distance, value of time
(Martín & Nombela, 2007)	gravity, logit (multinomial)	Spain	travel time, cost, frequency, distance, income, population
(Martínez et al., 2016)	data analysis	Spain	catchment area population with distance decay
(Miyoshi & Givoni, 2012)	formula fitting (logistic)	UK	air traffic demand, seat capacity, travel time
(Mizutani & Sakai, 2021)	regression (DID)	Japan	fare, distance, travel time, population, income
(Nelldal & Jansson, 2010)	Sampers, Samvips, Vips	Sweden	in-vehicle time
(Nurhidayat et al., 2018)	logit (binomial)	Indonesia	income, trip purpose, fare, travel time, mode to airport
(Nurhidayat et al., 2019)	logit (binomial)	Indonesia	fare, travel time
(Ortúzar & Simonetti, 2008)	regression (binary choice), logit (mixed, nested)	Chile	travel time, fare, comfort, service delay
(Outwater et al., 2010)	logit (multinomial, nested)	USA	employment, households, travel time, distance, travel cost, VOT, frequency, reliability, income
(Pagliara & Vassallo, 2012)	logit (multinomial)	Spain	cost, frequency
(Pagliara et al., 2015)	regression (logistic)	Spain	price, accessibility, frequency, safety, comfort
(Pan & Truong, 2020)	regression (logistic)	China	frequency, price, attitude
(Park & Ha, 2006)	logit (multinomial)	South Korea	fare, frequency, distance
(Qian et al., 2023)	logit (multinomial)	China	transit time
(Ren et al., 2020)	logit (binomial, multinomial)	China	distance, income
(Rich & Mabit, 2012)	logit (nested)	Europe	travel cost, travel time
(Rohr et al., 2010)	logit (nested)	UK	cost, travel time, frequency, wait time, interchanges, income
(Román & Martín, 2010)	logit (multinomial, nested)	Spain	travel time, travel cost, headway, access + egress time, waiting time, reliability
(Román et al., 2007)	logit (nested)	Spain	travel cost, travel time, access + egress time, headway, reliability, comfort, waiting time
(Román et al., 2010)	logit (nested)	Spain	access time, waiting time, in-vehicle time, egress time, travel cost, headway/frequency, reliability, comfort, income
(Sánchez-Borràs et al. 2010)	econometric	Europe	ticket prices
(Shilton, 1982)	gravity (adaption)	UK	population, distance, competing modes, socio-economics
(Strauss et al., 2021)	gravity	China	price, frequence, distance, population, GDP
(Su et al., 2020)	regression	China	frequency, distance
(Utomo et al., 2020)	logit (binomial)	Indonesia	fare, travel time
(Wan et al., 2016)	regression (DID)	China, Japan, South Korea	population, GDP/capita, distance
(Wang et al., 2018)	regression (DID)	China	speed, distance, population, income, tourism status, travel time, safety
(Wang et al., 2021)	vertical differentiation	China	fare, frequency, distance
(Wang, 2011)	logit (binomial)	Sweden	travel time, travel cost
(Wang, Jiang et al., 2020)	connectivity utility model	China	population, GDP

(Wang, Sun et al., 2020)	vertical differentiation	France	fare, frequency, travel time, seat cost, value of time
(Wardman, 2006)	regression (weighted least squares)	UK	GDP, travel time, travel costs, population
(Xia & Zhang, 2017)	vertical differentiation	China	access time, egress time, travel time, service
(Yang & Zhang, 2012)	competition model	China	access time, distance, speed, fixed cost, seat cost, value of time, operating hours
(Yang et al., 2022)	logit (binomial)	China	safety, comfort, fares, convenience
(Yang, Burghouwt, et al., 2018)	regression (panel)	China	frequency, ticket fare, travel time, population, GDP
(Yang, Dobruszkes, et al., 2018)	regression (linear, multiple)	China	GDP, population, distance
(Yao & Morikawa, 2005)	logit (multinomial)	Japan	travel cost, travel time (access / egress / line), (working) population, (non-)business attractiveness, industrial structure, GDP/capita, frequency
(Yu et al., 2021)	gravity (adaption)	China	GDP, population, industrial structure, distance, speed, income
(Zeng & Wang, 2020)	logit (adapted, improved)	China	fare, frequency, travel time, comfort
(Zhang & Lu, 2015)	vertical differentiation	USA	travel cost, travel time
(Zhang et al., 2014)	Lerner index	China	distance, number of air passengers, frequency, population, GDP
(Zhang et al., 2017)	regression (linear, multiple)	China	fare, frequency, distance, travel time, population, GDP
(Zhang et al., 2018)	regression (DID)	East Asia	population, GDP, access, distance, number of air passengers
(Zhang et al., 2020)	regression (panel)	China	distance, population, income, tourism factor
(Zhong et al., 2014)	data analysis	Worldwide	population, density, employment, income

Logit parameters

These were attained the following 57 different models by fifteen studies, and used to find good estimates of β^{TT} and β^{TC}. They were used as part of the results in section 4.1. Details about each of the models were stated here as well. The studies all originate from the long list of 100 studies in Table 0.2.

Table 0.3 Models and studies used for gathering logit parameters

Reference	Location	Org. units	#Obs	R^2	β^{TT} $[h]$	β^{TC} [2024 €]	VoT [€/h]	Note
(Behrens & Pels, 2012)	UK, France	min, 1995 £	9,470	-	-0.82800	-0.00152	543.84	business travellers
			18,536	-	-0.34800	-0.00574	60.61	leisure travellers
			9,470	-	-1.09200	-0.00162	672.11	business travellers
			18,536	-	-0.93600	-0.00408	229.59	leisure travellers
(Brand et al., 1992)	USA	h, 1990 $	-	-	-1.34440	-0.01724	77.98	business air travellers
			-	-	-1.72300	-0.02770	62.20	nonbusiness air travellers
			-	-	-0.56360	-0.01287	43.78	business car travellers
			-	-	-0.28170	-0.01460	19.29	nonbusiness car travellers
(Chen, 2010)	Sweden	min, 2006 SEK	12,048	0.483	-0.26760	-0.01832	14.61	
			12,048	0.530	-0.12600	-0.02205	5.72	
			12,048	0.558	-0.16860	-0.01919	8.79	
			12,048	0.607	-0.19560	-0.01316	14.86	
			12,048	0.611	-1.08600	-0.06170	17.60	
			12,048	0.234	-0.41364	-0.01927	21.47	
(de Bok et al., 2010)	Portugal	h, 2007 €	5,176	0.486	-0.01460	-0.08362	0.17	commuting travellers
			5,176	0.491	-0.01470	-0.08503	0.17	commuting travellers
			5,176	0.502	-0.01200	-0.01289	0.93	commuting travellers
			5,176	0.508	-0.01180	-0.01864	0.63	commuting travellers
			3,246	0.822	-0.01570	-0.05286	0.30	business travellers
			3,246	0.822	-0.01600	-0.05286	0.30	business travellers
			3,246	0.821	-0.01250	-0.02401	0.52	business travellers
			3,246	0.821	-0.01120	-0.03471	0.32	business travellers
			13,370	0.475	-0.00554	-0.05768	0.10	other travellers
			13,370	0.475	-0.00665	-0.05959	0.11	other travellers
			13,370	0.482	-0.00628	-0.03023	0.21	other travellers
			13,370	0.481	-0.00738	-0.04270	0.17	other travellers
(Fu et al., 2014)	Japan	h, 2005 (x100) $	901	-	-0.44270	-0.00007	6168.86	
(Jung & Yoo, 2014)	South Korea	min, 2012 ₩	3,534	-	-0.26400	-0.02459	10.74	business travellers, MNL
			3,534	-	-0.25200	-0.03200	7.88	nonbusiness travellers, MNL
			3,534	-	-0.16800	-0.02344	7.17	business travellers, ML
			3,534	-	-0.36000	-0.04576	7.87	nonbusiness travellers, NL
(Li & Sheng, 2016)	China	min	1,128	0.213	-0.96180			
			1,128	0.206	-0.83820			
			1,128	0.214	-0.73560			
(Lubis et al., 2019)	Indonesia	h, 2016 IDR	402	0.107	-0.71700			
			403	0.121	-0.71700			
			404	0.127	-0.71700			
			405	0.128	-0.71700			
(Mandel et al. 1994)	Germany	h, 1980 DM	62,982	0.314	-0.63900			
			62,982	0.492	-1.00600			
(Martín & Nombela, 2007)	Spain	h, 2010 €	143	0.510	-0.36450	-0.00241	151.19	
			40	0.292	-0.46060	-0.02478	18.59	
(Outwater et al., 2010)	USA	min, 2000 $	1,500	0.276	-3.60000	-0.12490	28.82	access models / business-commute
			2,724	0.365	-1.80000	-0.19984	9.01	access models / recreation-other
			1,466	0.075	-3.60000	-0.12490	28.82	egress models / business-commute
			2,668	0.231	-1.80000	-0.19984	9.01	egress models / recreation-other
			2,198	0.380	-1.08000	-0.02831	38.15	long trip / business-commute
			5,075	0.309	-0.66000	-0.05829	11.32	long trip / recreation-other
(Pagliara & Vassallo, 2012)	Spain	2010 €	1,011	0.334		-0.09322		
			1,011	0.340		-0.13796		
			1,011	0.343		-0.15001		
(Park & Ha, 2006)	South Korea	2003 ₩	829	0.203		-0.00010		
(Román & Martín, 2010)	Spain	min, 2004 €	2,917	-	-0.28200	-0.08615	3.27	
(Yao & Morikawa, 2005)	Japan	h, 2000 (x10000) JPY	18,798	0.495	-2.13600	-0.30413	7.02	business SP/RP
			18,798	0.495	-2.13600	-0.30413	7.02	business aggregate
			32,202	0.456	-0.56200	-0.22330	2.52	non-business SP/RP
			32,202	0.456	-0.56200	-0.22330	2.52	non-business aggregate

Ordinary Least Squares

In OLS, a linear relationship between the dependent variable Y and its factor of influence X is assumed. The method therefore considers a formula of the following shape to be fitted to the observed data (Malonda & Carles, 2003):

$$Y_i = \alpha + \beta X_i + e_i \tag{0.1}$$

Here, each unique observation is indicated by an index i. For simplicity, the equation above only considers one factor of influence. The α-factor represents the predicted value when the values of all impact factors is set to zero. β is the coefficient belonging to the impact factor and characterises the relationship between the impact factor and the variable it is attempting to forecast. The coefficient is multiplied by the value of that impact factor for an observation X_i. OLS ensures the average difference between observed and predicted values is set to zero. However, this gives no further information about the accuracy of this formula in absolute terms. In reality, the predicted outcome will deviate from the observed value. Therefore, an error term e_i is added in equation (0.1).

To approach reality, the value of these error terms for all observations must remain as small as possible. Let's take a closer look at this error term by rewriting equation (0.1):

$$e_i = Y_i - (\alpha + \beta X_i) \tag{0.2}$$

OLS minimises the sum of squared error terms. The necessity of squaring the terms comes from the fact that an error is seen as a nonnegative number; it is a measure of the distance between observed and predicted data (Malonda & Carles, 2003). This must be minimised. The squaring is performed to turn negative values into positives, thus seeing equivalence between a -3 and +3 error, as it is both squared to 9. Both α and β are chosen to minimise the sum of squares SS, as indicated in equation (0.3).

$$SS = \sum_{i=1}^{n} e_i^2 = (Y_i - \alpha - \beta X_i)^2 \tag{0.3}$$

Then, these values are substituted in for equation (0.3), which then becomes the linear trendline best fitting to the data. Quantifying the accuracy of this equation and its parameters commonly is done by the following few indicators, and also will be used to display calibration results in this project:

- **Standard deviation**¸ which can be calculated for each parameter estimated. It is the average deviation of the value from the mean, when the formula is applied to the data. Higher standard deviations indicate a greater uncertainty of the parameter's value. To be able to make fair comparisons between the different parameters, this is often scaled to the estimated value of the parameter by means of the t-test.
- **T-test**, which is calculated by dividing the parameters estimate by its standard deviation. The absolute value of the outcome is a measure to grasp the relative significance or magnitude of the parameter estimate in relation to its variability. Higer values of the t-test indicate a stronger significance, and increase the probability that the observed relationship is real and not due to random chance.
- **P-value**, which is calculated directly from the value of the t-test. The outcome is a number between 0 and 1. Lower values indicate a higher significance level; for example, a level of 95% is indicated by a p-value of 0.05. In studies, the this is usually considered most important when it comes to assessing the statistical significance of estimated parameters.
- To assess the fit of the trendline, $\boldsymbol{R^2}$ is considered the standard. It measures how well the model approximate the actual data. It is calculated by means of the following equation:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} \tag{0.4}$$

Here, Y_i and $\hat{Y}_i$ represent the observed and predicted value of observation i, respectively. The average of all these observations is indicated by $\bar{Y}$. The outcome R^2 will take values between 0 and 1, which represents the share of variation in the real data that is explained by the model (Coker, 1995). Thus, a higher R^2-value corresponds to a better model fit.

C. Air Passengers Data Set

Data sets used

Table 0.4 lists all Eurostat databases used for this project, along with their names. These data sets provided the air passenger data, needed to calibrate the model defined in chapter 3.

Table 0.4 Air passenger databases used (Eurostat, 2024a)

Name	Country	No. Flights	Years available
avia_par_at	Austria	291	1993-2022
avia_par_be	Belgium	319	1993-2022
avia_par_ba	Bosnia and Herzegovina	25	2021
avia_par_bg	Bulgaria	191	2007-2023
avia_par_hr	Croatia	254	2008-2023
avia_par_cy	Cyprus	157	2001-2023 excl. '02
avia_par_cz	Czechia	75	2002-2022
avia_par_dk	Denmark	293	1993-2022 excl. '00
avia_par_ee	Estonia	62	2001-2022
avia_par_fi	Finland	207	1997-2023
avia_par_fr	France	1915	1993-2022
avia_par_de	Germany	2160	1993-2022
avia_par_el	Greece	1015	1993-2022 excl. '01-'02
avia_par_hu	Hungary	129	2001-2022
avia_par_is	Iceland	83	2003-2023
avia_par_ie	Ireland	285	1993-2022
avia_par_it	Italy	2131	1993-2022
avia_par_lv	Latvia	90	2001-2023
avia_par_lt	Lithuania	165	2003-2023
avia_par_lu	Luxembourg	69	1993-2023
avia_par_mt	Malta	94	2001-2023
avia_par_me	Montenegro	50	2016-2022
avia_par_mk	North Macedonia	48	2015-2022
avia_par_no	Norway	439	1999-2023
avia_par_pl	Poland	758	2004-2022
avia_par_pt	Portugal	476	1993-2022
avia_par_ro	Romania	374	2001-2022
avia_par_rs	Serbia	84	2016-2022
avia_par_sk	Slovakia	109	2001-2023
avia_par_si	Slovenia	33	2004-2022
avia_par_es	Spain	1967	1993-2022
avia_par_se	Sweden	478	1993-2022
avia_par_ch	Switzerland	350	1993-2022
avia_par_nl	The Netherlands	502	1993-2023
avia_par_tr	Türkiye	1172	2012-2022
avia_par_uk	United Kingdom	2282	1993-2019

Scope selection

The table below lists all European countries and their rationale behind being included or not into the scope, which has been defined in section 1.5:

Table 0.5 Reasoning of area scope inclusion per country

European Country	ISO-3	Continental Europe	Included?	Reason of different choice
Albania	ALB	yes	yes	-
Andorra	AND	yes	yes	-
Austria	AUT	yes	yes	-
Belgium	BEL	yes	yes	-
Bosnia and Herzegovina	BIH	yes	yes	-
Bulgaria	BGR	yes	yes	-
Croatia	HRV	yes	yes	-
Czechia	CZE	yes	yes	-
Denmark	DNK	yes	yes	-
Estonia	EST	yes	yes	-
Finland	FIN	yes	yes	-
France	FRA	yes	yes	-
Germany	DEU	yes	yes	-
Greece	GRC	yes	yes	-
Hungary	HUN	yes	yes	-
Italy	ITA	yes	yes	-
Kosovo	XKK	yes	yes	-
Latvia	LVA	yes	yes	-
Liechtenstein	LIE	yes	yes	-
Lithuania	LTU	yes	yes	-
Luxembourg	LUX	yes	yes	-
Monaco	MCO	yes	yes	-
Montenegro	MNE	yes	yes	-
Netherlands	NLD	yes	yes	-
North Macedonia	MKD	yes	yes	-
Norway	NOR	yes	yes	-
Poland	POL	yes	yes	-
Portugal	PRT	yes	yes	-
Romania	ROU	yes	yes	-
San Marino	SMR	yes	yes	-
Serbia	SRB	yes	yes	-
Slovakia	SVK	yes	yes	-
Slovenia	SVN	yes	yes	-
Spain	ESP	yes	yes	-
Sweden	SWE	yes	yes	-
Switzerland	CHE	yes	yes	-
Turkiye	TUR	yes (partly)	yes (partly)	-
United Kingdom	GBR	no	yes	Eurotunnel connection
Vatican City	VAT	yes	yes	-
Armenia	ARM	no	no	-
Azerbaijan	AZE	partly	no	military conflict (Vision of Humanity, 2024)
Belarus	BLR	yes	no	military conflict
Cyprus	CYP	no	no	-
Georgia	GEO	partly	no	not connected to rest of scope
Iceland	ISL	no	no	-
Ireland	IRL	no	no	-
Kazakhstan	KAZ	partly	no	not connected to rest of scope
Malta	MLT	no	no	-
Moldova	MDA	yes	no	military conflict
Russia	RUS	partly	no	military conflict
Ukraine	UKR	yes	no	military conflict

Cities with multiple airports

The following cities are served by multiple airports. This includes only airports available in the data set that fulfil the scope requirements, and after preprocessing. These airports are assigned a new city-specific code, so that flight data serving the same origin and destination can be combined. Other airports keep their old code. This table is updated for the same year as the flight data used (2019).

Table 0.6 Cities served by multiple airports (2019)

#Airports	City	IATA	Airport Name	Code
6	London, England	LHR*	Heathrow Airport	C-LON
		LGW	Gatwick Airport	
		STN	London Stansted Airport	
		LTN	London Luton Airport	
		LCY	London City Airport	
		SEN	London Southend Airport	
3	Paris, France	CDG*	Charles de Gaulle Airport	C-PAR
		ORY	Orly Airport	
		BVA	Beauvais-Tillé Airport	
3	Stockholm, Sweden	ARN*	Stockholm Arlanda Airport	C-STO
		NYO	Stockholm Skavsta Airport	
		BMA	Stockholm Bromma Airport	
3	Milan, Italy	LIN	Milan Linate Airport	C-MIL
		MXP*	Milan Malpensa Airport	
		BGY	Orio al Serio International Airport	
3	Barcelona, Spain	GRO	Girona-Costa Brava Airport	C-BAR
		BCN*	Josep Tarradellas Barcelona-El Prat Airport	
		REU	Reus Airport	
3	Istanbul, Turkiye	IST*	Istanbul Airport	C-IST
		ISL	Istanbul Atatürk Airport	
		SAW	Istanbul Sabiha Gökçen International Airport	
2	Warsaw, Poland	WAW*	Warsaw Chopin Airport	C-WAR
		WMI	Warsaw-Modlin Airport	
2	Brussels, Belgium	BRU*	Brussels National Airport	C-BRU
		CRL	Brussels South Charleroi Airport	
2	Rome, Italy	CIA	Ciampino–G. B. Pastine International Airport	C-ROM
		FCO*	Leonardo da Vinci–Fiumicino Airport	
2	Turin, Italy	CUF	Cuneo International Airport	C-TUR
		TRN*	Turin Airport	
2	Venice, Italy	TSF	Treviso Airport	C-VEN
		VCE*	Venice Marco Polo Airport	
2	Oslo, Norway	OSL*	Oslo Airport, Gardermoen	C-OSL
		TRF	Sandefjord Airport, Torp	
2	Glasgow, Scotland	GLA*	Glasgow Abbotsinch Airport	C-GLA
		PIK	Glasgow Prestwick Airport	
2	Berlin, Germany	BER	Berlin Schönefeld Airport	C-BER
		TXL*	Berlin Tegel Airport	
2	Frankfurt, Germany	FRA*	Frankfurt Airport	C-FRA
		HHN	Frankfurt-Hahn Airport	
2	Hamburg, Germany	HAM*	Hamburg Airport	C-HAM
		XFW	Hamburg Finkenwerder Airport	
2	Munich, Germany	MUC*	Munich Airport	C-MUN
		FMM	Memmingen Airport	

*biggest airport serving the city, is used for coordinates

Matchings

The table below lists all 200 urban centres that could be matched with an airport for which flight data is available. The urban centres are originating from the full 726-long list, the airports are from the demand forecasting data set.

Table 0.7 Matching of airports with urban centres

IATA	UC	IATA	UCS	IATA	UC	IATA	UCS
AAL	Aalborg	CRA	Craiova	LCG	A Coruña	RIX	Riga
AAR	Aarhus	C-ROM	Rome	LCJ	Lodz	RMI	Rimini
ABZ	Aberdeen	C-STO	Stockholm	LEI	Almeria	RMU	Murcia
ACH	Sankt Gallen	C-TUR	Turin	LEJ	Leipzig	RNS	Rennes
AGH	Helsingborg	C-VEN	Venice	LGG	Liège	RTM	Rotterdam [The Hague]
AGP	Málaga	C-WAR	Warsaw	LIG	Limoges	RZE	Rzeszów
ALC	Alacant / Alicante	CWL	Cardiff	LIL	Lille	SBZ	Sibiu
AMS	Amsterdam	DEB	Debrecen	LIS	Lisbon	SCN	Saarbruecken
ANR	Antwerp	DRS	Dresden	LJU	Ljubljana	SCV	Suceava
AOI	Ancona	DSA	Sheffield	LNZ	Linz	SDR	Santander
ATH	Athens	DTM	Dortmund	LPI	Linköping	SJJ	Sarajevo
BCM	Bacău	DUS	Dusseldorf	LPL	Liverpool	SKG	Thessaloniki
BDS	Brindisi	EAP	Basel	LRH	La Rochelle	SKP	Skopje
BEG	Belgrade	EAS	Donostia / San Sebastián	LUX	Luxembourg	SOF	Sofia
BES	Brest	EDI	Edinburgh	LUZ	Lublin	SOU	Portsmouth
BGO	Bergen	EIN	Eindhoven	LYS	Lyon	SPU	Split
BHX	Birmingham	EMA	Nottingham	MAD	Madrid	STR	Stuttgart
BIO	Bilbao	ERF	Erfurt	MAN	Manchester	SVG	Stavanger
BIQ	Anglet	ETZ	Nancy	MME	Middlesbrough	SVQ	Seville
BLQ	Bologna	EXT	Exeter	MMX	Malmö	SXB	Strasbourg
BNX	Banja Luka	FDH	Constance	MPL	Montpellier	SZG	Salzburg
BOD	Bordeaux	FKB	Karlsruhe	MRS	Marseille	SZY	Olsztyn
BOH	Bournemouth	FLR	Florence	MST	Aachen	SZZ	Szczecin
BOJ	Burgas	FMO	Münster	NAP	Naples	TGD	Podgorica
BRE	Bremen	FNI	Nimes	NCE	Nice	TGM	Târgu Mureș
BRI	Bari	GDN	Gdansk	NCL	Newcastle upon Tyne	TIA	Tirana
BRQ	Brno	GIB	Algeciras	NRN	Nijmegen	TKU	Turku
BRS	Bristol	GNB	Grenoble	NTE	Nantes	TLL	Tallinn
BTS	Bratislava	GOA	Genoa	NUE	Nuremberg	TLN	Toulon
BUD	Budapest	GOT	Gothenburg	NWI	Norwich	TLS	Toulouse
BZG	Bydgoszcz	GPA	Patras	OPO	Porto	TMP	Tampere
BZR	Béziers	GRQ	Groningen	OSR	Ostrava	TRD	Trondheim
C-BAR	Barcelona	GRX	Granada	OST	Bruges	TRS	Triest
C-BER	Berlin	GRZ	Graz	OTP	Bucharest	TSR	Timișoara
C-BRU	Brussels	GVA	Geneva	OUL	Oulu	TUF	Tours
CFE	Clermont-Ferrand	HAJ	Hanover	OVD	Gijón	TZL	Tuzla
CFR	Caen	HEL	Helsinki	PAD	Paderborn	VAR	Varna
C-FRA	Frankfurt am Main	HUY	Hull	PEG	Perugia	VGO	Vigo
C-GLA	Glasgow	IAS	Iași	PGF	Perpignan	VIE	Vienna
CGN	Cologne	INI	Niš	PIS	Poitiers	VIT	Vitoria-Gasteiz
C-HAM	Hamburg	INN	Innsbruck	PLQ	Klaipėda	VLC	Valencia
C-IST	Istanbul	IOA	Ioannina	PMF	Parma	VLL	Valladolid
CLJ	Cluj-Napoca	KLU	Klagenfurt	PNA	Pamplona	VNO	Vilnius
C-LON	London	KRK	Krakow	POZ	Poznan	VRN	Verona
CMF	Chambéry	KSC	Košice	PRG	Prague	WRO	Wroclaw
C-MIL	Milan	KTW	Katowice	PRN	Pristina	XRY	Jerez
C-MUN	Munich	KUN	Kaunas	PSA	Pisa	ZAD	Zadar
C-OSL	Oslo	KUO	Kuopio	PSR	Pescara	ZAG	Zagreb
C-PAR	Paris	KVA	Xanthi	PUF	Pau	ZAZ	Zaragoza
CPH	Copenhagen	LBA	Leeds	REG	Reggio Calabria	ZRH	Zurich

D. Fare Setting Model

Proof that the revenue function always finds exactly one optimal solution

This is the original revenue model from section 3.3.1 can be shown by rewriting the model:

$$R_{ij} = \frac{\exp(V_{HSR,ij})}{\sum_{k \in \boldsymbol{K}} z_{k,ij} \cdot \exp(V_{k,ij})} \cdot k \cdot \frac{(P_i \cdot P_j)^{\alpha} \cdot (GDP_i \cdot GDP_j)^{\beta}}{(d_{ij})^{\gamma}} \cdot TC_{HSR,ij} \tag{0.5}$$

To find the optimal fare setting, that maximises the revenue, the maximum value of this equation must be found, by setting the derivate of the formula with respect to $TC_{HSR,ij}$ to zero.

$$TC_{HSR,ij} = \max_{TC_{HSR,ij}} (R_{ij}) \Rightarrow \frac{d}{dTC_{HSR,ij}}[R_{ij}] = 0 \tag{0.6}$$

Written out fully and simplified, it amounts to solving the following equation:

$$\frac{(GDP_i \cdot GDP_j)^{\beta} \cdot (P_i \cdot P_j)^{\alpha} \cdot k \cdot \exp(V_{HSR,ij}) \cdot \left[z_{HSR,ij} \cdot \exp(V_{HSR,ij}) + \left[\sum_{k \in \boldsymbol{K} \setminus \{HSR\}} z_{k,ij} \cdot \exp(V_{k,ij})\right] \cdot [\beta^{TC} \cdot TC_{HSR,ij} + 1]\right]}{(d_{ij})^{\gamma} \cdot \left(\sum_{k \in \boldsymbol{K} \setminus \{HSR\}} z_{k,ij} \cdot \exp(V_{k,ij})\right)^2} = 0 \tag{0.7}$$

This only holds when the numerator equals zero and the denominator does not. As the denominator will never equal zero in a practical situation (only when $z_{k,ij} = 0$ for all k), it can be solved by the following equation:

$$(GDP_i \cdot GDP_j)^{\beta} \cdot (P_i \cdot P_j)^{\alpha} \cdot k \cdot \exp(V_{HSR,ij}) \cdot \left[z_{HSR,ij} \cdot \exp(V_{HSR,ij}) + \left[\sum_{k \in \boldsymbol{K} \setminus \{HSR\}} z_{k,ij} \cdot \exp(V_{k,ij})\right] \cdot [\beta^{TC} \cdot TC_{HSR,ij} + 1]\right] = 0 \tag{0.8}$$

As the terms outside of square brackets will never become zero, solving this equation is equivalent to:

$$z_{HSR,ij} \cdot \exp(V_{HSR,ij}) + \left[\sum_{k \in \boldsymbol{K} \setminus \{HSR\}} z_{k,ij} \cdot \exp(V_{k,ij})\right] \cdot [\beta^{TC} \cdot TC_{HSR,ij} + 1] = 0 \tag{0.9}$$

After writing out the full $\exp(V_{HSR,ij})$ term and some rearrangement, the equation can be written as:

$$\frac{[\beta^{TC} \cdot TC_{HSR,ij} + 1]}{\exp(\beta^{TC} \cdot TC_{HSR,ij})} = -\frac{z_{HSR,ij} \cdot \exp(\beta^{TT} \cdot TT_{HSR,ij})}{\sum_{k \in \boldsymbol{K} \setminus \{HSR\}} z_{k,ij} \cdot \exp(V_{k,ij})} \tag{0.10}$$

The left side includes the variable of interest (HSR cost), while the right-hand side includes known values only. As this expression cannot be simplified further, it means that the optimal HSR fare setting $\widehat{TC}_{HSR,ij}$ cannot be found by means of a simple calculation. A more sophisticated, iterating solving strategy is required in order to approximate this value.

Now, let $C_0 = \beta^{TC} \cdot TC_{HSR,ij}$ and the right-hand side of equation (0.10) equal C_1. This is done to foresee the nature of the possible solutions. The equation can now be simply written as:

$$\frac{C_0 + 1}{\exp(C_0)} = C_1 \tag{0.11}$$

As costs are positive values and β^{TC} takes a negative value in literature, C_0 will always be negative as well. In practice, also C_1 will always be negative, as the numerator and denominator both will be positive as $\exp(x) > 0$ for all x, and $z_{HSR,ij}$ by definition is equal to 1 when this model is applied.

Plotting experiments in Desmos (2011) show that solving equation (0.11) for $C_0 < 0$ and $C_1 < 0$ always yields exactly one solution with regards to C_0. The smaller C_1 becomes, the smaller C_0 will be. This proves that the revenue model in equation (3.9) with HSR fare as a free variable will always allow to find one unique optimum point, and it belongs to the maximum revenue, our point of interest.

E. Train Acquisition Formula

Rewriting equation (2.6) into (2.7)

The equation of Belal et al. (2020) can be generalised by replacing the numerical values by constants:

$$n_{ij} = \frac{C \cdot D_{HSR,ij} \cdot (l_{ij} + v^{max})}{s \cdot v^{max}}$$

This can be written without brackets:

$$n_{ij} = \frac{C \cdot D_{HSR,ij} \cdot l_{ij}}{s \cdot v^{max}} + \frac{C \cdot D_{HSR,ij} \cdot v^{max}}{s \cdot v^{max}}$$

Showing individual parts:

$$n_{ij} = C \cdot \frac{D_{HSR,ij}}{s} \cdot \frac{l_{ij}}{v^{max}} + C \cdot \frac{D_{HSR,ij}}{s}$$

Now, dividing passenger numbers $D_{HSR,ij}$ by the number of seats s represents the needed frequency of operations f_{ij}. Also, dividing the line length l_{ij} by the operating speed v is a measure of travel time t_{ij}. Both newly introduced terms here can be substituted into the equation:

$$n_{ij} = C \cdot f_{ij} \cdot t_{ij} + C \cdot f_{ij} = C \cdot f_{ij} \cdot (t_{ij} + 1)$$

For simplicity, it is assumed that $t_{ij} + 1 \approx t_{ij}$, which transforms the equation to:

$$n_{ij} = C \cdot f_{ij} \cdot t_{ij}$$

Now, let's look at the units in this equation:

$$[trains] = [-] \cdot \left[\frac{trains}{hr}\right] \cdot [hr]$$

Which is correct. From logics, it can be deduced that the equation correctly estimates the number of trains needed to operate a line at a certain frequency:

If trains operate a line with length l_{ij} at a frequency f_{ij}, it means that at a certain station, f_{ij} trains depart per hour. The distance (in hours travel time) between each train equals $1/f_{ij}$. Each train travels back and forth between the line's terminal stations, meaning that they travel in a cycle with a duration of $2 \cdot t_{ij}$ hours. If we have a cycle of length $2 \cdot t_{ij}$ hours, and a distance between each successive train of $1/f_{ij}$ hours, it means that the number of trains needed equals:

$$n_{ij} = \frac{length}{distance\ between\ successive\ trains} = \frac{2 \cdot t_{ij}}{\frac{1}{f_{ij}}} = 2 \cdot t_{ij} \cdot f_{ij}$$

Thus, the value of C equals 2. The expression above must be rounded upwards, as the number of trains must be integer. If rounded downwards, the spacing between successive trains would become too large, meaning that the needed frequency cannot be satisfied. Rounding up a value in mathematics is typically done by usage of ceiling brackets.

$$n_{ij} = \lceil 2 \cdot t_{ij} \cdot f_{ij} \rceil$$

Which is equivalent to the new equation (2.7).

F. Construction Costs

Construction costs per country

Borgogno (2023) reports unit construction costs per km for 27 countries within the scope of this project:

Table 0.8 Construction costs per country (Borgogno, 2023)

ISO-3	Country	Surface cost [M€ / km]	Tunnel cost [M€ / km]
ALB	Albania	-	-
AUT	Austria	23	65
BEL	Belgium	26	74
BIH	Bosnia and Herzegovina	-	-
BGR	Bulgaria	15	43
HRV	Croatia	12	36
CZE	Czech Republic	19	54
DNK	Denmark	15	44
EST	Estonia	6	16
FIN	Finland	10	29
FRA	France	35	100
DEU	Germany	45	129
GRC	Greece	18	51
HUN	Hungary	12	34
ITA	Italy	36	106
XKO	Kosovo	-	-
LVA	Latvia	5	15
LTU	Lithuania	6	16
LUX	Luxembourg	21	60
MKD	North Macedonia	-	-
MCO	Monaco	-	-
MNE	Montenegro	-	-
NLD	Netherlands	34	98
NOR	Norway	20	58
POL	Poland	12	36
PRT	Portugal	17	50
ROU	Romania	15	42
SRB	Serbia	-	-
SVK	Slovakia	19	56
SVN	Slovenia	17	48
ESP	Spain	27	77
SWE	Sweden	17	48
CHE	Switzerland	29	84
TUR	Turkiye	-	-
GBR	United Kingdom	43	125

Of the 39 countries within this scope (Appendix C), 35 are represented in the database of 726 cities, as introduced in section 3.4.8. Thus, for eight of these 35, no data is available in the work of Borgogno (2023):

Table 0.9 Neighbouring countries of countries not reported by Borgogno (2023)

ISO-3	Country	#Neighbouring countries	Of which reporting
ALB	Albania	(4): Greece, Kosovo, North Macedonia, Montenegro	(1): Greece
BIH	Bosnia and Herzegovina	(3): Croatia, Montenegro, Serbia	(1): Croatia
XKO	Kosovo	(4): Albania, Montenegro, North Macedonia, Serbia	(0)
MKD	North Macedonia	(5): Albania, Bulgaria, Greece, Kosovo, Serbia	(2): Bulgaria, Greece
MCO	Monaco	(1): France	(1): France
MNE	Montenegro	(5): Albania, Bosnia and Herzegovina, Croatia, Kosovo, Serbia	(1): Croatia
SRB	Serbia	(8): Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Kosovo, Montenegro, North Macedonia, Romania	(4): Bulgaria, Croatia, Hungary, Romania
TUR	Turkiye	(8): Armenia, Azerbaijan, Bulgaria, Georgia, Greece, Iran, Iraq, Syria	(2): Bulgaria, Greece

All but Kosovo have neighbours with reported values. This means that attributing countries with data will be performed in two steps: firstly, all non-reporting countries except Kosovo will be attributed the average of their reporting neighbouring countries. Then, all of Kosovo's neighbours will have values, and the average of those is taken for Kosovo. The final table of construction costs is adjusted for inflation, and is then presented as:

Table 0.10 Construction costs per country, complete (Borgogno, 2023)

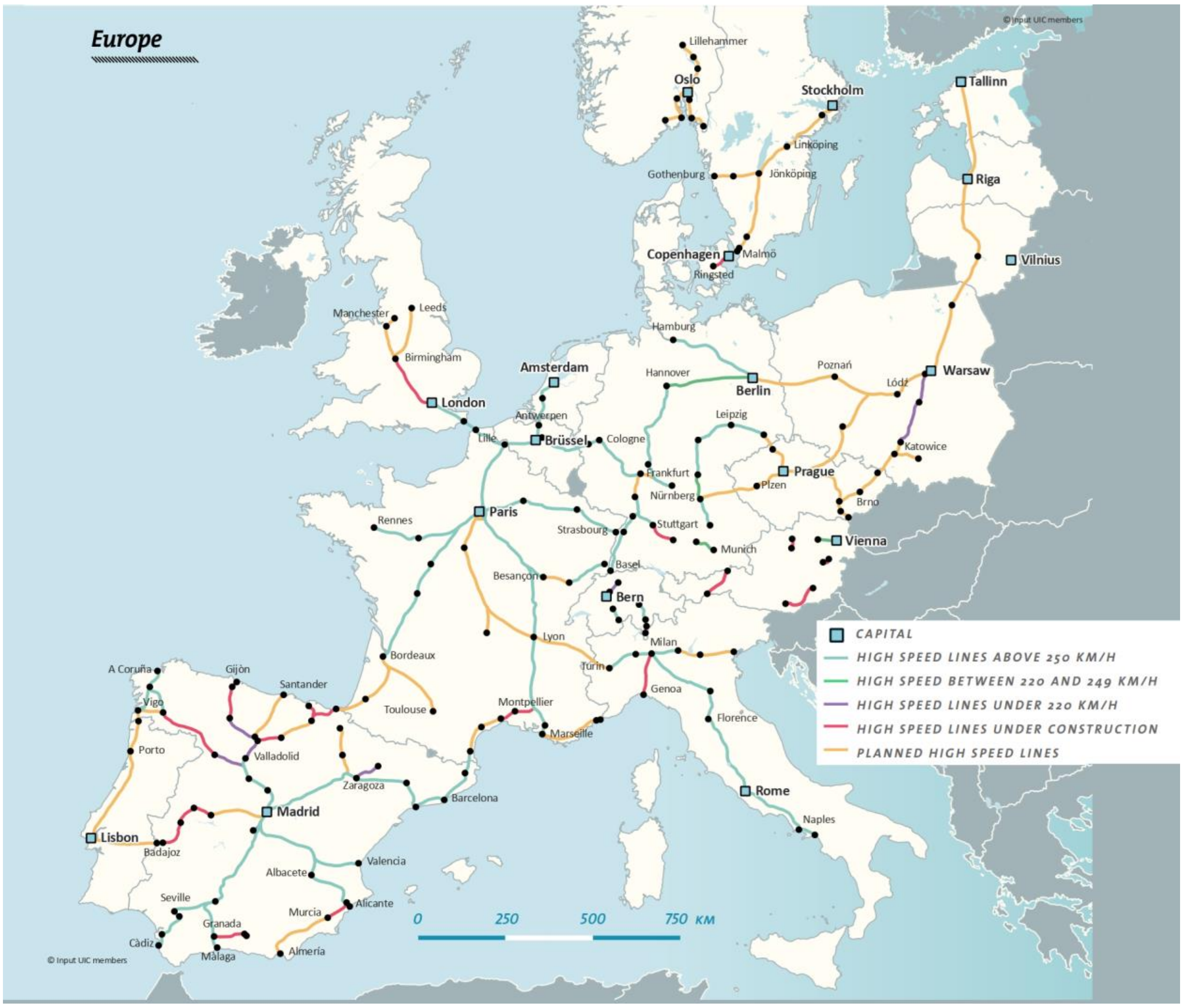
ISO-3	Country	Surface cost [M€ / km]	Tunnel cost [M€ / km]
ALB	Albania	22	64
AUT	Austria	29	81
BEL	Belgium	32	92
BIH	Bosnia and Herzegovina	15	45
BGR	Bulgaria	19	54
HRV	Croatia	15	45
CZE	Czech Republic	24	67
DNK	Denmark	19	55
EST	Estonia	7	20
FIN	Finland	12	36
FRA	France	44	125
DEU	Germany	56	161
GRC	Greece	22	64
HUN	Hungary	15	42
ITA	Italy	45	132
XKO	Kosovo	20	55
LVA	Latvia	6	19
LTU	Lithuania	7	20
LUX	Luxembourg	26	75
MKD	North Macedonia	21	59
MCO	Monaco	44	125
MNE	Montenegro	15	45
NLD	Netherlands	42	122
NOR	Norway	25	72
POL	Poland	15	45
PRT	Portugal	21	62
ROU	Romania	19	52
SRB	Serbia	17	49
SVK	Slovakia	24	70
SVN	Slovenia	21	60
ESP	Spain	34	96
SWE	Sweden	21	60
CHE	Switzerland	36	105
TUR	Turkiye	21	59
GBR	United Kingdom	54	156

G. Existent European Network

The figure below shows the map of the currently existing high-speed rail connections in Europe.

Map of existent network

Figure 0.1 Current European high-speed rail network (UIC, 2018)



List of upgradeable infrastructure

As mentioned in section 3.3.2, OD pairs with an average train operating speed exceeding 200 km/h will be given a 50% discount in HSR construction costs, since it would then come down to upgrading already existing infrastructure, rather than constructing new infrastructure. The table below lists all OD pairs for which this is the case. It should be noted that only the top 250 cities in population are taken into account for this list.

Table 0.11 OD pairs with upgradeable rail infrastructure

OD pair name	Average train travel speed
Bordeaux-Paris	282
Paris-Strasbourg	279
Madrid-Zaragoza	254
Barcelona-Madrid	250
Nancy-Paris	250
Córdoba-Zaragoza	250
Metz-Paris	238
Marseille-Paris	233
Lille-Lyon	232
Brussels-Paris	229
Le Mans-Paris	226
Barcelona-Zaragoza	224
Paris-Rennes	221
Lille-Paris	220
Sevilla-Zaragoza	216
Montpellier-Paris	216
Lyon-Paris	213
Bologna-Milan	213
Karlsruhe-Paris	211
Lille-London	210
Lille-Marseille	210
Paris-Toulon	207
Cologne-Wiesbaden	204
Málaga-Zaragoza	202
London-Paris	200
Mannheim-Stuttgart	200
Bielefeld-Dortmund	200

H. Optimal Network

This section presents extensive displays of the optimal network.

Served OD pairs

The table below shows data related to every served OD pair in the network, including market shares. "0.00%" means that the travel mode is unavailable (after HSR goes into operation).

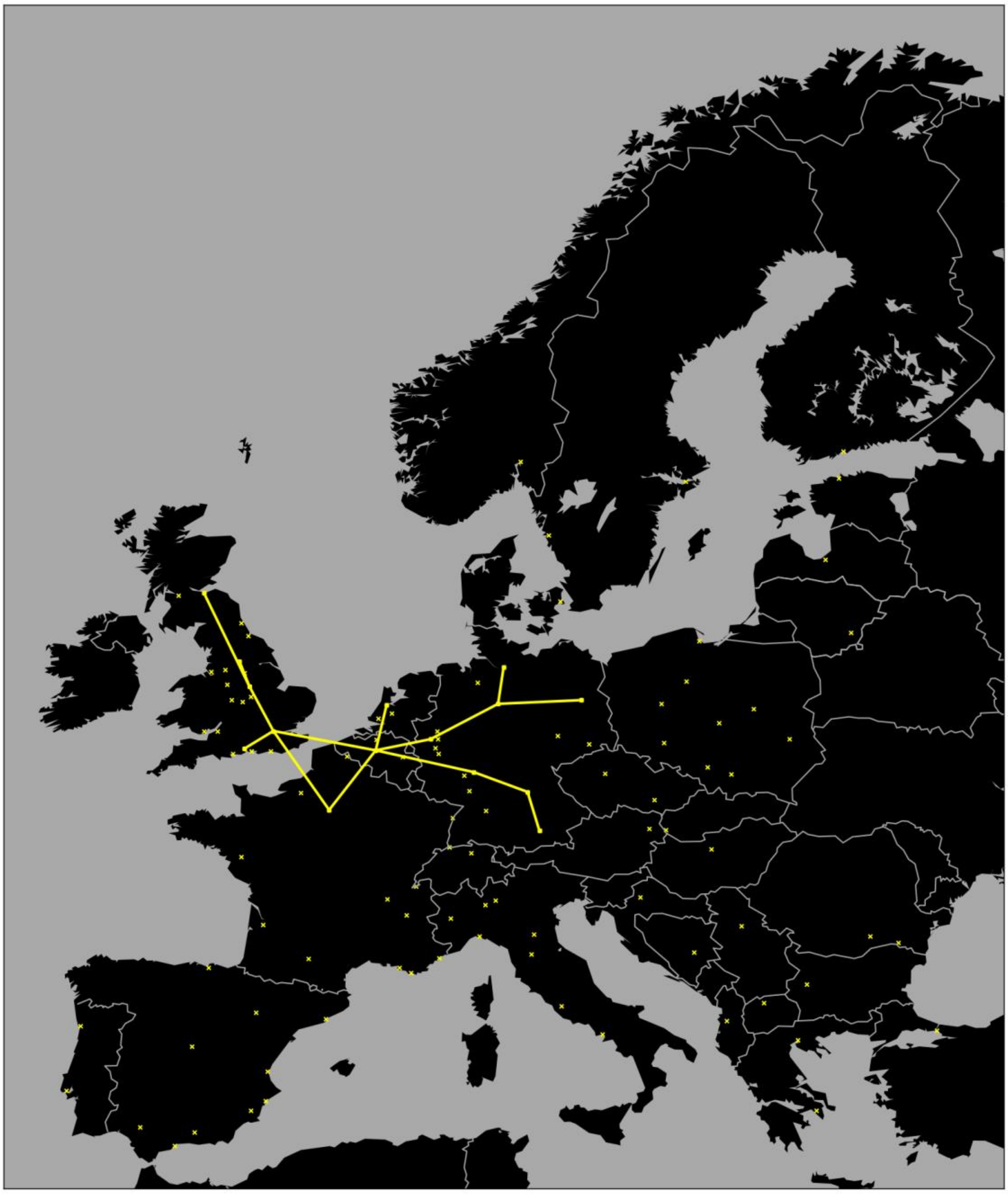
Table 0.12 Overview of served OD pairs

Name	Length [km]	Travel time [h]	First-year demand [pax]	Market share [%]			
				Plane	Train	Car	HSR
Amsterdam-London	575	1.99	20,224,504	4.88%	8.05%	1.47%	85.60%
London-Paris	464	1.56	12,936,155	11.02%	0.00%	4.00%	84.97%
Frankfurt am Main-London	761	2.58	7,750,188	2.81%	10.49%	0.38%	86.32%
Edinburgh-London	655	2.25	7,249,183	15.92%	0.23%	5.70%	78.15%
Düsseldorf-London	565	1.96	7,144,261	11.18%	3.17%	1.07%	84.58%
Düsseldorf-Paris	518	1.81	6,809,277	1.58%	2.50%	22.01%	73.91%
London-Munich	1156	4.00	4,079,705	10.93%	0.73%	0.06%	88.28%
Amsterdam-Paris	528	1.84	3,917,107	3.38%	6.32%	14.23%	76.08%
Brussels-London	364	1.24	3,894,770	1.00%	24.50%	0.70%	73.79%
Berlin-London	1135	3.93	3,181,802	13.26%	0.13%	0.02%	86.59%
Brussels-Paris	317	1.09	3,154,216	0.10%	26.92%	20.34%	52.65%
Berlin-Düsseldorf	570	1.97	2,019,844	2.35%	27.56%	5.93%	64.16%
Frankfurt am Main-Paris	714	2.43	2,009,311	0.53%	22.13%	7.24%	70.10%
Berlin-Paris	1088	3.78	1,973,123	15.86%	0.23%	0.92%	82.99%
Brussels-Düsseldorf	201	0.72	1,770,885	0.00%	13.10%	39.81%	47.09%
Leeds-London	326	1.20	1,749,705	0.66%	6.76%	27.64%	64.94%
Frankfurt am Main-Munich	395	1.42	1,642,057	0.19%	15.61%	23.44%	60.75%
Nottingham-Paris	670	2.30	1,573,034	0.05%	13.21%	0.82%	85.93%
Hamburg-London	997	3.50	1,451,245	13.49%	0.20%	0.11%	86.20%
Paris-Southampton	587	2.03	1,449,145	1.31%	10.52%	2.21%	85.96%
London-Nottingham	206	0.74	1,376,672	0.00%	24.26%	28.74%	46.99%
Leeds-Paris	790	2.76	1,307,810	11.25%	1.79%	0.46%	86.51%
Brussels-Frankfurt am Main	397	1.34	1,274,538	1.32%	25.24%	13.85%	59.59%
Düsseldorf-Hamburg	432	1.54	1,238,177	1.74%	0.91%	31.46%	65.89%
Berlin-Brussels	771	2.69	1,235,733	11.90%	2.44%	5.49%	80.17%
London-Southampton	123	0.47	1,129,688	0.00%	17.71%	42.71%	39.57%
Düsseldorf-Nottingham	771	2.70	1,117,338	9.49%	1.38%	0.79%	88.34%
Hannover-Paris	798	2.78	1,038,976	3.93%	1.13%	11.57%	83.36%
Brussels-Munich	792	2.76	1,012,084	4.46%	7.31%	6.83%	81.40%
London-Nuremberg	984	3.37	1,004,463	9.61%	2.36%	0.15%	87.88%
Düsseldorf-Southampton	688	2.43	938,313	9.80%	0.67%	1.65%	87.89%
Nuremberg-Paris	937	3.22	839,524	1.00%	10.35%	7.48%	81.18%
Hannover-London	845	2.93	700,371	13.31%	0.48%	0.20%	86.01%
Brussels-Leeds	690	2.44	692,411	12.04%	2.34%	0.47%	85.16%
Brussels-Nottingham	570	1.98	672,064	0.43%	15.17%	0.71%	83.69%
Düsseldorf-Leeds	891	3.16	671,342	13.38%	0.05%	0.12%	86.45%
Brussels-Southampton	487	1.71	660,309	0.08%	12.55%	2.55%	84.82%
Frankfurt am Main-Southampton	884	3.05	609,912	2.09%	6.31%	1.59%	90.01%
Frankfurt am Main-Nottingham	967	3.32	555,968	7.54%	3.47%	0.21%	88.78%
Amsterdam-Nottingham	781	2.73	537,745	9.14%	2.08%	0.64%	88.14%
Brussels-Hamburg	633	2.26	480,284	3.07%	12.29%	9.89%	74.75%
Frankfurt am Main-Leeds	1087	3.78	438,929	11.45%	0.22%	0.06%	88.27%
Frankfurt am Main-Nuremberg	223	0.79	391,279	0.56%	25.87%	31.47%	42.10%
Düsseldorf-Hanover	280	0.97	380,186	0.07%	32.15%	22.34%	45.45%
Brussels-Nuremberg	620	2.13	366,965	0.57%	12.45%	11.24%	75.74%
Hamburg-Nottingham	1203	4.24	286,732	9.50%	0.17%	0.15%	90.17%
Berlin-Southampton	1258	4.40	243,906	9.64%	0.04%	0.09%	90.24%
Brussels-Hanover	481	1.69	238,571	0.02%	17.13%	13.23%	69.62%

Edinburgh-Nottingham	449	1.51	228,162	1.51%	0.03%	23.57%	74.88%
Leeds-Nottingham	120	0.46	223,635	0.00%	35.96%	33.94%	30.10%
Nuremberg-Southampton	1107	3.84	204,078	4.31%	2.49%	1.40%	91.80%
Munich-Nuremberg	172	0.63	201,019	0.00%	8.87%	49.16%	41.97%
Munich-Nottingham	1362	4.74	188,554	9.59%	0.11%	0.02%	90.28%
Hamburg-Hanover	152	0.57	156,814	0.00%	28.38%	37.30%	34.32%
Munich-Southampton	1279	4.47	155,331	10.30%	0.06%	0.05%	89.59%
Berlin-Hanover	290	1.00	146,762	0.01%	42.71%	16.31%	40.97%
Leeds-Southampton	449	1.67	142,019	0.02%	0.70%	27.53%	71.74%
Hanover-Nottingham	1051	3.67	137,785	9.07%	0.40%	0.28%	90.26%
Hanover-Southampton	968	3.40	133,829	8.14%	0.27%	1.12%	90.48%
Edinburgh-Southampton	778	2.72	36,778	12.24%	0.07%	5.81%	81.88%

Figure of the optimal network

Figure 0.2 Optimal network design, full-size

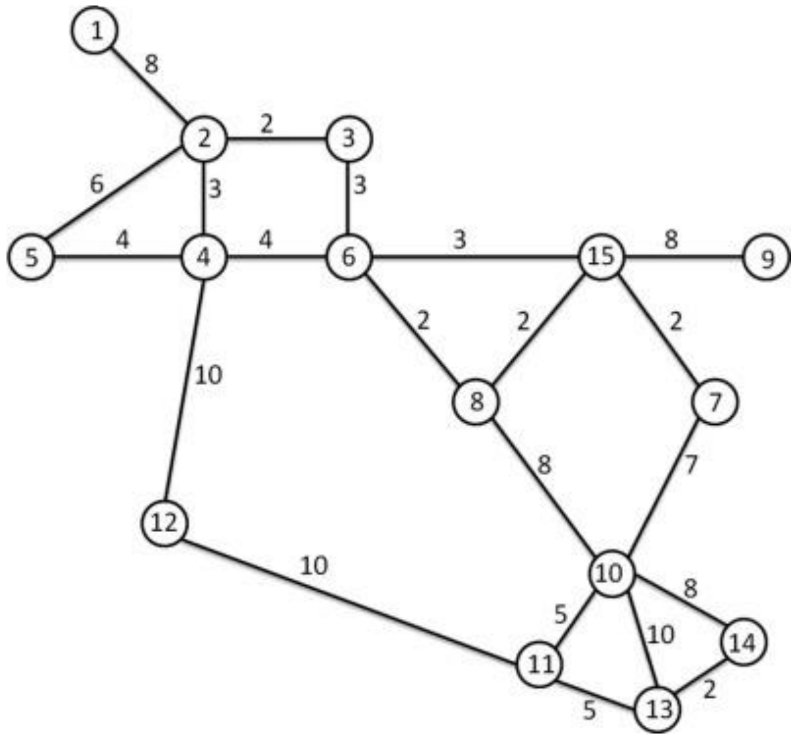


I. Benchmarking

Our optimisation model was *benchmarked* to the network of Mandl (1980). This section presents the underlying characteristics of this network.

Network

Figure 0.3 Mandl's network (Arbex & da Cunha, 2015)



Demand matrix

Table 0.13 Mandl (1980) network demand matrix

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	400	200	60	80	150	75	75	30	160	30	25	35	0	0
1	400	0	50	120	20	180	90	90	15	130	20	10	10	5	0
2	200	50	0	40	60	180	90	90	15	45	20	10	10	5	0
3	60	120	40	0	50	100	50	50	15	240	40	25	10	5	0
4	80	20	60	50	0	50	25	25	10	120	20	15	5	0	0
5	150	180	180	100	50	0	100	100	30	880	60	15	15	10	0
6	75	90	90	50	25	100	0	50	15	440	35	10	10	5	0
7	75	90	90	50	25	100	50	0	15	440	35	10	10	5	0
8	30	15	15	15	10	30	15	15	0	140	20	5	0	0	0
9	160	130	45	240	120	880	440	440	140	0	600	250	500	200	0
10	30	20	20	40	20	60	35	35	20	600	0	75	95	15	0
11	25	10	10	25	15	15	10	10	5	250	75	0	70	0	0
12	35	10	10	10	5	15	10	10	0	500	95	70	0	45	0
13	0	5	5	5	0	10	5	5	0	200	15	0	45	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

J. Scientific Paper

The paper starts at the next page

The Potential Profitability of a European High-Speed Rail Network

D.M.T. Mol [a]

Delft, The Netherlands, November 2024

[a] *Delft University of Technology, MSc Transport Infrastructure & Logistics*

ABSTRACT

Despite being a long-cherished EU ambition and a crucial key to achieving climate goals, there is still no European High-Speed Rail (HSR) network, with the few completed projects often facing disappointing demand resulting in unprofitability. In order to gain insights to profitable network design, this study develops a new formulation to the "Transport Network Design & Frequency Setting Problem" (TNDFSP), as current literature lacks one that can optimally solve the problem for instances of this size while also accounting for demand elasticity. Our model works with elastic and dynamic HSR demand calculated from optimal fare settings and observed travel mode characteristics from competing alternatives. The optimal solution is largely insensitive to fare changes, and in some measures outperform current state-of-the-art. Our model considers the 111 most populous European cities, along with all their origin-destination (OD) pairs, and finds the most profitable network design within a reasonable solution time frame. The results show HSR can be very profitable in Europe, but only when concentrated around a selected group of the largest cities in the western part of the continent. Despite being profitable and contributing significantly to set Green Deal goals, the viability of our network heavily depends on the willingness of several countries to invest significantly in infrastructure construction.

Keywords: High-speed rail, Network design, Frequency setting, Mathematical optimisation, Demand forecasting, Profitability estimation

Detailed report: An electronic version is available at https://repository.tudelft.nl/.

1 INTRODUCTION

1.1. Background & context

In Europe, the long-distance travel market (>700 km) has been dominated by planes, as it most often is the only practical option (Bleijenberg, 2020). The market showed continuous exponential growth in air passenger numbers of 6.0% yearly on average (Eurostat, 2019), and projected to double in passenger numbers by 2040 (Timperley, 2020), tripling its contribution to climate change between 2020 and 2050 (ICAO, 2019). This is incompatible with the active Climate Agreements (Gössling & Humpe, 2020; Bleijenberg, 2020), and therefore, the European Union is forced to look for greener travel alternatives, the most promising candidate being High-Speed Rail (HSR).

High-speed trains emit on average seven times less CO2 per passenger-km (Strauss et al., 2021), when compared to air or road alternatives. With commercial speeds reaching up to 350 km/h, relatively low waiting times and the ability to directly connect city centres (Martín et al., 2014), they have a competitive advantage for travel times up to four hours (UIC, 2018). Japan was the first country to develop HSR with the introduction of the Shinkansen in 1964. In recent decades, fuelled by HSR-backing governmental policy and subsidies, China has built a network comprising more than two-thirds of the global rail length (Chen, 2020), proving very successful by serving 2.4 million passengers in 2019 (Zhang, 2024), and decimating local airplane's market share (Bradsher, 2013). Following this example, it comes as no surprise that the EU has been pushing governments to develop international high-speed rail connections.

1.2. Problem definition

Even though Europe has an extensive conventional rail network, it was developed with national focus, complicating interoperability and efficiency when travelling internationally. Despite HSR-backing policy acting since the 1990s and investments of €23.7 billion into HSR development, the total transport-related greenhouse gas emissions have only increased since then (EEA, 2023). It can be concluded that the EU is not on track to meet its climate goals. As of today, still no European network exists (European Court of Auditors, 2018). This can be attributed to three leading problems:

1. HSR policy has proven to be ineffective, as the European Union has no legal powers in forcing member states to construct rail connections as envisioned, due to many national rules being still in place. These act as technical and administrative barriers when it comes to construction of cross-border connections (European Court of Auditors, 2018).
2. There is great uncertainty in the profitability of lines, putting their justifiability under scrutiny and causing governmental bodies to become reluctant when it comes to HSR development. Projects commonly have to deal with cost overruns and construction delays. In particular, the demand is hard to forecast. In Europe, three out of seven lines fail to have a sufficient number of passengers, while nine of out fourteen do not even have the potential number of passengers in the area to ever reach a sufficient amount (European Court of Auditors, 2018). These projects therefore rely on subsidies, raising major critiques whether this money could not have been better spent elsewhere.
3. As pointed out by Grolle et al. (2024), no HSR network design methods are currently available – a crucial literature gap. The great complexity of the "Transport Network Design & Frequency Setting Problem" (TNDFSP), primarily caused by demand elasticity, has led to the problem being primarily solved by (meta)heuristic algorithms, providing a good but not optimal solution. Due to the complexity of the problem, many assumptions and simplifications are made, putting the value of the found solutions under scrutiny. Prominent examples of these are the usage of fixed demand (despite its elastic nature) and network simplification (Cascetta & Coppola, 2012). Current literature lacks one that can optimally solve the problem for instances of this size while also accounting for demand elasticity.

1.3. Research goal

In order to address the problems mentioned above, this study aims to develop a model that is able to assess European HSR profitability through mathematical optimisation, finding the optimal configuration of connections, lines and their frequencies. Therefore, the goal is to answer the following main research question:

"Which European cities must be connected via High-Speed Rail, and how should these connections be served in order to lead to an (optimally) profitable network?"

1.4. Paper framework

The methodological framework required to answer this question consists of three parts: demand forecasting, profitability estimating and network design. Each of these parts return in the remaining sections of this paper.

Section 2 presents the highlights of our literature review. Following from that, section 3 formulates the methodology required to answer our main research question. Section 4 presents the results after implementation of the methodology. Section 5 derives policy implications from this and section 6 concludes by presenting the highlights among the findings. Finally, a discussion is presented in section 7.

2 LITERATURE REVIEW

2.1. Demand forecasting

The science behind long-distance travel demand forecasting has been covered many times by literature. Typically, it is performed by a forecasting model, each taking a number of demand-influencing factors into account. The relationships between these factors and the level of demand are quantified by means of calibration to real-life observed data. In order to create an overview of what is used in literature, 100 long-distance travel demand forecasting studies published between 1982 and 2023 were analysed, with special interest regarding model types, demand-impact factors and their combinations. In particular, great literature reviews are performed by Nurhidayat et al. (2023), Zhang et al. (2012) and Börjesson (2014), which delivered half of the found papers, while the rest was found by further snowballing through academic data bases such as *ScienceDirect*, *Scopus* and *Google Scholar*.

Impact factors

In total, these 100 studies make use of 38 distinguishable demand-impact factors, which can be divided into eight categories. We can present the following highlights among the findings:

- Travel time and travel cost are by far the most important impact factors, being used by 88% and 72% of studies, respectively.
- Economic factors are used by 44% of studies, with GDP and Income being the most popular examples in this group.
- Service frequency depends on the network's line design, and is included by 43% of studies.
- Population (34%) is most often expressed in terms of city population, but a far more accurate measure to represent the number of potential number of travellers would be the number of inhabitants living within the catchment area, as HSR stations might and should attract travellers from outside the city in which it is located (Martínez et al., 2016).
- Travel comfort factors are important, but hard to quantify and therefore are included by only 29% of studies, which a great number of different factors used, without a clear favourite.
- Destination attractiveness can be crucially influencing the level of demand, especially when it comes to touristic attraction (Mokhtarian & Salomon, 2001). However, due to its complexity of quantification, it is included by only 7% of studies.
- Together, these seven groups capture 89% of all factors used by studies to forecast demand.

Forecasting models

In total, these 100 studies make use of 21 distinguishable models, which can be divided into four types. We can present the following highlights among the findings:

- Models are destined to forecast either the market share of a travel mode, or the total passenger flow.
- Both linear and logistic regression are the two clear favourite model types among demand forecasting studies, used by 47% and 32%, respectively.
- Logistic regression is the only established model type able to forecast market shares. Most popular subtypes are both the multinomial (MNL) and nested (NL) logit.
- Typically, ordinary linear regression is used, but panel linear regression is commonly used as well.
- Relatively many studies (21%) experiment with alternative, non-established and self-invented models that do not fit within the previously mentioned groups.
- Despite being considered the most suitable method to estimate demand for new transit connections, especially when currently no direct service is available (Grosche et al., 2007), gravity models are used by only 6% of studies.
- Demand is not a fixed number, as its impact factors change over time (Cascetta & Coppola, 2011).

Model-factor combinations

The relative popularity of the eight found groups among established model types are presented in Figure 1 below.

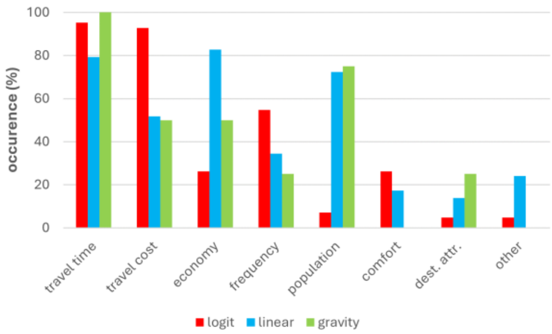


Figure 1 Demand-impacting factor usage among established models

The figure shows that some factors are very common to use in one model type, but not in another. The best example for this is 'population', which is commonly used in linear regression and gravity models, but practically non-existent in logit models. Travel time is the only factor popular among all model types. Travel costs are implemented in almost all logit models, but for other model types it is way less common. The 'economy' factors are mostly used in linear models. The analysis also shows that gravity models often are kept very simple. Comfort-related and 'other' factors are not implemented.

Demand evolution

As briefly mentioned in the "*Forecasting models"* section, the level of demand is prone to change over time. In order to forecast the total demand, it is commonly divided into the following parts, as listed by Cascetta & Coppola (2011):

- Diverted demand, which comes from people already travelling, changing their minds about which alternative to use.
- Induced demand, which occurs by people who were not travelling, but now choose to do so. The effects happen directly, or indirectly, when it requires a change in lifestyle, habit or land use.
- Demand growth, which is caused by an increase of mobility due to economic growth.

The forecasting model types assessed here often only take the directly occurring effects in account, which are called endogenous. Cascetta & Coppola (2011) advise to also look at exogenous (indirect) factors, since they can have a large impact on (future) demand.

2.2. Profitability estimation

In order to assess the profitability of a HSR project, each of its impacting cash-flows must be mapped out. Barrón et al. (2012) provide insights into this subject with great detail, applied to real European cases. This helps constructing an overview of all significant cash-flows that come into play when constructing and operating new high-speed rail lines:

- Ticket revenue, which is dependent on the fare setting and the total passenger flow. The latter factor itself also depends on the fare setting (May et al., 2022).
- Infrastructure construction, maintenance and operating costs. Here, the construction cost is an initial investment, while the other two cash-flows are recurring payments. Construction costs heavily depend on the line length, maximum operating speed, terrain, population density and the economic price level. Operating and maintenance costs also depend on the country's economic price level but varies much less from country to country. It mainly depends on the number of seat-km (Barrón et al., 2012).
- Rolling stock acquisition, maintenance and operating costs. Here, the acquisition cost is an initial investment, while the other two cash-flows are recurring payments. Acquisition costs depend on the choice and number of rolling stock, which tends to increase with faster trains with a higher capacity. Operating and maintenance costs depend on the number of seat-km.

To illustrate the magnitude of these cash-flows, Table 1 below lists it's indicative values, based on works of Barrón et al. (2012), de Rus et al. (2020), Nash (2010) and Fröidh (2006). Note that the reported values are all adjusted for inflation, based in 2024 prices.

Table 1 Indicative values of HSR cash-flows

Part	Types	Indicative values
Infrastructure	Construction	€26.5 million / km
	Operating and Maintenance	€100 thousand / km / year
Rolling stock	Construction	€37.5 million / train
	Operating and Maintenance	€0.03 / seat-km
Passengers		

2.3. Network design

The TNDFSP consists out of two distinctive problems: the TNDP, which relates to network design, and the FSP, which relates to line design. In order to assess the larger problem, both the smaller problems are analysed separately here at first.

2.3.1. Transport Network Design Problem (TNDP)

The definition of the problem is provided by Feremans et al. (2003): it consists of "finding the optimal subgraph of a graph, subject to side constraints". The subgraph is formed by the set of selected city-pairs (edges) and associated cities (nodes), while the graph encompasses all city-pairs (edges) and all cities (nodes). It is a well-studied transportation problem. In this project, the focus lies on one certain application of the TNDP: finding the optimal choice of adding new links.

Mathematical definition

The TNDP related to this project considers an undirected graph $G = (N, A)$, where N represents the set of nodes (cities) and A represents the set of arcs (connections) connecting these nodes. Each arc $a \in A$ has certain attributes, such as the travel time and/or cost related to traversing it. For each arc, the decision can be made about whether it must be built or not. For this reason, a binary decision variable y_a is included, defined for each arc $a \in A$. If the arc is built, y_a is set to 1 and zero otherwise. Most often, the choice in arc selection is not an entirely free choice – it usually comes with a number of constraints. The selection of arcs a form a new set $A^{sel} \subseteq A$. The nodes associated with the arcs $a \in A^{sel}$ also form a new set $N^{sel} \subseteq N$. Together, they form a new graph $G^{sel} = (N^{sel}, A^{sel})$, which then logically is a subgraph of the original graph $G = (N, A)$. Each possible graph $G^{sel} \subseteq G$ has an associated 'value' Z resulting from the objective function. The goal of the problem is to find the optimal network, maximising Z, while adhering to all constraints.

A promising recent development regarding optimal TNDP solving is the adaption of the Multi-Commodity Flow Problem's (MCFP) formulation. Marín & García-Ródenas (2009) already observed that this approach is required when considering passenger flows for each OD pair. However, with current practice, the formulation suits small instances of the problem only, due to the added complexity by OD pair specific flows (Gutiérrez-Jarpa et al., 2017), a problem overcome by last-mentioned authors by splitting the problem in two. First, they optimise for the network's topology only, based on the OD pair flows. Then, they optimise for the design of lines over this topology.

Objectives

Extensive reviews on TNDP studies are provided by Kepaptsoglou & Karlaftis (2009), encompassing 62 studies published between 1967 and 2007, and by Durán-Micco & Vansteenwegen (2021), for thirty studies ranging from 2009 to 2021. Analysis of these 92 TNDP studies shows that more recently, the focus has shifted towards optimisation from a user perspective rather than the operator's point of view. The user perspective is typically addressed by minimising total travel time or total user travel costs. The operator's perspective is addressed by maximising profit, or minimising costs.

Constraints

In the most basic formulation of the TNDP, it is constrained that lines can visit each node at most once and that all demand must be served (Kepaptsoglou & Karlaftis, 2009); (Guihaire & Hao, 2008). Constraints are related to either the network's performance or limited resources. Due to the complexity of the problem, studies attempting to find solutions generally include a low number of non-complex constraints (Fan & Machemehl, 2006a); (Guihaire & Hao, 2008). Demand is typically included as being 'fixed', despite naturally being elastic. Accounting for this into TNDP solving further complexifies an already very complex problem, thus the vast majority of studies choose to work with a fixed demand (Kepaptsoglou & Karlaftis (2009); Durán-Micco & Vansteenwegen (2021)).

Further examples of constraints are related to fleet size, operator costs and other operator-related budgets, maximum line length, capacity on lines and other constraints on the network's topology (Durán-Micco & Vansteenwegen, 2021).

Solving methods

Only 10% of studies reviewed by the authors use an exact method for solving. Typically, the size of the networks in these studies are very small. With major questionable simplifications, Gutiérrez-Jarpa et al. (2017) solve for a network of 108 nodes, 3,789 arcs and 360,000 passengers, and Borndörfer et al. (2007) is able to solve for a network with 410 nodes, taking several hours to do so.

Most often, TNDPs are solved by (meta)heuristics - known for finding a reasonably good solution relatively fast, while not being designed for a specific model. Hence their popularity over exact approaches. Guihaire & Hao (2008) classify TNDP-applied heuristics into four 'big families': neighbourhood search, evolutionary search, hybrid search and greedy heuristics. Even within these groups, the exact approach varies from study to study. The possible choices in approaches, premises and simplifications are endless. With again major simplifications, heuristics allow for solving of much larger network instances. For example, Oliker & Bekhor (2020) pre-define a set of passenger routes for each OD pair, allowing to solve a network of 903 nodes, 2975 arcs and 5394 OD pairs.

2.3.2. Frequency Setting Problem (FSP)

The TNDP does not address how the network must be operated in order to sufficiently handle the demand. Generally speaking, transport networks are served by lines, each having their own designed frequency. The optimal incorporation these two factors into a transit network design is known as the Frequency Setting Problem.

Mathematical definition

Similar to the TNDP definition, the frequency setting problem considers a directed graph $G = (N, A)$, where N represents the set of nodes (or vertices, in this context: cities) and A represents the set of arcs (or links, in this context: HSR connections) connecting these nodes. These represent the movement of trains.

Now, the frequency setting problem adds to this a set of lines L. Each line $l \in L$ consists of a set of adjacent arcs. Each arc has a passenger flow Q_a, defined for all $a \in A$, originating from the solution of the TNDP. The set L encompasses many lines, much more than eventually chosen, as the goal is to select the optimal subset $L^{sel} \subset L$. For this reason, an integer decision variable f_l is included which denotes the frequency, defined for each line $l \in L$. If the line is not chosen, f_l is simply set to zero.

Objectives

An extensive FSP literature review is provided by Durán-Micco & Vansteenwegen (2021), for thirty studies ranging from 2009 to 2021. Analysis of these show objectives vary much less for FSPs than for TNDPs. Most commonly, frequency setting problems optimise operator's and/or user's costs, as they are highly influenced by the set frequency (Durán-Micco & Vansteenwegen, 2021). As a one-sided optimal solution is not socially desirable, most studies formulate an objective function that includes both sides and call it 'social welfare'. The user's and operator's perspectives are addressed similarly as in TNDP, but the line design now allows for penalising transfers with a monetary value.

Constraints

In literature, frequency setting problem formulations impose a wide variation of constraints. But most prominently, all demand on the network must be served, as otherwise it would not be regarded as a valuable solution (Canca et al., 2018). The design of lines and setting of their associated frequencies is namely guided by the flow of passengers over the network (Kepaptsoglou & Karlaftis, 2009).

Constraints mainly act on lines, regarding their allowed length, number of stops, directness and degree of overlapping with other lines. Often, line frequencies are bounded by minimum and maximum values and must be integer.

Solving methods

Similarly to TNDPs, the exact approach is uncommon and used in approximately 10% of studies (Durán-Micco & Vansteenwegen, 2021), again due to its complexity. Therefore, exact approaches only suit small networks. Cancela et al. (2015) find a new Mixed-Integer Linear Programming (MILP) formulation for a bus network, and solves the problem successfully for relatively small networks. The authors emphasize that research on larger networks would benefit from solving with algorithms instead, as a network of 84 nodes and 143 arcs takes over four hours to solve. Ranjbari et al. (2020) solve a full TNDFSP by use of pre-defining candidate lines and stations.

Most of TNDFSP-related studies thus solve the problem by heuristics, the most popular being the evolutionary algorithm, allowing for solving larger instances of the problem.

2.4. Summary of literature gaps

Both the demand forecasting and profitability estimation parts are already well-researched. Within the network design part, however, an important literature gap was discovered.

Regarded as one of the most complex transportation problems to solve, the TNDP and TNDFSP are typically solved using heuristic methods (Guihaire & Hao, 2008). Studies make various compromises to keep their solving times within reasonable bounds, such as limiting the number of nodes, arcs or OD pairs. Demand elasticity is considered one of the main factors complexifying the problem (Jiang et al., 2014), and therefore not accounted for by studies when applying an exact mathematical approach to a medium-to-large sized network (Kepaptsoglou & Karlaftis, 2009; Durán-Micco & Vansteenwegen, 2021). This puts the scientific value of optimal transport network designs, found through mathematical optimisation, under scrutiny. It is evident that with current practices in literature, medium- to large-sized realistic and optimal transport networks cannot be designed with exact methods (Murray, 2003); (Guan et al., 2003), which is also evidenced by the literature review of Guihaire & Hao (2008). This project will attempt to bridge this gap, and account for demand elasticity in linear programming in larger instances of the problem.

3 METHODOLOGY

3.1. Foundational premises

These will form the foundation of our methodology, as they shape its overall framework and nature. Their values were carefully set and presented in Table 2 below:

Table 2 Methodological premises

Lifetime	40 years
Maximum design speed	350 km/h
Track type	double-track
Traffic type served	passenger traffic
Power supply	electric
Train set	name: CRH380CL
	year: 2011
	seats: 1,053
	homogeneously
Network & line design	integrated

A maximum design speed of 350 km/h was selected as it is the highest speed currently in operation. UIC High-Speed (2018) lists all train sets meeting this requirement, and after assessment of capacity and costs, type CRH380CL was chosen. During optimisation, network and line design will be integrated, as it produces higher-valued solutions.

3.2. Demand forecasting

Problem definition

In the last thirty years, demand forecasting knowledge has grown, but not the accuracy of forecasting models: demand is overestimated, on average by 106% (Flyvbjerg et al., 2005). This is reflected by European HSR lines: nine out of fourteen failed to generate sufficient demand in order to be successful (European Court of Auditors, 2018). This can be attributed to political causes, which have a substantial influence on rail projects: decision-makers generally ignored or downplay financial risks under the guise of social welfare or other variables that are impossible to measure accurately (Flyvbjerg et al., 2005). Börjesson (2014) identifies three other arising problems that complexes HSR passenger forecasting:

1. It depends on more factors than regional travel models. Therefore, HSR demand models need to be adapted.
2. The non-linear relationship between demand and travel time.
3. HSR models are harder to calibrate than regional travel models, due to scarcer data.

All three of these factors must be addressed in order to provide realistic HSR demand forecasts.

Model definition

Addressing the problems mentioned above, our model must incorporate more demand-impacting factors. As the most established and used method in practice, a logit model is used to forecast HSR demand. A direct consequence of this choice is that the HSR demand has to be estimated via its market share, by multiplying a total demand flow with the HSR market share (Sánchez-Borràs et al., 2010; Leng et al., 2015). The following model is constructed:

$$D_{AIR,ij} = \frac{\exp\left(V_{plane,ij}\right)}{\sum_{k \in K} z_{k,ij} \cdot \exp\left(V_{k,ij}\right)} \cdot k \cdot \frac{\left(P_i \cdot P_j\right)^{\alpha} \cdot \left(GDP_i \cdot GDP_j\right)^{\beta}}{\left(d_{ij}\right)^{\gamma}} \tag{1}$$

where:

$$V_{k,ij} = \beta^{TT} \cdot TT_{k,ij} + \beta^{TC} \cdot TC_{k,ij} \tag{2}$$

The meaning of these parameters are explained below in Table 3:

Table 3 Nomenclature of demand forecasting model

Parameter	Unit	Definition
$D_{AIR,ij}$	[pax]	air demand for city pair ij
$V_{k,ij}$	[-]	observed utility of alternative k for city pair ij
$z_{k,ij}$	[-]	presence of alternative k for city pair ij
β^{TT}	[util/h]	MNL coefficient for travel time
β^{TC}	[util/€]	MNL coefficient for travel cost
$TT_{k,ij}$	[h]	travel time of alternative k for city pair ij
$TC_{k,ij}$	[€]	travel cost of alternative k for city pair ij
k	[-]	intercept gravity coefficient
α	[-]	gravity coefficient for population
β	[-]	gravity coefficient for GDP
γ	[-]	gravity coefficient for distance
P_i	[pax]	population of city i
GDP_i	[€]	GDP of city i
d_{ij}	[km]	distance between city pair ij

This formulation provides a solution to all three complexing factors in HSR demand forecasting: it encompasses more factors of influence, the non-linearity between demand and travel time is addressed by the logit part and γ-parameter, and it is calibrated on widely available flight demand data rather than scarce HSR data.

The model defined includes the five most popular demand impacting factors from studies (see section 2.1), excluding 'frequency' which is addressed separately as the 'frequency setting problem'.

Data collection

In order to gather data to accurately calibrate and later implement the demand forecasting model, data must be gathered for all of its variables. Below is listed how:

Air demand: Arguably the most comprehensive data set on long-distance travel passenger data currently accessible is provided by Eurostat (2024a), the statistical office of the European Union. Data is updated on a monthly basis, for each EU member state. We neglect season trends, opting for a yearly basis of demand forecasting. For calibration, 2019 is the chosen year of interest, being the latest available year of normal operations before the pandemic. Passenger numbers are combined for outward and return flights, and for flights serving the same city pair. Only flights in continental Europe are considered. A figure of the scope is provided below.

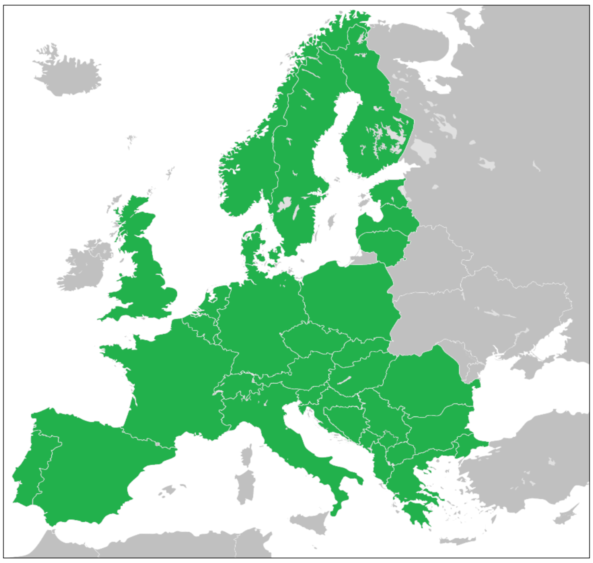


Figure 2 Project's scope

Population: Following from literature review, catchment area population is deemed the most accurate representation of population. Its size is determined by the chosen travel time, of which the figure has been debated in literature. Here, 120 minutes in car travel time is set as a maximum, matching the maximum allowed access time of most studies (Marcucci & Gatta, 2011). Seeking for an optimal fit, the model is calibrated for every increment of 15 minutes. The population living in catchment areas around the city centres are gathered with use of Smappen (2024).

Gross Domestic Product (GDP): As population data was attained in the previous section, the model will be calibrated on GDP and GDP/capita, as we seek for the best fit. For model calibration, the data will be acquired by usage of the following three databases from Eurostat:

1. met_10r_3gdp: Gross domestic product (GDP) at current market prices by metropolitan regions (Eurostat, 2024b)
2. tgs00003: Regional gross domestic product by NUTS 2 regions - million EUR (Eurostat, 2024d)
3. demo_r_d2jan: Population on 1 January by age, sex and NUTS 2 region (Eurostat, 2024c)

If the city is included, data set 1 finds us the city's total GDP directly. If not, is deducted by scaling the region's GDP (data set 2) down, based on the population of the region (data set 3) and the city itself. To find GDP per capita, population data from a data set by Florczyk et al. (2020) is used, which was created in name of the European Commission, reports the population for all European centres, based in 2015.

Distance: The great circle distance between city centres are used. The city's coordinates are used as input to calculate this. The coordinates again are found by ODS (2024). The great-circle distance is calculated by the Haversine formula (Agramanisti Azdy & Darnis, 2020). Using this formula is much less time-consuming than measuring distances in route planners, without sacrificing much in terms of accuracy.

Non-HSR travel time & travel cost: Rome2Rio (2024) provides great insights into all ways to travel from origin to destination, for plane, train, bus, ferry and car. For each way of travel, it provides the total door-to-door travel time (including waiting time) and travel costs. For the latter, it also looks at minimum and maximum values for travelling the route indicated, taking into account periodical price variations. It includes all possible travel modes, regardless of their likelihood to be used. Arguably the greatest benefit of using this source is the fact it shows real data that people weighing off their travel options would also use. Data is gathered through web scraping, for which a Python code is written, which makes use of the dedicated Selenium package.

HSR travel time & travel cost: Travel cost is a design choice, related to fare setting. The methodology behind that is explained in section 3.3. Travel time will be deduced from literature. In China, the average speed is 90% of the design speed v^{max}, which is the peak efficiency observed worldwide (Zhang & Zhang, 2021). Therefore, this value k^{eff} will be used in calculating travel times:

$$TT_{HSR,ij} = \frac{l_{ij}}{k^{eff} \cdot v^{max}} + t^{dwell} + t^{access} + t^{egress} \tag{3}$$

Here, l_{ij} is the line length between station i and j, while the t-parameters represent the dwelling, access and egress time, respectively. The dwelling time is set to 5 minutes (Grolle et al., 2024), while the access and egress time summed are set to 30 minutes (Sane, 2020).

Model calibration

Our complete model as presented in equation (1) cannot be calibrated, as observed market shares are scarcely reported. Added to this, numerous examples of long-distance travel studies have investigated this matter – unlike gravity model studies seeking for the ideal catchment area size. Only the air demand counts as observed data. It is for this reason that the model is split into two parts for calibration: logit and gravity.

Logit: The β^{TT} and β^{TC}-parameters can be estimated easily by statistical analysis of the MNL studies among the 100 demand studies introduced in section 2. The median of all reported values is taken here, and it will be checked if the attained Value of Time (VoT) is acceptable.

Gravity: To calibrate a gravity model, it is usually written in a log-linear form (Grosche et al., 2007):

$$\log(D_{HSR,ij}) = k + \alpha \log(P_i \cdot P_j) + \beta \log(GDP_i \cdot GDP_j) - \gamma \log(d_{ij}) \tag{4}$$

The model will be calibrated by the Ordinary Least Squares (OLS) method.

Bias elimination

The catchment area size is kept constant throughout the calibration process, despite greatly varying along with city and airport size (Lieshout, 2012; Grosche et al., 2007). This might induce a bias, where the demand related to large cities is underestimated and vice versa. In order to attain a non-biased forecasting model, the bias is eliminated by scaling the model output, based on the magnitude of the value itself. When plotting the log of predicted demand (here called: y) versus the log of observed demand (x), its trendline is described by $y = ax + b$. For an unbiased model, $a = 1$ and $b = 0$, which simplifies to $y = x$. In any other case, the model is biased and therefore lacks accuracy. This bias can be eliminated by transforming the y_{old} value, following the following formula:

$$y_{new} = \frac{y_{old} - b}{a} \tag{5}$$

Model validation

The model is validated by taking the same approach as by Belal et al. (2020): we take five of the connections with the highest estimated demand, while not selecting a city more than once and calculate how closely the model approaches the observed demand.

Demand evolution

In order to accurately forecast demand into the future, section 2.1 showed we have to take induced demand and economic demand growth into account. This study does so by calculating a factor for both influences, specific to each OD pair. The final demand forecast will then be a multiplication of both factors and the original demand forecast.

Induced demand: Preston (2013) mentions the level of induced demand for multiple European high-speed rail projects which are in operation, with most values around and above 20%. To avoid overestimation, which would result in a very optimistic and perhaps unrealistic network design, the multiplication factor is set to 1.20.

Economic demand growth: Trafikverket (2021) estimated a conservative value of 0.7 for GDP elasticity with respect to demand. Again, to limit the risk of demand overestimation, this value will be used here as well. World Bank (2023) reports yearly GDP growth for every country in the world between 1961 and 2023. For this project, the mean yearly growth rate p from the last 40 years is taken into account for each country within the scope. For a connection ij between two countries, the minimum yearly growth rate among them is considered normative for economy-based demand growth. This value p is used to compute the factor k_{ij}^{eco} as follows:

$$k_{ij}^{eco} = \frac{1}{T^{life} + 1} \sum_{t=0}^{T^{life}} \left(1 + \frac{p_{ij} \cdot e_{GDP}}{100}\right)^t \tag{6}$$

Here, T^{life} is the project's lifetime in years, with $t \in \{0,1,2, \dots, T^{life}\}$. p_{ij} is the minimum economic growth rate among the two countries related to city pair ij as a percentage, and e_{GDP} represents the demand-related GDP elasticity.

3.3. Profitability estimation

In order to estimate the profitability of a potential HSR line, each of the cash-flows introduced in Table 1 will be attributed a formula.

Ticket revenue: The revenue depends on the outcome of the product of the ticket fare and demand. The fare setting is a design choice. Operators generally set a price that maximises their passenger revenue (Qin et al., 2019). This study will follow the same approach, aided by Python library *SciPy.* It can be mathematically proven that optimising for maximum revenue will always yield exactly one, nonnegative, optimal fare setting.

Construction costs: It was determined that the total costs depend on the line length, and a unit cost for each km $k_{ij}^{X,infra}$. This unit cost depends on the location and difference in height. Borgogno (2023) quantifies these relationships and produces unit construction cost per km, for surface ($C^{surface}$) and tunnelling (C^{tunnel}) separately for European countries. It is assumed that $k_{ij}^{X,infra}$ is a result of a convex combination of $C^{surface}$ and C^{tunnel}, dependent on a normalised height difference. Thus, the maximum possible height difference between two cities in Europe is set to 1 and the minimum is set to 0, with linear interpolation in between. The formula calculating the total infrastructure construction cost is displayed as:

$$C_{ij}^{X,infra} = k_{ij}^{X,infra} \cdot l_{ij} \tag{7}$$

Acquisition costs: As mentioned, these depend solely on the number of train sets bought and a unit price. The minimum number of trains needed to operate a line is product of the frequency and the full round-trip time:

$$C_{ij}^{X,train} = k^{X,train} \cdot \lceil 2 \cdot f_{ij} \cdot t_{ij} \rceil \tag{8}$$

Infrastructure maintenance & operation costs: As stated, these are calculated based on yearly sum per km $k^{T,infra}$.

$$C_{ij}^{T,infra} = k^{T,infra} \cdot l_{ij} \tag{9}$$

Rolling stock maintenance & operation costs: These are calculated based on a value per seat-km $k^{T,train}$, and has to be multiplied with a number of factors in order to acquire the total yearly costs:

$$C_{ij}^{T,train} = k^{T,train} \cdot s \cdot H \cdot D \cdot \frac{n_{ij} \cdot l_{ij}}{t_{ij}} \tag{10}$$

The total profitability estimate of a HSR connection is a summation of all cash-flows, where the costs are negative. The meaning of all variables used in equations 7-10 are explained in Table 4 below.

Table 4 Nomenclature of profitability estimation model

Parameter	Unit	Definition
$C_{ij}^{X,infra}$	[€]	infrastructure construction costs between city i and j
$C_{ij}^{X,train}$	[€]	rolling stock acquisition costs between city i and j
$C_{ij}^{T,infra}$	[€/year]	infrastructure operation & maintenance costs between city i and j
$C_{ij}^{T,train}$	[€/year]	rolling stock operation & maintenance costs between city i and j
$k^{X,infra}$	[€/km]	unit infrastructure construction cost
$k^{X,train}$	[€/train]	unit rolling stock acquisition cost
$k^{T,infra}$	[€/km/year]	unit infrastructure operation & maintenance cost
$k^{T,train}$	[€/seat-km]	unit rolling stock operation & maintenance cost
s	[pax]	seats per train set
H	[h/day]	operating hours per day
D	[day/year]	operating days per year
v^{max}	[km/h]	maximum operating speed
T^{life}	[year]	project lifetime
l_{ij}	[km]	distance between city i and j
t_{ij}	[h]	travel time between city i and j
n_{ij}	[-]	trains to serve demand between city i and j

Upgradeable lines

In some cases, it makes no economic sense to build new costly high-speed rail infrastructure, and rather upgrade existing facilities. This does not apply to all infrastructure: according to UIC (2018), only dedicated lines having maximum speeds of at least 250 km/h could be further upgraded to 350 km/h. As the construction of 350 km/h HSR infrastructure is roughly twice as expensive as for 250 km/h (Preston, 2013), a 50% construction cost discount is applied, if the connection already provides a rail connection with an average travel speed exceeding 200 km/h, which can be calculated through previously garnered data through web scraping. In this case, the level of induced demand will be halved as well.

3.4. Network design

In order to create the final network design, based on demand forecasts and profitability estimates, a TNDFSP must be formulated, fitting the very nature of our field. Mathematical programming is chosen to find optimal solutions. Following the mentioned literature gap, it was decided to incorporate demand elasticity into the model, while passengers are assigned to the shortest path.

3.4.1. Model inspiration

Introduced as a promising formulation, the Multi-Commodity Flow Problem (MCFP), leading to an "efficient formulation" to handle "city-scale transit networks" (Ng et al., 2024). The philosophy behind the MCFP is explained by Magnanti & Wong (1984). The problem concerns finding the lowest cost of sending commodities (goods or people) through a network. A general formulation is provided in the equations below:

$$\min \sum_{a \in A} \sum_{k \in K} c_{ak} \cdot x_{ak} \tag{11}$$

$$\text{s.t.:} \quad \sum_{a \in A_i^{out}} x_{ak} - \sum_{a \in A_i^{in}} x_{ak} = \begin{cases} D_k & if\ i \in \boldsymbol{O(k)} \\ -D_k & if\ i \in \boldsymbol{D(k)} \\ 0 & otherwise\ \forall i \in \boldsymbol{N}, a \in \boldsymbol{A} \end{cases} \tag{12}$$

Here, c_{ak} is the unit cost related to transporting commodity k over arc a, while x_{ak} refers to the number of commodities k transported over arc a. The objective function sums for all $k \in \boldsymbol{K}$ and all $a \in \boldsymbol{A}$. The set of all arcs $\boldsymbol{A}$ is split into smaller sets $\boldsymbol{A_i^{in}}$ and $\boldsymbol{A_i^{out}}$, representing the set of arcs going in and out from node $i \in \boldsymbol{N}$, respectively. Equation (12) guarantees node continuity, meaning that the amount of a commodity going in a node equals the amount of that commodity going out of that node. Unless that node is the origin of that commodity k, being the set of $\boldsymbol{O(k)}$. Here the flow only goes outwards, meaning that the equation should equal the demand of commodity k, also named D_k. The same in the other direction holds for the destination of the commodity, defined by set $\boldsymbol{D(k)}$. The MCFP formulation will be adapted to the nature of this project, firstly to incorporate elements of network design (section 3.4.2), then of line design (section 3.4.3).

3.4.2. Network design elements

In order to construct an objective function and constraints, a number of elements must be introduced:

Nodes: We define the set of nodes $\boldsymbol{N}$, equivalent to the MCFP set of nodes. To indicate if node n is selected for our final network, we define a binary decision variable z_n.

Arcs: We also define the set of undirected arcs $\boldsymbol{A}$. Binary decision variables y_a are defined to indicate if arc a is selected for the final network. Each arc has been attributed with a travel time t_a, travel cost c_a, length d_a and lifetime cost f_a^{cost}.

OD pairs: The original MCFP set of commodities is translated to a set of OD pairs $\boldsymbol{P}$, as their meaning is equivalent for this problem.

OD flow routes: The main problem found with exact approaches was that they can only solve within reasonable running times for small to medium-sized networks. In the MCFP formulation, for each commodity k, decision variables are considered whether it is transported along arc a. Inspired by the work of Oliker & Bekhor (2020) and Liang et al. (2019), this can be reformulated without changing the intrinsic nature and optimal solution: for each commodity k, we can define a set of potential routes across the network instead. A route would then be defined as a valid sequence of arcs from origin to destination. The same is done for this project's model; to do so, we define the set of OD flow routes $\boldsymbol{R}$, inspired by the work of Arbex & da Cunha (2015). Each OD flow route r can be selected or not for our final network, which can be indicated by binary decision variable x_r. An algorithm must be written in order generate all possible OD flow routes. Each route is attributed with a travel time t_r, travel cost c_r and length d_r. Binary parameters are introduced in order to couple each OD flow route r to nodes, arcs and OD pairs:

- c_{nr}^{node}: whether OD flow route r covers node n (1) or not (0)
- c_{ar}^{arc}: whether OD flow route r covers arc a
- c_{pr}^{odpair}: whether OD flow route r covers OD pair p
- m_{pr}^{odpair}: whether OD flow route r matches OD pair p

Demand: Since the relationship between travel time and demand was found to be non-linear by our forecasting model, it has to be linearised for this mathematical programming model:

$$D_{HSR,p} = \alpha_p + \beta_p \cdot t_p + \gamma_p \cdot c_p \tag{13}$$

Here, the demand $D_{HSR,p}$ is estimated by a 'max demand' α_p for OD pair $p \in \boldsymbol{P}$ (when travel time and costs are zero), an associated 'time decay' β_p and 'cost decay' γ_p, indicating by how much the demand would reduce for respectively one hour extra travel time t_p, or one euro in extra travel cost c_p. The yearly flow on route r is referred to as q_r^{year}, calculated by:

$$q_r^{year} = \alpha_r + \beta_r \cdot t_r + \gamma_r \cdot c_r \tag{14}$$

A trendline will be fitted for the interval $t_r \in [t_r,\ k^{detour} \cdot t_r]$; $c_r \in [c_r,\ k^{detour} \cdot c_r]$. In this regard, k^{detour} is the maximum allowed detour factor, which is set at 1.25, matching the value used in European HSR network design by Grolle et al. (2024).

Ticket revenue: Recall that the objective of the formulation is a maximisation of lifetime profitability, and that revenue was found to be a component of that. We define the lifetime revenue f_r^{rev} for route r as the product of its demand per year q_r^{year}, fare price c_r and the lifetime of the project T^{life} in years.

Now all variables and parameters are defined, the constraints can be constructed.

(1) Node selection

A node n is selected if a selected OD flow route r flows over it (15). Also, a node n cannot be selected if no selected OD flow route r flows over it (16).

$$z_n \geq x_r \cdot c_{nr}^{node} \qquad \forall n \in \boldsymbol{N}, \forall r \in \boldsymbol{R} \tag{15}$$

$$z_n \leq \sum_{r \in \boldsymbol{R}} (x_r \cdot c_{nr}^{node}) \qquad \forall n \in \boldsymbol{N} \tag{16}$$

(2) Arc selection

An arc a must be selected if a selected OD flow route r flows over it (17). As arc selection automatically induces costs, a constraint oriented to the opposite such as (16) can be disregarded – it is illogical for the model to build arcs if no flow exists on it, as it maximises for profitability.

$$y_a \geq x_r \cdot c_{ar}^{arc} \qquad \forall a \in \boldsymbol{A}, \forall r \in \boldsymbol{R} \tag{17}$$

(3) OD pair selection

An OD pair p may be served by at most one selected OD flow route r (18).

$$\sum_{r \in \boldsymbol{R}} (x_r \cdot m_{pr}^{ODpair}) \leq 1 \qquad \forall p \in \boldsymbol{P} \tag{18}$$

(4) Minimum node separation

Having too short distances between selected nodes undermines HSR's rationale (Rodrigue, 2017). Therefore, a minimum distance l^{min} between nodes is introduced. To impose the constraint, firstly a set of all node pairs $(i,j) \in \boldsymbol{N}$ that would violate l^{min} is introduced: $\boldsymbol{N^{close}}$. Following up on this, we can define the following constraints to ensure that the minimum distance between nodes is respected:

$$z_i + z_j \leq 1 \qquad \forall (i,j) \in \boldsymbol{N^{close}} \tag{19}$$

(5) Non-crossing arcs

Selected arcs are not allowed to cross each other. To constrain this, we defined the set of all crossing arc pairs (i,j) as $\boldsymbol{A^{cross}}$, with $a, b \in \boldsymbol{A}$ and $a \neq b$. Then, the following constraints should hold:

$$y_i + y_j \leq 1 \qquad \forall (i,j) \in \boldsymbol{A^{cross}} \tag{20}$$

(6) Decision variables

As stated before: x_r, z_n and y_a are all binary:

$$x_r \in \{0,1\} \qquad \forall r \in \boldsymbol{R} \tag{21}$$
$$y_a \in \{0,1\} \qquad \forall a \in \boldsymbol{A} \tag{22}$$
$$z_n \in \{0,1\} \qquad \forall n \in \boldsymbol{N} \tag{23}$$

3.4.3.Line design elements

In order to construct an objective function and constraints, a number of elements must be introduced:

Operating frame: the number of operating hours per day H is set to 18, and the number of operating days per year D is set to 365.

Valid operating lines: each route r will be attributed a binary value l_r, which equals 1 if it is considered a valid operating line and 0 if not. An OD flow route r is considered an invalid operating line if it has no intermediate stops, and if the total travel time along the entire route is more than 9 hours.

Line frequencies: a nonnegative integer decision variable w_r denoting the frequency of route r is introduced, along with a universal maximum frequency W^{max}, here set to 12 trains per hour, matching minimum headway rules for the same 350 km/h trains in the Chinese network (Tian & Zhang, 2024).

Number of trains: This is related to the operated frequency w_r on route r. The associated nonnegative integer decision variable is defined as n_r^{train}. The seat capacity of each train is denoted by s.

Peak hour demand: This is calculated by calculating the average demand per operating hour throughout the year, while multiplying with a peak hour factor k^{peakhr}, set to 1.25. This factor is a conversion factor between peak hour flow and average hourly flow. The peak hour flow q_r^{peak} is then calculated as follows:

$$q_r^{peak} = \frac{k^{peakhr}}{2 \cdot D \cdot H} \cdot q_r^{year} \tag{24}$$

The division by two is performed in order to convert to flow per direction, as the flow q_r^{year} combines flow in both directions.

Now all variables and parameters are defined, the constraints can be constructed.

(1) Frequency setting

An operating line r cannot be selected if the corresponding OD flow route r is not selected, or when it is not considered a valid operating line. Also, maximum operating frequencies W^{max} universally apply.

$$w_r \leq W^{max} \cdot l_r \cdot x_r \qquad \forall r \in \boldsymbol{R} \tag{25}$$

(2) Serve all demand

Together, the selected valid operating lines r must serve all demand in entire network.

$$s \cdot \sum_{r \in \boldsymbol{R}} (w_r \cdot c_{ar}^{arc}) \geq \sum_{r \in \boldsymbol{R}} (x_r \cdot c_{ar}^{arc} \cdot q_r^{peak}) \qquad \forall a \in \boldsymbol{A} \tag{26}$$

(3) Rolling stock acquisition

Acquire the correct number of trains n_r^{train} for each operating line r. It should equal $\lceil 2 \cdot w_r \cdot t_r \rceil$ (terms corrected for the definitions in this section). This can be linearly defined:

$$n_r^{train} - 1 \leq 2 \cdot w_r \cdot t_r \qquad \forall r \in \boldsymbol{R} \tag{27}$$
$$n_r^{train} \geq 2 \cdot w_r \cdot t_r \qquad \forall r \in \boldsymbol{R} \tag{28}$$

Transfer penalties

In order to address the user's perspective in the objective function as well, as monetary penalty is added for every passenger who has to transfer. de Keizer et al. (2015) provide an added perceived travel time of 22.63 minutes if a trip is not without transfer. The real transfer time is included in this value as well, allowing for calculation of monetary valued penalties for every passenger who has to transfer. Wardman et al. (2012) provide a common European VoT of €14.80 per hour. The penalty per passenger having to transfer thus equals $(22.63/60) \cdot €14.80 \approx €5.58$.

To include this into the objective function, a binary decision variable T_p, which indicates whether the passengers from OD pair p are served with (1) or without (0) transfer. Given the definitions of this model, the following situations must both occur for an OD pair p in order for a transfer penalty to be imposed:

3. The OD pair p must be served. This means that any OD flow route r **exactly matching** the OD pair p **must be selected**. We therefore introduce a binary decision variable u_p indicating whether this statement is true (1) or false (0).
4. The OD pair p may not be served directly. This means that any OD flow route r **covering** that OD pair p **may not be selected**. We therefore introduce a binary decision variable v_p indicating whether this statement is true (0) or false (1)

Note the reverse order of values to the Booleans, as this allows for construction of the following equation:

$$T_p = u_p - v_p \qquad \forall p \in \boldsymbol{P} \tag{29}$$

We can check that this equation is formulated correctly: it only equals 1 if an OD pair p is served, but not directly. As non-active OD pairs p cannot be served by selected operating lines it is constrained that $u_p \geq v_p, \forall p \in \boldsymbol{P}$. This way, it is also constrained that the outcome of $u_p - v_p$ is always binary, making the separate addition of decision variable T_p unnecessary. Both u_p and v_p are correctly defined by the constraints below:

(4) Served OD pairs

An OD pair p is served if the OD flow route r exactly matching that OD pair p is selected (31). If not, the OD pair is not served (30).

$$u_p \leq \sum_{r \in \boldsymbol{R}} (x_r \cdot m_{pr}^{ODpair}) \qquad \forall p \in \boldsymbol{P} \tag{30}$$
$$u_p \geq x_r \cdot m_{pr}^{ODpair} \qquad \forall p \in \boldsymbol{P}, \forall r \in \boldsymbol{R} \tag{31}$$

(5) Served OD pairs, without transfer

An OD pair p is served directly if any OD flow route r covering that OD pair p is selected (33). If none of these routes r are selected, the OD pair is not served directly (32). In the second constraint below, M is an arbitrary large constant.

$$v_p \leq \sum_{r \in \boldsymbol{R}} (w_r \cdot c_{pr}^{ODpair}) \qquad \forall p \in \boldsymbol{P} \tag{32}$$
$$M \cdot v_p \geq \sum_{r \in \boldsymbol{R}} (w_r \cdot c_{pr}^{ODpair}) \qquad \forall p \in \boldsymbol{P} \tag{33}$$

(6) Decision variable boundaries

As stated before, u_p and v_p are binary, while w_r and n_r^{train} are nonnegative and integer:

$$w_r \in \mathbb{R}_{\geq 0} \qquad \forall r \in \boldsymbol{R} \tag{34}$$
$$n_r^{train} \in \mathbb{R}_{\geq 0} \qquad \forall r \in \boldsymbol{R} \tag{35}$$
$$u_p \in \{0,1\} \qquad \forall p \in \boldsymbol{P} \tag{36}$$
$$v_p \in \{0,1\} \qquad \forall p \in \boldsymbol{P} \tag{37}$$

3.4.4. Notation & formulation

The full model can now be presented as:

$$
\begin{aligned}
max \quad & \sum_{r\in\boldsymbol{R}}(x_r \cdot f_r^{rev}) - \sum_{a\in\boldsymbol{A}}(y_a \cdot f_a^{cost}) \\
& - \left[k^X \cdot \sum_{r\in\boldsymbol{R}}(n_r^{train})\right] - \left[T^{life} \cdot k^T \cdot s \cdot H \cdot D \cdot \sum_{r\in\boldsymbol{R}}\left(\frac{n_r^{train} \cdot d_r}{t_r}\right)\right] \\
& - \left[T^{life} \cdot k^{transfer} \cdot \sum_{p\in\boldsymbol{P}}\left[(u_p - v_p) \cdot \sum_{r\in\boldsymbol{R}}(c_{pr}^{ODpair} \cdot q_r^{year})\right]\right] \qquad (38)
\end{aligned}
$$

$$
\begin{aligned}
s.t. \quad & z_n \geq x_r \cdot c_{nr}^{node} && \forall n \in \boldsymbol{N}, \forall r \in \boldsymbol{R} && (39)\\
& z_n \leq \sum_{r\in\boldsymbol{R}}(x_r \cdot c_{nr}^{node}) && \forall n \in \boldsymbol{N} && (40)\\
& y_a \geq x_r \cdot c_{ar}^{arc} && \forall a \in \boldsymbol{A}, \forall r \in \boldsymbol{R} && (41)\\
& \sum_{r\in\boldsymbol{R}}(x_r \cdot m_{pr}^{ODpair}) \leq 1 && \forall p \in \boldsymbol{P} && (42)\\
& z_i + z_j \leq 1 && \forall (i,j) \in \boldsymbol{N}^{close} && (43)\\
& y_i + y_j \leq 1 && \forall (i,j) \in \boldsymbol{A}^{cross} && (44)\\
& w_r \leq W^{max} \cdot l_r \cdot x_r && \forall r \in \boldsymbol{R} && (45)\\
& s \cdot \sum_{r\in\boldsymbol{R}}(w_r \cdot c_{ar}^{arc}) \geq \sum_{r\in\boldsymbol{R}}(x_r \cdot c_{ar}^{arc} \cdot q_r^{peak}) && \forall a \in \boldsymbol{A} && (46)\\
& n_r^{train} - 1 \leq 2 \cdot w_r \cdot t_r && \forall r \in \boldsymbol{R} && (47)\\
& n_r^{train} \geq 2 \cdot w_r \cdot t_r && \forall r \in \boldsymbol{R} && (48)\\
& u_p \leq \sum_{r\in\boldsymbol{R}}(x_r \cdot m_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (49)\\
& u_p \geq x_r \cdot m_{pr}^{ODpair} && \forall p \in \boldsymbol{P}, \forall r \in \boldsymbol{R} && (50)\\
& v_p \leq \sum_{r\in\boldsymbol{R}}(w_r \cdot c_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (51)\\
& M \cdot v_p \geq \sum_{r\in\boldsymbol{R}}(w_r \cdot c_{pr}^{ODpair}) && \forall p \in \boldsymbol{P} && (52)\\
& u_p \geq v_p && \forall p \in \boldsymbol{P} && (53)\\
& x_r \in \{0,1\} && \forall r \in \boldsymbol{R} && (54)\\
& y_a \in \{0,1\} && \forall a \in \boldsymbol{A} && (55)\\
& z_n \in \{0,1\} && \forall n \in \boldsymbol{N} && (56)\\
& w_r \in \mathbb{R}_{\geq 0} && \forall r \in \boldsymbol{R} && (57)\\
& n_r^{train} \in \mathbb{R}_{\geq 0} && \forall r \in \boldsymbol{R} && (58)\\
& u_p \in \{0,1\} && \forall p \in \boldsymbol{P} && (59)\\
& v_p \in \{0,1\} && \forall p \in \boldsymbol{P} && (60)
\end{aligned}
$$

The twenty-three equations above form a Mixed-Integer Linear Program (MILP). Equation (38) is the objective function, which maximises profit while also taking the user perspective into account. It consists out of four parts, here listed in the order of how they are presented above:

1. Ticket revenue is simply the sum of the revenue of selected lines;
2. Infrastructure costs are simply the sum of the lifetime costs of selected arcs;
3. Rolling stock costs split in acquisition costs (the first part) and operating & maintenance costs (second part);
4. Transfer penalties, only added if decision variables u_p and v_p have the correct values.

A nomenclature to the formulation mentioned above is provided in Table 5.

3.4.5. Pre-processing

This aims to reduce the sizes of sets that are input for the optimisation model. These are the set of nodes N, arcs A, OD pairs P ('network simplification') and OD pair flow routes R ('route generation').

Network simplification: Frei et al. (2010) cites multiple European long-distance travel studies, who all set 100 km as a minimum threshold regarding arc length. In Europe, the longest distance between two stations is 253 km (European Court of Auditors, 2018). Therefore, considering all potential arcs is unnecessary, as many of them are not within the boundaries of workable length. To have a safe margin, it is decided to set 500 km as the maximum allowed arc length. This also matches the maximum distance over which European high-speed rail generally is competitive with other transport modes, according to last-mentioned authors. The 100 km minimum bound by Frei et al. (2010) will be used for this project as well. Arcs with lengths outside of these boundaries will be deleted from the arcs set, and the sets of nodes and OD pairs are updated accordingly.

Table 5 Nomenclature of optimisation model

	Sets	
$\boldsymbol{N}$	$n \in \boldsymbol{N}$	Set of nodes
$\boldsymbol{A}$	$a \in \boldsymbol{A}$	Set of arcs
$\boldsymbol{P}$	$p \in \boldsymbol{P}$	Set of OD pairs
$\boldsymbol{R}$	$r \in \boldsymbol{R}$	Set of OD flow routes (and potential operating lines)
$\boldsymbol{N}^{close}$	$(i,j) \in \boldsymbol{N}^{close}$	Set of node pairs too close together (< 100 km)
$\boldsymbol{A}^{cross}$	$(i,j) \in \boldsymbol{A}^{cross}$	Set of arc pairs crossing each other
	Decision variables	
x_r	[-]	Whether OD flow route $r \in \boldsymbol{R}$ is selected
y_a	[-]	Whether arc $a \in \boldsymbol{A}$ is selected
z_n	[-]	Whether node $n \in \boldsymbol{N}$ is selected
n_r^{train}	[trains]	Number of trains acquired to operate line $r \in \boldsymbol{R}$
w_r	[trains / hr]	Operating frequency of line $r \in \boldsymbol{R}$
u_p	[-]	Whether OD pair $p \in \boldsymbol{P}$ flows over the network
v_p	[-]	Whether OD pair $p \in \boldsymbol{P}$ is served without transfer
	Parameters	
f_r^{rev}	[€]	Lifetime revenue for OD flow route $r \in \boldsymbol{R}$
f_a^{cost}	[€]	Lifetime cost for arc $a \in \boldsymbol{A}$
k^T	[€ / (paxkm)]	Unit rolling stock operating and maintenance cost
k^X	[€ / train]	Unit rolling stock acquisition cost
$k^{transfer}$	[€ / pax]	Unit transfer penalty
T^{life}	[years]	Project lifetime
H	[hours / day]	Operating hours per day
D	[days / year]	Operating days per year
s	[pax / train]	Seat capacity
W^{max}	[trains / hr]	Maximum allowed frequency
c_{nr}^{node}	[-]	Whether OD flow route $r \in \boldsymbol{R}$ covers node $n \in \boldsymbol{N}$
c_{ar}^{arc}	[-]	Whether OD flow route $r \in \boldsymbol{R}$ covers arc $a \in \boldsymbol{A}$
c_{pr}^{ODpair}	[-]	Whether OD flow route $r \in \boldsymbol{R}$ covers OD pair $p \in \boldsymbol{P}$
m_{pr}^{ODpair}	[-]	Whether OD flow route $r \in \boldsymbol{R}$ matches OD pair $p \in \boldsymbol{P}$
q_r^{year}	[pax / year]	Yearly demand OD flow route $r \in \boldsymbol{R}$ (both directions)
q_r^{peak}	[pax / hr]	Peak hour demand OD flow route $r \in \boldsymbol{R}$ (per direction)
t_r	[hr]	Travel time along route $r \in \boldsymbol{R}$
d_r	[km]	Length of route $r \in \boldsymbol{R}$
l_r	[-]	Whether OD flow route $r \in \boldsymbol{R}$ is a valid operating line
M	[-]	Arbitrarily large constant

Route generation: The goal of this algorithm is to find all valid routes for each OD pair's passengers. What makes a route 'valid' or 'invalid' will be described later in this section. Following the previously mentioned reasoning of Magnanti & Wong (1984), the potential number of routes is extremely large, increasing exponentially with the size of the network. An efficient algorithm must be used in order to find all routes. The Python package 'NetworkX' will be used, which is known for efficiently analysing the structure of networks. Our algorithm will iterate over the OD pairs $p \in P$, and follow the steps below for each p.

- Define set N_p as all nodes the flow of OD pair p can traverse while keeping within detour boundaries.
- Define the set A_p as all arcs that connect two nodes in N_p. Then, for OD pair p, the algorithm only considers graph $G_p = (N_p, A_p)$, which can be considerably smaller than $G = (N, A)$.

Using breadth-first strategy (BFS), which is more efficient when attempting to find shortest paths (Rocha & Ferreira, 2018), all possible routes are explored. A path is not explored any further if:

- It visits a node for the second time;
- The added node moves closer to the path's origin node;
- The angle of deviation between two consecutive arcs exceeds a half turn (90 degree);
- The cumulative angle of deviation exceeds a three-quarter turn (135 degrees);
- The route fails to generate a positive level of demand

The outcome for each OD pair p is a list of routes, defined as sequences of arcs. The algorithm will be ran over all $p \in \boldsymbol{P}$ to produce a complete set of routes $\boldsymbol{R}$.

3.4.6. Optimisation

The model formulation for network and line design will be written in Python 3.7.13 code language, with a loaded in Gurobi Optimizer 9.5.2. The mathematical problem related to this project will be solved on a Lenovo laptop with 2.3GHz Intel i7-11800H CPU and 32GB of RAM, and 64-bit Windows 11 as OS.

Since it will be unknown a priori how the running time of the optimisation depends on the number of cities taken into account, it is decided to rank them based on their population. At first, the top two cities in terms of population make up the set of nodes. One city will be added at a time, as long as the running time of the model stays within a reasonable duration of six hours. It is assumed that after some point, no new cities will be added to the optimal network as their size in population has become too small to make an impact. As discussed earlier, the Chinese HSR network is considered a leading example in HSR network potential. This network aims to connect all cities with over 500,000 inhabitants (China Daily, 2020). It is aimed for to take all European cities meeting this criterion at the least. Literature review offered no findings of a ‘minimum population’. Therefore, we hope to include more cities, lowering this inhabitant bound by as much as possible, as it increases the result’s scientific value.

3.4.7.Model validation

The workings of the model will be validated by means of a model stability analysis and benchmarking.

Model stability analysis: This will be performed by randomly varying the fare settings of a smaller instance of the network, by a maximum of 10% from the original value. This test is run six times, and possible changes from the original optimal solution will be assessed.

Benchmarking: To compare the outcomes of our model with state-of-the-art, it will be benchmarked to the network of Mandl (1980), which is the most widely known benchmark problems (Kechagiopoulos & Beligiannis, 2014). Benchmarking comes down to validating the optimisation model. The network encompasses 15 nodes (the same as our optimal solution) and 21 arcs, resembling a number of cities in Switzerland. Similar to our problem, demand is exerted bidirectionally, with peak hour demand already provided. In total, this demand equals 15,570.

4 RESULTS

4.1. Demand forecasting

4.1.1.Data collection for calibration

Air demand: After pre-processing, demand data was found for 514 city pairs, encompassing 71 unique cities and 320 million passengers, with the yearly demand per city pair varying from 34,000 to 7.2 million.

Demand-impacting factors: The data shows great variation among all variables, indicated by the standard deviation being larger than the mean value. This is great news, as the calibrated model should be able to cope with a wide range of population catchments $P_{i,t}$. The same can be said about the other variable that relates to people: the number of air passengers $D_{AIR,ij}$. GDP (city total GDP_i or per capita $GDPCAP_i$) and distances d_{ij} vary much less, which does not impose a problem, since the choice of locations is spread well across the continent. Multiple scenarios regarding the catchment area size and GDP measure enter the calibration stage, as we seek for the best model fit. Therefore, all possible scenarios are included in Table 6 below, representing the key characteristics of the data used for model calibration.

Table 6 Model calibration characteristics

Var.	Unit	N	MIN	MEDIAN	MAX	AVG	STD
$P_{i,15}$	[pax]	71	923	253,994	2,342,971	442,383	474,620
$P_{i,30}$	[pax]	71	4,133	708,944	9,974,841	1,265,734	1,548,360
$P_{i,45}$	[pax]	71	12,388	1,113,145	13,417,545	1,916,266	2,332,799
$P_{i,60}$	[pax]	71	36,904	1,515,685	14,474,396	2,504,137	2,835,263
$P_{i,75}$	[pax]	71	43,916	1,952,954	15,108,562	3,121,031	3,270,876
$P_{i,90}$	[pax]	71	59,127	2,580,338	17,432,372	3,820,950	3,768,015
$P_{i,105}$	[pax]	71	68,970	3,298,628	20,158,470	4,618,074	4,424,464
$P_{i,120}$	[pax]	71	72,672	3,646,519	23,084,990	5,473,373	5,280,943
GDP_i	[M€]	71	1,178	35,447	757,630	84,055	127,604
CAP_i	[K€]	71	5.980	58.562	607.256	83.758	84.513
d_{ij}	[km]	514	1,001	1,517	3,364	1,623	482
$D_{AIR,ij}$	[pax]	514	34,660	309,341	7,209,728	622,587	871,726

N = number of values; MIN = minimum; MAX=maximum; AVG=average; STD=standard deviation; K€ = thousand euros; M€ = million euros; CAP=GDP/capita

4.1.2.Model calibration

This section presents the calibration results for both the gravity and logit part of the model.

Gravity model

The gravity model was calibrated for various catchment area sizes, as well as for both total GDP and GDP per capita, to see what GDP indicator explains the data best. The R^2 fit for both models at various catchment area sizes are plotted in Figure 3.

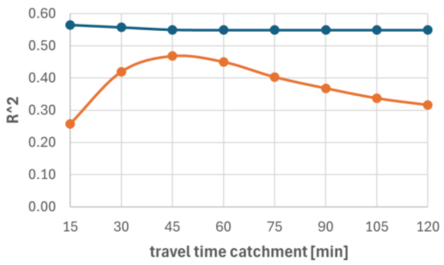


Figure 3 Model fit for varying catchments and GDP measures

In terms of R^2, the GDP model outperforms the GDP/cap model for all catchment area sizes. However, this seems due to overfitting, as most calibrated parameters are not showing a decent statistical significancy, and some are of the wrong sign. The likely cause is the intercorrelation between two of the GDP model’s variables: population and total GDP. The GDP/cap model’s fit varies with the choice of catchment area size, with an optimal fit at 45 minutes. This aligns with the findings of Martínez et al. (2016): 80% of HSR users live within a 30 min travel. Even though the GDP/cap model has a poorer fit than the total GDP model, most of its calibrated parameters are showing great statistical significance, all of the correct sign. Due to the significant parameters, the model is likely to represent true relationships in the data, making it easier to interpret and trust. This also enhances predictive power, as the model likely captures real effects. For these reasons, the GDP/cap model is chosen, as it balances a good fit with great parameter significance. Below, Table 7 shows the calibration results for the chosen catchment area size and GDP measure.

Table 7 Calibration results for optimal model

Parameter	Coeff.		Std	t-stat	p-value
k	-2.524	**	0.606	-4.16	3.74×10^{-5}
α	0.564	**	0.028	20.31	0
β	0.382	**	0.046	8.22	1.55×10^{-15}
γ	0.139		0.13	1.03	3.02×10^{-1}

Est. = estimated value; Std = standard deviation; Significant at conf. level: 95% (*), 99% (**)

All parameters’ estimates of the expected sign, and all but distance parameter γ are showing extreme significance. This means that in the data, the relationship between distance and demand cannot be picked up as easily as for other model-assumed relationships. The parameter is still significant at a 70% confidence level.

The model fits the data reasonably well, reaching a R^2 value of 0.468. However, it is biased in the sense that it overestimates demand involving small cities and vice versa, as expected in section 3.2. When plotting observed data against model-predicted data, we attain a trendline $y = ax + b$ with $a = 0.468$ and $b = 2.926$. After bias elimination, the model fit is increased to a respectable 0.751, and the plot can be presented as:

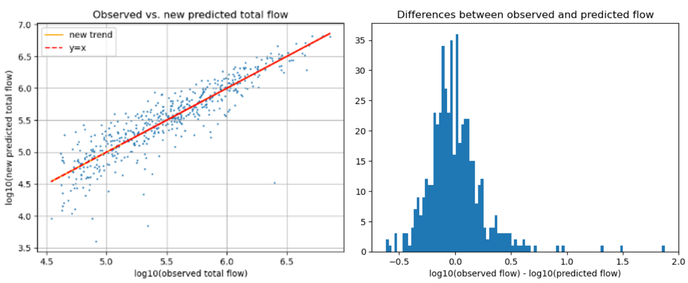


Figure 4 Left: observed vs. un-biased predicted total flow. Right: accompanying error histogram (N=514)

Logit model

The logit parameters in our model were attained by statistical analysis of 57 MNL models, originating from 17 case studies in our previously mentioned 100-study literature review. Their methodologies underline the necessity of choosing not to calibrate this project's MNL model to actual data: these studies use extensive surveys to estimate parameters, with the number of reported observations summing up to almost half a million. It is indicative of the complexity involved that estimating these parameters constitutes a distinct scientific discipline in its own right. Studies were applied in the UK, France, Germany, Spain, Portugal, Sweden and a few other countries outside of Europe, with the number of observations ranging from 40 to 63,000, and the r-squared model fit ranging between 0.075 and 0.822.

The median reported values of the 53 travel time and 47 travel cost beta parameters were -0.4606 and -0.0311, respectively. This results in an effective Value of Time of €14.80 per hour, closely approximating commuting values attained by an European-wide applied study by Wardman et al. (2012), when adjusted for inflation.

4.1.3. Model validation

After pre-processing, a validation data set of 180 million passengers, 187 cities and 890 connections remain. A statistical summary of this data set is provided in the table below.

Table 8 Characteristics of validation data

Var.	Unit	N	MIN	MEDIAN	MAX	AVG	STD
$P_{i,45}$	[pax]	187	4,191	4,129,149	29,518,637	6,225,502	6,358,880
CAP_i	[K€]	187	5,980	48,534	607,256	72,364	68,010
d_{ij}	[km]	890	1,000	1,473	3,120	1,527	377
$D_{AIR,ij}$	[pax]	890	20,114	99,356	3,195,192	202,426	293,041

N = number of values; MIN = minimum; MAX=maximum; AVG=average; STD=standard deviation; K€ = thousand euros; M€ = million euros; CAP=GDP/capita

The validation approach as taken by Belal et al. (2020) will be used here as well: we take five interesting connections and calculate how closely the model approaches the observed demand. Five air connections with more than 1 million in yearly observed demand are chosen. The table below shows the accuracy of the demand forecasting model for these connections.

Table 9 Validation results

Connection	Observed demand		Predicted demand		Difference (%log)
	Ordinary notation	log	Ordinary notation	log	
London-Faro	3,195,192	6.50	54,448	4.74	-27.2
Istanbul-Dusseldorf	2,253,703	6.35	4,475,305	6.65	+4.7
London-Krakow	1,775,915	6.25	1,696,741	6.23	-0.3
London-Porto	1,646,479	6.22	1,486,021	6.17	-0.7
London-Valencia	1,189,040	6.08	1,181,056	6.07	+0.0

The results show that the model predicts demand reasonably well, with the absolute deviation in terms of 10-log staying within 5%. The only exception here is the London to Faro connection, which is severely underestimated by the model. This indicates that the model is not able to estimate trips that have a heavy touristic character. This comes as no surprise, since the model only takes population, GDP and distance into account. The correlation between observed and predicted demand, when expressed in base 10 logarithm, equals a R^2 value of 0.337. Without taking connections with a touristic character into account, this value would reach 0.792.

4.1.4. Demand evolution

The mean yearly GDP growth between 1964 and 2023, among the 39 countries within this project's scope, vary between 1.38% (Greece) and 5.04% (Turkiye). When inserting the possible values into equation (6), it turns out that the factor k_{ij}^{eco} varies between 1.22 and 2.17.

As mentioned before, the level of induced demand is the same for every connection, unless it already has an upgradeable high-speed rail line.

4.2. Profitability estimation

As our model is straightforward and relies on rules of thumb, it is not destined for validation.

4.3. Network design

4.3.1. Data collection

This section presents the results of data collection as input for the optimisation model.

Set of nodes

The data set by Florczyk et al. (2020) provides data on 160 metrics for 13,135 urban centres (hereafter referred to as cities). After pre-processing, it was found that 726 cities, originating from 35 countries, comply with all scope requirements. Together, they form the set of potential nodes N, sharing 263,175 possible connections among them. Below, Figure 5 shows their distribution across the continent.

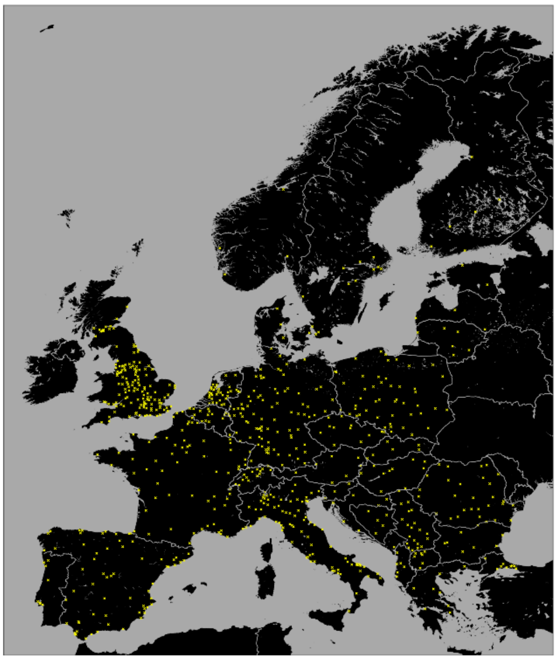


Figure 5 Cities within scope

These 726 cities encompass a wide range of values in all characteristics. With all settlements of population over 50,000 represented, there is confidence that the set encompasses all potential HSR stations. The required data was successfully gathered regarding all cities. Below, a statistical summary of the city-related model input data is given:

Table 10 Statistical summary of nodes set

Var.	Unit	MIN	MEDIAN	MAX	MEAN	STD
POP	[pax]	50,054	101,361	14,111,242	285,165	844,314
$P_{i,45}$	[pax]	195,736	1,243,560	13,417,545	1,845,190	1,786,929
CAP_i	[€]	84	15,772	51,478	16,105	6,749
H	[m]	-3	80	909	142	161

POP=population; CAP=GDP/capita; H=mean elevation

Set of arcs

Now the information regarding the set of nodes N is gathered, it is used to find all necessary arc- and OD pair-related data by means of web scraping. This was performed for all 31,125 city pairs for the 250 most populated cities, as this already took more than three days to complete. The task of performing such large amounts of searches regularly caused errors due to not being able to load web pages in time. The web scraping code had to be adapted to this, to automatically search arc-related data again if the load errors occurred. A complete database encompassing the needed data for all 31,125 city pairs was made successfully. In Table 11, a statistical summary of this data base is given. Note that all characteristics are based on the direct arc between origin and destination.

Table 11 Model input characteristics

Var.	Unit	Min	Med	Max	Mean	Std
Demand forecasting						
l_{ij}	[km]	16	1,423	5,041	1,507	832
TT_{plane}	[h]	0.0	6.6	26.9	6.6	2.3
TT_{train}	[h]	0.0	15.4	100.0	17.9	12.7
TT_{car}	[h]	0.2	13.5	59.6	14.5	8.1
TT_{HSR}	[h]	0.1	4.6	16.1	4.9	2.6
TC_{plane}	[€]	0	220	1,355	228	90
TC_{train}	[€]	0	268	1,341	272	158
TC_{car}	[€]	4	336	1,453	355	196
TC_{HSR}	[€]	37	170	734	177	71
$D_{HSR,ij}$	[pax]	14	24,237	10,263,681	108,926	313,918
$\bar{v}_{train}$	[km / h]	0	83	282	83	29
α	[pax]	35	61,874	22,732,344	234,875	662,871
β	[pax / h]	-974,438	-2,515	-1	-11,355	32,557
γ	[pax / €]	-66,572	-171	0	-772	-2,213
R^2	[-]	0.975	0.997	1.000	0.997	0.003
Profit estimation						
$k^{X,infra}$	[M€ / km]	7.0	47.6	116.1	48.9	16.7
f_a^{rev}	[M€]	0.1	139.6	64,577.4	521.5	1,540.6
f_a^{cost}	[B€]	0.8	71.5	314.4	75.7	42.6

Since all cities are located in continental Europe, travelling between them by car or the potential new HSR line should always be possible. This is reflected in the data, as the minimum travel time of these modes is larger than zero. A relatively low share city pairs are not connected by plane (4.5%) or train (3.9%), indicating the existence of infrastructure related to all modes. Most of the city pairs can be traversed by plane within a day, while the median train travel time already exceeds 15 hours. A comparison to the median 4.6 hours of potential HSR infrastructure is indicative to the opportunity of HSR development in continental Europe. This is even further accentuated by comparing travel costs: when instructed to maximise revenue, high-speed rail can be significantly cheaper than competing modes. This does not mean that HSR infrastructure should be developed everywhere across the continent. As the data shows, in many cases the demand is low, also indicated by a median of only 24,000 passengers per year. Only 1.6% of all city pairs would be able to generate a first-year demand of more than 1 million. Considered solely, only eleven (0.03%) of them would be profitable. The 'average train speed' was calculated in order to find upgradeable infrastructure, for which the value must be at least 200 km/h. A total of 29 arcs meet this demand.

4.3.2.Pre-processing

The largest potential network size our laptop could solve for within six hours considers the 111 most populated cities within the scope, which are all cities with a population exceeding 315,000. Recall that model pre-processing was performed by means of network simplification and route generation.

Network simplification: Without pre-processing, our optimisation model would consider 111 nodes, connected by 6,105 arcs and the same number of OD pairs. The size of the network and mainly its related set *A* was reduced significantly: only 109 nodes, 589 arcs and 5,886 OD pairs. This is due to the fact that our network simplification process primarily imposes constraints to arc lengths.

Route generation: The NetworkX route generating algorithm proved to be significantly faster than the optimisation process. Considering the final potential network size of 111 cities, it found 77,062 valid potential routes in approximately eight minutes. Even though the number of routes averages to 13 for each OD pair, the distribution of routes among the OD pairs is extremely uneven. In fact, the top 10% OD pairs with the most routes account for over 90% of all routes found. This can be attributed to the fact that a relatively low number of city pairs are far apart and are able to generate demand, exponentially increasing the number of potential routes.

4.3.3.Optimisation

Optimising for a network of 111 cities (effectively: 109, as explained earlier), 589 arcs, 2,269 OD pairs and 77,067 routes resulted in construction of a model with 243,671 integer decision variables (of which 89,537 are binary) and 1,035,732 constraints. The optimal solution was found after a little under six hours, an optimal lifetime profitability of €222.8 bn was reported, which can be broken down into the cash-flows present in the objective function:

Table 12 Profitability breakdown in optimal network (B€)

Revenue		Costs	
Ticket revenue	€ 655.413	Infrastructure construction	€ 193.809
		Infrastructure maintenance & operation	€ 15.876
		Rolling stock acquisition	€ 5.062
		Rolling stock operation & maintenance	€ 194.134
		Transfer penalty	€ 12.739
Total revenue	**€ 644.413**	**Total cost**	**€ 421.620**
			Total profit: € 222.793

Topology

The optimal configuration consists out of 15 cities, connected by 15 arcs and is displayed in Figure 6. The yellow dots not connected by lines, are cities that the model considered, but did not add to the network. Most of the arcs will be newly built, as only two are currently in high-speed operation: Brussels-Paris (average speed: 229 km/h) and London-Paris (200 km/h).

It heavily focuses on north-western Europe. Since our model accounts for already existent HSR infrastructure, it becomes evident that it's not worth the investment of upgrading domestic lines in the networks of France, Spain and Italy, as their current quality and coverage is sufficient. Simultaneously, it shows that mainly Germany and Great Britain are in dire need of more border-crossing HSR infrastructure. The network avoids mountainous terrain, indicating that this might affect the viability of HSR operation. Some connections are drawn as if they would cross water, but HSR travel times are determined by the shortest distance over land.

The network covers the largest cities in the remaining western countries: Great Britain, Germany, Belgium and the Netherlands. One can notice a clear triangle-structure, with extending arms in multiple directions to large cities such as Berlin, Amsterdam, Hamburg, Munich and Edinburgh, which all show great air demand with at least one of the cities in the triangle. Having one of the largest airports in the world, the city of Frankfurt is also part of the network. Some medium-to-large cities function as intermediate stops between larger cities, as they benefit from their location, while still adding a sufficient number of passengers to the network. Examples of these cities are Dusseldorf, Hannover, Stuttgart, Leeds, Nottingham and Nuremberg. Perhaps the most surprising addition to the network is the city of Southampton, which is located closely to London, and shows sufficient demand levels with the latter city and Paris.

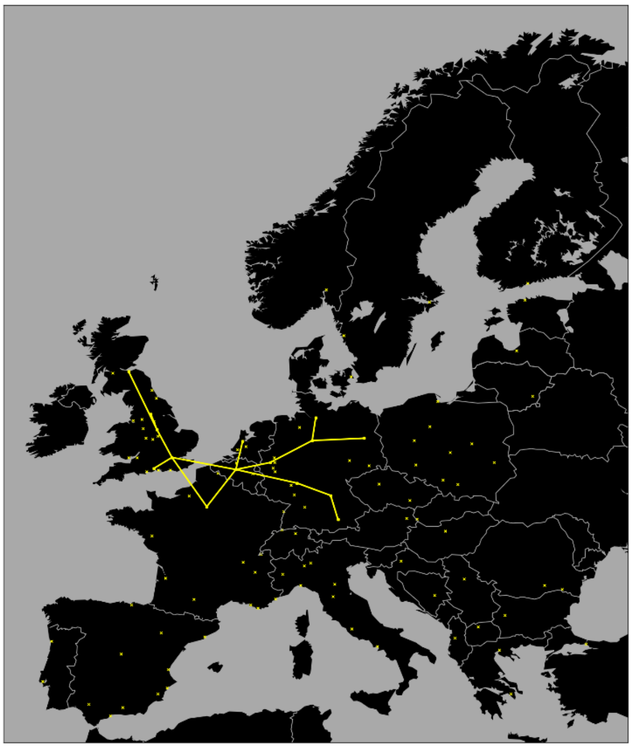


Figure 6 Optimal network topology; loose dots are considered though not selected cities

The combined length of all arcs equals 3,969 km, with lengths of individual arcs varying between 120 and 464 km. The network would serve close to 600 thousand passengers per day, on average over its lifetime. Focussing purely on ticket revenue and infrastructural costs, it appears that almost two-thirds (63%)

of all profit originates from the London-Paris and London-Brussels arcs. Another remarkable finding is that for a maximally profitable network, not all individual arcs have to be profitable on their own: 5 out of 15 arcs are not, which are all situated at an end point of the network. The table below shows the profitability for each selected arc, only considering ticket revenue and infrastructural costs, as rolling stock-related costs depend on the design of lines, which will be addressed in the next section.

Table 13 Profitability of connections

Connection name	Length [km]	Flow [pax/ day]	Revenue [B€]	Costs [B€]	Profit [B€]
Brussels-London	364	131,680	199.943	17.513	**182.430**
London-Paris	464	42,105	106.965	13.023	**93.941**
Brussels-Dusseldorf	201	67,474	54.182	9.890	**44.292**
Amsterdam-Brussels	211	56,345	45.245	9.563	**35.683**
Brussels-Paris	317	45,726	40.057	6.934	**33.122**
Brussels-Frankfurt	397	46,779	46.443	21.436	**25.007**
London-Nottingham	206	42,771	33.721	12.233	**21.488**
Dusseldorf-Hanover	280	33,654	26.042	17.371	**8.671**
London-Southampton	123	13,021	9.125	7.323	**1.803**
Edinburgh-Nottingham	449	17,155	26.800	26.017	**0.783**
Leeds-Nottingham	120	11,931	7.490	7.646	**-0.155**
Frankfurt -Nuremberg	223	22,566	16.473	18.474	**-2.000**
Munich-Nuremberg	172	16,618	12.131	14.557	**-2.426**
Berlin-Hanover	290	20,093	14.375	17.865	**-3.490**
Hamburg-Hanover	152	8,249	5.420	9.840	**-4.420**
TOTAL	**3,969**	**576,167**	**644,413**	**209,685**	**434,728**

OD pairs

60 out of 105 potential OD pairs within the network are served. Travel times, costs and distance vary between 0.46 - 4.74 hours, €43 – 326 and 120 – 1,362 km with means of 2.34 hours, €168 and 673 km, respectively. First-year passenger demand (accounted for induced demand) varies from 36,000 to 20.2 million. HSR market share per served OD pair varies between 30 and 92%, with a mean of 75% per OD pair. The market share of the entire network equals 79% (with a roughly equal share for the other modes).

Line design

The line design ensures direct connections for 52 out of 60 (87%) of OD pairs and 95% of passengers. All OD pairs are served with at most one transfer. Brussels can be considered a main hub, being associated with nine out of eleven lines, while having a direct connection with all but one of the other cities. Serving 95% of passengers directly, the inclusion of transfer penalties resulted in a well-balanced line design, which appears to "care" about the number of transfers passengers make, but does not overdo it in the sense that every OD pair is served by a separate line. Together, the lines serve all arcs, most often with the minimum required frequency. The topology displayed in Figure 6 will be served by eleven lines and a fleet of 81 trains. They are listed in Table 14. Below, Figure 7 shows the optimal line map.

Table 14 Line design

#	Stops	Length [km]	Trip time [h]	Freq [h^{-1}]	Fleet [-]
1	(2): London, Paris	464	1.55	3	10
2	(5): Berlin, Hannover, Dusseldorf, Brussels, Paris	1,088	3.78	1	8
3	(3): Amsterdam, Brussels, Paris	528	1.84	1	4
4	(3): Amsterdam, Brussels, London	575	1.99	2	8
5	(5): Frankfurt am Main, Brussels, London, Nottingham, Leeds	1,087	3.78	1	8
6	(3): Dusseldorf, Brussels, London	565	1.96	1	4
7	(6): Hamburg, Hannover, Dusseldorf, Brussels, London, Nottingham	1,203	4.24	1	9
8	(4): Nuremberg, Frankfurt am Main, Brussels, Paris	937	3.22	1	7
9	(3): Edinburgh, Nottingham, London	655	2.25	1	5
10	(6): Munich, Nuremberg, Frankfurt am Main, Brussels, London, Southampton	1,279	4.47	1	9
11	(6): Berlin, Hannover, Dusseldorf, Brussels, London, Southampton	1,258	4.40	1	9
TOTAL		9,639	-	-	81

4.3.4.Model stability analysis

The variations are applied to an optimisation model version considering the 77 most populous cities in Europe. In total, our model was ran six times, which generated the optimal network designs in running times varying between 1200 and 2300 seconds. This result already indicates that our fare variations created problems with a large variation in complexity. However, the number of nodes and arcs was the same for all simulations, and the optimal network stayed exactly the same in three out of six simulations. The other half showed minor changes compared to the original network: in three cases, a connection London-Lille was added, and in one case Nottingham-Edinburgh was deleted. The number of lines varied between 10 and 12, and the fleet size between 70 and 87. Even though the fares were changed by a maximum of 10%, the maximum deviation in profit value equals only 2.5%. Therefore, the results show that the inclusion of arcs often is insensitive to relatively small changes in fare setting, which builds confidence in our model's predictions, which is important in order to be able to make trustworthy recommendations.

4.3.5.Benchmarking

Our model finds an optimal solution consisting of 31 lines and a fleet of 134 vehicles, with the length of lines varying between two and eight nodes. The solving time of our model equals 20 seconds on average, which is considerably lower than algorithms proposed by Kechagiopoulos & Beligiannis (2014). Our model outperforms state-of-the-art algorithms in terms of total travel time and level of service, therefore providing a solution which is heavily focused on user-friendliness. However, this does come at the expense of operator costs, as our solution requires much more lines and a substantially larger fleet size than solutions by state-of-the-art approaches.

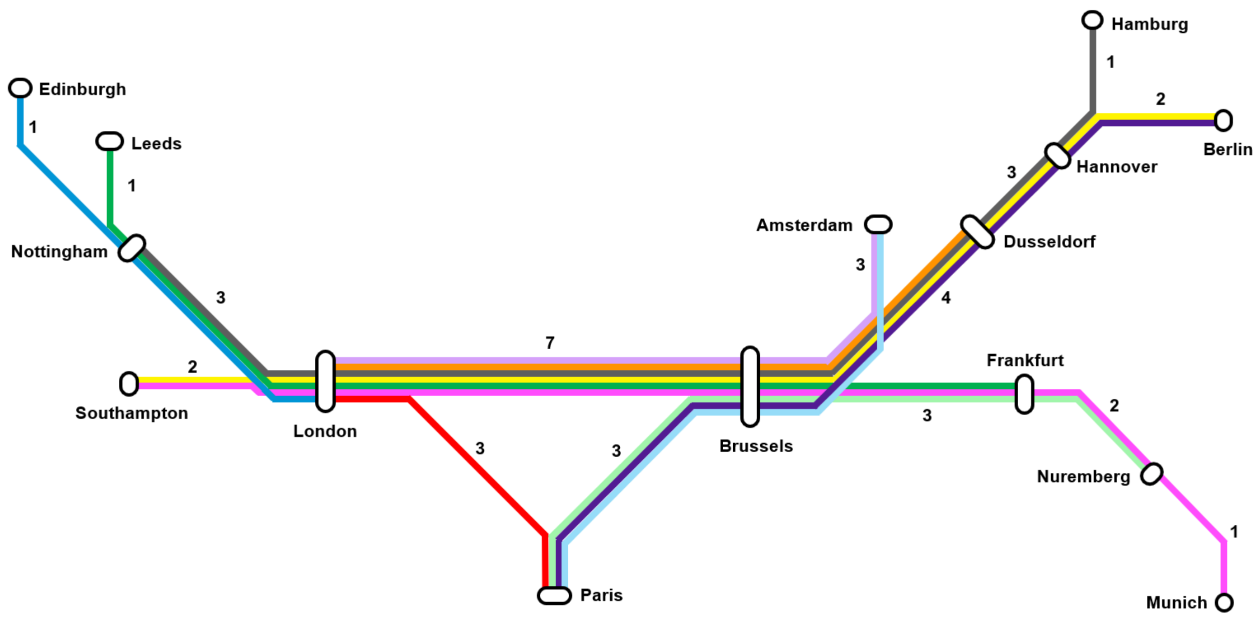


Figure 7 Line design map; the numbers denote the joint frequency per arc

The balance between operator and user costs is determined by the choice of parameters such as transfer penalty per passenger and unit maintenance & operating costs. Given the experimental character, addressing these parameters and their interaction regarding this network is considered outside of this project's scope and saved for further research.

5 POLICY IMPLICATIONS

This section translates our findings from the previous section into the lessons implications they have for policy, and recommendations that can be made regarding them. We shine our light on the implications from three perspectives: international cooperation, funding & financing, and environmental sustainability.

International cooperation

Cross-border operations are vital for the vitality of our network, as 84.8% of its passengers travel internationally. For this reason, it's of crucial importance that national and regional governmental bodies are coordinated. The heap of national and regional laws and regulations should be lifted, and new regulation should be made in order to make it easier for countries to cooperate. A special task force serving this exact purpose could be greatly beneficial.

Funding & financing

As stated before, high-speed rail infrastructure requires substantial investments. Our optimal network is able to underline this statement with numbers: for a total revenue of €644.4 bn and lifetime profit of €222.8 bn, approximately €421.6 bn must be invested in the coming 40 years. A substantial part of €193.8 bn is invested initially for construction and acquisition, which is 46% of all expenses during the project's lifetime. The rest will be spent for maintenance and operation for both the infrastructure and rolling stock, and amounts to a yearly expense of €5.69 bn.

Our network pays itself back in 19 years. Given the significant expenses, subsidiary help is needed from not only the European Union, but also from national and/or regional governmental bodies. The money would be primarily spent on infrastructure construction (42% in international connections).

Analysis of the outcomes of our profitability model shows that many lines would be profitable with a subsidy of no more than €3 bn. These lines are located most often in Poland or the Baltic States. Therefore, tactical investments could make the dream of a continent-wide HSR network come true.

Environmental sustainability

Our network would serve 94.3 million passengers on average per year during its lifetime, which would increase the total HSR traffic volume in the EU by 72%, from 131 (Statistica, 2018) to 225 billion km travelled on a yearly basis. Among these passengers, 42 million used to travel by plane and 21.3 million by car. The mentioned shift alone would significantly contribute to that, reducing emissions of the entire European transport sector by 33% by 2050. This percentage is based on an average of 200g CO_2/paxkm for air travel, 170g for car travel and 4g for HSR travel (UK Government, 2022), an average travel distance of 1,000 km for both HSR and air and 500 km for car travel, and 714 Mt CO_2 total transport-related emissions in 1990 (EEA, 2023). It is assumed that HSR operations would start in 2030.

6 CONCLUSIONS

The Transport Network Design Problem (TNDP) model developed in this study effectively overcomes the limitations of current methods, particularly in solving medium-to-large network designs with demand elasticity. By adapting the Multi-Commodity Flow Problem (MCFP) formulation to incorporate OD flow routes, the model retains the intrinsic logic of MCFP while significantly reducing computational demand. This allows the model to achieve optimality for networks with up to 111 nodes, or all European cities with a population over 315,000. It represents a significant step forward in high-speed rail (HSR) network design in this field.

The application of this model demonstrated that a profitable HSR network in Europe would not span the entire continent but would rather be concentrated in the north-western region, with a primary focus on major cities in the UK and Germany, and Brussels as the central hub. The model suggests that upgrading existing infrastructure is generally not cost-effective. The network would consist of 15 cities, 15 arcs (of which two upgraded from existing infrastructure) and would serve 60 OD-pairs. Eleven lines would serve the network, most with operating frequencies of one train per hour, increasing to a maximum of three for the busiest connections. The addition of transfer penalties makes up for a seemingly well-thought line design, which -besides appearing to be complex- serves almost all connections with the minimum frequency needed.

Moreover, the proposed network design aligns with the European Green Deal's climate objectives, potentially increasing HSR traffic volume by 72% and reducing emissions from the transport sector by 33% by 2050. However, achieving these outcomes will depend heavily on international cooperation and the provision of financial support, particularly for cross-border connections during the initial stages of infrastructure development.

7 DISCUSSION

This study addresses the limitations in demand forecasting, profitability estimation, and network design for high-speed rail (HSR) in Europe, contributing new insights into the Transport Network Design Problem (TNDP) under elastic demand. While the demand forecasting model used was functional, its simplifications—such as the exclusion of certain travel modes (e.g. bus) and impact factors—indicate room for further accuracy improvements, particularly through incorporating mixed logit or dynamic gravity models. The inability to fully capture inter-modal competition limits the precision of the demand forecasts, primarily when working with trips with a touristic character, but these were not the primary focus of the research.

The profitability estimation model, though simplified, provides reasonable insights into HSR viability under current assumptions. However, it overlooks critical factors such as fare competition and inflation, which could significantly change the results over a 40-year project span. Additional research could explore greener policies for optimising HSR demand, looking beyond only the maximisation of revenue.

The network design model performs strongly, efficiently solving for larger networks while taking demand elasticity into account. However, the model's reliance on 'smart restrictions' to reduce the number of OD flow routes and the inherent assumptions about fare sensitivity reveal areas where further work, particularly on line design robustness, could enhance its applicability to real-world scenarios. Similarly to the profitability estimation model, the network design model's focus on profitability leaves untouched potential for considering non-monetary benefits such as sustainability gains.

In conclusion, while the TNDP model successfully advances HSR network design, further studies are essential to refine demand forecasts, enhance profitability models, and explore robust design approaches.

ACKNOWLEDGEMENTS

This paper was written as part of a thesis project in partial fulfilments for the title Master of Science (MSc) in Transport, Infrastructure & Logistics (TIL) at Delft University of Technology, under the supervision of prof. dr. O. Cats, dr. F. Schulte and dr. A. Bombelli, with dr. Y. Zhu functioning as an additional examiner.

REFERENCES

Agramanisti Azdy, R., & Darnis, F. (2020). Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites. *Journal of Physics: Conference Series*, *1500*(1), 012104.

Arbex, R. O., & da Cunha, C. B. (2015). Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, *81*, 355–376.

Barrón, I., Campos, J., Gagnepain, P., Nash, C., Ulied, A., & Vickerman, R. (2012). *Economic Analysis of High Speed Rail in Europe* (G. de Rus, Ed.). Fundacion BBVA.

Belal, E. M., Khalil, A. A., & El-Dash, K. M. (2020). Economic investigation for building a high-speed rail in developing countries: The case of Egypt. *Ain Shams Engineering Journal*, *11*(4), 1001–1011.

Bleijenberg, A. (2020). *Air2Rail: Reducing CO2 from intra-European aviation by a modal shift from air to rail* (p. 21).

Borgogno, F. (2023). *Roadmap Towards a Unified European High-Speed Rail Infrastructure*.

Börjesson, M. (2014). Forecasting demand for high speed rail. *Transportation Research Part A: Policy and Practice, 70*, 81–92.

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2007). A Column-Generation Approach to Line Planning in Public Transport. *Transportation Science*, *41*(1), 123–132.

Bradsher, K. (2013, September 23). Speedy Trains Transform China . *The New York Times*.

Canca, D., José Luis Andrade-Pineda, De, A., & Calle, M. (2018). The Railway Rapid Transit frequency setting problem with speed-dependent operation costs. *Transportation Research. Part B: Methodological/Transportation Research. Part B, Methodological*, *117*(A), 494–519.

Cancela, H., Mauttone, A., & Urquhart, M. E. (2015). Mathematical programming formulations for transit network design. *Transportation Research Part B: Methodological*, *77*, 17–37.

Cascetta, E., & Coppola, P. (2011). High speed rail demand: empirical and modelling evidences from Italy. *Proceedings of European Transport Conference 2011*. Association for European Transport.

Cascetta, E., & Coppola, P. (2012). An elastic demand schedule-based multimodal assignment model for the simulation of high speed rail (HSR) systems. *EURO Journal on Transportation and Logistics*, *1*(1-2), 3–27.

Chen, F. (2020, August 24). *China sets railway building spree in high-speed motion*. Asia Times.

China Daily. (2020). *Chinese cities with a population of 500,000 to have high-speed rail in 15 years*. Chinadaily.com.cn.

de Keizer, B., Kouwenhoven, M., & Hofker, F. (2015). New Insights in Resistance to Interchange. *Transportation Research Procedia*, *8*, 72–79.

de Rus, G., Campos, J., Socorro, M., Valido, J., & Ortuño, A. (2020). Cost-benefit analysis of railway projects: high-speed lines and suburban lines. *Independent Authority for Fiscal Responsibility(AIReF)*.

Durán-Micco, J., & Vansteenwegen, P. (2021). A survey on the transit network design and frequency setting problem. *Public Transport*, *14*, 155–190.

EEA. (2023, October 24). *Greenhouse gas emissions from transport in Europe*.

European Court of Auditors. (2018). *Special report: A European high-speed rail network*. Op.europa.eu.

Eurostat. (2019). *Eurostat Air Transport Statistics* (pp. 1–2).

Eurostat. (2024a). *Data browser - avia_pa*. Europa.eu.

Eurostat. (2024b). Gross domestic product (GDP) at current market prices by metropolitan regions. *Europa.eu*. met_10r_3gdp.

Eurostat. (2024c). *Population on 1 January by age, sex and NUTS 2 region*. demo_r_d2jan.

Eurostat. (2024d). Regional gross domestic product by NUTS 2 regions - million EUR. *Europa.eu*. tgs00003.

Feremans, C., Labbé, M., & Laporte, G. (2003). Generalized network design problems. *European Journal of Operational Research*, *148*(1), 1–13.

Florczyk, A., Corbane, C., Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2020). *GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A*.

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2005). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association*, *71*(2), 131–146.

Frei, A., Kuhnimhof, T., & Axhausen, K. W. (2010). Long distance travel in Europe today. *89th Annual Meeting of the Transportation Research Board 2010*, *611*.

Fröidh, O. (2006). *Modelling operational costs of a future high-speed train*. Royal Institute of Technology (KTH).

Gössling, S., & Humpe, A. (2020). The global scale, distribution and growth of aviation: Implications for climate change. *Global Environmental Change*, *65*(102194), 102194.

Grolle, J., Donners, B., Annema, J. A., Duinkerken, M., & Cats, O. (2024). Service design and frequency setting for the European high-speed rail network. *Transportation Research Part A: Policy and Practice*, *179*(103906).

Grosche, T., Rothlauf, F., & Heinzl, A. (2007). Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, *13*(4), 175–183.

Guan, J. F., Yang, H., & Wirasinghe, S. C. (2003). Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B*, *40*(10), 885–902.

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273.

Gutiérrez-Jarpa, G., Laporte, G., Marianov, V., & Moccia, L. (2017). Multi-objective rapid transit network design with modal competition: The case of Concepción, Chile. *Computers & Operations Research*, *78*, 27–43.

ICAO. (2019). *Trends in Emissions that affect Climate Change*. Icao.int.

Jiang, X., Zhang, L., & Michael) Chen, X. (2014). Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C: Emerging Technologies*, *44*, 110–127.

Kechagiopoulos, P. N., & Beligiannis, G. N. (2014). Solving the Urban Transit Routing Problem using a particle swarm optimization based algorithm. *Applied Soft Computing*, *21*, 654–676.

Kepaptsoglou, K., & Karlaftis, M. (2009). Transit Route Network Design Problem: Review. *Journal of Transportation Engineering*, *135*(8), 491–505.

Leng, N., Nie, L., Guo, G., & Wu, X. (2015). Passenger Flow Forecasting for Chinese High Speed Rail Network. *Proceedings of the 2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering*.

Liang, J., Wu, J., Gao, Z., Sun, H., Yang, X., & Hong Kam Lo. (2019). Bus transit network design with uncertainties on the basis of a metro network: A two-step model framework. *Transportation Research Part B: Methodological*, *126*, 115–138.

Lieshout, R. (2012). Measuring the size of an airport's catchment area. *Journal of Transport Geography*, *25*, 27–34.

Magnanti, T. L., & Wong, R. T. (1984). Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, *18*(1), 1–55.

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, *5*(6), 396–404.

Marcucci, E., & Gatta, V. (2011). Regional airport choice: Consumer behaviour and policy implications. *Journal of Transport Geography*, *19*(1), 70–84.

Marín, Á., & García-Ródenas, R. (2009). Location of infrastructure in urban railway networks. *Computers & Operations Research*, *36*(5), 1461–1477.

Martín, J. C., Román, C., García-Palomares, J. C., & Gutiérrez, J. (2014). Spatial analysis of the competitiveness of the high-speed train and air transport: The role of access to terminals in the Madrid–Barcelona corridor. *Transportation Research Part A: Policy and Practice*, *69*, 392–408.

Martínez, H. S., Moyano, A., Coronado, J. M., & Garmendia, M. (2016). Catchment areas of high-speed rail stations: a model based on spatial analysis using ridership surveys. *European Journal of Transport and Infrastructure Research*, *16*(2).

May, M., J., S., & Bart, A. (2022). *2.2. Modeling Revenue, Costs, and Profit*. Slu.edu.

Mokhtarian, P. L., & Salomon, I. (2001). How derived is the demand for travel? Some conceptual and measurement considerations. *Transportation Research Part A: Policy and Practice*, *35*(8), 695–719.

Murray, A. T. (2003). A Coverage Model for Improving Public Transit System Accessibility and Expanding Access. *Annals of Operations Research*, *123*(1/4), 143–156.

Nash, C. (2010). *nvironmental and Other Co-benefits of Developing a High Speed Rail System in California: A Prospective Vision 2010–2050*. Center for Environmental Public Policy.

Ng, M. T. M., Mahmassani, H. S., Verbas, Ö., Cokyasar, T., & Engelhardt, R. (2024). *Redesigning large-scale multimodal transit networks with shared autonomous mobility services*.

Nurhidayat, A. Y., Widyastuti, H., Sutikno, & Upahita, D. P. (2023). Research on Passengers' Preferences and Impact of High-Speed Rail on Air Transport Demand. *Sustainability*, *15*(4), 3060.

ODS. (2024). *Airports code*. Opendatasoft.com.

Oliker, N., & Bekhor, S. (2020). An infeasible start heuristic for the transit route network design problem. *Transportmetrica A: Transport Science*, *16*(3), 388–408.

Preston, J. (2013). *The economics of investment in high speed rail: Summary and conclusions*. Www.econstor.eu.

Qin, J., Qu, W., Wu, X., & Zeng, Y. (2019). Differential Pricing Strategies of High Speed Railway Based on Prospect Theory: An Empirical Study from China. *Sustainability*, *11*(14), 3804.

Ranjbari, A., Hickman, M., & Chiu, Y.-C. (2020). A network design problem formulation and solution procedure for intercity transit services. *Transportmetrica A: Transport Science*, *16*(3), 1156–1175.

Rocha, M., & Ferreira, P. G. (2018). Graphs: Concepts and Algorithms. *Elsevier EBooks*, 275–288.

Rodrigue, J.-P. (2017, December 28). *B.4 – High Speed Rail Systems | The Geography of Transport Systems*. The Geography of Transport Systems.

Rome2Rio. (2024, March 26). *Rome2rio: discover how to get anywhere*.

Sánchez-Borràs, M., Nash, C., Abrantes, P., & López-Pita, A. (2010). Rail access charges and the competitiveness of high speed trains. *Transport Policy*, *17*(2), 102–109.

Sane, Y. (2020). *Multi-Criteria Analysis of the proposed Hyperloop transport project in Northern Holland* [Thesis].

Smappen. (2024). *Map*. Smappen.com.

Statistica. (2018). *EU-28: high-speed rail transport 2000-2018 | Statista*. Statista; Statista.

Strauss, J., Li, H., & Cui, J. (2021). High-speed Rail's Impact on Airline Demand and Air Carbon Emissions in China. *Transport Policy*, *109*, 85–87.

Tian, C., & Zhang, S. (2024). Study on the pattern of train arrival headway time in high-speed railway. *Railway Sciences*, *3*(3), 344–366.

Timperley, J. (2020, February 19). *Should we give up flying for the sake of the climate?* BBC.

Trafikverket. (2021). *Travel Demand Review: A study performed within the government assignment regarding New Main Lines for high-speed trains 2020/2021*.

UIC. (2018). *High Speed Rail: Fast Track To Sustainable Mobility*.

UIC High-Speed. (2018). *World High Speed Rolling Stock*.

UK Government. (2022). *Greenhouse gas reporting: conversion factors 2022*. GOV.UK.

Wardman, M., Chintakayala, P., de Jong, G., & Ferrer, D. (2012). *EUROPEAN WIDE META-ANALYSIS OF VALUES OF TRAVEL TIME: FINAL REPORT TO THE EUROPEAN INVESTMENT BANK* .

World Bank. (2023). *GDP growth (annual %)*. World Bank Open Data. ZG

Zhang, J., & Zhang, J. (2021). Comprehensive Evaluation of Operating Speeds for High-Speed Railway: A Case Study of China High-Speed Railway. *Mathematical Problems in Engineering*, *2021*, 1–16.

Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A. (2012). Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review. *Transport Reviews*, *32*(4), 399–433.

Zhang, W. (2024, January 3). *China: passenger transport volume of highspeed rail*. Statistica.