Universitat Politècnica de Catalunya (UPC)

Mathematical Programming Models to Design and Analyse Efficient and Robust Railway Freight Transport Networks

A Dissertation presented

by

Francisca Rosell Camps

to

Departament d'Estadística i Investigació Operativa (DEIO)

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Estadística i Investigació Operativa

Thesis Advisor: Professor Esteve Codina Sancho

Barcelona, September 2022.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

This research was supported by Grant TRA2016-76914-C3-1-P of the Spanish Ministerio de Economía y Competitividad, Spanish R+D Program PID2020-112967GB-C31 of Spanish Ministerio de Ciencia e Innovación, Grant DI-071 2019, AGAUR, of Generalitat de Catalunya.

para Ángel, Blanca, Núria y Olga

Abstract

Searching to achieve an ambitious reduction in greenhouse gas emissions, the European Union has set as a goal a modal shift in freight transport of 30% by rail or waterborne for the near future. The increasing efforts of many governments to intensify rail freight transport often must face the difficulties involved in improving both infrastructure and rail operations. Moreover, infrastructure management and business operations usually correspond to different entities with highly contradictory economic interests. Making progress on the reliability of the railway network is one of the main factors to be taken into account to make the use of the train more attractive as a means of transport for industry. Also, focusing on shippers' response to road and rail competition and the role of different rail undertakings competing with each other may help boost the use of rail for freight transport. Seeking to reinforce these two goals, this thesis introduces two independent mathematical optimisation models, which may also be complementary, and which have been developed under a common conceptual framework of data structures and variables to guarantee their compatibility.

The first model is a mathematical programming-based design model for evaluating the impact on a mixed railway network from proposals for infrastructure improvement and capacity expansion that are oriented mainly toward increasing freight transportation. The model has been applied to extend elements of an existing mixed railway network, perform relatively less costly actions on the network, and enhance capacity by adding new blocking/control systems at specific locations. These aspects are usually not taken into account in models for regional planning. Rather than a model whose sole focus is on railway capacity expansion, this approach combines capacity-expansion with network design. Because the way investments generate returns to the freight transportation system is of utmost relevance for these types of problems, this model is based on the efficient frontier between investment and operating costs.

The second model is a combined model for jointly evaluating the modal split road-rail, and the resulting railway freight flows on the railway network. This combined modal split-traffic assignment model is addressed to the case when a modal split based on a random utility model is available, and some of its coefficients may present a non-negligible variability. To this end, after the initial deterministic formulation, a robust counterpart of the model is developed. The model, formulated as a non-linear integer programming problem, is oriented to a multi-carrier environment and includes constraints to consider the interactions between the different types of flows on the railway network, allowing a detailed evaluation of the cost types of the carriers and the network capacity. An algorithmic solution based on the outer approximation method is shown to provide accurate solutions in a reasonable computational time for the robust and non-robust versions of the model.

Examples centred on a section of the Trans-European Transport Network, the TEN-T Core network corridors, are reported to test the applicability of the models. Results show the effectiveness of both models. The design model can be a helpful tool for analysing the impact infrastructure investments may have on operating costs, where (implicit) capacity limitations in the scenarios to be evaluated may necessarily be taken into account. At the same time, it can be complemented with the combined modal split-traffic assignment model by assessing the possible shippers' response to the different railway carriers' services competing with each other and the road.

Resumen

Tratando de lograr una significativa y ambiciosa reducción de las emisiones de gases de efecto invernadero, la Unión Europea se ha marcado como objetivo que los modos de transporte de mercancías alternativos a la carretera, como el ferrocarril o la navegación fluvial, alcancen una cuota del 30 % sobre el total de mercancías transportadas por tierra en Europa en los próximos años. Los crecientes esfuerzos que llevan a cabo los diferentes gobiernos se enfrentan con demasiada frecuencia con las dificultades que suponen mejorar de forma simultánea infraestructura y operaciones ferroviarias. Además, la gestión de la infraestructura y el desarrollo de las operaciones habitualmente corresponden a entes diferentes con intereses económicos enfrentados. Mejorar la fiabilidad de la red ferroviaria es uno de los principales factores a tener en cuenta para hacer más atractivo el uso del tren como medio de transporte para la industria. Por otro lado, centrarse en los criterios que pueden llevar a las empresas a elegir entre carretera o tren, y en el papel que juegan las diferentes compañías ferroviarias en esta elección, compitiendo entre sí, puede ayudar a incrementar el uso del tren para el transporte de mercancías. Con la idea de reforzar estos dos objetivos, este trabajo de tesis presenta dos modelos matemáticos de optimización, independientes pero a la vez complementarios, y desarrollados bajo un marco conceptual de estructuras de datos y variables común para garantizar su compatibilidad.

El primer modelo es un modelo de diseño basado en programación matemática para evaluar el impacto que pueden tener, sobre una red ferroviaria de uso mixto, propuestas de mejora de la infraestructura y de ampliación de la capacidad dirigidas principalmente a incrementar el uso del tren para el transporte de mercancías. El modelo se ha orientado a la modificación de elementos de una red ferroviaria de uso mixto existente, proponiendo intervenciones en la red relativamente poco costosas, y aumentando la capacidad añadiendo nuevos sistemas de bloqueo y control en ubicaciones específicas. Son aspectos que no acostumbran a ser tenidos en cuenta en modelos de planificación a nivel regional. Más que un modelo centrado únicamente en incrementar la capacidad en la red ferroviaria, nuestro enfoque combina la ampliación de capacidad con el diseño de la red.

Para este tipo de problemas, es de la máxima relevancia la manera en que las inversiones generan retornos al sistema de transporte ferroviario. Por eso, este modelo está basado en el óptimo equilibrio entre la inversión y los costes de operación.

El segundo modelo es un modelo combinado para evaluar de forma conjunta el reparto modal entre carretera y tren, y los flujos de mercancías en la red ferroviaria resultantes. Este modelo combinado de reparto modal y asignación de flujos está enfocado hacia aquellas situaciones en que hay un modelo de utilidad aleatoria disponible, pero algunos de sus coeficientes pueden presentar una variabilidad que no puede ser ignorada. Con esta finalidad, tras la formulación inicial del modelo determinístico se presenta una versión robusta de la formulación. El modelo, formulado como un problema de programación no lineal entera, está enfocado hacia un entorno en el que conviven (y compiten) diferentes compañías ferroviarias. Incluye restricciones que permiten reflejar las interacciones entre los diferentes tipos de flujos en la red ferroviaria. Se detalla un algoritmo para resolver el modelo, basado en el método de aproximaciones externas, que permite obtener soluciones precisas con un tiempo computacional razonable, tanto para la versión determinística como para la versión robusta.

Se presentan diferentes ejemplos basados en una sección de la Red Trans-Europea de Transporte (TEN-T por sus siglas en inglés), la red de corredores del núcleo de la TEN-T, que permiten validar la aplicabilidad del modelo. Los resultados reportados muestran la eficacia de ambos modelos. El modelo de diseño puede ser una herramienta útil para analizar el impacto que las inversiones en infraestructura pueden tener en los costes de operación, teniendo en cuenta las limitaciones de capacidad (de forma implícita) que existen en los escenarios evaluados. De la misma forma, se puede complementar este análisis con el modelo combinado de reparto modal y asignación de flujos, en el que se puede comprobar la posible respuesta de las empresas que requieren transportar sus productos ante los diferentes servicios ofrecidos por las compañías ferroviarias compitiendo entre si, y compitiendo con la carretera.

Agradecimientos

A punto ya de cerrar esta etapa del doctorado, son muchas las sensaciones que me acompañan. La primera es la de satisfacción: es un largo proceso que no tenía muy claro que pudiese llegar a finalizar. También la de alivio, por poder disponer de tiempo que permita plantearme nuevos retos y enfrentarme a ellos. Pero la más destacada es la sensación de plenitud, por todo lo que he aprendido en estos años, por todas las nuevas amistades que he podido cosechar, por todo el acompañamiento que me han brindado las personas más queridas.

En primer lugar, quiero agradecer al director de la tesis, Esteve Codina, todo el esfuerzo que ha dedicado para que esta tesis saliera adelante. Él me abrió las puertas de la aplicación de la investigación operativa en el transporte, y luego me ha apoyado en mi interés y pasión por el mundo ferroviario, con todas sus dificultades añadidas. Su experiencia y su vasto conocimiento han sido una gran motivación para mi. También quiero agradecer a Lídia Montero su apoyo, y su ayuda en ese momento crítico en los que no sabía cómo manejar los datos de los que disponía para poder ilustrar los modelos.

Agradezco de todo corazón el soporte recibido en el Departamento de Estadística e Investigación Operativa de la UPC, donde se ha desarrollado gran parte de esta tesis hasta que nos llegó la pandemia. A Toni, Klaus, Lupe, Erik, José Antonio, Jordi y demás miembros del departamento, por hacerme sentir parte de este pequeño mundo, y de forma muy especial a Sonia, por su ayuda en todo momento, y todos esos cafés y comidas compartidas, ¡que bien nos sentaban!.

Un recuerdo muy importante es el que tengo de mis compañeros de doctorado, Cecilia, Yovanina, Grace, Jordi, David, Marta, la mayoría de ellos ya flamantes doctores, por todas las risas, los momentos difíciles, y esas tabletas de chocolate para subir los ánimos.

Quiero agradecer a mis amigas Lali, Míriam y Carmina simplemente por estar ahí cuando hace falta, y por abrir mi mente a nuevas experiencias, y a Isidoro e Ismael con los que compartimos tantas fantásticas veladas, con un recuerdo muy especial para mis apasionantes discusiones con Brian, quien lamentablemente nos dejó demasiado pronto.

Durante todo este tiempo he contado con el apoyo incondicional de mis personas más queridas. El de mis padres, Janín y Toni, que siempre han creído en mi y me han dado esa libertad para crecer, el de mis hermanos, Toni y Lini, mi prima Catalina y toda la gran familia que formamos, y mis cuñados Àngels y Jordi, siempre interesados en la evolución de este proyecto.

Por último, mi mas profundo agradecimiento a mi marido Ángel, por su apoyo incondicional a esta aventura que supuso volver a estudiar y después insistir con el reto de la investigación y el doctorado. Y a mis hijas Blanca, Núria y Olga: me conformo con que estén la mitad de orgullosas de su madre de lo que estoy yo de ellas, aunque cuando empecé no sabían explicar en el colegio a qué me dedicaba.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Freight transport is a critical element of the world economy and plays an essential role in maintaining industrial and service activity. The economic value produced by the transport sector (including both passenger and freight) represented in 2019 roughly 5% of the European Gross Domestic Product (GDP) (European Commission (2021b)), and nearly 6% of the USA GDP (U.S. Department of Transportation, Bureau of Transportation Statistics (2021)). Projections show that total transport activity will be more than double by 2050 compared to 2015, being the expected growth of freight transport 2.6-fold. This growth is slower than previously estimated, before the impact of the pandemic Covid-19, where the expected increase was trifold (ITF (2021)).

However, transport has also a negative impact on the environment. In 2019, the transport sector was responsible for nearly 26% of the total Greenhouse Gas (GHG) emissions and 31% of the total CO_2 emissions in the EU-27. Moreover, it is the unique sector in the EU-27 with continuous growth in pollutant emissions in the last thirty years (European Commission (2021a)). Current climate-change reduction policies in transport are insufficient to achieve the goal of the Paris Agreement to limit global warning to 1.5°C. CO_2 emissions from transport will increase by 16% in 2050 even if all current commitments to reduce transport emissions are fully implemented (ITF (2021)). The expected growth in transport demand will reduce the improvements resulting from these policies.

Furthermore, freight transport has a massive impact on society. Road freight transport contributes substantially to the high congestion levels in urban areas. More importantly, different studies point out that freight transport is responsible for high societal costs due to premature deaths, mainly caused by pollution and road accidents.

The promotion of more efficient and sustainable transport methods, particularly rail freight, has been a key part of EU policy for the last 25 years. The EU White Paper on Transport fixed ten goals for a competitive and resource-efficient transport system, intending to achieve a reduction of at least 60% of GHGs by 2050 with respect to 1990 (European Commission (2011, p. 9)). The third goal is that

> *Thirty per cent of road freight over 300 km should shift to other modes such as rail or waterborne transport by 2030, and more than 50 % by 2050, facilitated by efficient and green freight corridors. To meet this goal will also require appropriate infrastructure to be developed.*

However, the share of rail freight transport in Europe has remained steady since 2008, even decreasing in 2019, according to Eurostat (Eurostat (2021)). In 2019, road freight transport accounts for 73.2% of the total inland freight transport (based on tonne-kilometres performed), while rail transport share was 16.9% (Spain with a 4.6% is one of the countries with the lowest share). In contrast, in 2019, GHG emissions share from road freight transport and rail transport[1] were roughly 5.5% and 0.1%, respectively, the same share that for CO_2 emissions.

Despite rail transport being more environmentally friendly and safer, road freight transport remains the most used. Today, European railways have not reached their real potential, and in consequence, the European authorities pursue how to revert the situation in the sector, its efficiency and its share in the European transport market, and how to open it to greater competition, improving the quality of cross-border services.

The EU's policy objectives for shifting goods from road to rail have been translated into a series of EU legislative measures aiming to open the market, ensuring non-discriminatory access and promoting interoperability and safety (European Court of Auditors (2016)). Consequently, the EU followed the strategy of vertical separation of activities, which means that formerly integrated railway companies have been separated into national infrastructure managers and railway undertakings, and the rail freight market was fully open to competition by 1 January 2007. Other countries, such as Japan or some South American countries, have opted for horizontal separation. In other words, the network has been divided into local monopolies. The concession for its exploitation has been granted to a private company through a tender for an extended period of time. In countries such as the United States and Canada, there are vertically integrated private companies - that is, they have their own trains and tracks - that compete

[1]Road freight transport GHG and CO_2 share is taken from the heavy-duty trucks and buses' share data. Rail transport share includes both passenger and freight transport

with each other by setting their own prices. The vertical separation followed by the EU favours market integration by allowing the same company to operate traffic in different countries. Nevertheless, at the same time, it makes it challenging to assign international paths since the collaboration of different infrastructure managers from different countries is required, and the path assignment requires supranational coordination.

The interest in providing solutions to promote rail transport for freight is the origin of this thesis. Making progress on the reliability of the railway network is one of the main factors to be taken into account to make the use of the train more attractive as a means of transport for industry. Also, focusing on shippers' response to road and rail competition and the role of different rail undertakings competing with each other may help boost the use of rail for freight transport.

This thesis introduces two independent mathematical optimisation models, which may also be complementary, motivated by the aforementioned EU objectives and policies.

Firstly, a design model where new infrastructure enhancements are proposed to increase rail freight transport operation. Rail infrastructure investments require large amounts of capital. They do not only involve new elements (usually very expensive) that determine the railway network topology but also actions taken on existing network components to increase the capacity and efficiency of railway traffic. Therefore, strategic decisions are not sufficient, and several tactical and operational aspects must also be considered. Rather than a model of railway capacity expansion, our design approach combines capacity-expansion with network design. The primary purpose of our design model is to ascertain how and which investments may generate returns to the transportation system.

Secondly, a tactical model that relies on the infrastructure (the existing or the proposed one) focuses on shippers' requirements when choosing between road and train transport and among different rail carriers when competing with each other. Every day, thousands of tonnes of goods are transported across the EU to factories, warehouses, or final customers. Rail freight is in direct competition with the road: shippers regularly compare both when deciding which mode of transport to use. They naturally choose the one which best suits their needs, taking mainly into account: reliability, price, customer service, frequency, and transport time. Also, recently, the impact the transport mode has on greenhouse gas emissions is crucial for making a choice. The tactical model corresponds to a combined modal-split/traffic assignment model for evaluating the train-road modal share in future scenarios, or also to the case when a modal split model based on random utilities is available, and some of its coefficients may present a non-negligible range of variability. The assignment of railway flows considers its various components in a multi-carrier environment, including explicit constraints when interactions occur between the different types of flows on the railway network, allowing a good evaluation of the various cost types of the carriers and the network capacity. The main purpose of our tactical model is to analyse how different competition levels among railway carriers may help or not increase the rail freight share with respect to road freight transport, under shippers' criteria.

1.1. Contributions

The main contributions of this work can be summarised as follows:

- The novel approach of the design model seeks to determine an optimal trade-off between infrastructure investments and the operating costs for using rail freight transport. Classical decision variables for enhancing or not the infrastructure are conditioned by the prospective volume of merchandise to be transported by rail.
- As far as we know, it is the first time a rail freight model combines a modal split and a traffic assignment in a unique model. Our tactical model may help to have a joint view of the more probable shippers' modal choice criteria while highlighting the most attractive paths for freight transport.
- Both models incorporate constraints to reflect the distinctiveness of the rail freight transport, conditioned by infrastructure features and operational characteristics.
- The design model and the tactical model complement each other: the resulting infrastructure from the design model can be analysed with the tactical model to test its capacity to attract shippers' interest in rail freight transport.

In particular, our intention is that the resulting models of this thesis could be of application within the context of the technical problems that arise in the increasing of the rail freight transportation share.

1.2. Publications and Conferences

Some of the results of this thesis have been published in international journals or presented at conferences or workshops. The publications and conference participations are listed below.

Publications

- Rosell, F., Ubalde, L., & Saurí, S. (2018). *Potenciació del transport multimodal al Corredor del Mediterrani des de la implementació de models multiclient i multiproducte eficients.* Primera Ed. – Papers de l'Observatori de la Indústria;4. Catalunya: Departament d'Empresa i Coneixement.
- Rosell, F., & Codina, E. (2020). A model that assesses proposals for infrastructure improvement and capacity expansion on a mixed railway network. *Transportation Research Procedia, 47*, 441–448. https://doi.org/10.1016/j.trpro.2020.03.119
- Rosell, F., Codina, E.,& Montero L. (2022). A Combined and Robust Modal-split/Traffic Assignment Model for Rail and Road Freight Transport, *European Journal of Operational Research, 303* (2), 688-698, https://doi.org/10.1016/j.ejor.2022.03.008

Conferences

- Rosell, F., Codina, E., Saurí, S. (2018),*A Mathematical Programming Model for the Design of Railway Freight Transport Networks*, in XXXVII Congreso Nacional de Estadística e Investigación Operativa (SEIO), 2018, Oviedo, Spain
- Rosell, F., Codina, E. (2019) *A railway network design model for the joint expansion and improvement of freight railway infrastructures*, in International Conference on Railway Operations Modelling and Analysis - RailNorrköping, 2019, Norrköping, Sweden.
- Rosell, F., Codina, E. (2019) *Model for assessing proposals for infrastructure improvement and capacity expansion on a mixed railway network*, in 22nd Euro Working Group in Transportation Meeting - EWGT2019, 2019, Barcelona, Spain.

1.3. Structure of the thesis

The remainder of the book is organized as follows: Chapter 2 explains the main topics on rail freight transport related to this work, while Chapter 3 shows an overview of the existing literature on Operational Research focused on aspects of rail freight transport. Chapter 4 introduces the mathematical elements required to represent the rail network and the requirements due to the particular characteristics of the rail freight transport. Chapter 5 develops the design model as a mathematical programming-based model to evaluate the impact of infrastructure improvements and capacity expansion on rail freight transport. The tactical model is detailed in Chapter 6. Chapter 7 introduces the scenarios used to test the validity and the usability of the model. Chapter 8 presents a summary of the computational tests carried out on both models as well as a discussion of their results. Finally, Chapter 9 concludes this thesis with some remarks and proposals for future research.

Chapter 2

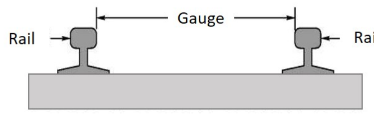
Rail Freight Transport Planning

This chapter develops the main topics on rail freight transport that will be used later when building the models. Talking about rail transport, both for freight or passengers, means to enter a world governed by rules very different from those we are used to when we think of road transport. It is not only obvious that the train requires tracks to move while both trucks and cars need roads, but also the infrastructure and the way trains are operated confer rail transport its particular conditions. Also, even if the basics elements are standard worldwide, there are relevant differences in rail management among Europe, North America, Asia or Australia. Because of that, trying not to overextend this chapter, we will focus mainly on how rail freight transport is managed in Europe.

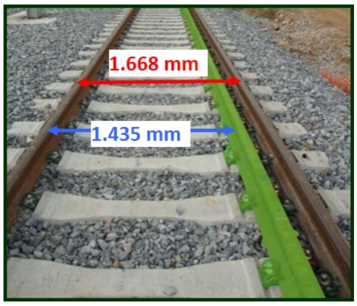
2.1. Railway infrastructure main topics

Track gauge, the electrification system, loading gauges, or the maximum lengths of the trains are some of the infrastructure elements that most affect the interoperability of trains in Europe.

The **track gauge** is the distance between the inner sides of the railheads of a track (Figure 2.1a). The most widely **track gauge** used in the world is known as UIC or standard gauge, with a separation of 1,435mm between rails. However, some countries have traced their national rail networks with other track gauges. These changes of track gauges among countries (or even inside the same country) directly affect interoperability, as it affects rail traffic from one country to another. For instance, different track gauges can be found in Europe. Most of its railway network in Spain has the "Iberian gauge" (1,668mm), but the standard gauge is used on high-speed lines. To deal with this complexity, in some tracks, a mix of both two gauges is used (Figure 2.1b). Finland and



(a) Diagram of track gauge. Source: author



(b) Track with mixed gauge. Source: Rosell et al. (2018)

Figure 2.1: Track gauges

Russia, and most of the countries under the influence of the former USSR, also have different track gauges, in this case of 1,520mm (1,524mm in the case of Finland).

Each country also determines the **electrical voltage** of the railway network. Moreover, in the case of Europe, few countries coincide: for example, in Spain, Italy and Belgium is standard the use of 3 kV DC, while France and the Netherlands use 1.5 kV DC, France also partially uses 25kV AC, Portugal uses 25 kV AC, or Germany, Austria and Switzerland use 15kV AC. There is an international agreement that recommends the use of 25kV AC for European corridors. As a curiosity, carriers need a special train engine to connect to three different voltages to traverse the Perthus Tunnel, a railway tunnel that connects France and Spain under the Eastern Pyrenees by the high-speed line since December 2013: inside the tunnel, the locomotive requires a 25kV AC connection, while just on the French side, the voltage is 1.5kV DC and on the Spanish side, the voltage is 3kV DC.

The **loading gauges** are the dimensions of height and width which must not be exceeded by a rail vehicle or its load to ensure that they can pass safely through tunnels and under bridges, and keep clear of trackside buildings and structures. It directly impacts the cargo. For instance, it can limit the possibility of transporting double-stack containers, a common practice in USA intermodal shipments but hardly applied in Europe due to its more restricted loading gauges and weight train limits (Figure 2.2a). Also it can limit the future expansion of the rail motorways, where the truck (with or without the tractor unit) is loaded directly onto the train (Figure 2.2b).

The maximum **train length** is usually limited by the size of passing loops and refuge sidings as well as the placement of signals. For instance, in Spain, in most of the rail network, the maximum lengths allowed are between 400m and 500m, while in Germany or Belgium, the trains can measure 750m, or even in France, some freight trains of up

(a) Maersk's double-stack train in India.

(b) A train loaded with semi-trailers, running on a rail motorway.

Figure 2.2: Loading gauges. Source: railfreight.com (ProMedia Group (2021)).

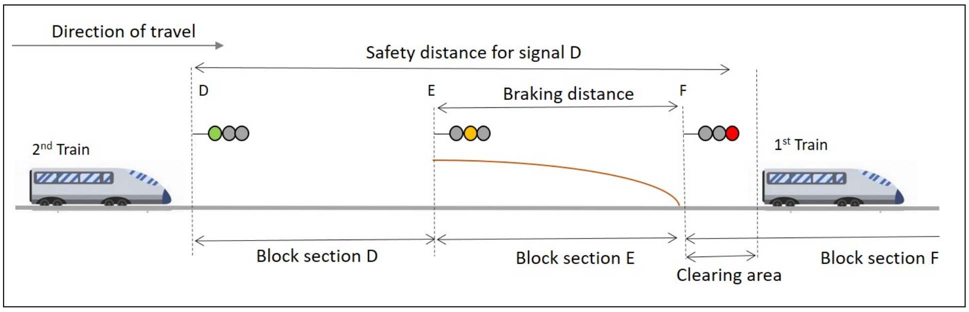
to 850m are running. To compare with, in North America freight trains can arrive to measure 5,000m. As happens with the track gauge or the electrification systems, the differences in countries of maximum train length affect interoperability, yet trains should be decomposed or assembled when crossing different countries.

Finally, the **maximum tonnage** a train can carry over depends mainly on the locomotive, the couplers and the slopes to be found on the route, as well as axle load limits. Sometimes, the train can be assembled with two or more locomotives, distributed along the train or two at the beginning to increase the maximum load, if the country's regulations allow it.

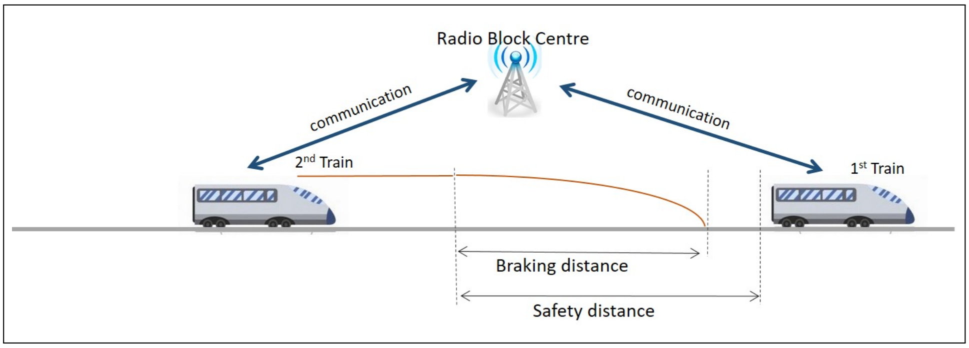
2.2. Control and signalling systems

Control and signalling are two of the essential components of the many which make up a railway system. Since it is very complex, we will focus on a very specific aspect: how a signalled operation with a block system works. Signalling with fixed lineside signals is the most usual. However, there is also an increasing use of cab signal systems, especially for high-speed lines where lineside signals cannot be watched safely.

A **fixed block system** is a block system using fixed block sections in which the tracks are divided and which are protected by signals (lineside or cab signals). Only one train at a time can be in each block section. Figure 2.3a shows the basic principle of the fixed block system. The block occupied by the first train (block section F) is protected by the red signal behind it at the block entrance. The block behind (block section E) is clear of trains, but a yellow signal provides advanced warning of the red signal ahead. This block provides a safe braking distance for the second train. The block behind, block section D, is also clear of trains, and a green signal will allow the second train to enter this block while maintaining the maximum allowed speed over this line until the driver



(a) Schematic of fixed block system.



(b) Schematic of a moving block system.

Figure 2.3: Schematic of fixed and moving block systems. Source: author.

can see the next signal. Note that in this case, the safety distance is longer than the sum of the distance of the block sections. The primary purpose of the extension of the safety distance is to provide additional protection in case a train is overrunning a stop signal by a short distance. To clear a signal, the entire control length must be clear and safe. This way of operating is common in European railways. Other systems do not add extra-distance, and then the control length of a signal is equal to the block section, as in railways in North America and Russia.

On the contrary, in a **moving block system** trains are continuously controlled and kept at a braking distance from each other. Figure 2.3b shows a diagram of the moving block system. In this case, a train clears the track behind its rear according to the tracking intervals of train location, based on real-time information from the Radio Block Centre.

As with electrical voltages, each country has its **railway signalling and control system**. The European Rail Traffic Management System (ERTMS) is a European initiative for management and control systems that aims to be a standard in Europe (Unife

(2021)). The European Train Control System (ETCS) and the GSM-R are the two main elements that compose the ERTMS. The ETCS is the standard train protection system introduced to replace the current national automatic train protection systems. GSM-R is a radio system based on standard GSM to communicate between the track or the radio block centre and the train.

The different uses of ERTMS as a train control system are classified into three levels, varying from Level 1 to Level 3, where Level 1 and Level 2 are based on a fixed block system, while Level 3, still in a development phase, will be based on moving block system. As detailed in Unife (2021), for Level 1 compliance, ETCS is installed on the lineside (possibly superimposed with legacy systems) and on-board. There is a spot transmission of data from track to train (and vice versa) via Eurobalises or Euroloops (the European standard technical solutions for the system of communication from ground to board). Level 2 is as Level 1, but Eurobalises are only used for the precise train position detection. The continuous data transmission via GSM-R with the Radio Block Centre (RBC) gives the driver the required signalling information. There is additional lineside equipment needed, i.e., for train integrity detection. Level 3 will improve Level 2: train location and train integrity supervision will rely completely on continuous communication via the RBC, and no additional information on trackside equipment will be required. Consequently, ERTMS Level 3 will further increase ERTMS potential by introducing a "moving block" technology and reducing trackside equipment for train detection.

2.3. Rail yards

Railyards can be defined as "a system of tracks within defined limits provided for making up trains, storing cars, and other purposes, over which movements not authorised by time table or by train-order may be made, subject to prescribed signals and rules, or special instructions" (University of Birmingham and Network Rail (2011)). Several types of rail yards exist, depending on the main operational processes they execute. Shunting yards are characterised by the disassembling and reassembling procedures of trains via a system of tracks and switches. In general, loads are not moved from one train to another: instead, the loaded railcar can be decoupled from one train and assigned to a different train. Rail-road terminals are mainly dedicated to point-to-point rail transport, although not exclusively. They mainly focus on standardised loading units, taking advantage of the intermodal transport, where the first and last mile transport is performed by trucks, and the long-distance corresponds to rail transport. Load movements are performed by gantry cranes, and trains themselves are not composed nor decomposed in the terminal.

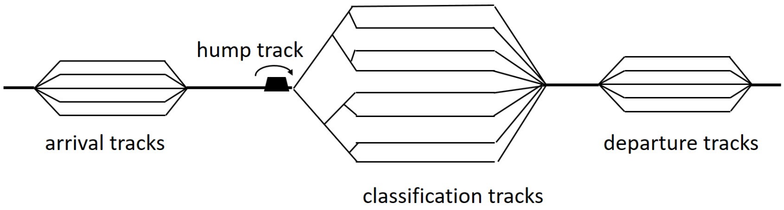


Figure 2.4: Outline of a hump shunting yard. Source: author.

Modern rail-rail transhipment yards also use gantry cranes for moving loading units. They are designed to quickly exchange loading units between trains: equipped with an automated ground sorting system to help work cranes more efficiently. The models we present in this thesis will focus on train movements and not on load movements, then shunting yards, and the operations related to composing and decomposing trains will play a significant role.

In general, a shunting yard is made up of three main areas, and all of them consists of a set of parallel tracks: the receiving area where the inbound trains arrive, the classification area, where railcars are rearranged, and the departure area, where the outbound trains wait until they are allowed to depart. On the yards, railcars are usually moved by specific shunting engines. The operations where usually railcars spend most of the yard time are the classification and the assembly into an outbound train.

There are three main types of shunting yards, depending on their physical characteristics: hump yards, flat yards and gravitational yards. Hump yards have the hump between receiving and classification area, while at gravity yards, the whole yard has a slight decline to ease switching. Both seek to take advantage of the gravitational pull to minimize the resource demand, but gravity yards require more staff for setting brake shoes. Thus, they are less common in Europe. Flat yards have almost the same layout as hump yards, but in this case, the gravitational advantage has to be replaced by switch engines. Figure 2.4 shows a schema of a hump shunting yard.

2.4. Railway operation topics

Rail freight transport requires the involvement of various stakeholders, as detailed in European Court of Auditors (2016). Figure 2.5 shows the main actors and how they interact. The **Railway Undertakings** or Railway Carriers are rail freight operators who provide the service of transporting goods. Their trains use infrastructure, which the Infrastructure Managers manage. **Infrastructure Managers** are in charge of the

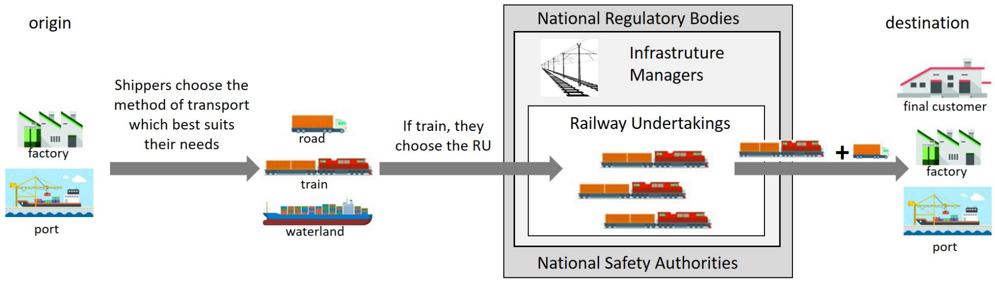


Figure 2.5: Main rail freight transport stakeholders. Source: author, based on European Court of Auditors (2016).

infrastructure and are responsible for allocating capacity on the infrastructure to Railway Undertakings and setting the associated costs. Also, the **National Regulatory Bodies** corresponds to ensuring that all Railway Undertakings have fair and non-discriminatory access to the rail network. The **National Safety Authorities**, in cooperation with the European Railway Agency, are responsible for the rail vehicles authorisations and the safety certificate release for Railway Undertakings.

Shippers may use rail transport in different ways, depending on their needs. The client may reserve a full or block train if they have enough goods to fill a train or contract as many wagons, intermodal or combined rail-road transport as required on a shared (and usually scheduled) train.

In some cases, railway carriers need to provide to the origin of the shipper a rail-owned empty railcar to begin loading. After the loaded railcar is delivered to the shippers' destination and emptied, the railcar is released back to the custody of the railway company and the cycle begins. The imbalance in the supply and demand of goods becomes a movement of repositioning multiple railcars to various origins from different destinations. Moreover, there are many types of freight railcars and not every one of them may be used to transport all types of goods. As a consequence, the flow of empty cars not only depends on the number of railcars required and delivered, but also on its type. Empty transport represents a significant part of the total workload in the railway network, and then it must be considered when rail transport is analysed in detail.

2.5. Rail versus road

Rail freight is in direct competition with the road. Shippers compare both and usually analyse various offers from different carriers when deciding how to transport their goods, taking mainly into account: price, frequency, transport time, reliability, and quality of service to the customer. Recently, the impact that transport mode has on greenhouse gas emissions is also crucial for making a choice.

As European Court of Auditors (2016) points out, traffic management procedures do not quite fit the needs of rail freight transport, even within rail freight corridors. The level of accessibility across all Member States for the road transport makes it difficult for the train to compete with it in end-to-end supply chain logistics. In most countries, railway transport has to pay a canon for the use of infrastructure while trucks use roads for free (or at least, paying a lower price than the actual cost). Also, rail freight needs the support of trucks for what is known as first-mile/last-mile deliveries, that is, the freight transport to/from rail terminals from/to client. Furthermore, transport external costs that most impact society, like traffic accidents, pollution, noise or climate change impact are currently not internalised into transport prices, being road transport the main cause of these damages. External costs are valued by 4.2€-cents per ton-km for heavy good vehicles vs 1.3€-cents per ton-km for rail freight, as reflected in European Commission (2019).

Furthermore, according to Rail Freight Forward (2019), rail freight transport consumes six times less energy per ton-km than the road. The lower friction from operating steel wheels on steel rail compared with rubber wheels on tar roads is the main reason for this difference in energy consumption. Also, a train set with up to 40 wagons offers lower air resistance than the corresponding fleet of trucks running on the road. Other direct costs, as drivers, also balance in favour of rail: the need for more training in the case of train drivers is compensated by requiring fewer drivers to transport the same load.

The recent crisis of the COVID-19 and the shortage of trucks highlighted the relevance rail freight transport has on the economy. Rail assured safe transportation of goods without being hardly affected by the lockdowns and the lack of drivers while being sustainable and efficient.

2.6. Rail freight transport versus rail passenger transport

Freight transport is clearly different from passenger transport. The main difference can be found in the way trains operate, e.g., there is a large variety in commodity types, the time scale is much longer, and intermediate storage is an option. Logistics deals with decisions on shipment size, inventory policy, warehousing, etc. Note that the logistic perspective also includes empty rides, as freight transport is not usually balanced in both directions.

When talking about passengers, a timetabling system is mandatory. Nevertheless, in addition to the timetabling service, rail freight transport usually requires an on-demand service. In fact, in North America, all freight trains run on-demand, and freight trains only depart when they are full. In Europe, a timetable is the most usual way to operate with the single wagon or intermodal load, or even full periodic trains, but operating on-demand is also a common practice.

Another difference lies in users preferences. Passengers value frequency, connectivity and travel time. In contrast, freight train clients are more worried about reliability, that is, the load must be delivered on time (which does not always mean in the shortest time) and assuring the quality standards the cargo needs, than the choice of route or the speed of the transport mode.

Regarding the infrastructure, although passenger and freight trains usually share tracks and sometimes terminals, passenger trains have their own passenger stations. Some tracks are specially designed for passenger trains, as high-speed tracks. Passenger trains are lighter than freight trains, and passenger locomotives can reach higher speeds than freight locomotives. For instance, in Spain, the average speed for freight trains in 2018 was 52 km/h, while for long-distance passenger trains (excluded high-speed trains), the average speed was 140 km/h (Ministerio de Fomento (2019)).

2.7. Environmental impact methodologies

Two main reasons move governments to increment the rail freight transport by reducing the road freight transport: road congestion and pollution. While changing trucks by trains clearly imply a reduction of road congestion, the impact of this change on pollution requires defining standard methodologies to measure energy consumption and pollutants emissions for different means of transport. The EN-16258 standard, “Methodology for the calculation and declaration of energy consumption and greenhouse gas emissions in transport services (transport of goods and passengers)” (CEN-European Committee for Standardization (2012)), is the first international standard to harmonize and standardize the procedures for the calculation and reporting of emissions and energy for the transport sector. This standard has been fully accepted among European transport companies. The methodology proposed in the EN-16258 standard analyses the fuel life cycle, known as Well-to-Wheels (WTW). The WTW analysis include the Well-to-Tank (WTT) emissions, that is, the impacts of the extraction of the raw materials, transportation, transformation and distribution of the fuel to the service station, plus the Tank-to-Wheels (TTW) emissions, that is, the generated impact by the energy con-

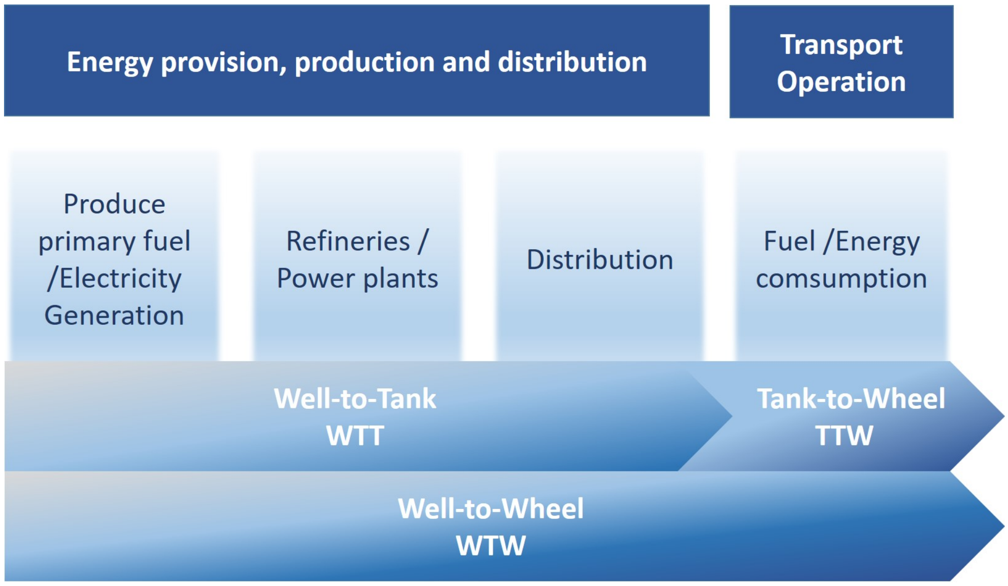


Figure 2.6: Well-to-Wheel main processes. Source: IVE et al. (2021).

sumption in the vehicle. Figure 2.6 shows a schematic of the main processes related to the Well-to-Wheel system.

The tool EcoTransITWorld, Ecological Transport Information Tool for Worldwide Transports (EcoTransIT World) is a free of charge internet application that shows the environmental impact of freight transport for any route in the world and any transport mode. This tool analyses the impact of a single shipment by comparing different transport chains, thus allowing the user to choose that with the lowest impact. This tool gives an excellent example of how transport emissions and energy consumption can be calculated following the EN-16258 standard guidelines.

In short, the transport service can be split into different elements of the transport chain, each of them corresponding to a transport mode covering part of the total trip. For each transport chain component, the energy consumption and vehicle emissions per transport have to be calculated, based on the metric weight of the shipment and the trip distance, in tons per kilometre. The environmental parameters covered are energy consumption, carbon dioxide, the sum of all greenhouse gases (measured as CO_2 equivalents) and air pollutants, which are mainly responsible for acidification, eco-toxicity, human toxicity and summer smog, such as nitrogen oxides, sulphur dioxide, non-methane hydrocarbons and particulate matter.

The process has two steps: in the first step, the final energy consumption (quantity of fuel or electricity) of each component of the transport chain has to be calculated, while in a second step, these values have to be translated into standardised energy consumption (MJ) and CO_2 equivalent emissions (kg CO_2 equivalent), differentiating between Tank-to-Wheels and Well-to-Tank consumption and emission. The Well-to-Wheel total consumption and emission results from the sum of both components (TTW and WTT) total consumption and emission, respectively. For more details, see the methodological report IVE et al. (2021).

Chapter 3

Strategic Rail Freight Transport in Operational Research

For a long time, rail freight has not been studied as intensively as rail transport passengers. As reflected in the previous chapter, requirements towards freight transport models are complex: the need to include various modes of transportation, the variability and seasonality of supply and demand of goods, different players and levels of decision, with infrastructures that require long implementation delays and huge investments. In this chapter, the primary studies on Operational Research that are centred on solving the rail freight transport needs are summarised.

3.1. Railway systems planning

Perhaps the most visible element in rail transport planning is the train schedule. However, a whole set of previous tasks need to be carried out before defining the schedules, each in different periods, some developed simultaneously and others following a given sequence. In general, thinking about the level of impact, all these tasks can be classified into three levels: strategic, tactical and operational. In Crainic and Laporte (1997), the authors present a complete review of the main issues in freight transportation planning and operations, and the different classes of Operational Research models and methods to address them. The following is a summary of the most significant aspects directly related to rail freight transport.

Each of the different activities implies performing significant planning tasks, in which the use of operational research models plays an important role. Figure 3.1 shows a diagram of the main activities at each of the groups, for which many efforts have been

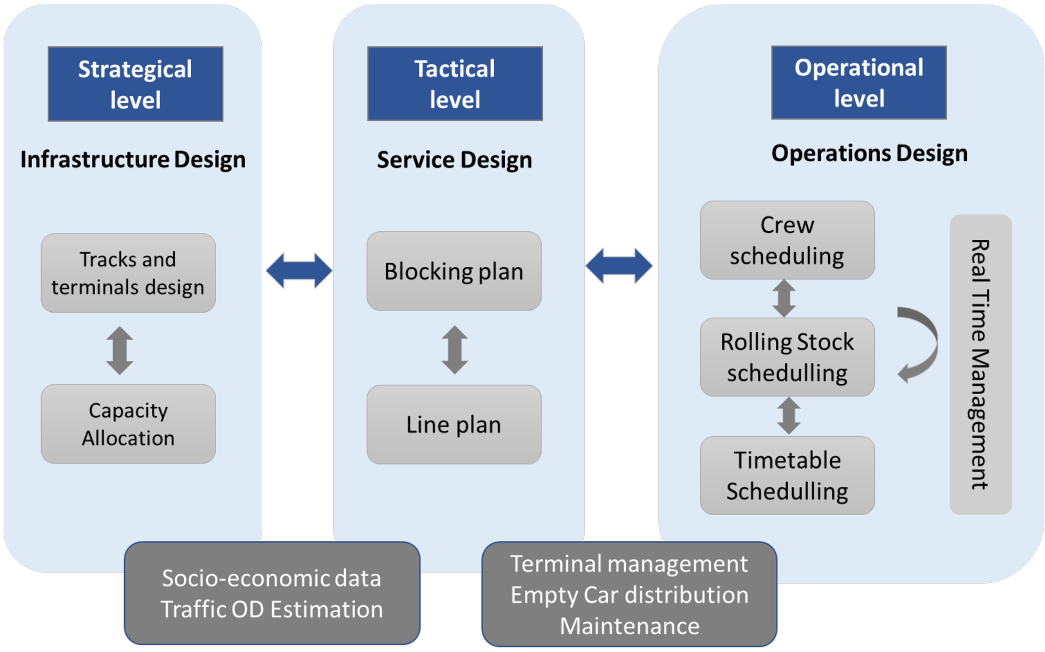


Figure 3.1: Schema of main steps in railway systems planning. Source: author.

dedicated and continue to be made from an operational research point of view.

Network design models are usually considered models at a **strategic level**, i.e., decisions that affect the transportation system during significant periods – and which also involve large investments. Some of the main strategic decisions concern network design: locating the facilities and their characteristics (loading and unloading terminals, centres of consolidation, sidings), which lines to improve and which lines to abandon, among others. A key element for planning at any level is estimating the railway capacity of tracks and classification terminals.

On the **tactical level**, the tasks are carried out in the medium term. Their primary goal is the efficient and rational distribution of existing resources to improve the performance of the entire system. Here can be included activities such as the design of the services in the network, that is, the possible routes and the type of service that will operate, the traffic flow assignment with the available resources, the replenishment of resources (empty vehicles, for instance), as well as the blocking plan models, which try to solve the decision problem of how to optimally group the shipments from their origins to their destinations, using different routes with alternative marshalling yards during the trip.

The **operational level** refers to short-term activities in a dynamic environment where the time factor is critical. The main operational research problems at this level are timetabling and scheduling, rolling-stock allocation and crew scheduling, and models to solve decision problems concerning real-time management, like the time scheduling reassignment as a consequence of a disruption.

Other support activities also appear, which may correspond to different levels, depending on the degree of information available. It is the case of data analysis, essential to estimate flows and origin-destination matrices, necessary both at a **strategic** and **tactical** level, and even at a **operational** level. Also, maintenance tasks, terminal management or, for instance, the distribution of empty vehicles can correspond to both a **tactical** and **operational** phase.

A good perspective about freight transportation models commercially available is offered in Friesz and Kwon (2007). The analysis of **strategic** freight network planning models presents a list of main criteria that an ideal freight planning model should correctly address. They analyse five key commercial models and how they deal relative to these criteria. As a result, the authors recommend dedicating more research efforts to some essential aspects not correctly treated in their understanding. These key aspects are, in short, the simultaneous treatment of shippers and carriers, the necessity of integrating computable general equilibrium models with network models, the inclusion of back-hauling and fleet constraints, considering an imperfect competition, including validation of data in the process, and taking into account the revenue management.

3.2. Traffic assignment models

Usually, network design models are complemented with a multimodal and multi-product regional approach that allows them to include the impact of demand (current and future) and its distribution throughout the network. The multicommodity network design model is a basic Operational Research model used in many aspects of transportation planning. Multiple commodities such as goods, data, or people must be routed between different points of origin and destination on the available arcs (Magnanti and Wong (1984)).

An assignment problem is the distribution of traffic in a network connected by paths, considering a demand between origins and destinations and the transport supply of the network. Assignment methods allocate the flows on the paths in the network according to a set of constraints, mainly related to capacity limits (on links, nodes, and even on vehicles), travel time or travel cost, among others. Determining the flows on these paths involves a solution to a demand/performance equilibrium problem. A performance function is defined independently related to its utility on each path. The demand for transport is associated with the users' behaviour and is not defined for each path separately. Instead, it represents how users choose among the different paths connecting each origin and destination pair. For this reason, neither paths nor origin-destination pairs can be analysed individually.

One of the most used techniques for solving the traffic assignment problems when the capacity limits on the network are not considered is the "All-or-Nothing" (AoN). An AoN algorithm computes the minimum weight path between each pair of origin and destination, so the total demand between these nodes is transported using this unique path. However, this result seems not to be realistic since the flow of transport between two given nodes is usually distributed over different routes. The existence of capacity limits on the network and the fact that the perception of costs can differ from one user to another are the main reasons that explain the distribution of flows among different paths.

The first Wardrop's principle of route choice, known as "User Equilibrium", states that an equilibrium solution for the traffic assignment problem is reached when no user can improve their utility by unilaterally changing routes. This type of problem can be solved by using optimisation methods. The user equilibrium methods require cost functions related to the flow on the network. They apply for the case with capacity constraints, but it is assumed that users choose the least cost path from their origin to their destination. In this case, congestion effects are considered by including penalty costs when the flow on a link or a node is close to its maximum capacity. The usual techniques for solving these types of models were heuristic methods such as the incremental assignment, the well-known method of successive averages (MSA) and the Frank-Wolfe algorithm (Frank and Wolfe (1956)). It should be noted that the models based on the concept of user-equilibrium used for the planning of road and urban transport networks have experienced a huge development, as well as the algorithmic methods to solve them (in contrast to railway networks with a more moderate development). After the Frank-Wolfe algorithm a large number of algorithms for nonlinear optimisation and variational inequalities can be quoted, that have been successful when applied to traffic networks of very large dimensions. It should be noted, among other methods, the restricted simplicial decomposition of Hearn et al. (1987), the algorithm in Bar-Gera (2002) , or the Frank-Wolfe Biconjugate Algorithm by Mitradjieva and Lindberg (2013).

When users behaviour does not follow the mere minimisation of the generalised costs as an undisputable deciding factor, the stochastic multiflow models are the most suitable. These models are based on the concept of utility maximisation and random utility by including a random component in users' perception of the costs. They can help to improve the difficulties that user equilibrium techniques present for distributing flows in different routes in a multi-flow multimodal assignment problem for a large network, as Jourquin (2006) shows. Users can potentially choose any path between their origin and destination, so a multiflow algorithm to obtain a set of realistic alternative routes for each origin-destination pair is required.

A different approach shows the optimisation model developed in Maia and do Couto (2014), where the authors present a support tool based on a strategic traffic assignment model designed to model macro networks with a high aggregation level, being exclusively designed for freight traffic. The model contemplates road and rail transport modes and considers two types of commodities: intermodal cargo, which usually is transported in containers, and general cargo, which does not accept intermodality. The goal is to analyse the impact that new links or the improvement of some existing links have on rail freight share. The model is solved through a local search heuristic, where an AoN algorithm is applied for intermodal cargo, and a stochastic multi-flow method is applied for general cargo. Capacity limits on tracks are imposed by removing from the network the links that have reached their maximum capacity in subsequent iterations. The disadvantage of this method is that once a flow is assigned to a link, it is no anymore possible to remove it for its reassignment to another link.

The model presented in Fernández L. et al. (2004) is oriented to the analysis the performance of freight rail networks. It is formulated as a variational inequality and considers a prioritising treatment for commodities and the distribution of empty rail cars, jointly with the assignment over the rail network of products to be transported. A very detailed representation of rail freight operations at yards is included. The authors suggest applying a method based on strategies to obtain equilibrium flows for each route section. The fact that a non-linear system of equations must be solved to get the solution finally ends up conditioning the use of heuristics to solve the model. Furthermore, calibrating the parameters of the cost functions for the different arcs makes the model hard to use for real scenarios.

3.3. Normative multimode multiproduct network assignment models

At the **strategic** level, the earlier multimode-multiproduct traffic assignment model by Guelat et al. (1990) states a normative model for the distribution of freight flows through various interacting modes by applying user equilibrium-based concepts. This section summarises the main features of this model, which is the basis of the interactive-graphic planning tool STAN (today outdated) for national or regional strategic analysis and planning of freight transportation. It evaluated alternatives through disjointed scenarios that work over huge regional networks. Its relevance relies on the impact of this model on the interurban freight network transport system modelling. The multimode multiproduct model is formulated in a general way to represent easily the distinct transport modes and their inter-connections that compose a multimode transport system. However, given the data aggregation level, the models do not include shippers' or carriers' preferences. Instead, the least cost path and link congestion are the criteria to assign to each product the transport mode among the modes that the product is allowed to use.

The basic idea is to create virtual network structures with specific costs for a particular use of the infrastructure, a similar framework as proposed in Sheffi (1985) under the concept of "supernetworks" and transfer links between modal networks. The network representation relies on three main elements: nodes, links and modes. Each mode corresponds to a mean of transportation that has its own characteristics, such as vehicle type, specific infrastructure or capacity limits. Each link is defined as a triplet made by the origin node, the destination node and the mode allowed in the link. Parallel links represent the situation where different modes may transport goods between two adjacent nodes. Nodes and links represent transfers between modes: each node where transfers occur will be expanded by adding as many nodes as arcs enter and exit the node, and connecting those added nodes by transfer links. As the authors highlight, this network representation facilitates the path definition in the network, as it consists of a sequence of directed links. A transfer from one mode to another mode is also a sequence of directed paths. Moreover, transfers can be limited to certain nodes, and the flows of certain products may be restricted to subsets of modes.

Costs and flows are associated to links, which include transfer links. The generalised cost of a link depends on the flow of the arcs. They suggest calculating the cost for each link as the weighted sum of an operating cost function, a delay function and an energy function. Products are transported from origins to destinations, and it is assumed that each product, when transported from one origin to one destination, has allowed only a subset of modes. Both demand and the corresponding subsets of mode choice are determined exogenously. The optimisation model minimises the total cost of the flows of all products over the multimodal network and over the set of flows that satisfy the conservation of the flow and the non-negativity constraints. An algorithm that takes advantage of the natural decomposition by product is detailed in the paper. They apply the classical method of descent direction for each product, while all the flows corresponding to the other products are kept at their previous value. The descent direction is obtained by applying the shortest path method, and the algorithm stops when the flows barely change within a previously set margin.

From the model presented in Guelat et al. (1990), the particular framework for rail freight transport is detailed in Crainic et al. (1990). Here, the concept of mode may be extended to different track gauges, rail companies, or other particularities on infrastructure, so parallel links joining the same origin and destination nodes, sharing infrastructure and capacity, can be defined. Costs of rail freight mainly depend on trains and railcars. Then, an approximation to convert tons of goods into trains and railcars is described, where it is assumed that each train transports one unique type of product. Limitations on capacities are not included but instead appear indirectly through the non-linearities of the delay functions, in which capacity is treated as a parameter that can be exceeded by flows (soft capacities). Empty railcars are defined as a separate product (one product for each type of railcar) to have empty and loaded railcars simultaneous assigned. Then, a previous OD matrix for empty railcars movements must be calculated. They propose a gravity model based on observed flows. When no information is available, the total product supplies and demands, considering the average weight per product and railcar, are used as the base data.

3.4. Empty railcar distribution

Managing empty railcars also demands special attention since it is a factor that can determine train composition or the order in which products are transported. The impact empty railcars distribution has on operating and capital cost is analysed in Joborn et al. (2004). The authors present a model in which the cost structure for repositioning empty railcars includes economies of scale explicitly. In addition to the usual cost proportional to the number of railcars, there is a cost related to car-handling operations at yards, which mainly depends on the number of trains required for transporting the empty railcars.

As detailed above, when talking about the framework for rail freight transport in Crainic et al. (1990), the software STAN includes empty railcars flow by treating them as a separate product. The only required information is the total product supplies and demands to estimate the OD matrix for empty railcars movements. This approach assumes that railcars run along the network as it was a closed circuit, and all railcars required for transporting the demand must be inside the system. Then, empty and loaded railcars are simultaneously assigned to trains. The same structure follows the model of freight operations for rail transport system presented in Fernández L. et al. (2004), where the distribution model for empty car trips is calculated from the OD matrices by loaded car type, and loaded and empty railcars are then assigned to trains.

At a **tactical** level, a model framework is proposed in Gorman (2015) when information about railcar attributes and customer requirements are available. A railcar can be defined by its permanent attributes (railcar type, track gauge, maximum tonnage, among others) and ephemeral attributes, such as the next available date or location. A customer order also has attributes: specific requirements on railcar type, origin, destination, priority, acceptable earliness or lateness, among others. The proposed model searches for matching the maximum number of customer orders with available railcars at minimum cost. The model is formulated as a min-cost transportation problem.

3.5. Capacity on rail sections

The capacity in a railway network does not have a precise definition because, rather than being a simple upper bound on flows, it is a consequence of multiple factors. An excellent review of capacity concepts and evaluation methodologies is reported in Abril et al. (2008) and Pachl (2015). When talking about track capacity, some factors are defined at a **strategic** level, such as the type of track (double or single), the loops, their size, type of signalling, or turnouts and crossings. Other factors emerge at the **tactical** level, such as the type of train that can run on tracks (depending on average speed and type of signalling, among others), the number of services, or the train composition. Finally, some factors are determined at the **operational** level, such as timetabling and the order in which trains and railcars are handled. Concerning shunting yards, their capacity is mainly defined by the number and size of tracks dedicated to incoming trains, shunting new trains (reclassification of railcars), and inspection and departure operations.

The leaflet prepared by UIC (2013) presents a method for enabling infrastructure managers to carry out capacity calculations from a timetable by following standard definitions, criteria and methodologies from an international standpoint. The approach is based on the compression method, which defines how to calculate capacity consumption for a line, node or corridor. After determining the sections for evaluation and timetable boundaries, the method calculates the capacity consumption, a percentage representing the degree of infrastructure utilisation. The available capacity results from analysing the capacity consumption values from the representative line sections and filling up the train path line sections with additional train paths until a specific capacity consumption is reached. Also, the method proposed facilitates the identification of bottlenecks.

Given the importance and complexity of the capacity problem, many operational research studies that take various approaches can be found in the literature. Most of them present their results from a timetabling basis when the maximum of the information is available (Meirich and Nießen (2016), Jamili (2018), Bešinović and Goverde (2018)). The work developed in Cacchiani et al. (2010) searches for introducing in a railway network with a prescribed timetable as many non-scheduled freight trains as capacity limits permit. They present an integer linear programming formulation, where the input data are scheduled trains and their timetables, and track capacity constraints that limit train circulations are imposed. In Harrod (2009), the author presents an experimental study to analyse the impact of introducing on a track railway saturated with a base of homogeneous (single speed) train paths, priority train paths with significantly higher speed. A network model based on a discrete-time hypergraph results in a utility maximising problem where side constraints remark the operational interaction of separate trains.

To obtain capacity estimates for different components of a railway network, in a previous stage of having the timetable while considering the mutual train interactions, a group of simplified methods have been developed in recent years. According to the different types of trains, these are based on the occupancy of the necessary time slots in a limited time period. As these approaches can be considered a relaxation of the problem, these methods provide an upper bound of the capacity. The approach presented in Rotoli et al. (2016) offers an analytical method based on the UIC proposals, with a schematisation of stations and line segments to be applied in case of lack of more detailed data. The method proposes the evaluation of the capacity and utilisation of each element of the system. In particular, the authors present an analytical expression for the capacity of a line section, distinguishing between single-track and double-track lines, based on the total run of trains classified by categories (high-speed trains, long-distance passenger trains, local-regional passenger trains, freight trains).

Other methods are based on nonlinear mathematical programming problems that are solved heuristically. In Burdett (2015), the author develops various multi-objective mathematical models for analysing the absolute capacity of railway networks. Each of them is focused on different levels of competition for using the railway network: train services, train types or railway corridors. The capacity calculation presented in Mussone and Wolfler Calvo (2013) is based on determining the number of possible circulating trains on a railway system. They take into account knock-on delays indirectly by classifying trains as regular and irregular: regular trains circulate without incurring any conflict, while irregular trains are those incurring a conflict somewhere along the path. As input to the model, the maximum capacity (in number of trains) has to be calculated for each single element of the railway system (lines, complex nodes and station tracks). Due to the non-linearities in the model's constraints, a heuristic is applied to solve the optimisation problem.

It is worth mentioning the effect that the signalling system has on defining capacity, as shown in UIC (2008), which is a generic study supported by the UIC to analyse the influence on capacity consumption that results from using different levels of the European Train Control System (ETCS). Conventional signalling is addressed in Burdett (2016), where the author develops a model for expanding capacity, given a fixed budget, by using track subdivisions and optimal locations of blocking systems, taking into account how the network topology influences capacity limits.

3.6. Classification yards

It is essential to consider operations at classification terminals because shunting operations may take up a significant portion of the total train travel time. They depend mainly on the physical characteristics of the yard and the traffic handled. Classification operations and assembling railcars to an outbound train are the most time-consuming tasks in the yard. The review of models for rail transportation by Assad (1980) dedicates a section to the yard models, where they are classified into two groups, queuing models and simulation models. Most recently, the review by Boysen et al. (2012) analyses the literature on the operational processes at shunting yards, with special attention to the papers focused on the performance of sorting strategies. The author first presents recent works related to single-stage sorting, that is, when a railcar only moves in forward direction through the yard. Next, a detail review of multi-stage sorting papers, where railcars need to be reclassified several times, are detailed.

For the estimation of the level of congestion of marshalling yards, in Petersen (1977a) a simplified queueing model is developed where the major operations performed within a yard are analysed in detail. In Petersen (1977b), the author focus on the impact the physical configuration of the yard and traffic intensities has on the yard delays. Also, within the queueing theory and following the former papers, Turnquist and Daskin (1982) analyse classification and rearrangement operations and present a railcar-based queueing model to deal with delay times at yard. At an operational level, optimal control-based approaches to increase the efficiency of the railcar classification process have been investigated in Shi and Zhou (2015).

3.7. Combined models for modal choice

Traditional approaches to transport planning are based on the well-known “four-step” travel forecasting paradigm: trip generation, trip distribution, mode split and trip assignment. Previous base-year data is the input for the first of the four steps, and then each stage results in the input for the following. However, it is generally recognised that performing one unique iteration of this process is inadequate for predicting travel flows in a congested, multimodal transport network. The solution usually adopted is to iterate the four-step procedure until the link flows, the generalised travel costs and the corresponding origin-destination-mode choices reach an equilibrium, and the results of the four steps are mutually coherent.

In urban passenger’s transport, a better approach which combine modal-split/traffic assignment models have been developed long ago (Evans (1976), Dafermos (1976), Florian (1977), Florian and Nguyen (1978), Abdulaal and Leblanc (1979)). The model proposed in Florian (1977) considers simultaneously demand functions for travel by each mode, route choice equilibrium conditions and flow interactions on the links. Two broad classes of demand models are considered: the first postulates demand for each origin-destination-mode triple to be dependent of flows between the origin-destination pair for all modes. The second class of demand models considers the demand to be dependent on flows from the origin, for all destinations and all modes. This second class postulates that choices of destination and mode are assigned simultaneously. The author proposes as an example for the demand functions a multinomial logit for both model classes. The model is defined as an equilibrium model, where travel time is flow-dependent, and the demand functions determine flows conservation. A particular form of this last model is the model proposed in Florian and Nguyen (1978). Here, entropy type distribution models for each mode and linked together are included. The authors demonstrate that this equilibrium model based on Wardrop’s first principle, with several entropy type distribution models linked with each other, and with shortest route choice for public transport trips is equivalent to a minimisation problem, and a logit model implicitly gives the mode choice.

In Boyce et al. (1994), the authors compare the performance and results of four different iteration procedures based on the four-step paradigm with the model and the algorithm presented in Evans (1976), which combines trip distribution and assignment models in one unique model. The analysis shows that not all iteration procedures work well and that the combined modal-split/traffic assignment is a valid alternative that converges to the desired equilibrium.

These combined modes have been continuously adapted to different scenarios in the urban passenger transportation context. Thus, recently a combined modal split and assignment model with deterministic travel demand is proposed in Li et al. (2009) for intercity bus and train modes for economically related cities. Also, in Hou et al. (2020) a combined modal split/traffic assignment is developed, taking into account park-and-ride facilities.

However, up to our knowledge, no similar models have been adapted or extended to the case of freight transportation: the freight transportation planning models use separate modal split and assignment. In the normative models in Guelat et al. (1990) and Crainic et al. (1990), although multimodal, the mode choice is exogenous, and a subset of permitted origin-destination-modes matrix corresponds to each product. The choice of the paths is determined by the congestion and the functional form of the cost structure. Other authors, such as Maia and do Couto (2012) or Jourquin (2006), which models are based on these normative models, also follow the same pattern.

3.8. Demand models for rail freight transport

International freight transport demand models need to include freight flows between countries and internal flows in the countries, involving data from different sources, most of them not available for third parties. As a consequence, one of the significant challenges for building freight transportation models is the quality of data (Meersman et al. (2016), Friesz and Kwon (2007), de Jong et al. (2015)).

For long-distance freight transport demand modelling, two different data model approaches can be found: aggregate and disaggregate data (Nuzzolo et al. (2013)). Aggregate models are mainly related to the relationship between freight demand flows and socio-economic and level-of-service variables. They are relatively simple to estimate, and, usually, aggregate data is available via periodic statistical publications by governments and other public entities. In the modal choice model proposed in Crisalli et al. (2013), the utility parameters are estimated by using observed data from the Italian Ministry of Transport. In this work, the authors develop a freight mode-service choice model to simulate the competition among road, train and sea transport for national and interregional freight trips in Italy.

However, the aggregate approach appears to be inadequate to reflex the complexity associated with freight transport. Disaggregate models are based on individual behaviour, and, in consequence, they may reflect more accurately the transport flows. But obtaining good quality data for freight transport can be challenging because many critical variables are considered sensitive information, such as transport cost or travel times, and thus, are difficult (or impossible) to obtain. In Feo-Valero et al. (2016) the authors try to avoid the lack of prior information of unknown parameters when a stated preference survey is launched by carrying out the survey in two stages. This method results in a more realistic questionnaire with a better adjustment of the cut-offs. In Arencibia et al. (2015) discrete choice models are applied to analyse the main factors that determine modal choice in freight transportation, focusing on Spain and Europe's flows. They use a stated preference survey where the studied population is limited to the shipper (or receiver) companies. Transport cost, travel time, frequency, and delays are factors determining the utility of the alternatives.

In between the aggregate and the disaggregate models, new modelling approaches propose to simulate disaggregate behaviour of the stakeholders involved in the freight transport process. In Jourquin (2016), the authors present a methodology for validation and calibration of freight transport models adapted to the case of limited and heterogeneous sources of information, especially suited for the Trans-European Network projects.

The route choice model is a critical component of disaggregate models and can be deterministic or stochastic. The deterministic case corresponds to the situation where users select the path that minimises the total generalised costs from their origin to their destination. When congestion appears, not all the paths chosen are those that minimise generalised costs: users are diverted to other paths less congested. But real scenarios show that not all users opt for the least cost path. The stochastic models for route choice deal with this situation, trying to offer an explanation for all these cases that divert from the deterministic selection. The multinomial logit model (MNL) is the most extended model proposed to reflect users' criteria for selecting alternative routes on a transport network. An excellent and comprehensive description of discrete choice methods can be found in the book by Train (2009), and their application to transport modelling in the book by de Dios Ortúzar and Willumsen (2011).

3.9. International projects

In recent years, several public sector national, international, and regional freight transport models have been developed and improved to increase the understanding of the impacts of transport policies on shippers, forwarders, carriers, drivers, the environment, and ultimately the whole society. These models rely on different methodologies and approaches in the scientific literature that deal with the complexity and needs of freight transportation models (Meersman et al. (2016)), such as including various modes of transportation and the variability and seasonality of supply and demand of goods. The increasing use of freight transport models by public authorities for transport planning purposes reflects that freight transport is on the governments' agenda (de Jong et al. (2012)). Many efforts have been devoted to evaluate policies and actions implemented by governments and authorities through the applications of freight planning models. One example can be seen in Crisalli et al. (2013), where authors evaluate the purposes of the Italian National Plan in long-distance freight transport. They present a methodology for assessing the impact that the offer of new services or incentives may have on long-distance freight transport and the mode choice. Also, in Abate et al. (2019), authors present a disaggregated stochastic model of transport chain and shipment size choice, which is compared with the existing Swedish national model, based on disaggregated data but deterministic. The detailed analysis about freight transport chains presented in Jensen et al. (2019) also highlights the importance that freight transport models have for the EU.

In 2011, the EU Commission launched the "White Paper on Transport: Roadmap to a Single European Transport Area - Towards a competitive and resource-efficient transport system" (European Commission (2011)). The goal for transport was to reduce GHG emissions to around 20% below their 2008 level by 2030. Following this objective, the Commission set a target that by 2030, 30% of inland freight transport being transported further than 300 km should be carried out by rail or waterborne, and by year the 2050, that percentage should be 50%. In line with these ambitious objectives, different initiatives and projects have been launched from the private and public sectors in the last decades.

Examples of projects promoted in the last decade are **iFreightMED-DC** project (2008-2013), which pursues the creation of regional Intermodal Freight Services Development Committees, and its continuation, **TRAILS** project (2019-2021), to promote the modal shift in the regions between Catalonia and Occitania on the Mediterranean corridor. Also, the **MARATHON** project (2011-2014) was developed to test the feasibility and effectiveness of introducing longer and heavier trains on a selected high-volume Trans European freight corridor. Other initiative inside the EU is the **United Nations Economic Commission for Europe (UNECE)**, an intergovernmental body dealing with the development of appropriate methodologies and terminology for the harmonization of statistics as well as the collection of data from member States and the dissemination of these data, which has a section on transport statistics.

Shift2Rail, a public-private partnership, is the first European rail initiative which provides a shared platform for the stakeholders of the European rail system, seeking to implement comprehensive and coordinated research and innovation strategy. **Shit2Rail** helps boost the rail supply industry's competitive edge by funding and promoting innovative rail product solutions and technologies to complete the Single European Transport Area. Under the **Shit2Rail** umbrella can be found the **SMART-RAIL** project (2015-2018). This project looks at the European rail freight system, searching for innovating and optimising its operations. Also, the **MOVINGRAIL** project (2018-2020) aims at identifying operational procedures and testing methods for Moving Block signalling, as well as analysing the impacts of Virtual Coupling on different segments of the railway market. More recently, the **FR8RAIL-IV** project (2020-2023) has the goal to develop technologies relevant to the rail freight sector for a more significant share of transport demand to be taken up by the rail sector over the next few decades.

On the initiative of several European railway Infrastructure Managers and Allocation Bodies, **RailNetEurope** (RNE) started in January 2004 to establish a common, Europe-wide organisation to facilitate their international business. Today RNE counts 38 Full Members from over 30 different countries and 11 Associate Members (the Rail Freight Corridors). The **Path Coordination System** is an example of the results of this initiative: it is an international path request coordination system for Path Applicants, e.g. Railway Undertakings, Infrastructure Managers, Allocation Bodies and Rail Freight Corridors. The internet-based application optimises international path coordination by ensuring that all involved parties harmonise path requests and offers. Another project owned by RNE is the **Rail Facilities Portal**, created initially by the European Commission (DG MOVE). The **Rail Facilities Portal** provides quick access to information on all kinds of rail facilities, in particular rail freight facilities, e.g. for the planning of rail services.

Other private associations created to improve rail freight transportation and industrial competitiveness are **FERRMED**, **Rail Freight Forward** or the **Community of European Railway and Infrastructure Companies (CER)**. **FERRMED** is a non-profit multisectoral association founded in Brussels in August 2004 which promotes studies and cooperation among rail stakeholders in Europe (and recently, in Asia via the Euro-Asian Corridor). Currently, they are developing the Study of Traffic and Modal Shift Optimisation in the EU. **Rail Freight Forward** is a coalition of European rail freight companies committed to incrementing the use of rail freight transport and reducing the increasing negative externalities that freight transport expected growth would have on society. **CER** is a European association with members from the entire railway system, founded in 1988, representing the interest of its members by actively providing input to EU policy, in particular, to support an improved business and regulatory environment for European railway undertakings and infrastructure managers.

Finally, it is worth mentioning the web **EcoTransIT World**, previously discussed in Section 2.7, which gives access to software for automatic calculations of energy consumption, carbon emissions and air pollutants. Independent scientific institutes provide the methodology. **EcoTransIT World** calculates, for every global transport chain and different transport modes, the trip distances, energy consumption, carbon dioxide emissions, greenhouse gases emissions as CO_2-equivalent, and air pollutants such as sulphur oxide emissions, nitrogen oxide emissions, non-methane hydrocarbons and particulate matter from vehicles and energy production and provision (power plants, refineries, sea transport of primary energy carriers). The freight transport mode includes air, sea, water inland, rail and road.

Chapter 4

Strategic Modelling for Rail Freight Flows

This chapter introduces the tools required to mathematically represent the rail freight transport characteristics. First, there is a description of the essential elements composing the rail network and how they are represented: tracks, yards, and paths. Then, the different requirements due to the particular characteristics of rail freight transport, which have been detailed in Chapter 2, are translated to their equivalent mathematical expressions.

4.1. Elements of the models

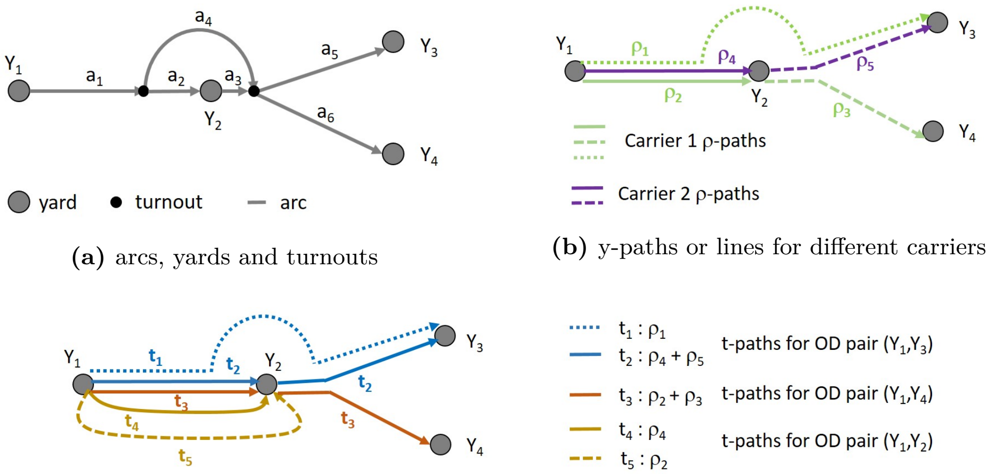
The railway network will be modelled using an undirected graph $\bar{G} = (N, E)$, in which edges $e \in E$ in the graph have a direct correspondence with tracks, and yards and diverting/crossing points will be represented by nodes $i \in N$. Because of the bi-directionality of rails, this will be equivalent to working with a directed graph $G = (N, A)$ where, for each edge $e \in E$ two links $a = (i, j)$ and its opposite $-a = (j, i)$ will exist in A. References to either $\bar{G}$ or to G will be made as convenient. The subset of links $A^- = \{a \in A; a = (i, j), i < j\}$ should also be considered. Each arc $a \in A$ is assumed to have similar physical characteristics along its length, i.e., we can assume that trains run on each of them at constant speed.

Yards are considered highly relevant elements where railcars may be re-classified to assemble outgoing trains using incoming ones. The compatibility of track gauges determines the configuration of these facilities. In addition, it is assumed that suitable facilities exist for unloading incoming railcars and reloading other railcars capable of using the outgoing tracks. Terminals are considered a particular type of yard in which a net amount of products enters or leaves from the "external world" and thus, new trains must be formed or, on the contrary, railcars are left empty and available to form new trains. Terminals usually consist of a port or an intermodal connection point. The set of yards will be denoted by Y, while the set of terminals by $T \subseteq Y$. Finally, diverting/crossing points consist simply of points at which several tracks may merge, where two opposite trains can cross, or where overtaking between trains may occur.

The (inelastic) demand for products for the time horizon is assumed to be known in advance and will be associated to the set of possible origin-destination terminal pairs (OD pairs) on the network. When describing multicommodity flows a multiple superscript $\omega = (\mathfrak{o}(\omega), \mathfrak{d}(\omega), \mathfrak{p}(\omega))$ will be used, where $\mathfrak{o}(\omega)$, $\mathfrak{d}(\omega)$ and $\mathfrak{p}(\omega)$ are the origin $\mathfrak{o}$, destination $\mathfrak{d}$ and product type $\mathfrak{p}$ respectively in triple ω. By W it will be designated the set of all these triples, and P will be the set of all products. Special type of products named "priority products", which require a shorter delivery time, will be considered. This subset of products will be denoted by P^H. The amount of demand of products of type $\mathfrak{p}(\omega)$ to be transported from origin $\mathfrak{o}(\omega)$ to destination $\mathfrak{d}(\omega)$ during a given period $\mathcal{T}$ (for instance, yearly) will be denoted by χ^ω. For a given OD-pair, the set of paths on the network joining the origin terminal $\mathfrak{o}(\omega)$ with the destination terminal $\mathfrak{d}(\omega)$ will be denoted by $R(\omega)$ and the set containing all the paths will be $R = \cup_{\omega \in W} R(\omega)$. Paths between origin terminals and destination terminals will be referred to as t-paths.

Given that rail freight transport is, in general, a deregulated competitive market in most countries, it is necessary to consider the effect of several railway undertakings or carriers competing under possibly some regulated conditions. Let $\mathcal{O}$ be the set of carriers. Each carrier $o \in \mathcal{O}$ operates a set of corridors, which are composed of lines. Each line can be run in both ways and can be decomposed into two directed lines. Each directed line will be referred to as a y-path and denoted by symbol ρ. It has a yard as origin and another different yard as destination. Each y-path is composed of a subset of arcs from the set A that continuously connect both two yards. Because it is possible to have different ways to connect two different yards, it is necessary to previously define exactly the subset of arcs that composes each y-path. The set of y-paths for each carrier will be denoted by $\Gamma(o)$, and $\Gamma = \cup\Gamma(o)$ is the disjoint union of all carrier y-paths. Also, let $Y(o) \subseteq Y$ be the subset of yards which are origin or destination of at least one y-path operated by carrier $o \in \mathcal{O}$. Given an arc $a \in A$, and a carrier $o \in \mathcal{O}$, let $\Gamma(a, o)$ be the subset of y-paths $\rho \in \Gamma(o)$ containing the arc a.

Figure 4.1a illustrates with an example the basic elements of the network: arcs, yards and turnouts. Besides, Figure 4.1b depicts two carriers and their y-paths. Path ρ_1 is made from arcs a_1, a_4, a_5, paths ρ_2, ρ_4 are made from arcs a_1, a_2, while paths ρ_3 and ρ_5 are made from arcs a_3, a_5 and a_3, a_6, respectively. Also, Carrier 1 owns y-paths ρ_1, ρ_2, ρ_3



(a) arcs, yards and turnouts

(b) y-paths or lines for different carriers

(c) t-paths for connecting OD-pairs

Figure 4.1: Railway network elements representation. Source: author.

and Carrier 2 manages y-paths ρ_4, ρ_5.

Let $R(o)$ be the set of t-paths between terminals offered by carrier o. Given a t-path $r \in R(o)$, it can be considered as composed by a subset of y-paths owned by the carrier o, $\rho \in \Gamma(o)$, that connect the t-path origin terminal to the t-path destination terminal in a continuous way (having different stops, defined by yards that are part of each y-path). Figure 4.1c shows different t-paths connecting Y_1, Y_2, Y_3 and Y_4 terminals. t-path t_1 connects terminal Y_1 and Y_3, and is made by the y-path ρ_1. t_2, made by ρ_4 and ρ_5, also connects terminals Y_1 and Y_3, but it has one stop at yard Y_2. Terminals Y_1 and Y_2 are connected by t_4 and t_5, made by ρ_4 and ρ_2, respectively. And terminals Y_1, Y_4 are connected by t_3, which is made by ρ_2, ρ_3. Each t-path is managed by a single carrier: Carrier 1 owns t_1, t_3 and t_5, while Carrier 2 manages t_2 and t_4.

Also, $R(\rho)$ will denote the set of t-paths containing ρ as part of their composition. y-paths between yards are assumed to have homogeneous characteristics accordingly to the types of rolling stock allowed on them, mainly differences in track gauges, loading gauges or electrification. Then, the characteristics on a y-path is assumed to be homogeneous on the track segments composing a y-path. This condition does not necessarily apply to t-paths between terminals. Given a yard $i \in Y$, Γ_i^+, Γ_i^- will denote the set of y-paths outgoing or incoming into i, respectively.

Let $\mathcal{V}$ be the set of railcar types. It will be assumed that each type of product can be transported by only a subset of railcar types. Let $P(v)$ ($P^H(v)$) be the set of products (priority products) compatible with the type of railcars v. For a given y-path, $\mathcal{V}(\rho)$ will denote the set of railcar types $v \in \mathcal{V}$ that can run on ρ, reflecting the possibility that different gauges may be operating on the network. For instance, a track segment can be prepared for only one gauge or more than one, or incompatible track segments may operate in a yard or a terminal. Also, railcars within a railcar type may circulate on more than one gauge. $\Gamma(v, o)$ represents the subset of y-paths compatible with v-type railcar and operated by carrier $o \in \mathcal{O}$, while $\Gamma_i^+(v, o)$, $\Gamma_i^-(v, o) \subseteq \Gamma(v, o)$ represent the same, but restricted to y-paths emergent from $i \in Y$ or incident to $i \in Y$, respectively.

Rail freight transport uses different types of locomotives, with different characteristics, as maximum speed, maximum weight, weight/speed ratio, track gauge or voltage. Let K_M be the set of freight locomotive types (which identify the different type of trains), and $K_M^H \subseteq K_M$ the subset of freight locomotives suitable to transport priority products. So, as in the case of railcars, not all locomotives are compatible with all types of tracks. Let $K_M(\rho) \subseteq K_M$ be the subset of $k \in K_M$ which are compatibles with $\rho \in \Gamma$.

Finally, the road network, the alternative transport mode, is represented by additional links that directly connect each OD-pair's origin and destination.

The list of variables and parameters will be introduced in the following sections, as they are required. Table A.1, Table A.2 and Table A.3 in Appendix A summarize the list of sets, parameters and variables, respectively.

4.2. Demand balance equations

Shippers need their products to be transported from origin to destination, having two options. On the one hand, rail carriers provide rail freight transportation service competing with each other, and on the other, different carriers offer road freight transportation services. Let h_r^ω be the total demand or flow during period $\mathcal{T}$ transported by train for OD-pair and product ω using t-path r, and let $h^\omega = \sum_{r\in R(\omega)} h_r^\omega$ and $\tilde{h}^\omega$ be the total demand transported by train and by truck, respectively, for OD-pair and product. Variables h_r^ω and $\tilde{h}^\omega$ can be grouped in flow vectors $\mathbf{h}^\omega = (..., h_r^\omega, ...)^\top \in \mathbb{R}^{|R(\omega)|}$, $\omega \in W$, $\mathbf{h} = (..., \mathbf{h}^\omega, ...)^\top \in \mathbb{R}^{|W|}$ and $\tilde{\mathbf{h}} = (..., \tilde{h}^\omega, ...)^\top \in \mathbb{R}^{|W|}$, for convenience (here $\top$ denotes transpose).

When competition between rail and road is considered, demand will be carried from origin to destination by train or by truck, as equation (4.1) shows:

$$\chi^\omega = \sum_{r\in R(\omega)} h_r^\omega + \tilde{h}^\omega \qquad \forall \omega \in W. \quad (4.1)$$

However, when the model focuses on rail transport and the competition among different rail carriers, demand will be carried from origin to destination only by train, following the rail paths that carriers provide. Then, equation (4.2) reflects the demand balance:

$$\chi^{\omega} = \sum_{r \in R(\omega)} h_r^{\omega} \qquad \forall \omega \in W. \quad (4.2)$$

4.3. Rolling stock conditions the rail freight transport

In rail transportation, products are transported in railcars, and locomotives pull railcars, forming trains that run along the rail network and are composed and decomposed at yards. This section describes the equations that determine how trains, railcars and locomotives condition rail freight transport.

4.3.1. Products and rolling stock relationship

Usually, a railcar is loaded at origin and unloaded at the destination; that is, there is no product manipulation during the trip. Taking this into account, there is a relationship between the amount of freight of type $\mathfrak{p}(\omega)$ transported from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$, moved on a given directed line ρ operated by carrier $o \in \mathcal{O}$, $\rho \in \Gamma(o)$, and the necessary flow of railcars compatibles with ρ. Equation (4.3) reflects this connection, where variable $f_{\rho}^{v,\omega}$ represents the loaded railcars of type v that run on ρ and go from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$, and $\alpha^{v,\mathfrak{p}(\omega)}$ represents the average capacity of units of freight of type $\mathfrak{p}(\omega)$, for a railcar of type v:

$$\sum_{r \in R(\rho)} h_r^{\omega} \leq \sum_{v \in \mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_{\rho}^{v,\omega} \qquad \forall \omega \in W,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}. \quad (4.3)$$

Carrier conditions for the transport of products are mainly related to physical characteristics of rail transportation: track features, trains' length and weight, capacity limits on tracks and yards, among others. Transportation services provided by carriers also influence shippers' choice. Next equations gather rail freight transportation characteristics and conditions with carrier flow requirements. Equation (4.4) below defines the total flow, F_{ρ}^{v}, of railcars of type v that run on line ρ. Variables $f_{\rho}^{v,\emptyset}$ represent empty railcars, while variables $f_{\rho}^{v,\omega}$ represent loaded railcars. Clearly, railcars of type v should not run on line ρ if they are not compatible (because of track gauge, for instance).

$$F_\rho^v = \begin{cases} f_\rho^{v,\emptyset} + \sum_{\omega \in W} f_\rho^{v,\omega} & \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall o \in \mathcal{O}, \\ 0 & v \text{ and } \rho \text{ incompatible.} \end{cases} \tag{4.4}$$

Ideally, the aim is that railcars should run along the entire network, and yards have no spare units of railcars. Then, equation (4.5) sets a balance on yards: at each yard and for each carrier, the entrance flow of railcars of type v belonging to carrier o must be equal to the exit flow of that railcars.

$$\sum_{\rho \in \Gamma_i^-(v,o)} F_\rho^v = \sum_{\rho \in \Gamma_i^+(v,o)} F_\rho^v \qquad \forall i \in N,\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}. \tag{4.5}$$

Let m_ρ^k be the variable for the flow of trains of type k on line ρ. Tracks and yards on a directed line determine the maximum train length, while the locomotive characteristics and track features condition maximum train weight. Let ℓ^v, α^v be the length and tare of railcars of type v, respectively. Also, let $\alpha^{v,\mathfrak{p}(\omega)}$ be the average weight of product $\mathfrak{p}(\omega)$ transported on railcars of type v, while $\bar{\ell}_\rho$ represents maximum train length allowed on y-path ρ. Finally, let $\bar{\alpha}_\rho^k$ be the maximum weight allowed for locomotives of type k on y-path ρ. Then, the following two constraints, (4.6) and (4.7), state limitations for flows of railcars and trains due to the maximum length and weight, respectively:

$$\sum_{v \in \mathcal{V}(\rho)} \ell^v F_\rho^v \leq \bar{\ell}_\rho \sum_{k \in K_M} m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O} \tag{4.6}$$

$$\sum_{v \in \mathcal{V}(\rho)} (\alpha^v F_\rho^v + \sum_{\omega} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega}) \leq \sum_{k \in K_M} \bar{\alpha}_\rho^k m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in O. \tag{4.7}$$

Let h_ρ^L be the variable for the amount of priority products transported by non-priority trains on y-path ρ. The following constraints need to be applied jointly with a penalty factor for the variables h_ρ^L that will be applied in the objective function, thus ensuring that, due to capacity limitations, priority products will be the first option for filling faster trains.

$$\sum_{v \in \mathcal{V}(\rho)} \sum_{\mathfrak{p}(\omega) \in P^H(v)} \alpha^{v,\omega} f_\rho^{v,\omega} - \sum_{k \in K_M^H} \bar{\alpha}_\rho^k m_\rho^k \leq h_\rho^L \qquad \forall \rho \in \Gamma. \tag{4.8}$$

4.3.2. Trains formation

A railcar may be part of different convoys during the trip if the t-path has some intermediate yards distinct from the origin and the destination. Each additional stop and composition or decomposition of trains increases the total travel time and the total costs. So, it is important to know where trains are mounted and dismounted. Let us define the variable $\theta_{i',i}^{k,o}$ as the number of trains of type k operated by carrier o that are mounted at yard $i' \in Y(o)$ and dismounted at yard $i \in Y(o)$. Also, let $m_{\rho,i}^{k}$ be the variable for the number of trains of type k that run on each line ρ operated by carrier o, with destination yard i. The total number of trains of type k running on line ρ, m_{ρ}^{k}, holds (4.9), while the relationship between $m_{\rho,i}^{k}$ and $\theta_{i',i}^{k,o}$ will be given by the balance equation (4.10):

$$m_{\rho}^{k} = \sum_{i \in Y(o)} m_{\rho,i}^{k} \qquad \forall k \in K_M,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O} \tag{4.9}$$

$$\sum_{\rho \in \Gamma_{i'}^{+}(o)} m_{\rho,i}^{k} - \sum_{\rho \in \Gamma_{i'}^{-}(o)} m_{\rho,i}^{k} = \theta_{i',i}^{k,o} \qquad \forall k \in K_M,\ \forall i', i \in Y(o), i' \neq i,\ \forall o \in \mathcal{O}. \tag{4.10}$$

4.3.3. Rolling stock requirements

Rolling stock is expensive. Carriers try to adjust the number of railcars and locomotives to their service needs. Actually, this corresponds to a tactical model, which requires information that is perhaps not available at the moment the network is being designed. Models for assigning fleet units to line services are widely used in railway passenger transportation, but the integration of these models would probably lead to an intractable problem. Instead, an estimation based on Little's law (see, e.g. Little (1961)) can be applied to state lower bounds for the number of railcars and locomotives a carrier needs to provide the service. This approach is based on considering that railcars run along the network as if it was a closed tour, and that all vehicles needed to transport all the demand must remain inside the system.

Let $\bar{\mathcal{T}}$ be a parameter for the effective time that rolling stock may be running during the period $\mathcal{T}$. For instance, if $\mathcal{T}$ corresponds to one year, $\bar{\mathcal{T}}$ may represent the total working days in one year. Parameter t_ρ is the average run time for one train in line ρ, taking into account necessary layovers: t_ρ could be calculated from the average travel time among the different locomotive types, plus an extra time for the waiting time related to the arrival to destination yard, or the exit from the origin yard. Variable $\lambda^{v,o}$ represents the minimum number of v-type railcars that a carrier o needs to provide the service. Following Little's law, the first inequality on the equation (4.11) states that the average number of railcars needed to perform the service is at least the sum of the number of wagons that run on a line multiplied by the average time spent on that line. The second inequality allows limiting the maximum number of railcars available by fixing a parameter $L^{v,o}$ as the maximum number of v-type railcars that a carrier o may require. Equation (4.12) is the equivalent of (4.11) for locomotives, being the parameter $\hat{L}^{k,o}$ the maximum number of k-type locomotives at carrier o disposal.

$$\frac{1}{\bar{\mathcal{T}}} \sum_{\rho \in \Gamma(o)} t_\rho F_\rho^v \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (4.11)$$

$$\frac{1}{\bar{\mathcal{T}}} \sum_{\rho \in \Gamma(o)} t_\rho m_\rho^k \leq \hat{L}^{k,o} \qquad \forall k \in K_M,\ \forall o \in \mathcal{O}. \quad (4.12)$$

Notice that $t_\rho / \bar{\mathcal{T}}$ corresponds to the fraction of the effective time $\bar{\mathcal{T}}$ expended to run along the y-path ρ.

4.4. Carrier-Shipper relationship

In the tactical model, where the focus is on rail and road competition and also the competition among carriers, it is essential to consider the cost-effectiveness of carriers. Equation (4.13) below states that a carrier has no losses, or on the contrary, it is left out and carries out no transportation of goods.

The left hand side of inequality (4.13) represents the total import paid by shippers to a carrier, while the right hand side of the inequality corresponds to direct costs associated with rail transport. These direct costs are composed of three terms: the first term captures train composition and decomposition costs, the second is the running time cost, and the third is for renting/maintenance costs of railcars.

A *big-M* component is added to this equation with a new binary variable $\hat{y}_0$ also included into an additional constraint (4.14). Due to the modal choice characteristics of the tactical model none of the modes has the possibility of capturing entirely the demand of a product per O-D pair. Then, a constraint that forces the viability of carriers transportation may cause infeasibility on the model (for instance, when an OD pair is served by only one carrier and this carrier cannot reach enough demand to be competitive). M_o is a constant greater than the maximum carrier cost, and $\bar{\chi}^\omega$ is a small fraction of the total demand χ^{ω}:

$$\sum_{\omega \in W} \sum_{r \in R(\omega,o)} U_r^{\omega} h_r^{\omega} \geq \sum_{k \in K_M} \sum_{i \in Y(o)} \sum_{\substack{j \in Y(o) \\ j \neq i}} (C_{i'}^{k,o} + \tilde{C}_i^{k,o}) \theta_{i',i}^{k,o} + \sum_{k \in K_M} \sum_{\rho \in \Gamma(o)} \hat{C}_{\rho}^{k} m_{\rho}^{k} + \sum_{v \in \mathcal{V}} D^{v,o} \lambda^{v,o} - M_o(1 - \hat{y}_o) \quad \forall o \in \mathcal{O} \quad (4.13)$$

$$\sum_{r \in R(\omega,o)} h_r^{\omega} \leq \bar{\chi}^{\omega} + \chi^{\omega} \cdot \hat{y}_o \quad \forall o \in \mathcal{O},\ \forall \omega \in W. \quad (4.14)$$

The rest of the parameters are: U_r^{ω} is the price per unit for transporting $\mathfrak{p}(\omega)$ from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$ using carrier path r; $C_{i'}^{k,o}, \tilde{C}_i^{k,o}$ are the cost for train formation at yard i' and train decomposition at yard i, respectively, for k-type train and carrier o; $\hat{C}_{\rho}^{k}$ is the cost of a k-type train running on ρ. $D^{v,o}$ is the cost for renting/maintenance of v-type railcars, for carrier o. Obviously, these are not the unique costs associated with rail transport. So, to avoid the lack of information for other costs, it is advisable to apply a percentage of increment on these costs when the model is applied.

4.5. Approximation of the network capacity

As stated in Section 3.5, estimating capacity is one of the greatest difficulties that network railway problems must cope with. This section details the capacity analysis for the design model, developed in Chapter 5, which is different from that used for the tactical model, developed in Chapter 6. Section 4.6 presents the capacity approach for the tactical model. In the design model, capacity analysis is based on recommendations for the line sections' capacity analysis from UIC (2013). The occupancy time of a single block section definition and the concept of compression detailed in the former constitute the main elements for building the design model's capacity constraints. These consider the headway between two consecutive trains and, in this case, the approach in Rotoli et al. (2016) is applied to calculate the average headway between two consecutive trains, based on the blocking signal distance when a fixed block system is the signalling system used on the network.

4.5.1. Decomposing tracks into block sections

As stated in Section 4.1, when describing the set of arcs A, each arc $a \in A$ is assumed to have similar physical characteristics (slope, curvature and others) all along the arc. This condition allows assuming that each type of train has uniform speed on each arc

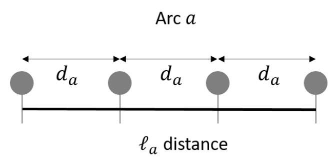


Figure 4.2: Arc decomposition into signal blocks. Source: author.

a. A section of rail should be split into equidistant subsections in order to increase the maximum theoretical capacity (Burdett (2016, Lemma2)). Then, each arc a is divided into equidistant signal blocks, whose length will be represented by the variable d_a on the model, while the variable b_a corresponds to the total number of signals on the arc, and ℓ_a will be the length of arc a, which is previously known. Figure 4.2 shows an example of an arc split up into three equidistant signal blocks.

Following the guidelines published by the rail authorities, such as the technical report Railway Group Standard (2015), which establishes the rules for calculating the minimum distance between two line-side signals, a lower bound for each d_a can be fixed. It depends on the braking distance, and it can be calculated from the authorized speed for the block and track slope, among other things. Let ℓ_a^{min} the lower bound for d_a. The next equations (4.15) - (4.17) show the relationship between the block signal distance and length of the arc:

$$b_a d_a = \ell_a \qquad \forall a \in A \quad (4.15)$$

$$d_a \geq \ell_a^{min} \qquad \forall a \in A \quad (4.16)$$

$$d_a \in \mathbb{R}^+, b_a \in \mathbb{N}, b_a \geq 1 \qquad \forall a \in A. \quad (4.17)$$

4.5.2. The blocking-time components

Afterwards, it is necessary to define the blocking time components: part of them are constant and part of them depends on the train characteristics. Let t_a be the safety blocking time for each block section of arc a for a given type k of train. Here, train class index k will be omitted for simplicity. Figure 4.3 shows the components of the safety blocking time, t_a, defined as:

$$t_a = t_s + t_r + t_{at} + t_{ot} + t_{clr} + t_s,$$

where t_s is the switching time (before accessing to block interval and after leaving it), t_r is the driver reaction time, t_{at} is the approaching time, t_{ot} is the physical occupation

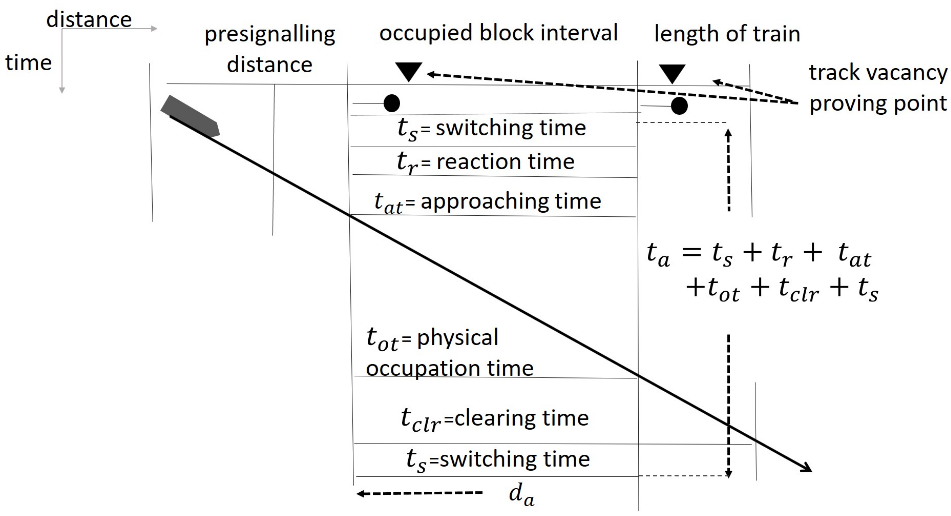


Figure 4.3: Blocking time components for a single block section. Source: UIC (2013).

time and t_{clr} is the clearing time. Notice that t_s, t_r have fixed values and are known in advance. Also, t_{clr} depends on the speed and length of the train, while t_{ot} and t_{at} depend on train speed and block length. Let σ_a be the average speed for the k-type train on arc a. It can be assumed that the length of the train is the maximum train length allowed on track a, represented by ℓ_a^{max}. Then, $t_{clr} = \ell_a^{max}/\sigma_a$. t_{ot} is the physical occupancy time. The block distance is represented by the variable d_a. As the train speed is assumed to be constant on all the arc a, then $t_{ot} = d_a/\sigma_a$. Finally, t_{at}, the approaching time, depends on the block length, and, as it is shown in Rotoli et al. (2016), it can be assumed as equal to the physical occupation time, to guarantee the not disrupted circulations, avoiding unnecessary acceleration/deceleration phases. So, $t_{at} = d_a/\sigma_a$ Then, the blocking time results as follows:

$$t_a = 2t_s + t_r + \frac{2d_a + \ell_a^{max}}{\sigma_a} \qquad \forall a \in A. \quad (4.18)$$

4.5.3. Line-segment capacity

This approach follows the approximation presented in Rotoli et al. (2016). The International Union of Railways (UIC) in the first edition of its leaflet 405R proposes the next formula to calculate the capacity of a line-segment:

$$P_a = \frac{\mathcal{T}_a}{\underline{\Delta}_a + t_{ru} + t_{zu}}, \quad (4.19)$$

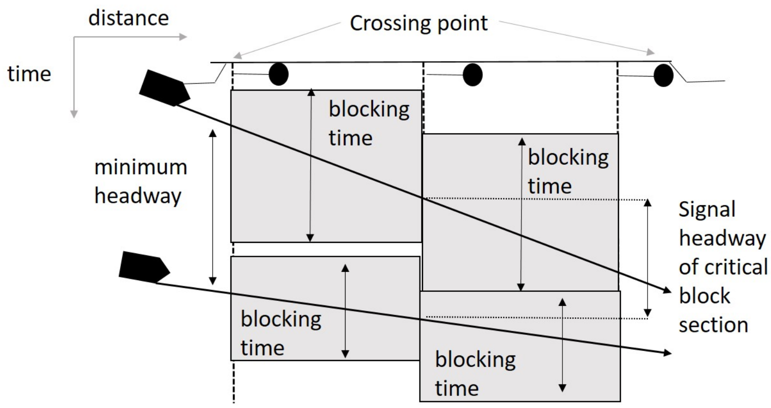


Figure 4.4: Blocking time between two trains. Source: UIC (2008)

where P_a is the capacity of the line-segment a in number of trains in period $\mathcal{T}_a$, $\underline{\Delta_a}$ is the average minimum headway of the line segment a, t_{ru} is for the running time margin and t_{zu} is for the added time. Both t_{ru} and t_{zu} are previously fixed values, added for safety.

The average minimum line headway, when a timetable is unknown, can be calculated by using the minimum line headway between two consecutive trains, train k' following train k, represented by $\Delta_a^{k,k'}$, and the relative frequency of the combination of train k' following train k (Pachl (2015, Chapter 5)):

$$\underline{\Delta_a} = \sum_{k,k'} \Delta_a^{k,k'} \frac{n_a^k n_a^{k'}}{\sum_k n_a^k} , \tag{4.20}$$

where n_a^k is the total number of trains of type k that runs on arc a. The headway between two consecutive trains can be calculated as a function of the blocking signal distance. Figure 4.4 shows different block sections for two trains on a track, where each rectangle corresponds to one signal block and the security lag time required by a train. To determine the minimum headway time, the blocking-time sequence of the second train is shifted until it touches the graph of the preceding train. The minimum headway time is represented by $\Delta_a^{k,k'}$.

But, if applied to the design model, equation (4.20) will result in a set of non-linear constraints, involving the product of continuous variables. To avoid this extra-difficulty, the simplification proposed in Rotoli et al. (2016) can be adopted, where the average minimum headway for each line segment is calculated by using a weighted average of the minimum headway between two consecutive trains of the same type, as expressed in equation (4.21). The proportion of each type of train $n_a^k / \sum_{k \in K_M} n_a^k$ is the weight applied, and Δ_a^k is the variable that represents the minimum headway between two consecutive trains of the same type k on arc a:

$$\underline{\Delta}_a = \frac{1}{\sum_{k \in K_M} n_a^k} \sum_{k \in K_M} n_a^k \Delta_a^k \qquad \forall a \in A. \quad (4.21)$$

As $\sum_{k \in K_M} n_a^k \leq P_a$, that is, the total trains that runs on arc a must be lower than the capacity of the arc, the following equation (4.22) states:

$$\sum_{k \in K_M} n_a^k \leq \frac{\mathcal{T}_a}{\underline{\Delta}_a + t_{ru} + t_{zu}} \qquad \forall a \in A. \quad (4.22)$$

From equations (4.21) and (4.22), and taking into account that the total number of trains n_a^k can be expressed in terms of m_ρ^k (the number of trains on ρ):

$$n_a^k = \sum_{\rho \in \Gamma(a)} m_\rho^k \qquad \forall k \in K_M, \forall a \in A \quad (4.23)$$

$$\sum_{k \in K_M} \sum_{\rho \in \Gamma(a)} m_\rho^k \Delta_a^k + \sum_{k \in K_M} \sum_{\rho \in \Gamma(a)} m_\rho^k (t_{ru} + t_{zu}) \leq \mathcal{T}_a \qquad \forall a \in A. \quad (4.24)$$

Notice that equation (4.24) is also non-linear, but it results from the product of an integer variable by a continuous variable. In this case, a well-known standard techniques facilitate its linearisation.

For double-tracks, the minimum headway between two consecutive trains of the same type corresponds to the blocking-time, that is, $\Delta_a^k = t_a^k$. From this equivalence and (4.18), the minimum headway between two trains of the same type can be expressed as in equation (4.25):

$$\Delta_a^k = \frac{2d_a + \ell_a^{max}}{\sigma^k} + 2t_s + t_r \qquad \forall a \in A, \forall k \in K_M. \quad (4.25)$$

For single-track line segments, the approach in Rotoli et al. (2016) assumes that only one train can occupy the whole line segment per time, independently from its running direction. Although this assumption holds in the case of two trains in the opposite direction, operating two or more trains in the same directions may be possible under some safety rules. This is not always implemented because it is a standard practice to balance traffic. Then, for single-track line segments, the authors propose calculating the minimum line headway as the result of the travel time with constant speed, plus the acceleration and deceleration times. However, when the line segments are long enough, both acceleration and deceleration times can be assumed to be included when calculating the average speed of a train on the line segment. Then, the only additional condition that should be added for single-track line segments is that there is only one block section in the line segment, represented by:

$$b_a = 1 \qquad \forall a \in A_s, \quad (4.26)$$

where $A_s \subseteq A$ is the subset of single-track arcs.

4.5.4. Capacity limit on yard tracks

The capacity of a yard track layout depends on the number of tracks and the dwell times of trains served in the terminal. Also, the required number of tracks depends on the inbound and outbound traffic flow the yard has to handle (Pachl (2015, Chapter 5)). Usually, the inbound traffic flow should equal, on average, the outbound traffic flow. Then, the total number of trains that run on y-paths will depend on the maximum service rate capacity M_i^0 in number of trains per track in the origin and destination yards $i \in Y$. Let $\bar{g}_i, i \in Y$ be the variable for the size of the yard i, in number of tracks. Then, the following equations (4.27) and (4.28) state the capacity limits on yards:

$$\sum_{o \in \mathcal{O}} \sum_{\substack{j \in Y(o) \\ j \neq i}} \sum_{k \in K_M} \theta_{i,j}^{k,o} \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y \quad (4.27)$$

$$\sum_{o \in \mathcal{O}} \sum_{\substack{j \in Y(o) \\ j \neq i}} \sum_{k \in K_M} \theta_{j,i}^{k,o} \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y. \quad (4.28)$$

4.6. External and fixed capacity limits

Previous approximation to estimate capacity limits on the network relies on design decisions considered in the optimisation model. This kind of approximation can be applied to a design model where infrastructure improvements analysis is the goal of the model. However, for a tactical model, infrastructure design is usually exogenous, and infrastructure managers are in charge of allocating capacity on tracks and yards for carriers' disposal. In this situation, capacity limits typically are expressed in terms of the maximum number of trains that can run along a track or can be handled on a yard.

4.6.1. Track capacity limits based on slots

Consequently, rail carriers buy the slots that interest them among those available. These slots limit the maximum number of trains a carrier can operate in each directed line.

Let N_ρ be a fixed value that represents carriers' capacity allocation on ρ-path. Then, constraint (4.29) imposes a limit on the maximum number of trains on that directed line. Likewise, the tracks capacity limit, N_a, restricts the maximum number of trains that run on track a; this is represented by (4.30):

$$\sum_{k \in K_M} m_\rho^k \leq N_\rho \qquad \forall \rho \in \Gamma(o), \forall o \in \mathcal{O} \quad (4.29)$$

$$\sum_{o \in \mathcal{O}} \sum_{\rho \in \Gamma(a,o)} \sum_{k \in K_M} m_\rho^k \leq N_a \qquad \forall a \in A. \quad (4.30)$$

4.6.2. Fixed capacity limit on yards

The following constraints limit the total incoming and outgoing flow of trains that can be dismounted and mounted, respectively, on a yard $i \in Y$, accordingly to the yard's capacity in the number of trains (parameter $\tilde{N}_i$):

$$\sum_{o \in \mathcal{O}} \sum_{i' \in Y(o)} \sum_{k \in K_M} \theta_{i,i'}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y \quad (4.31)$$

$$\sum_{o \in \mathcal{O}} \sum_{i' \in Y(o)} \sum_{k \in K_M} \theta_{i',i}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y. \quad (4.32)$$

4.7. Dealing with additional traffic

External or traversal traffic flows on the area under study (passenger train flows, traverse freight train flows, or maintenance trains) have been included in the model as constant values by reducing the period available for running the freight trains that take part in the model (the $\mathcal{T}_a$ period parameter on equation (4.24)). Another option is to define a new set of types of trains, the extra-trains K_E, and defining $K := K_M \cup K_E$. Let $N_a^k, k \in K_E$ be the total extra-trains of type k that run on arc a, previously fixed. Then, (4.24) should be changed by the following equation (4.33):

$$\sum_{k \in K_M} \sum_{\rho \in \Gamma(a)} m_\rho^k \Delta_a^k + \sum_{k \in K_E} N_a^k \Delta_a^k + \sum_{k \in K} \sum_{\rho \in \Gamma(a)} m_\rho^k (t_{ru} + t_{zu}) \leq \mathcal{T}_a \qquad \forall a \in A. \quad (4.33)$$

being $\mathcal{T}_a$ the period available for running freight trains and extra-trains in this case. Note that to have an exact value for the parameter $N_a^k,\ k \in K_E,\ a \in A$ to run the model could be a challenging task. For this reason, we have opted for working with the first option (equation (4.24)).

4.8. Location/Allocation conditions

The usual location/allocation constraints should be included in a design model. Let g_i, z_a, r_a be binary variables. g_i represents if yard i is built or not, z_a represents if track a is built or not, and r_a permits to differentiate if track a is to build double ($r_a = 1$) or single ($r_a = 0$):

$$z_a = z_{-a},\ r_a = r_{-a} \qquad \forall a \in A^{\text{-}} \quad (4.34)$$

$$r_a \leq z_a \qquad \forall a \in A \quad (4.35)$$

$$z_a \leq g_i,\ z_a \leq g_j \qquad \forall a = (i,j) \in A \quad (4.36)$$

$$z_a = 0 \implies \sum_{k \in K} \sum_{\rho \in \Gamma} \epsilon_{a,\rho} \cdot m_\rho^k = 0 \qquad \forall a \in A. \quad (4.37)$$

Equation (4.34) guarantees that one track exists if and only if its opposite exists, and that the condition of single or double track is the same for each track and its opposite. Equation (4.35) assures that the condition of being double applies only to existing tracks. Constraint (4.36) sets that no track exists if either the origin or the destination yards do not exist. Being $\epsilon_{a,\rho} = 1$ if arc a is part of y-path ρ, and $\epsilon_{a,\rho} = 0$ on the contrary, equation (4.37) ensures that trains run only on existing tracks.

4.9. Domain of the variables

To complete the model, the domain of the variables is as follows:

- Demand products, in tons:

$$h_r^\omega \in \mathbb{R}^+ \qquad \forall r \in R(\omega),\ \forall \omega \in W \quad (4.38a)$$

$$\tilde{h}^\omega \in \mathbb{R}^+ \qquad \forall \omega \in W. \quad (4.38b)$$

- Rolling stock, in number of railcars or trains:

$$f_\rho^{v,\omega} \in \mathbb{Z}^+ \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall \omega \in W,\ \forall o \in \mathcal{O} \quad (4.39a)$$

$$f_\rho^{v,\emptyset} \in \mathbb{Z}^+ \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall o \in \mathcal{O} \quad (4.39b)$$
$$m_{\rho,j}^k \in \mathbb{Z}^+ \qquad \forall \rho \in \Gamma(o), \forall j \in Y,\ \forall k \in K_M,\ \forall o \in \mathcal{O} \quad (4.39c)$$
$$\lambda^{v,o} \in \mathbb{Z}^+ \qquad \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (4.39d)$$
$$\theta_{i',i}^{k,o} \in \mathbb{Z}^+ \qquad \forall i, i' \in Y, i \neq i',\ \forall k \in K_M,\ \forall o \in \mathcal{O}. \quad (4.39e)$$

- Infrastructure decision variables ($\bar{g}_i$ is expressed in number of tracks):

$$g_i \in \{0,1\} \qquad \forall i \in Y \quad (4.40a)$$
$$\bar{g}_i \in \mathbb{R}^+ \qquad \forall i \in Y \quad (4.40b)$$
$$z_a \in \{0,1\} \qquad \forall a \in A \quad (4.40c)$$
$$r_a \in \{0,1\}, \qquad \forall a \in A. \quad (4.40d)$$

- Signalization system decision variables:

$$b_a \in \mathbb{N}, b_a \geq 1 \qquad \forall a \in A \quad (4.41a)$$
$$d_a \in \mathbb{R}^+ \qquad \forall a \in A. \quad (4.41b)$$

- Auxiliary variables for cost-effectiveness of carriers business:

$$\hat{y}^o \in \{0,1\} \qquad \forall o \in \mathcal{O}. \quad (4.42)$$

Chapter 5

Rail Network Design Model

This chapter develops a mathematical programming-based model to evaluate the impact of infrastructure improvements and capacity expansion, specifically when applied to a mixed rail network with multiple carriers offering their services. The model here introduced is an evolution of the model presented in the 22nd Euro Working Group on Transportation Conference (2019):

> Rosell, F., & Codina, E. (2020). A model that assesses proposals for infrastructure improvement and capacity expansion on a mixed railway network. *Transportation Research Procedia, 47, 441–448.* doi:10.1016/j.trpro.2020.03.119

From now on, we will refer to the model introduced in this chapter as **DCECC**-model (for Design and Capacity Expansion with multiple Carriers). The **DCECC**-model will be built from the structure introduced in Chapter 4.

5.1. Model Description

The problem presented is a multi-objective minimisation problem, where each of the objectives corresponds to: a) cost related to infrastructure investments and maintenance; and b) operating costs. The main decision variables of the model are related to the construction of new tracks (double/single), the length of the blocking sections on the tracks for the fixed block system, and the capacity on terminals/yards. The structure of the problem relies on three blocks of constraints. First, balance equations for products and rolling stock define one set of constraints. Second, a set of hard capacity constraints are defined, given that a rail network is usually of shared use by both passengers and freight. In this case, they are based on blocking distance criteria, as explained in Section 4.5. Finally, the third block evaluates the necessary rolling stock, i.e., how railcars (full and empty) and locomotives run throughout the network and how many are needed to satisfy the transportation demand.

5.1.1. Life-cycle costs involved

The life-cycle costs associated with investments in building, maintenance and operations are formulated in this section. It must be noted that, because the model is oriented toward the design/expansion of a freight transport system, the effects on the passenger network are taken into account only in the constraints, which guarantees that the passenger transport network performs within a set of acceptable bounds. The following components are taken into account:

Capital investment. It includes the initial investment and maintenance of the classification yards, tracks, and fixed block systems. Regarding the yards, costs can vary depending on their size, so in this case, we apply a factor related to the number of tracks. The first group of addends of (5.1) corresponds to yard costs, while the second group is for tracks and the third is for signalling in the fixed block system:

$$\sum_{i\in N}\left(S_i \cdot g_i + \bar{S}_i \cdot \bar{g}_i\right) + \sum_{a\in A^-}\left(\dot{S}_a \cdot z_a + \ddot{S}_a \cdot r_a\right) + \sum_{a\in A}\hat{S}_a \cdot b_a \; . \tag{5.1}$$

Binary variables g_i and z_a denote respectively, whether or not yard i and track a are to be built, while $\bar{g}_i$ is a non-negative integer variable for representing the size of yard $i \in Y$ in terms of the number of tracks. The binary variable r_a represents if track a is double ($r_a = 1$) or single ($r_a = 0$). Finally, b_a represents the number of block signal sections on track a. To estimate the monetary cost of each component, the following parameters are defined: S_i and $\bar{S}_i$ correspond (respectively, fixed and per track) the building or maintenance costs for yard i; $\dot{S}_a$ and $\ddot{S}_a$ are (respectively, fixed and additional - if double) the building or maintenance costs of a track on link $a \in A^-$; $\hat{S}_a$ is the maintenance cost per blocking signal on link $a \in A$. Notice that in the case of tracks, the set $A^- = \{a = (i,j) \in A; i < j\}$ is used to avoid duplicity in track costs, due to that arcs a and $-a$ share the same physical infrastructure. However, signalling may be different in each direction.

Operating costs. They are reflected in (5.2) and each of the components are described below following the order they appear in the equation:

a) preparation/reclassification costs in the terminals/yards, depending on the number of trains and the type of trains (monetary and/or waiting time and/or delay costs);

b) railway cost for goods transport on the tracks, depending on the travel distance and the type of locomotive;

c) investing or maintenance costs for railcars; and

d) a penalty cost for tons of priority products transported on non-fast trains.

$$\sum_{\substack{k\in K_M,\\ o\in O}} \sum_{\substack{i,j\in Y(o)\\ i\neq j}} (C_i^{k,o} + \tilde{C}_j^{k,o})\theta_{i,j}^{k,o} + \sum_{\substack{k\in K_M,\\ o\in O}} \sum_{\rho\in\Gamma(o)} \hat{C}_\rho^k m_\rho^k + \sum_{\substack{v\in\mathcal{V},\\ o\in O}} D^{v,o}\lambda^{v,o} + \sum_{\rho\in\Gamma} \tilde{D}_\rho h_\rho^L \,. \tag{5.2}$$

Parameters $C_i^{k,o}/\tilde{C}_j^{k,o}$ are the inbound/outbound costs at yards i, j, respectively, for a train composed at yard i and decomposed at yard j, represented by the variable $\theta_{i,j}^{k,o}$. $\hat{C}_\rho^k$ is the travel cost for k-type trains on y-path ρ, represented by the variable m_ρ^k. $D^{v,o}$ is the investment or maintenance cost for the total railcars of type v owned/hired by carrier o, represented by the variable $\lambda^{v,o}$. $\tilde{D}_\rho$ is a penalty for the total tons of priority products h_ρ^L transported on non-priority trains on y-path ρ. Notice that the travel cost for train usually includes the investment/maintenance cost for locomotive, expressed as import per kilometre.

5.1.2. Constraints of the model

Three main keystones define the structure of the constraints of the model. The first concerns how goods are transported throughout the railway network, which gives rise to equations for products and rolling stock. Second, because the rolling stock is expensive, including an approximation of the total number of wagons and locomotives required for composing the trains is essential. And third, the capacity limits on tracks must be considered since most of the paths share tracks, and most tracks are also used for passenger transportation. Also, the total demand needs to be transported from each origin to its destination, so a demand balance must be considered. All the constraints have been introduced previously in Chapter 4, but we reproduce here some of them for the reader's convenience.

Demand balance equations. The main purpose of the introduced model in this chapter is to ascertain how and which railway investments may generate returns to the rail freight transportation system. Thus, road transport is not considered, and the sole traffic analysed is rail traffic. Under this assumption, variables $\tilde{h}^{\omega}$ (and the flow vector $\tilde{\mathbf{h}} = (..., \tilde{h}^{\omega}, ...)^{\top} \in \mathbb{R}^{|W|}$) will not be used in the **DCECC**-model. Therefore, the demand balance equation to apply in the model is equation (4.2), which distributes the demand for each triple origin-destination-product among the different rail t-paths connecting the origin and destination (c.f. Section 4.2).

Relationships for carriers' rail transport flows. Because all the products are transported by following t-paths, it is required railcars and trains to perform the transport flow. The conditions on rail transport flows will be given by:

- the relationship between tons of products and railcars (equation (4.3));
- the balance equations on yards (equations (4.4) and (4.5));
- train length and maximum weight limits (equations (4.6) and (4.7));
- priority products will be the first option for faster trains (equation (4.8));
- trains composition and decomposition on yards (equations (4.9) and (4.10)).

Rolling stock. Section 4.3.3 introduces the equations for estimating a lower and an upper bound of the total of railcars and locomotives required for each carrier to carry out the services. Equation (4.11) establishes the lower and upper bound for railcars, while equation (4.12) sets the upper bound for locomotives.

Capacity limits on the railway network. Section 4.5 describes in detail the approach made in the **DCECC**-model for applying capacity limits on the railway network. To sum up, capacity limits in the **DCECC**-model will be given by:

- the decomposition of tracks into block sections (equations (4.15)-(4.16));
- the line-segment capacity limits based on the average minimum headway between two consecutive trains (equations (4.24), (4.25) and (4.26)); and
- the capacity limits on yard tracks conditioned by the number of tracks on yard (equations (4.27) and (4.28)).

Location-allocation constraints. As introduced in Section 4.8, the usual location / allocation constraint should be considered to guarantee the coherence of the resulting infrastructure. That is, a link exists if and only if its opposite also exists, and the condition of being single or double applies to a link and its opposite (equations (4.34)).

Only existing tracks can be double (equation (4.35)). If a yard does not exist, neither links exiting nor arriving at the yard exist (equation (4.36)). And finally, trains run only on existing tracks (equation (4.37)).

Domain of the variables. To complete the model, let us summarize the domain of the variables, as they are introduced in Section 4.9:

$$h_r^{\omega} \in \mathbb{R}^+ \,, \tag{4.38a}$$

$$f_{\rho}^{v,\omega}, f_{\rho}^{v,\emptyset}, m_{\rho,j}^{k}, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+ \,, \tag{4.39}$$

$$g_i, z_a, r_a \in \{0,1\}, \bar{g}_i \in \mathbb{R}^+ \,, \tag{4.40}$$

$$b_a \geq 1, \in \mathbb{N},\ d_a \in \mathbb{R}^+ \,. \tag{4.41}$$

The explicit and detailed formulation of the **DCECC**-model appears fully compiled in Appendix B.1.

5.1.3. Pareto efficiency analysis

A good way to find out how investments generate returns to the freight transport costs is by means of analysing points in the efficient frontier between investment costs and operating costs.

Let X be the feasible set defined by the sets of constraints of the **DCECC**-model, that is, equations (4.2)-(4.12), (4.15)-(4.16),(4.24)-(4.28), (4.34)-(4.37), (4.38a), (4.39)-(4.41), while x represents the vector of variables of the **DCECC**-model.

In the **DCECC**-model a bi-objective function $f = (f_0, f_1)$ is considered, where f_0 is comprised of infrastructure investments given in (5.1), and f_1 is made up of summation of the operating cost as expressed in (5.2). Functions f_0, f_1 must be rescaled using the elements in the *trade-off* table:

	f_0	f_1
x_0^*	$f_0(x_0^*) = \underline{f}_0$	$f_1(x_0^*) = \hat{f}_1$
x_1^*	$f_0(x_1^*) = \hat{f}_0$	$f_1(x_1^*) = \underline{f}_1$

where x_0^* is an optimal solution of the optimisation problem $\min_{x\in X} f_0(x)$, and x_1^* is an optimal solution of the optimisation problem $\min_{x\in X} f_1(x)$. The rescaled bi-objective problem results as follows in (5.3), with $0 \leq \gamma \leq 1$:

$$Min_{x \in X} \;\; \gamma \frac{f_0(x) - \underline{f}_0}{\hat{f}_0 - \underline{f}_0} + (1-\gamma) \frac{f_1(x) - \underline{f}_1}{\hat{f}_1 - \underline{f}_1} \,. \tag{5.3}$$

Then, after a reduction of the constant term in (5.3), the efficient points will be found by solving the following bi-objective problem, with $0 \leq \gamma \leq 1$:

$$Min_{x \in X} \;\; \gamma \frac{f_0(x)}{\hat{f}_0 - \underline{f}_0} + (1-\gamma) \frac{f_1(x)}{\hat{f}_1 - \underline{f}_1} \,. \tag{5.4}$$

Obtaining the complete set of efficient points can be a very hard task because the problem (5.4) to solve are of mixed-linear integer type. Instead, a limited subset of values for γ is chosen in the analysis of the case study in Chapter 8. Graphically, the Pareto (or efficient) frontier will be represented approximately for both values of f_0 and f_1 for the chosen set of values for γ.

5.2. Replacing nonlinearities

The capacity constraint (4.24) contains non-linearities. Also, railcar bound constraint (4.11) and locomotive bound constraint (4.12) may be non-linear. This section details how to deal with these non-linearities for easy-solving the optimisation problem.

5.2.1. Capacity limits on tracks

Constraint (4.24) limits the maximum number of trains that can run on a track, and the first addend is non-linear, as shown in Section 4.5.3:

$$\sum_{k \in K_M} \sum_{\rho \in \Gamma(a)} m_\rho^k \Delta_a^k + \sum_{k \in K_M} \sum_{\rho \in \Gamma(a)} m_\rho^k (t_{ru} + t_{zu}) \leq \mathcal{T}_a \qquad \forall a \in A. \tag{4.24}$$

The non-linear part results from the product of an integer variable ($n_a^k = \sum_{\rho \in \Gamma(a)} m_\rho^k$) and a continuous variable (Δ_a^k). There are standard techniques to reformulate the product of a bounded integer variable and a continuous variable as linear constraints using additional binary variables. It can be assumed that n_a^k, which represents the number of k-type trains that runs on track a, is bounded. Let $c_a^k \in \mathbb{N}$, such as $n_a^k \leq 2^{c_a^k}$. Then n_a^k can be decomposed as a sum of products of power-of-two terms multiplied by a binary variable, as follows:

$$n_a^k = \sum_{i=0}^{c_a^k} 2^i \phi_{a,i}^k \qquad \forall k \in K_M,\ \forall a \in A \quad (5.5)$$

$$\phi_{a,i}^k \in \{0,1\} \qquad i = 0,\ldots,c_a^k,\ \forall k \in K_M,\ \forall a \in A. \quad (5.6)$$

Let us define $\zeta_{a,i}^k := \Delta_a^k \phi_{a,i}^k$. The next conditions state:

$$\sum_{k \in K_M} \sum_{i=0}^{c_a^k} 2^i \zeta_{a,i}^k + \sum_{k \in K_M} n_a^k (t_{ru} + t_{zu}) \leq \mathcal{T}_a \qquad \forall a \in A \quad (5.7)$$

$$\phi_{a,i}^k = 0 \implies \zeta_{a,i}^k = 0 \qquad \forall i = 0,\ldots,c_a^k,\ \forall k \in K_M,\ \forall a \in A \quad (5.8)$$

$$\phi_{a,i}^k = 1 \implies \zeta_{a,i}^k = \Delta_a^k \qquad \forall i = 0,\ldots,c_a^k,\ \forall k \in K_M,\ \forall a \in A \quad (5.9)$$

$$\zeta_{a,i}^k \in \mathbb{R}^+, \qquad \forall i = 0,\ldots,c_a^k,\ \forall k \in K_M,\ \forall a \in A. \quad (5.10)$$

Equations (5.5)-(5.10) will replace equation (4.24) when solving the **DCECC**-model.

5.2.2. The impact of congestion on yards

Also, constraints (4.11) and (4.12) may be non linear, depending whether t_ρ value is deterministic or depends on trains congestion on yards:

$$\frac{1}{\mathcal{T}} \sum_{\rho \in \Gamma(o)} t_\rho \cdot F_\rho^v \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (4.11)$$

$$\frac{1}{\mathcal{T}} \sum_{\rho \in \Gamma(o)} t_\rho \cdot m_\rho^k \leq \hat{L}^{k,o} \qquad \forall k \in K_M,\ \forall o \in \mathcal{O}. \quad (4.12)$$

In this case, notice that variable t_ρ, which represents the y-path ρ run time, can be decomposed on three terms:

$$t_\rho := \tau_\rho + \hat{\tau}_{\mathfrak{o}(\rho)} + \hat{\tau}_{\mathfrak{d}(\rho)},\ \forall \rho \in \Gamma, \quad (5.11)$$

where τ_ρ represents the run time along the tracks, and $\hat{\tau}_{\mathfrak{o}(\rho)}, \hat{\tau}_{\mathfrak{d}(\rho)}$ correspond to the waiting time on origin and destination yards, represented by $\mathfrak{o}(\rho)$ and $\mathfrak{d}(\rho)$, respectively, for the departure and the arrival. It can be assumed τ_ρ to be deterministic, as there is no congestion on tracks. As explained in Section 4.3.3, τ_ρ could be calculated from the average travel time weighted by type of locomotive. However, the waiting time on yards may be more dependent on congestion, which in turn depends on the number of trains

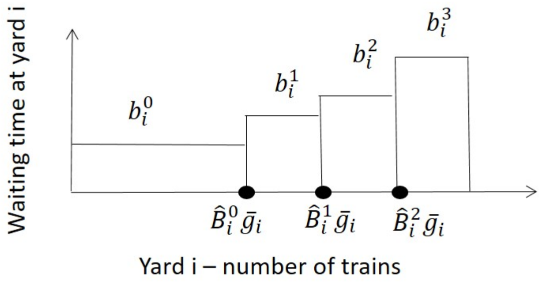


Figure 5.1: Yard congestion approximation. Source: author.

that arrives or departs from the yard.

The approach to estimate each $\hat{\tau}_i, i \in Y$ as congestion-dependent is made using an approximation by an increasing stepwise function, depending on the total trains being handled and the yard capacity. Figure 5.1 shows an example for the yard i, where $\hat{B}_i^0, \hat{B}_i^1, \hat{B}_i^2$ are the total number of handled trains per hour and track, while $\bar{g}_i$ is the integer variable that represents the size of the yard in terms of the number of tracks. Values $b_i^0, b_i^1, b_i^2, b_i^3$ are the time congestion levels (expressed in hours). These levels can be established by using, for example, simulation or surveys. Then, $\hat{\tau}_i$ can be decomposed as a sum of products (one for each level of congestion) of one binary variable multiplied by the corresponding level of congestion and thus, for each of the new products resulting, the previous technique can be applied in order to linearise the product of a binary variable and a continuous variable.

For instance, from the example depicted on Figure 5.1, the linearisation of (4.11) will be as follows:

$$\hat{\tau}_i = \sum_{j=0}^{3} b_i^j x_i^j \qquad \forall i \in Y \quad (5.12)$$

$$\hat{m}_i = \frac{1}{\bar{B} \cdot \mathcal{T}} \sum_{k \in K_M} \Big(\sum_{\rho \in \Gamma_i^+} m_\rho^k + \sum_{\rho \in \Gamma_i^-} m_\rho^k \Big) \qquad \forall i \in Y \quad (5.13)$$

$$x_i^0 = 1 \Rightarrow \hat{m}_i \leq \hat{B}_i^0 \bar{g}_i \qquad \forall i \in Y \quad (5.14)$$

$$x_i^1 = 1 \Rightarrow \hat{B}_i^0 \bar{g}_i \leq \hat{m}_i \leq \hat{B}_i^1 \bar{g}_i \qquad \forall i \in Y \quad (5.15)$$

$$x_i^2 = 1 \Rightarrow \hat{B}_i^1 \bar{g}_i \leq \hat{m}_i \leq \hat{B}_i^2 \bar{g}_i \qquad \forall i \in Y \quad (5.16)$$

$$x_i^3 = 1 \Rightarrow \hat{m}_i \geq \hat{B}_i^2 \bar{g}_i \qquad \forall i \in Y \quad (5.17)$$

$$\sum_{j=0}^{3} x_i^j = 1 \qquad \forall i \in Y \quad (5.18)$$

$$x_i^j \in \{0,1\},\ j = 0,1,2,3 \qquad \forall i \in Y. \quad (5.19)$$

Equation (5.12) decomposes $\hat{\tau}_i$, while equation (5.13) defines $\hat{m}_i$ as the total trains handled on yard i by hour, being $\bar{B}$ the yard daily-average opening-hours and $\mathcal{T}$ the analysed period. Equations (5.14)-(5.17) state the level of congestion on yard. Equation (5.18) guarantees that only one level of congestion is reached, and equation (5.19) determines the nature of the binary variables added.

Replacing (5.11) and (5.12) into (4.11) we obtain:

$$\frac{1}{\mathcal{T}} \sum_{\rho \in \Gamma(o)} (\tau_\rho F_\rho^v + \sum_{j=0}^{3} b^j_{\mathfrak{o}(\rho)} x^j_{\mathfrak{o}(\rho)} F_\rho^v + \sum_{j=0}^{3} b^j_{\mathfrak{d}(\rho)} x^j_{\mathfrak{d}(\rho)} F_\rho^v) \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in O. \quad (5.20)$$

Then, each product $x_i^j F_\rho^v$ should be replaced by a continuous variable $\hat{x}_{i,\rho}^{j,v}$ and a new set of constraints will appear, as it was done previously for the linearisation of the capacity constraints on tracks:

$$\frac{1}{\mathcal{T}} \sum_{\rho \in \Gamma(o)} (\tau_\rho F_\rho^v + \sum_{j=0}^{3} (b^j_{\mathfrak{o}(\rho)} \hat{x}^{j,v}_{\mathfrak{o}(\rho),\rho} + b^j_{\mathfrak{d}(\rho)} \hat{x}^{j,v}_{\mathfrak{d}(\rho),\rho})) \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (5.21)$$

$$x_i^j = 0 \Rightarrow \hat{x}_{i,\rho}^{j,v} = 0 \qquad \forall j = 0,\ldots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (5.22)$$

$$x_i^j = 1 \Rightarrow \hat{x}_{i,\rho}^{j,v} = F_\rho^v \qquad \forall j = 0,\ldots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O} \quad (5.23)$$

$$\hat{x}_{i,\rho}^{j,v} \in \mathbb{R}^+ \qquad \forall j = 0,\ldots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}. \quad (5.24)$$

Constraints (5.13)-(5.19) and (5.21)-(5.24) will replace (4.11). A similar linearisation could be made for (4.12).

The explicit and detailed formulation of the **DCECC**-model after the constraints linearisation appears fully compiled in Appendix B.2.

5.3. Previous versions of the DCECC-model

An initial version of the **DCECC**-model was presented during the 8th International Conference on Railway Operations Modelling and Analysis (RailNorrköping2019), held at Linköping University, campus Norrköping, in Sweden on June 17th – 20th, 2019. An enhanced version was presented during the 22nd Euro Working Group on Transportation Meeting (EWGT 2019), held in Barcelona (Spain) on September 18th - 20th, 2019 (Rosell and Codina (2020)). We will refer to it as the **DCEM**-model (for Design and Capacity Expansion in a Monopolistic environment). For the **DCEM**-model, it is assumed that traffic flows obey centralized decisions and that a single operator establishes which services will be appropriate for transporting specific freight flows. There are two main differences between the **DCEM**-model and the **DCECC**-model: a) how they deal with capacity limits on tracks, and b) the introduction of carriers in the **DCECC**-model, which in turn impact how rolling stock requirements are calculated. Below both differences will be detailed.

The explicit and detailed formulation of the **DCEM**-model appears fully compiled in Appendix B.3.

5.3.1. The capacity limits in the DCEM-model

The approach to the capacity limits in the **DCEM**-model has two significant issues: the (possible) lack of enough data to use the model in real scenarios, particularly in medium/large rail networks, and the excess of simplification applied to single-tracks.

First, we need to introduce a new concept within the set of types of trains. Here the set of types of trains includes also all external traffic, as traversal freight trains and passenger trains, i.e., $K = K_M \cup K_E$. Trains are supposed to run in blocks or groups, with each group containing only one type of train, $k \in K$. $\Theta(K)$ will denote the ordered set of all groups of trains, and $\varsigma \in \Theta(K)$ represents a group of trains. Different groups may be made up by trains of the same type, that is, if $k(\varsigma)$ denotes the train type for the group of trains ς, then it is possible that another group $\varsigma' \in \Theta(K)$ exists so that $k(\varsigma) = k(\varsigma')$. These groups are assumed to follow a predetermined order during the running period. To simplify, the same order will be assumed on all tracks. It is easy (but hard in notation) to extend the formulation for the case of a particular order on each track, by associating each track with its particular order for the groups of trains that run on it. Let $g(k) \subseteq \Theta(K)$ be the subset of all groups in $\Theta(K)$ which share the same train type $k \in K$.

The total number of each type of train that runs all along each track is split into all the groups containing the same type of trains. Let π^{ς} be the percentage that represents the portion of trains of the type $k(\varsigma)$ that run in ς group, and it is fixed in advance. For single tracks, it is assumed that all trains run in one direction first, followed by all trains running in the opposite direction. These hypotheses are rather restrictive and allow us to obtain only an upper bound for theoretical capacity on tracks. Then, by trying to better adjust to realistic cases and in following recommendations of UIC (2013), a

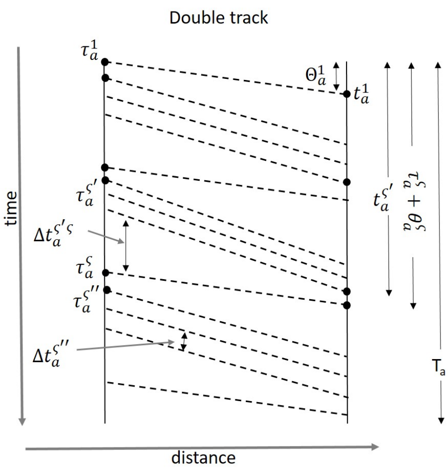


Figure 5.2: Train flows along a double track. Source: author.

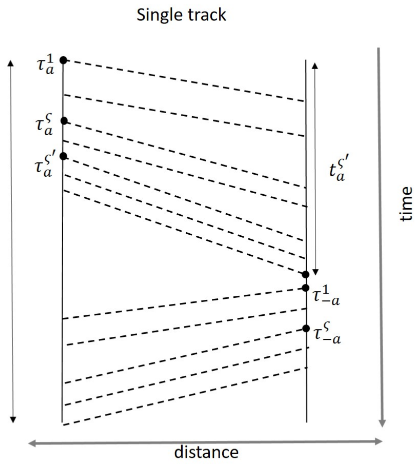


Figure 5.3: Train flows along a single track. Source: author.

factor will be applied to the maximum theoretical capacity in order to reduce it. So, the maximum capacity allowed will be a percentage of the theoretical capacity calculated.

Figure 5.2 shows an example of how these train groups share a track on a double-track link, whereas Figure 5.3 provides an example of a distribution of train groups applied to a track in a single track link. Figures 5.2 and 5.3 also show how the running time of train groups conditions the occupancy time. Following the compression method defined in UIC (2013), the idea is to fit on each track all the train groups that have at least one train running on it. Therefore, train groups run consecutively, leaving only the minimum safety time between one group and the next. The same rule applies to consecutive trains that run on the same group. $\varsigma \rightarrow \varsigma'$ will denote the relationship between two consecutive train groups, where group $\varsigma' \in \Theta(K)$ runs just after group $\varsigma \in \Theta(K)$. For each group of trains $\varsigma \in \Theta(K)$, let τ_a^ς, t_a^ς and Δt_a^ς be, respectively, the instant of initial running time, the instant of ending running time and the minimum headway time between two consecutive trains of type $k(\varsigma)$. Next, $\theta_a^{k(\varsigma)}$ will be a (constant) value for the running time of class $k(\varsigma)$ trains, depending on its average speed and track slope. Finally, let $\Delta t_a^{\varsigma,\varsigma'}$ be the minimum headway between train groups ς and ς', where group ς' runs just after group ς.

Using these definitions, the following capacity constraints (5.27) to (5.36) can be built as a function of trains groups and applied to track links. Notice that a decision binary variable r_a is included so that these constraints can apply to both single and double track, with ς_0 and ς_s being the first and last trains groups.

$$\Delta_a^k = \frac{2d_a + \ell_a^{max}}{\sigma^k} + 2t_s + t_r \qquad \forall k \in K,\ \forall a \in A \quad (5.25)$$

$$\Delta_a^{k,k'} = \frac{d_a}{\sigma^{k'}} + \frac{d_a + \ell_a^{max}}{\sigma^k} + 2t_s + t_r \qquad \forall k, k' \in K, k \rightarrow k'\ \forall a \in A \quad (5.26)$$

$$n_a^\varsigma \leq \pi^\varsigma \sum_{\rho \in \Gamma_a} m_\rho^{k(\varsigma)} \qquad \forall \varsigma \in \Theta(K), k(\varsigma) \in K_M, \forall a \in A \quad (5.27)$$

$$\sum_{\varsigma \in g(k)} n_a^\varsigma = \sum_{\rho \in \Gamma_a} m_\rho^k \qquad \forall k \in K_M, \forall a \in A \quad (5.28)$$

$$\tilde{n}_a^\varsigma = \max(0, n_a^\varsigma - 1) \qquad \forall \varsigma \in \Theta(K), \forall a \in A \quad (5.29)$$

$$\tau_a^\varsigma + \delta_a^\varsigma \cdot \theta_a^{k(\varsigma)} + \Delta t_a^\varsigma \cdot \tilde{n}_a^\varsigma \leq t_a^\varsigma \qquad \forall \varsigma \in \Theta(K),\ \forall a \in A \quad (5.30)$$

$$\tau_a^{\varsigma'} \geq \tau_a^\varsigma + \Delta t_a^\varsigma \cdot \tilde{n}_a^\varsigma + \Delta t_a^{\varsigma,\varsigma'} \qquad \forall \varsigma, \varsigma' \in \Theta(K), \varsigma \rightarrow \varsigma',\ \forall a \in A \quad (5.31)$$

$$\tau_a^{\varsigma'} + \delta_a^{\varsigma'} \cdot \theta_a^{\varsigma(\varsigma')} \geq t_a^\varsigma + \Delta t_a^{\varsigma,\varsigma'} \qquad \forall \varsigma, \varsigma' \in \Theta(K), \varsigma \rightarrow \varsigma',\ \forall a \in A \quad (5.32)$$

$$\delta_a^\varsigma = 0 \Rightarrow \sum_{\rho \in \Gamma_a} m_\rho^{k(\varsigma)} = 0 \qquad \forall \varsigma \in \Theta(K), \forall a \in A \quad (5.33)$$

$$r_a = 0 \Rightarrow t_a^{\varsigma_s} \leq \tau_{-a}^{\varsigma_0} \qquad \forall a \in A^- \quad (5.34)$$

n_a^ς has a fixed value, previouly known $\qquad \forall \varsigma \in \Theta(k), k(\varsigma) \in K_P, \forall a \in A \quad (5.35)$

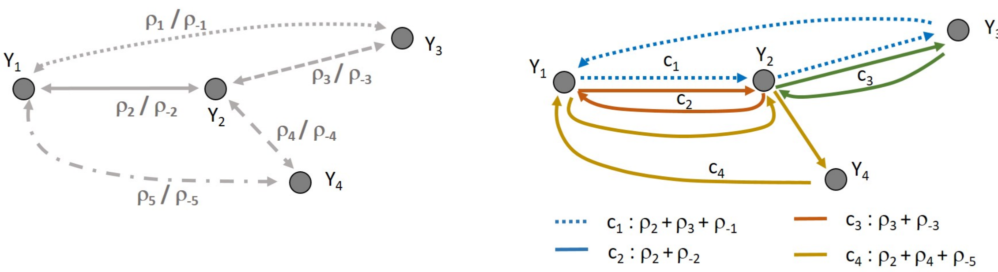
$$\delta_a^\varsigma, r_a \in \{0,1\}, \quad \tau_a^\varsigma, t_a^\varsigma \in \mathbb{R}^+,\ \tau_a^\varsigma \leq \mathcal{T},\ t_a^\varsigma \leq \mathcal{T}\,. \quad (5.36)$$

Notice that equation (5.25) which defines the minimum headway time between two consecutive trains of the same group, is equivalent to equation (4.25). However, equation (5.26) for defining the minimum headway time between two consecutive trains of a different group changes slightly with respect to equation (4.25). The difference consists in that the approaching time depends on the average speed of the second train, while the occupation time and the clearing time depend on the first train. Parameters σ^k and $\sigma^{k'}$ correspond to the average speed of trains of group k and k', respectively. Equation (5.27) applies the portion of trains that corresponds to group ς to bound from above the maximum number of trains of groups ς that runs on track a (the n_a^ς variable). Equations (5.28) ensures that the sharing out of the k-type train groups on a track is coherent with the total k-type trains that run on the track. Equations (5.29) and (5.30) link the initial time of a train group with its ending time and the total number of trains that run on it. Equations (5.31) and (5.32) define the relationship between two consecutive train groups. Also, equation (5.32) ensures that initial and ending times are equal if no trains run on the group. Equation (5.33) guarantees that no train runs on the group if there is no difference between the initial and ending times. Equation (5.34) applies only for single tracks, in order to verify that trains run in one direction first before running in the opposite. Equation (5.35) allows to include external flows (passenger or other freight trains) that are currently running (or are expected to run). Finally, equation (5.36) fixes the domain of variables.

5.3.2. The concept of cycles for rolling stock estimation on the DCEM-model

The **DCEM**-model assumes that traffic flows obey centralized decisions and that a single operator establishes which services will be appropriate for transporting specific freight flows. This assumption conditions the way rolling stock runs on the rail network. The approach in the **DCEM**-model differs from the one made in the **DCECC**-model. Both suppose that railcars run on the rail network following a closed tour, and the difference appears in the scope of each closed tour. For the **DCECC**-model, each carrier has its own paths, and it is supposed they can get their railcars to run along all their paths. That is, paths managed by the carrier delimit the scope for closed tours for this carrier's railcars. Instead, by having a single carrier for all the rail network, it is necessary to establish some consistently closed tours. Here appears the concept of cycles.

For a better approximation, a new graph $\mathcal{G} = (Y, \Gamma)$ is defined in order to properly state the relationship between the required rolling stock and the sojourn time in the cycles in $\mathcal{G}$. Thus, we can presume that railcars run by following some predefined cycles inside the network. Figure 5.4 shows an example of how cycles are introduced in the **DCEM**-model. On the left, Figure 5.4a represents four yards, Y_1, Y_2, Y_3, Y_4 and the y-paths that connect them, in one direction and the opposite. For instance, ρ_1 goes from Y_1 to Y_3, while ρ_{-1} goes from Y_3 to Y_1. Also, ρ_2 goes from Y_1 to Y_2, and ρ_{-2} goes from Y_2 to Y_1. Having only one carrier, each two yards are connected by at most one y-path and its opposite. Figure 5.4b depicts some of the cycles trains may cover among these yards. For instance, c_1 is a cycle between Y_1, Y_2 and Y_3, and it is made up by y-paths ρ_2, ρ_3 and ρ_{-1} . c_2 and c_3 are round trips which join Y_1 and Y_2, and Y_2 and Y_3, respectively. c_2 is made up by ρ_2 and ρ_{-2}, while c_3 is made up by ρ_3 and ρ_{-3}. Obviously, not all the cycles that are feasible make sense: the cycle made up by y-paths ρ_2, followed by ρ_4, then ρ_{-4}, ρ_3 and ρ_{-1} is unlikely to be run by a train. Let C be the subset of cycles in (Y, Γ) that railcars are supposed to follow more likely, and let $C(\rho)$ be the subset of cycles in C containing the y-path ρ. Following the example in Figure 5.4b, $C(\rho_2) = \{c_1, c_2, c_4\}$. $\mathcal{V}(c)$ will denote the subset of railcar types that are compatible with all y-paths that form c. It will be assumed that the subset of cycles in



(a) Paths connecting yards

(b) Some cycles among yards

Figure 5.4: Example of cycles on graph $\mathcal{G} = (Y, \Gamma)$.

$\mathcal{G}$ is such that $C(\rho) \neq \emptyset$, $\forall \, \rho \in \Gamma$.

The variables used are as follows: λ_c^v represents the total flow of v-class railcars for each cycle c and for the total period $\mathcal{T}$; w_c^v is the minimum number of v-type railcars required for covering cycle c; τ_c represents the running time on cycle c; and μ^v is the minimum total number of v-class railcars required to cover all services. Then, given a railcar class and a y-path, equation (5.37) connects railcars that run on cycles with full and empty railcars that run on the y-path. Equation (5.38) defines a lower bound for the railcars of each class that are needed to cover a cycle. Equation (5.39) groups all railcars of the same class and estimates the total number of railcars of each class that are needed to cover all services, including a maintenance factor η^v, which is previously known.

$$\sum_{q \in P} f_\rho^{v,q} + f_\rho^{v,\emptyset} = \sum_{c \in C(\rho)} \lambda_c^v \qquad \forall \rho \in \Gamma, \; \forall v \in \mathcal{V} \quad (5.37)$$

$$w_c^v \geq \frac{1}{\mathcal{T}} (\tau_c \cdot \lambda_c^v) \qquad \forall c \in C, \; \forall v \in \mathcal{V} \quad (5.38)$$

$$\sum_{c \in C} w_c^v = \mu^v \cdot \eta^v \qquad \forall v \in \mathcal{V} \quad (5.39)$$

$$\lambda_c^v, \tau_c, w_c^v, \mu^v \in \mathbb{R}^+ \,. \quad (5.40)$$

Chapter 6

A Modal-Split/Traffic Assignment Model

This chapter develops a combined modal-split/traffic assignment model for evaluating the train-road modal share in future scenarios, or also for the case when a modal split model based on random utilities is available, and some of its coefficients may present a non-negligible range of variability. What follows in this chapter mainly corresponds to the article published in the European Journal of Operational Research:

Rosell, F., Codina, E.,& Montero L. (2022). A Combined and Robust Modal-split/Traffic Assignment Model for Rail and Road Freight Transport, *European Journal of Operational Research, 303*, (2), 688-698, https://doi.org/10.1016/j.ejor.2022.03.008.

We will refer to the model introduced in this chapter as the **MINLP**-model (for Mixed-Integer Non-Linear Problem). The **MINLP**-model will be built from the structure introduced in Chapter 4.

6.1. Model description

The aim is to obtain consistent train-road modal splits when a modal split model based on random utilities is available to assess changes in the railway system. Still, its predictive capabilities can be affected because of the limited quality of the input data, mainly for reasons like:

- They may come from heterogeneous sources or aggregate values, e.g., from official statistics or databases supplemented with aggregated data. As a result, all or part of the model coefficients may present considerable uncertainty, this being contributed by the extremely low use of rail in some countries.

- It has been obtained through stated preference surveys where possible capacity limitations are not reflected, and it cannot be established a priori where they will take place. However, it may subsequently occur due to technical deficiencies in implementing future actions on the railway network.

The predictive role to which the model is addressed can be contrasted taking into account its robust version where it is allowed that a limited number of elements of the utilities can deviate from their nominal values simultaneously. The model does not strictly impose competition between operators, but it does exclude those who do not obtain benefits. Railway carriers' offer of services depends on its profitability, although limited by the characteristics of the infrastructure and the availability of rolling stock.

6.1.1. General form of the model

Let us remember some of the notation presented in Chapter 4. Variables h_r^ω, $\tilde{h}^\omega$ are the total flow transported by train using t-path $r \in R(\omega)$ and truck, respectively, for an OD-pair and product represented by the triplet ω. Also, we have defined the flow vectors $\mathbf{h}^\omega = (..., h_r^\omega, ...)^\top \in \mathbb{R}^{|R(\omega)|}$, $\omega \in W$, $\mathbf{h} = (..., \mathbf{h}^\omega, ...)^\top \in \mathbb{R}^{|W|}$ and $\tilde{\mathbf{h}} = (..., \tilde{h}^\omega, ...)^\top \in \mathbb{R}^{|W|}$. For an initial description of the model, the remaining variables related to flow of railcars, locomotives, flow of trains between yards and other auxiliary variables will be assumed to be comprised in a generic vector of variables $\mathbf{y}$ lying in a specific domain $\mathcal{Y}$. Variables $\mathbf{h}$ and $\mathbf{y}$ will be related each other by some binding constraints $g_\ell(\mathbf{h}, \mathbf{y}) \leq 0$, $\ell = 1, ..., m$.

Let $\tilde{u}^\omega, u_r^\omega$ be the generalized cost or disutility, for OD-pair and product ω, when transported by truck and by train using t-path $r \in R(\omega)$, respectively. The model is formulated as the following non-linear optimisation problem:

$$\min_{\mathbf{h},\tilde{\mathbf{h}},\mathbf{y}} \quad F(\mathbf{h}, \tilde{\mathbf{h}}) = \sum_{\omega \in W} \left[\sum_{r \in R(\omega)} u_r^\omega h_r^\omega + \int_0^{\tilde{h}^\omega} G_\omega^{-1}(s) ds \right] \tag{6.1}$$

$$\text{s.t.} \quad \sum_{r \in R(\omega)} h_r^\omega + \tilde{h}^\omega = \chi^\omega \qquad \omega \in W \tag{6.1a}$$

$$g_\ell(\mathbf{h}, \mathbf{y}) \leq 0 \qquad \ell = 1, .., m \tag{6.1b}$$

$$\mathbf{y} \in \mathcal{Y} \tag{6.1c}$$

$$h_r^\omega \geq 0 \qquad r \in R(\omega),\ \omega \in W \tag{6.1d}$$

$$\tilde{h}^\omega \geq 0 \qquad \omega \in W. \tag{6.1e}$$

As expressed in (6.1a) the total freight demand (χ^ω) must be equal to the sum of the demand on the railway system (first term) plus the demand on the truck system (second term). It is aimed to reflect shippers priorities, as well as the modal split behaviour. Then, following the models for user equilibrium with variable demand developed in Sheffi (1985, chapter 6), the problem is expressed using the excess-demand formulation, being $G_\omega^{-1}(\cdot)$ the inverse demand function, as expressed in (6.2) :

$$G_\omega^{-1}(\tilde{h}^\omega) = \tilde{u}^\omega + \log\left(\frac{\tilde{h}^\omega}{\chi^\omega - \tilde{h}^\omega}\right), \quad \omega \in W, \tag{6.2}$$

corresponding to the direct demand function $G_\omega(\cdot)$ for the road mode of transportation, that provides the amount of flow $\tilde{h}^\omega$ transported by road:

$$\tilde{h}^\omega = G_\omega(u^\omega) = \chi^\omega \left\{1 + \exp(\tilde{u}^\omega - u^\omega)\right\}^{-1} . \tag{6.3}$$

In (6.1), the first component corresponds to total generalized cost for OD-pair and product, and the second component represents the total generalized cost for the excess demand transported by truck, for OD-pair and product. The excess-demand component of (6.1) can be developed as shown next:

$$\begin{aligned}\int_0^{\tilde{h}^\omega} G_\omega^{-1}(s)ds = \int_0^{\tilde{h}^\omega}\left(\tilde{u}^\omega + \log\left(\frac{s}{\chi^\omega - s}\right)\right)ds = \tilde{u}^\omega\tilde{h}^\omega + \int_0^{\tilde{h}^\omega}\log\left(\frac{s}{\chi^\omega - s}\right)ds = \\ = \tilde{u}^\omega\tilde{h}^\omega + \tilde{h}^\omega\log(\tilde{h}^\omega) + (\chi^\omega - \tilde{h}^\omega)\log(\chi^\omega - \tilde{h}^\omega) - \chi^\omega\log(\chi^\omega)\,.\end{aligned} \tag{6.4}$$

The expression of the excess-demand as appears in (6.4) corresponds to a convex function. Thus, the objective function (6.1) results in a convex function and can be expressed as follows:

$$F(\mathbf{h}, \tilde{\mathbf{h}}) \equiv \sum_{\omega\in W}\sum_{r\in R(\omega)} u_r^\omega h_r^\omega + \sum_{\omega\in W}\tilde{u}^\omega\tilde{h}^\omega + \sum_{\omega\in W}\int_0^{\tilde{h}^\omega}\log\left(\frac{s}{\chi^\omega - s}\right)ds\,. \tag{6.5}$$

6.1.2. Modal choice properties of the model

The modal choice properties of the model derive from the following first order conditions of problem (6.1) with respect the variables h_r^ω, $\tilde{h}^\omega$:

$$\frac{\partial F}{\partial h_r^\omega} = \quad u_r^\omega = \vartheta^\omega - \gamma_r^\omega + \xi_r^\omega, \qquad \xi_r^\omega \geq 0, \;\; \xi_r^\omega h_r^\omega = 0 \tag{6.6a}$$

$$\frac{\partial F}{\partial \tilde{h}^\omega} = \quad \tilde{u}^\omega + \log\left(\frac{\tilde{h}^\omega}{\chi^\omega - \tilde{h}^\omega}\right) = \vartheta^\omega + \eta^\omega, \qquad \eta^\omega \geq 0,\ \eta^\omega \tilde{h}^\omega = 0, \tag{6.6b}$$

where ϑ^ω, ξ_r^ω and η^ω are the Lagrange multipliers of constraints (6.1a), (6.1d) and (6.1e), respectively. Also, γ_r^ω result from the Lagrange multipliers ζ_ℓ of constraints (6.1b) as

$$\gamma_r^\omega = \sum_{g_\ell(\mathbf{h},\mathbf{y})=0} \zeta_\ell \frac{\partial g_\ell}{\partial h_r^\omega}\,. \tag{6.7}$$

From (6.6a) and being $\xi_r^\omega \geq 0$:

$$\vartheta^\omega \leq u_r^\omega + \gamma_r^\omega,\ \ \forall r \in R(\omega),\ \forall \omega \in W. \tag{6.8}$$

It must be remarked that, because Lagrange multipliers η^ω must be finite, any solution of problem (6.1) must verify that $\tilde{h}^\omega > 0$ and $\tilde{h}^\omega < \chi^\omega$. Thus, from (6.1a), $\sum_{r\in R(\omega)} h_r^\omega > 0,\ \forall \omega \in W$, and then:

$$\begin{array}{lr} \forall \omega \in W,\ \sum_{r\in R(\omega)} h_r^\omega > 0 \text{ and } h_r^\omega \geq 0\ \ \forall r \in R(\omega) & \Longrightarrow \\ \forall \omega \in W,\ \exists r \in R(\omega),\ h_r^\omega > 0 \text{ and } \xi_r^\omega = 0 & \Longrightarrow \\ \forall \omega \in W,\ \exists r \in R(\omega),\ \vartheta^\omega = u_r^\omega + \gamma_r^\omega\,. & \end{array} \tag{6.9}$$

Then, from (6.8) and (6.9):

$$\vartheta^\omega = \min_{r\in R(\omega)} \{u_r^\omega + \gamma_r^\omega\},\ \ \forall \omega \in W. \tag{6.10}$$

From (6.6b), taking into account that $\tilde{h}^\omega > 0 \implies \eta^\omega = 0,\ \forall \omega \in W$, the modal split following a logit model can be derived after some calculation:

$$\frac{\tilde{h}^\omega}{\chi^\omega} = \{1 + \exp(\tilde{u}^\omega - \vartheta^\omega)\}^{-1}\,. \tag{6.11}$$

It must be noted that in logit-like expression (6.11) for the fraction of products shipped by road, the utilities of the rail alternative appear now to be ϑ^ω, i.e., the initially stated utilities u_r^ω are modified by multipliers γ_r^ω corresponding to constraints (6.1b), having an effect of explicit or implicit capacities. In the next section it will be shown the functional form of the constraints (6.1b) which will turn out to be linear. Including side capacity constraints in equilibrium models can also be found in Larsson and Patriksson (1994), where the corresponding Lagrange multipliers are interpreted as additional delays or costs. We also interpret here the inclusion of the side constraints (6.1b) when evaluating the modal share as a correction for the previously evaluated utilities u_r^ω that may come from a random utility model in which these constraints associated to flows could not be taken properly into account when estimating the random utility model.

6.1.3. Rail freight transport constraints

Variables $\mathbf{y}$ in problem (6.1) represent the variables related to the flow of railcars, locomotives, the flow of trains between yards, while the set $\mathcal{Y}$ corresponds to the constraints that rail freight transport characteristics determine. Three groups of constraints can be found: first, those which reflect how goods are transported using the rail network; second, an approximation to the total number of railcars and locomotives required by carriers to perform the transport; and third, the capacity limits on tracks and yards. The first two groups result in a common approach with the **DCECC**-model. In contrast, the capacity limits approach differs from the **DCECC**-model because of the tactical characteristic of the **MINLP**-model.

Relationships for carriers' rail transport flows. The conditions on rail transport will be given by:

- the balance equations on yards (equations (4.4) and (4.5));
- train length and maximum weight limits (equations (4.6) and (4.7));
- trains composition and decomposition on yards (equations (4.9) and (4.10)).

Notice that equations (4.3), which link products and railcars, are not included in this group and will be added later.

Rolling stock. Also as in the **DCECC**-model, Section 4.3.3 introduces the equations for estimating a lower and an upper bound of the total of railcars and locomotives required for each carrier to carry out the services. Equation (4.11) establishes the lower and upper bound for railcars, while equation (4.12) sets the upper bound for locomotives.

Capacity limits on the railway network. Here, the approach differs from the one presented in the **DCECC**-model. Because the **DCECC**-model is a design model and tracks and signalisation infrastructure are related to capacity limits, its capacity limits depend directly on the proposed design and then are part of the transport conditions.

However, in the **MINLP**-model, infrastructure design is exogenous, and infrastructure managers are in charge of allocating capacity on the infrastructure so it can be made available to carriers. Consequently, rail carriers buy the slots that interest them among those available. These slots limit the maximum number of trains a carrier can operate in each directed line. Equations (4.29), which impose a limit on the maximum number of trains on that directed line, and equations (4.30), which restrict the maximum number of trains that run on a track, represent the capacity limits on tracks.

Also, in the **DCECC**-model, capacity limits on yards depend on the number of classification tracks available to the yard, which may be part of the decision variables of the model. In contrast, in the **MINLP**-model, infrastructure on yards is fixed, and infrastructure managers determine the maximum number of trains a yard can handle during the period under analysis. Equations (4.31) and (4.32) limit the total incoming and outgoing flow of trains that can be dismounted and mounted, respectively, on a yard, accordingly to the capacity of the yard in the number of trains.

6.1.4. Relationship between demand and railway transport characteristics

The biding constraints relate demand with the flow of trains, railcars, and the other variables that reflect rail transport characteristics, and they are represented by equations (6.1b) in problem (6.1). They correspond to two groups of constraints. On the one hand, equation (4.3) links tons of products with railcars. On the other hand, constraints (4.13) and (4.14) state the viability of carriers business, by assuring that a carrier has no losses or, on the contrary, they carry out no transportations of goods, without compromising the feasibility of the model.

6.1.5. Domain of the variables

To complete the model, let us summarize the domain of the variables, as they are introduced in Section 4.9:

$$h_r^\omega \in \mathbb{R}^+ \tag{4.38a}$$

$$\tilde{h}^\omega \in \mathbb{R}^+ \tag{4.38b}$$

$$f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_{\rho,j}^k, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+ \tag{4.39}$$

$$\hat{y}_o \in \{0,1\}. \tag{4.42}$$

6.1.6. Deterministic version of the optimisation problem

In previous subsections, we have detailed the elements of the model related to flows of trains and railcars with flows of products on the network defining new variables and constraints. Particularly, the vector of variables $\mathbf{y}$ in problem (6.1) comprises: $\mathbf{y} \equiv f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_\rho^k, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+, \hat{y}_o \in \{0,1\}$, while the vector of variables $\mathbf{x}$ comprises $\mathbf{x} \equiv h_r^\omega, \tilde{h}^\omega$. Then, problem (6.1) can be stated as the following mixed integer non-linear problem **MINLP-D**:

$$\textbf{(MINLP-D)} \tag{6.12}$$

$$\begin{aligned} \min_{h,\tilde{h},f,m,\lambda,\theta,y_o} \quad & \sum_{\omega\in W}\sum_{r\in R(\omega)} u_r^\omega h_r^\omega + \sum_{\omega\in W} \tilde{u}^\omega \tilde{h}^\omega + \sum_{\omega\in W} \int_0^{\tilde{h}^\omega} \left(\log \frac{x}{\chi^\omega - x}\right) dx \\ \text{s.t.} \quad & (6.1a), (4.3)-(4.7), (4.9)-(4.14), (4.29)-(4.32), \\ & (4.38a), (4.38b), (4.39), (4.42) \end{aligned}$$

where the "-D" stands for "deterministic". The explicit and detailed formulation of the **MINLP-D** problem appears fully compiled in Appendix B.4. Following, the robust counterpart of problem **MINLP-D** is developed taking into account the uncertainty in the parameters u_r^ω, $\tilde{u}^\omega$

6.2. Robustness on Utility Function

In this section the uncertainty in the generalized costs, or disutilities, u_r^ω, $\tilde{u}^\omega$ for transporting products $\mathfrak{p}(\omega)$ from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$ is dealt with, developing a model under the scope of robust optimisation. It can be assumed that the systematic component of the disutilities for the rail and road transport modes are given by affine functional forms of m and $\tilde{m}$ explanatory variables as in (6.13) and (6.14).

$$u_r^\omega = \beta_0^\omega + \sum_{j=1}^{m} \beta_j^\omega u_{r,j}^\omega \qquad \forall r \in R(\omega),\ \forall \omega \in W. \tag{6.13}$$

$$\tilde{u}^\omega = \tilde{\beta}_0^\omega + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^\omega \tilde{u}_j^\omega \qquad \forall \omega \in W. \tag{6.14}$$

Typically these explanatory variables are travel time, price, distance, GHG (Greenhouse Gas) emissions, among others. A critical step is to obtain a good estimation of the parameters β and $\tilde{\beta}$ by using generalized linear model regression, for instance. It is not too hard to obtain reasonable estimates for the mean value and confidence intervals of the parameters of the utility function, and then, the uncertainty of parameters β and $\tilde{\beta}$ may be expressed as follows:

$$\beta_j^\omega \in [\beta_j^{-,\omega} - \beta_j^{+,\omega}, \beta_j^{-,\omega} + \beta_j^{+,\omega}], \qquad j = 0, \dots, m, \quad \forall \omega \in W. \tag{6.15}$$

$$\tilde{\beta}_j^\omega \in [\tilde{\beta}_j^{-,\omega} - \tilde{\beta}_j^{+,\omega}, \tilde{\beta}_j^{-,\omega} + \tilde{\beta}_j^{+,\omega}], \qquad j = 0, \dots, \tilde{m}, \quad \forall \omega \in W. \tag{6.16}$$

Our approach is based on the robustness concept developed in the work by Koster et al. (2013). Following this work, the number of parameters that may take its worse value is restricted to a (small) number H, and in this way, it is possible to set the level of conservatism and robustness of the solutions. The greater the value for H, the greater the uncertainty on parameters and the robustness of the model solutions. Here, we will analyse the case for the uncertainty in u_r^ω. Given the similarity between u_r^ω and $\tilde{u}^\omega$ affine functional forms, the results for the former can be applied to the latter. For simplicity, we remove the r and ω indices.

Let us define the uncertain set as follows:

$$\mathcal{B}(H) = \{\beta \in \mathbb{R}^{m+1} : \beta_j = \beta_j^- + \beta_j^+ z_j, \ j = 0, \dots, m, \ z \in \mathcal{Z}(H)\}, \tag{6.17}$$

$$\mathcal{Z}(H) = \{z \in \mathbb{R}^{m+1} : |z_j| \leq 1, \ j = 0, \dots, m, \ \sum_{j=0}^{m} |z_j| \leq H\}. \tag{6.18}$$

The following equivalences can be stated:

$$\begin{aligned} uh &\Leftrightarrow (\beta_0 + \sum_{j=1}^{m} \beta_j u_j)h, \ \beta \in \mathcal{B}(H) \\ &\Leftrightarrow \max_{z \in Z(H)} \{(\beta_0^- + \beta_0^+ z_0 + \sum_{j=1}^{m} (\beta_j^- + \beta_j^+ z_j) u_j)h\} \\ &\Leftrightarrow (\beta_0^- + \sum_{j=1}^{m} \beta_j^- u_j)h + \max_{z \in Z(H)} \{(\beta_0^+ z_0 + \sum_{j=1}^{m} \beta_j^+ u_j z_j)h\}. \end{aligned} \tag{6.19}$$

Let $\mathcal{A}(H, h) \equiv \max_{z \in Z(H)} \{(\beta_0^+ z_0 + \sum_{j=1}^{m} \beta_j^+ u_j z_j)h\}$. Given $h^\star$, Koster et al. (2013) shows that $\mathcal{A}(H, h^\star)$ is equivalent to the optimisation problem:

$$\begin{aligned} \max_{z} \quad & (\beta_0^+ z_0 + \sum_{j=1}^{m} \beta_j^+ u_j z_j) \cdot h^\star \\ \text{s.t.} \quad & \sum_{j=0}^{m} z_j \leq H, \quad 0 \leq z_j \leq 1, \ j = 0, \dots, m. \end{aligned} \tag{6.20}$$

The dual of problem (6.20) can be stated as follows:

$$\begin{aligned}\min_{\pi,p} \quad & H\pi + \sum_{j=0}^{m} p_j \\ \text{s.t} \quad & \pi + p_j \geq \beta_j^{+} u_j h^{\star} \\ & \pi + p_0 \geq \beta_0^{+} h^{\star} \\ & \pi \geq 0 \\ & p_j \geq 0, \quad j = 0, \ldots, m.\end{aligned} \tag{6.21}$$

Then, we can reformulate the problem **MINLP-D** by replacing each element $u_r^\omega h_r^\omega$ in the objective function by a new expression as it appears in (6.22), and adding the new set of constraints (6.23), where π, p are non-negative continuous auxiliary variables required for the reformulation:

$$\min_{h,\pi,p\geq 0} \quad \sum_{\omega} \sum_{r\in R(\omega)} [\ (\beta_{r,0}^{-,\omega} + \sum_{j=1}^{m} \beta_{r,j}^{-,\omega} u_{r,j}^{\omega}) h_r^\omega + H\pi_r^\omega + \sum_{j=0}^{m} p_{r,j}^{\omega}\ ] \tag{6.22}$$

$$\pi_r^\omega + p_{r,j}^\omega \geq \beta_{r,j}^{+,\omega} u_{r,j}^\omega h_r^\omega, \qquad \forall j,\ \forall r \in R(\omega),\ \forall \omega \in W \tag{6.23}$$

$$\pi_r^\omega + p_{r,0}^\omega \geq \beta_{r,0}^{+,\omega} h_r^\omega, \qquad \forall r \in R(\omega),\ \forall \omega \in W \tag{6.24}$$

$$\pi_r^\omega \in \mathbb{R}^+ \tag{6.25}$$

$$p_{r,j}^\omega \in \mathbb{R}^+\,. \tag{6.26}$$

Analogously, the same methodology can be applied to each term $\tilde{u}^\omega \tilde{h}^\omega$:

$$\min_{\tilde{h},\tilde{\pi},\tilde{p}\geq 0} \quad \sum_{\omega} [\ (\tilde{\beta}_0^{-,\omega} + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^{-,\omega} \tilde{u}_j^{\omega}) \tilde{h}^\omega + H\tilde{\pi}^\omega + \sum_{j=0}^{\tilde{m}} \tilde{p}_j^{\omega}\ ] \tag{6.27}$$

$$\tilde{\pi}^\omega + \tilde{p}_j^\omega \geq \tilde{\beta}_j^{+,\omega} \tilde{u}_j^\omega \tilde{h}^\omega \qquad \forall j,\ \forall \omega \in W \tag{6.28}$$

$$\tilde{\pi}^\omega + \tilde{p}_0^\omega \geq \tilde{\beta}_0^{+,\omega} \tilde{h}^\omega, \qquad \forall \omega \in W \tag{6.29}$$

$$\tilde{\pi}^\omega \in \mathbb{R}^+ \tag{6.30}$$

$$\tilde{p}_j^\omega \in \mathbb{R}^+\,. \tag{6.31}$$

Let us define:

$$\mathbf{u}_r^\omega \triangleq \beta_{r,0}^{-,\omega} + \sum_{j=1}^{m} \beta_{r,j}^{-,\omega} u_{r,j}^\omega, \quad \tilde{\mathbf{u}}^\omega \triangleq \tilde{\beta}_0^{-,\omega} + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^{-,\omega} \tilde{u}_j^\omega,$$

to maintain a common nomenclature. The robust counterpart of problem **MINLP-D** results in the problem **MINLP-R**:

$$\textbf{(MINLP-R)} \tag{6.32}$$

$$\begin{aligned} \min \quad & \sum_{\omega\in W}\sum_{r\in R(\omega)} \mathbf{u}_r^\omega h_r^\omega + \sum_{\omega\in W} \tilde{\mathbf{u}}^\omega \tilde{h}^\omega + \sum_{\omega\in W} \int_0^{\tilde{h}^\omega} (\log \frac{x}{\chi^\omega - x}) dx \\ & + \sum_{\omega\in W}\sum_{r\in R(\omega)} \left(H\pi_r^\omega + \sum_{j=0}^{m} p_{r,j}^\omega \right) + \sum_{\omega\in W} \left(\tilde{H}\tilde{\pi}^\omega + \sum_{j=0}^{n} \tilde{p}_j^\omega \right) \\ \text{s.t.} \quad & (6.1a), (4.3) - (4.7), (4.9) - (4.14), (4.29) - (4.32), \\ & (4.38a), (4.38b), (4.39), (4.42), (6.23) - (6.26), (6.28) - (6.31) \end{aligned}$$

The explicit and detailed formulation of the **MINLP-R** problem appears fully compiled in Appendix B.5.

6.3. Solution Algorithm

MINLP will be used to refer indistinctly to both **MINLP-D** and **MINLP-R** if such is the context. We apply an outer-approximation algorithm for mixed integer non-linear optimisation problems detailed in Floudas (1995), which is extracted from the work by Duran and Grossmann (1986). A summarised, simpler version of the notation is used for making easier the description of the algorithm.

- $\mathbf{y}$: Vector of integer variables whose components are: $f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_\rho^k,\ \lambda^{v,o},\ \theta_{i',i}^{k,o} \in \mathbb{Z}^+, \hat{y}_o \in \{0,1\}$ for **MINLP**.
- $\mathbf{x}$: Continuous variables. $\mathbf{x}$ corresponds to $h_r^\omega, \tilde{h}^\omega$ for **MINLP-D**, while $\mathbf{x}$ corresponds to $h_r^\omega, \pi_r^\omega, p_{r,j}^\omega, \tilde{h}^\omega, \tilde{\pi}^\omega, \tilde{p}_j^\omega$ for **MINLP-R**.
- $\mathcal{Y}$: Set containing values for $\mathbf{y}$ that verify constraints (4.4)-(4.7), (4.9)-(4.12), (4.29)-(4.32), (4.39) plus (4.42) for **MINLP**. Actually, all constraints that involve only trains and railcars.
- X set. Constraints involving only continuous variables. (6.1a), (4.38a) and (4.38b) for **MINLP-D**, also includes (6.23)-(6.26), (6.28)-(6.31) for **MINLP-R**.
- $g_\ell(\mathbf{x}, \mathbf{y}) \leq 0,\ \ell = 1, 2, 3$. Relationship between rail freight demand flows and flows of railcars and trains, i.e., constraints (4.3) for $\ell = 1$, (4.13) for $\ell = 2$ and (4.14) for $\ell = 3$, for **MINLP**.
- The objective function for **MINLP** will be denoted by $F(\mathbf{x})$.

The algorithm decomposes the original problem **MINLP** into one non-linear primal subproblem **NLPP** (6.33) :

$$\textbf{(NLPP)} \qquad \min_{\mathbf{x}\in X} \; F(\mathbf{x}) \tag{6.33}$$

$$g_\ell(\mathbf{x}, \mathbf{y}^{(\mathbf{s}}) \le 0, \quad \ell = 1, 2, 3 \tag{6.33a}$$

$$\Rightarrow \; \mathbf{x}^* \longrightarrow \mathbf{x}^{(s}\,,$$

and one mixed-integer linear master problem **MLMP** (6.34):

$$\textbf{(MLMP)} \quad \min_{\mathbf{x},\mathbf{y},z} \; z \tag{6.34}$$

$$z \ge F(\mathbf{x}^{(s)}) + \nabla F(\mathbf{x}^{(s)})(\mathbf{x} - \mathbf{x}^{(s)}), \quad \forall s \tag{6.34a}$$

$$g_\ell(\mathbf{x}, \mathbf{y}) \le 0, \; \forall \ell, \quad \mathbf{x} \in X, \quad \mathbf{y} \in \mathcal{Y} \tag{6.34b}$$

$$z^{(s} \le z < UBD \tag{6.34c}$$

$$\Rightarrow \quad z^*, \mathbf{x}^*, \mathbf{y}^* \longrightarrow z^{(s+1}, \hat{\mathbf{x}}, \mathbf{y}^{(s+1}\,.$$

MLMP includes two type of cuts induced by $\mathbf{x}^{(s}$, $\forall s$: a linearization around $\mathbf{x}^{(s}$ of the convex functions $F(\mathbf{x})$:

$$z \ge F(\mathbf{x}^{(s)}) + \nabla F(\mathbf{x}^{(s)})(\mathbf{x} - \mathbf{x}^{(s)}), \; \forall s, \tag{6.35}$$

and also around the convex function $g_\ell(\mathbf{x}, \mathbf{y})$:

$$0 \ge g_\ell(\mathbf{x}^{(s}, \mathbf{y}) + \nabla g_\ell(\mathbf{x}^{(s}, \mathbf{y})(\mathbf{x} - \mathbf{x}^{(s)}), \; \forall s, \; \ell = 1, 2, 3\,. \tag{6.36}$$

In this case, $g_\ell(\mathbf{x}, \mathbf{y})$ are all linear functions, so cuts of the type (6.36) are equivalent to $g_\ell(\mathbf{x}, \mathbf{y}) \le 0$.

Subproblem **NLPP** generates an upper bound on the **MINLP** solution by fixing the value of the integer variables and solving the resulting **MINLP**, while solving **MLMP** allows to obtain a lower bound. **MLMP** results as an outer linearization of the **MINLP** non-linear objective function and constraints at $\mathbf{x}^{(s}$. **NLPP** provides a temptative value $\mathbf{x}^{(s}$ for the continuous variables $\mathbf{x}$ at iteration $(s$, while the solution of **MLMP** provides the new values for the integer variables, to solve a new iteration of **NLPP**. Observe that solution $\mathbf{x}^*, \mathbf{y}^*$ verifies equation (6.34b), that is, **NLPP** is feasible. The algorithm for solving **MINLP** is shown in Algorithm 1, at the end of this chapter.

The algorithm for solving **NLPP** is described in next Section 6.3.2, while **MLMP**, being a linear problem, can be solved by using any commercial solver. Also, the algorithm requires to obtain an initial feasible solution. We have opted by directly solving

MINLP without any objective function, under the condition that a minimum demand is transported by rail. This condition avoids the simplest solution, with all demand transported by truck, which is not a valid solution for the problem **MINLP**.

6.3.1. Method convergence

The convergence of the method relies on the characteristics of the functions and sets which define the problem (Duran and Grossmann (1986, Theorem 3)). In short, these conditions are the following:

- X is a non-empty compact and convex set;
- the function $F(\mathbf{x})$ is convex and once continuously differentiable, and functions $g_\ell(\mathbf{x}, \mathbf{y}), \ell = 1, 2, 3$ are convex in $\mathbf{x}$ and once continuously differentiable,
- the set $\mathcal{Y}$ is a finite discrete set, and finally
- a constraint qualification holds for each **NLPP** problem, given that $g_\ell(\mathbf{x}, \mathbf{y}), \ell = 1, 2, 3$ and constraints that defines X are all linear.

6.3.2. Solving the non-linear subproblem

NLPP is a problem with non-linear objective function and linear constraints. It can be observed that, were it not for constraints $g_2 \leq 0$ in equation (6.33a) (that is, equation (4.13) evaluated at $\mathbf{y}^{(\mathbf{s})}$), **NLPP** could be decomposed by $\omega \in W$. The objective function terms are also separable by ω. This condition allows decomposing **NLPP** into a set of non-linear subproblems, one for each $\omega \in W$, after applying a Lagrangian relaxation on constraints $g_2 \leq 0$ in (6.33a). Equation (6.37) shows the Lagrangian relaxation expression, where $\mu = (\mu_o; o \in O)$ is the vector of Lagrange multipliers for the relaxed constraints $g_2 \leq 0$ in (6.33a) (that is, (4.13) evaluated at $\mathbf{y}^{(s)}$) and, for simplicity, $D_o^{(s}$ represents its right hand side.

$$\mathcal{L}(\mathbf{x}, \mathbf{y}^{(s}, \mu) := F(\mathbf{x}) + \sum_{o \in \mathcal{O}} \mu_o \left(D_o^{(s} - \sum_{\omega \in W} \sum_{r \in R(\omega, o)} U_r^\omega h_r^\omega \right) . \tag{6.37}$$

The new term in (6.37), added to $F(\mathbf{x})$ (the **NLPP** objective function), can be expressed as a sum of terms depending on ω plus a term depending only on μ variables.

The Dantzig Cutting-Plane algorithm decomposes the original problem **NLPP** into one non-linear primal subproblem, the Lagrangian relaxation of **NLPP** as appears below in $\mathcal{L}$-**NLPP** (6.38):

$$(\mathcal{L}\text{-}\mathbf{NLPP}) \quad \phi(\mu^{(k}, \mathbf{y}^{(s)}) \triangleq \min_{\mathbf{x}\in X} \mathcal{L}(\mathbf{x}, \mathbf{y}^{(\mathbf{s}}, \mu^{(k)}) := \tag{6.38}$$
$$\min_{\mathbf{x}\in X} F(\mathbf{x}) - \sum_{\omega\in W}\sum_{o\in\mathcal{O}}\sum_{r\in R(\omega,o)} (\mu_o^{(k} U_r^{\omega}) h_r^{\omega} + \sum_{o\in\mathcal{O}} \mu_o^{(k} D_o^{(s}$$
$$\text{s.t. } (6.33\text{a}),$$

and one linear master problem, **DZLP** (6.39), given by:

$$(\mathbf{DZLP}) \quad \max_{z,\mu} z \tag{6.39}$$
$$z \le F(\mathbf{x}^{(k)}) + \sum_{o\in\mathcal{O}} \mu_o \left(D_o^{(s} - \sum_{\omega}\sum_{r\in R(\omega,o)} U_r^{\omega} h_r^{\omega(k} \right) \Bigg| \; \alpha_k \quad, \forall k$$
$$z, \mu \in \mathbb{R}, \; \mu \ge 0,$$

where α_k are the non-negative dual variables associated to the constraints. Let $(k$ be the iteration number for the Dantzig Cutting-Plane algorithm. Note that $(s$ is the iteration number drawn from Algorithm 1 (for soliving the **MINLP**-problem), which reflects the fixed value for $\mathbf{y}$ variables. The Dual-Lagrangian function ϕ, corresponding to the lagrangian (6.37), for a point $\mathbf{y}^{(s}$ evaluated at $\mu^{(k}$, $\phi(\mu^{(k}, \mathbf{y}^{(s)})$, is evaluated at each non-linear primal subproblem $\mathcal{L}$-**NLPP**.

The problem $\mathcal{L}$-**NLPP** can be decomposed into ω-subproblems. Let $\mathbf{x}_\omega$ be the subset of variables $\mathbf{x}$ that corresponds to ω, $\forall \omega \in W$. Subproblem $\mathcal{L}$-**NLPP**-ω (6.40) corresponds to each ω-decomposition of $\mathcal{L}$-**NLPP**.

$$(\mathcal{L}\text{-}\mathbf{NLPP}\text{-}\omega) \quad \min_{\mathbf{x}_\omega} \mathcal{L}_\omega(\mathbf{x}_\omega, \mathbf{y}^{(s}, \mu^{(k)}) := \tag{6.40}$$
$$\min_{\mathbf{x}_\omega} F(\mathbf{x}_\omega) - \sum_{o\in\mathcal{O}}\sum_{r\in R(\omega,o)} (\mu_o^{(k} U_r^{\omega}) h_r^{\omega}$$
$$\text{s.t. subset of (6.33a) that corresponds to } \omega.$$

After solving $\mathcal{L}$-**NLPP**, a feasible solution can be obtained from the dual variables α_k of the last master problem **DZLP** solved: $\mathbf{x}^* = \sum_k \alpha_k \mathbf{x}^{(k}$. Moreover, the Dantzig Cutting-Plane's algorithm requires an initial feasible solution: in this case, a good option is to take advantage that the solution of the problem **MLMP** is feasible for the problem **NLPP**. The algorithm for solving **NLPP** is summarized in next Algorithm 2.

Algorithm 1: Solving **MINLP**

initialization
  Find a feasible point $\hat{\mathbf{x}}, \mathbf{y}^{(0}$
  Let $s = 0$
  Fix ϵ
while *not STOP* **do**
  Solve **NLPP**($\mathbf{y}^{(s)}$), with $\hat{\mathbf{x}}$ as initial solution: $\rightarrow \mathbf{x}^{(s}$
  UBD $= F(\mathbf{x}^{(s)})$
  Solve **MLMP** $\rightarrow z^{*}, \mathbf{x}^{*}, \mathbf{y}^{*}$
  if MLMP *has no feasible solution* **then** STOP
  else
    LBD $= z^{*}$
    if *(UBD-LBD)/LBD* $< \epsilon$ **then** STOP
    else
      $s + 1 \rightarrow s$
      $\mathbf{y}^{*} \rightarrow \mathbf{y}^{(s}$
      $\mathbf{x}^{*} \rightarrow \hat{\mathbf{x}}$
return $\mathbf{x}^{(s}, \mathbf{y}^{(s}, F(\mathbf{x}^{(s)})$

Algorithm 2: Solving **NLPP**($\mathbf{y}^{(s)}$)

Input: $\mathbf{y}^{(s}, \hat{\mathbf{x}}$
Output: $\mathbf{x}^{*}, F(\mathbf{x}^{*})$
initialization
  $\hat{\mathbf{x}} \rightarrow$ **NLPP** initial feasible point $\mathbf{x}^{(0}$
  Let $k = 1$
  Fix $\hat{\epsilon}$
while *not STOP* **do**
  Solve **DZLP** $(\mathbf{x}^{(k-1)})$: $z^{\star}, \mu^{\star} \rightarrow z^{(k}, \mu^{(k}$ dual variables of **DZLP** $\rightarrow \alpha_i$, $i = 1, \ldots, k$
  Solve $\mathcal{L}$-**NLPP** by decomposition in $\mathcal{L}$-**NLPP**-ω subproblems
  $\mathbf{x}^{\star}_{\omega}, \ \forall \omega \rightarrow x^{\star} \rightarrow x^{(k}$
  $\mathcal{L}_{\omega}(\mathbf{x}^{\star}_{\omega}, \mathbf{y}^{(s}, \mu^{(k)}), \ \forall \omega \rightarrow \mathcal{L}(\mathbf{x}^{\star}, \mathbf{y}^{(s}, \mu^{(k)})$
  if $(z^{\star} - \mathcal{L}(\mathbf{x}^{\star}))/\mathcal{L}(\mathbf{x}^{\star}) < \hat{\epsilon}$ **then** STOP
  else $k + 1 \rightarrow k$
NLPP-feasible solution
  $\mathbf{x}^{*} = \sum_{i=1}^{k} \alpha_i \mathbf{x}^{(i}$
  Calculate $F(\mathbf{x}^{*})$
return $\mathbf{x}^{\star}, F(\mathbf{x}^{\star})$

Chapter 7

Scenarios

The mathematical models introduced in previous chapters can be analysed individually, as they have been presented, although also both models can be seen as complementaries. This chapter will introduce the common scenario used to test the validity and usability of the models. The scenarios consist of a rail network definition, the carriers' corridors where trains allocation will occur, a set of products to be transported, with their corresponding origin and destination terminals and the data required to feed the models.

7.1. Rail network infrastructure

The EU Member States have traditionally developed transport infrastructures. With decision 1692/96/EC of the European Parliament and the Council, the EU adopted guidelines for developing the Trans-European Transport Network (**TEN-T**). The **TEN-T** programme, led by the European Climate, Infrastructure and Environment Executive Agency (CINEA), consists of hundreds of projects whose ultimate purpose is to ensure the cohesion, interconnection and interoperability of the trans-European transport network and access to it. The **TEN-T** planning framework consists of a comprehensive network as the primary layer and a core network, overlaying the latter and representing the strategically most important part of the trans-European transport network. The comprehensive network directly reflects the relevant and existing planned infrastructure in the Member States, while the core network represents the backbone of a European integrated transport system. A "corridor approach" was adopted for the core network, and nine corridors were selected. The objective was to improve coordination between different stakeholders regarding traffic management, access to infrastructure and invest-

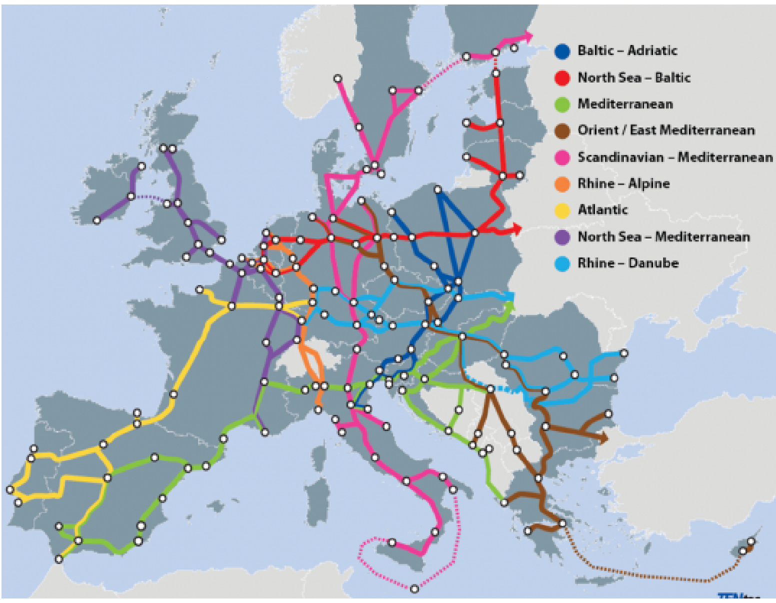


Figure 7.1: TEN-T Core Network Corridors. Source: European Comission DG MOVE TENtec Information System

ments in rail infrastructure and improve the continuity of traffic throughout the Member States, focusing on giving sufficient priority to rail freight traffic. The case studies are based on the **TEN-T** core network corridors. Figure 7.1 shows a schematic of the nine core network corridors.

The first step to build the rail network for the scenarios is to define the region in which the models are to be tested. Then, a macro-level rail network will be depicted, with the main terminals and their basic rail connections. The region selected covers the rail transport connections between the South-East of Spain, and Central Europe, including the Eastern area of Western Europe. It goes from Valencia and Zaragoza (Spain), to Malaszewicze (Poland), and from Marseilles (France) and Milan (Italy) to Rotterdam (the Netherlands) and Hamburg (Germany). The rail network has 17,406 km and 34 yards. Figure 7.2 depicts the main yards and their rail connections.

From the macro-level, the rail network should be decomposed into the elements of the graph: edges and nodes. In turn, each edge should be decomposed into two arcs, one for each direction. Arcs are the lower level of the network, and it is supposed each arc has homogeneous physical characteristics, so that it could be assumed trains run on each of them at constant speed. Arcs join each other via nodes, and nodes could be terminals, yards, or diverting/crossing points. Also, extra-arcs should be introduced to represent direct connections to by-pass yards. Thus, the schematic representation of the network

Figure 7.2: Schematic map of the rail network used as test. Source: author.

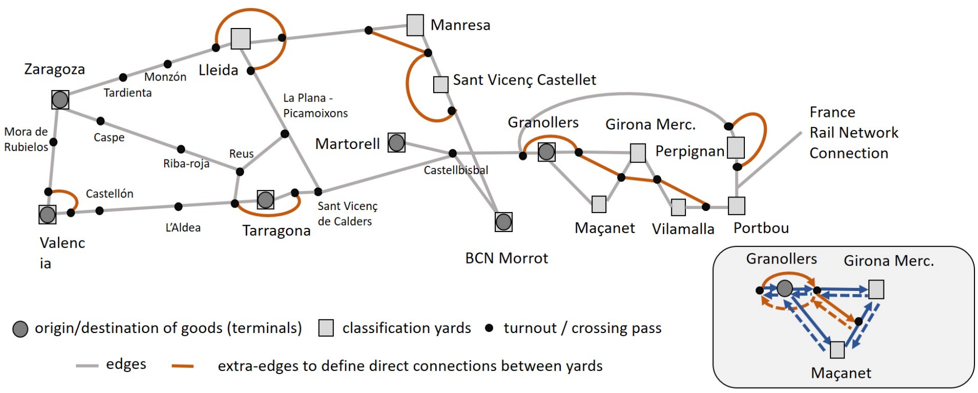


Figure 7.3: Network decomposition example. Source: author.

represented in Figure 7.2 results in an enlarged network with all the arcs and nodes required. Figure 7.3 shows an example how to decompose the Spanish section of the rail network in edges and nodes, where there are classification yards (represented by squares), terminals as origin or destination of goods (dark circles) and turnout/crossing pass (black points). Edges connect all these nodes, and the extra-edges added to define direct connections between non-consecutive yards are drawn in red. Decomposing into arcs and nodes results after replacing each edge by two directed arcs, one for each direction. The grey box on the right shows an example with arcs and nodes, corresponding to connections between Granollers, Girona Merc and Maçanet yards.

7.2. Data sources

Data comes from free databases, public reports, academic and official studies, and web pages. This section will detail the sources and how the data is applied to the models.

7.2.1. Infrastructure data

Jointly with the structure of arcs and nodes, a database is built to gather all the data required to run the models: double or single track, track length, track gauge, electrical voltage, loading gauges, maximum train length, block distance, among others. Infrastructure characteristics are defined from public data offered by infrastructure managers: ADIF (2020), SNCF Réseau (2020) or Deutsche Bahn AG (2020).

Also, the **DCECC**-model requires data for calculating the infrastructure costs for new or upgrading developments or maintenance on tracks and yards. Rail infrastructure projects vary from a wide range of imports, depending on the types of works included (construction of new tracks, upgrading, tunnels, bridges, signalling and electrification), the types of infrastructure elements (permanent way, equipment), or the topography, among others (Trabo et al. (2013), Attina et al. (2018)). Baumgartner (2001) is one of the most renowned studies regarding the costs of rail infrastructure. This study provides the average median value and two extreme values for costs for a wide range of elements covering nearly all the aspects of the rail network. Also, Schroten et al. (2019) is a study developed within the EU-project "Sustainable Transport Infrastructure Charging and Internalisation of Transport Externalities" which gives an overview of transport infrastructure costs (total, average and marginal) in the EU28 Member States and some other Western countries, for 2016. The data for infrastructure applied in tests intends to be a realistic approximation from orders of magnitude, and in no case can be taken literally. The costs that have been included in the test are detailed below. All costs have to be annualised, so the amortisation years-term has to be considered:

- Costs per track. All of them are defined per kilometre of track and mainly depend on the volume of traffic, type of traffic (only freight, or freight and passengers) and topography complexity. They result in parameters $\dot{S}_a, \ddot{S}_a$.
  - Annual amortisation of investment for new construction of a single track.
  - Annual amortisation of extra investment for new construction of a double track.
  - Annual maintenance.
  - Electrification for non-electrical tracks.
  - Conversion to a mixed-gauge track.
- Signalisation costs. Defined per block signal. They result in parameter $\hat{S}_a$.
  - Annual amortisation of investment.
  - Annual maintenance.
- Yards. Mainly depends on size, topography and type of terminal, it can be summarized in level of complexity. They result in parameter $S_i, \bar{S}_i$.
  - Annual amortisation for new construction.
  - Annual maintenance.
  - Annual amortisation for new track.

7.2.2. Operation costs

Different works about costs on rail freight transportation (Guinot (2008), Institut Cerdà (2019)), jointly with public information from infrastructure managers about prices for their services (ADIF (2020), DB Cargo AG (2019), SNCF Réseau (2020), Deutsche Bahn AG (2020), RailTech (2021)) are the basis to estimate the cost for rail freight transport. In this section we will detail the costs we have applied in the tests, and their relationship with the parameters on the constraints (4.13) and the **DCECC**-model objective function.

- Costs per km-train. Defined per track, based on its length, and from tracks, assigned to each y-path as a sum of costs. They result in parameter $\hat{C}'^{k}_{\rho}$:

- Annual driver salary divided by the average total amount of km per driver.
- Annual locomotive maintenance divided by the average total amount of km per locomotive.
- Annual amortisation, financial and assurance costs: all of them divided by the average total amount of km per locomotive.
- Energy supplied by the Infrastructure Manager. It is usually charged per km.
- Fee for use of the infrastructure. It is usually charged per km, sometimes depending on other factors, as the type of track or type of train.

▪ Costs per train. Defined per track and from tracks, assigned to each y-path as a sum of costs. They are added to parameter $\hat{C}_{\rho}^{k}$ when corresponds:

- Fee for running on particular tracks, as can be the Perthus tunnel, on Spain-France border.
- Fee for changes on locomotives or railcars, when traversing zones of different track characteristics or electrification, for instance.

▪ Costs per railcar. Defined as an annual amount per railcar. They correspond to the parameter $D^{v,o}$:

- Annual maintenance.
- Annual amortisation and financial costs.

▪ Costs for composition and decomposition of trains. Defined per train and yard.

- Train composition, includes shunting movements and corresponds to parameter $C_i^{k,o}$.
- Train decomposition, includes shunting movements and train control before departure, and corresponds to parameter $\bar{C}_i^{k,o}$

7.2.3. Demand data

Both models work on the basis that there are products to be transported from their origins to their destinations by road or train. Thus, the first point is to identify the amount of each product to be transported and the origin and destination yards.

From Datacomex (Ministerio de Industria, Comercio y Turismo (2020)), the statistics web page for Spanish Foreign trade, the criteria was to select origin and destination pairs and the products transported between them, which had rail as one of their modes of

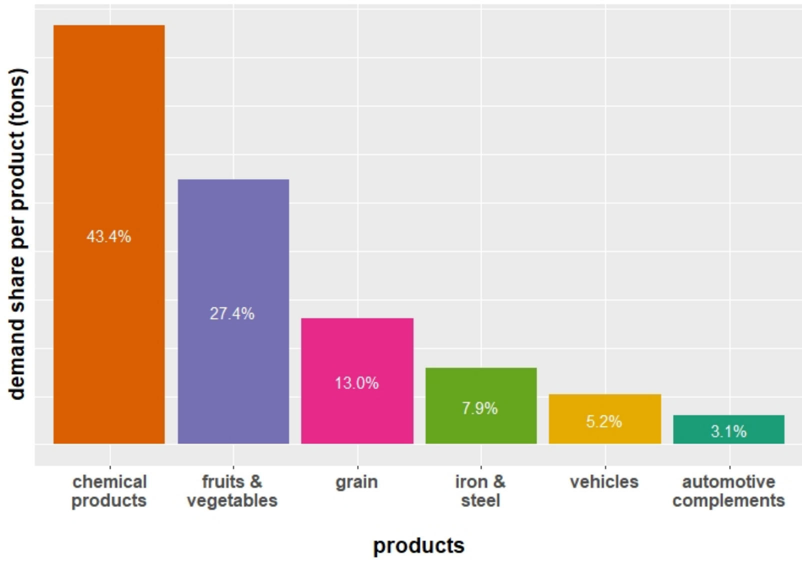


Figure 7.4: Demand share per products. Source: author.

transport during last years (2015-2018), and with origin or destination in Catalunya, Comunidad Valenciana or Aragón. The countries selected were France, Germany, Italy, the Netherlands, Belgium, and Poland. Not all the volume of products reflected on the statistics follows the rail network defined. The volume transported was calculated as a fraction of the total demand transported by train and truck, during 2018, for the products and regions selected. The fraction varies from 20% and 70%, depending on the product and its origin or destination. The total volume transported in the period is $7,764,760$ tons, distributed in Chemical products (43.4%), Fruits and Vegetables (27.4%), Grain (13.0%), Steel and Iron (7.9%), Vehicles (5.2%), and Automotive complements (3.1%). Figure 7.4 shows the distribution by products of the total demand used in the test cases.

7.2.4. Definition of rail corridors

Once the rail network is built and the OD pairs defined, the next step is to describe the rail corridors that the unique or various carriers offer to shippers for freight transport. These corridors are made up of lines connecting yards, where trains can be composed or decomposed, or railcars are loaded or unloaded, depending on transport needs. Figure 7.5 shows a representation of the lines used to computationally test the models.

As it was pointed out in Section 4.1, each line may be decomposed into two directed lines, one for each direction. Also, the ordered set of directed arcs that composes each directed line is unique and defined previously. The definition of their rail corridors,

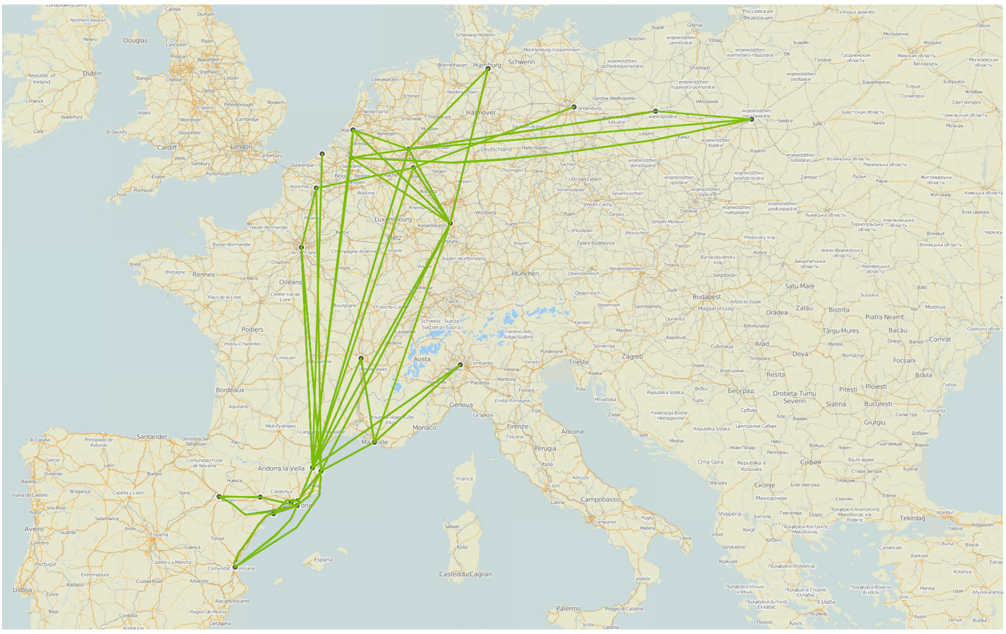


Figure 7.5: Lines used for tests. Source: author.

connections and frequency are based on public information about rail freight corridors in Europe. However, these examples do not match exactly with any specific carrier offer. Figure 7.6 shows a schematic of the corridors for five carriers used in the tests, while Figure 7.7 depicts the corridors for the case when there is no carrier competition, with a unique carrier acting as a monopoly. Each direct connection between two yards corresponds to two y-paths, one for each direction.

In these examples, three different track types are highlighted because running from one type to another may require a locomotive change, or at least, to stop at the yard to make some adaptations on the train. Differences are track gauge and electrical voltage. To make tests easier to follow, we have assumed that there are only two types of electric voltage, 3 kV for the Spanish regions and another generic and common for the European region. However, Central European countries do not currently share the same electrical voltage. On the contrary, the assumptions on track gauges are conformed to reality. The **UIC** gauge corresponds to the standard track gauge, with a separation of 1,435mm between tracks, while the **IBE** is for the "Iberian gauge", that is, with 1,668mm of separation between tracks.

As can be seen in the schematic of the lines for multi-carriers scenario (Figure 7.6), most of the y-paths are of shared-use among different rail carriers. This situation has a direct impact on carriers competition, and the options each carrier has to attract the

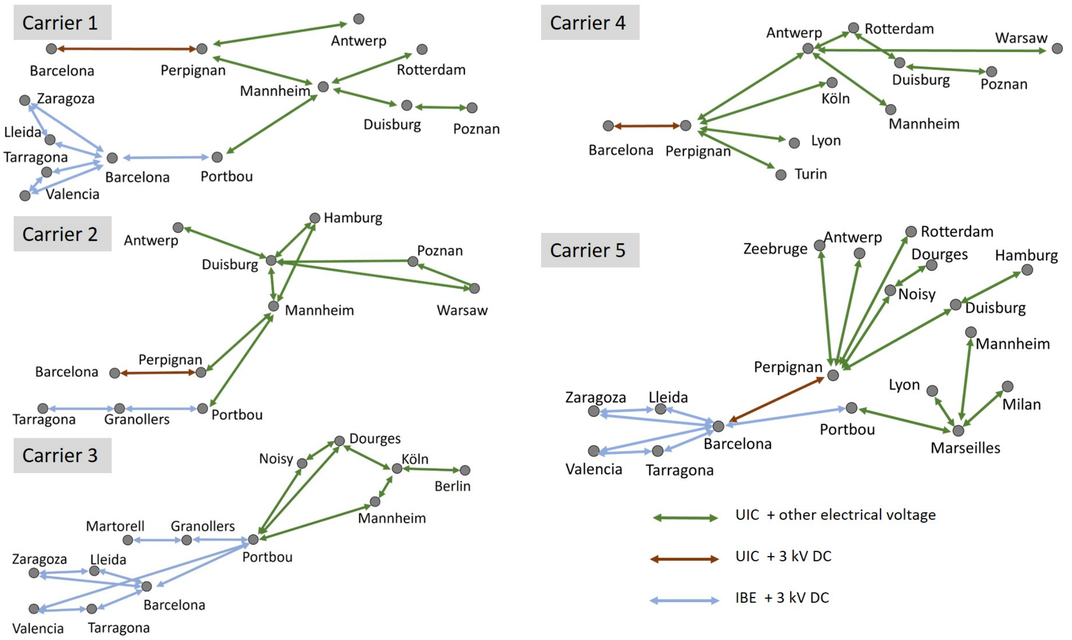


Figure 7.6: Schema of the corridors for each carrier used for the tests under carriers competition. Source: author.

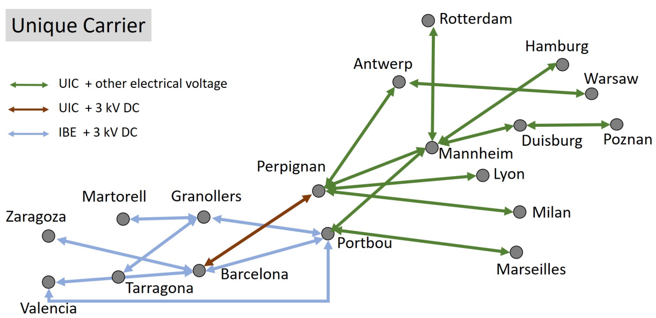


Figure 7.7: Schema of the corridors for a unique carrier used for the tests when no carriers competition is applied. Source: author.

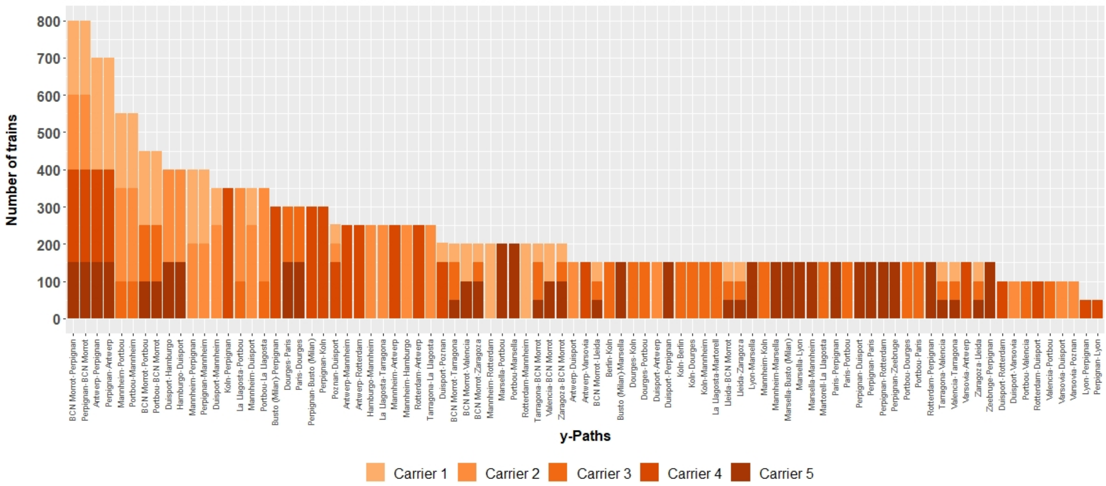


Figure 7.8: Carrier slots for each y-path. Source: author.

needs of transport for the shippers. The level of competition among carriers is shown in next Figures 7.8 and 7.9.

Figure 7.8 depicts for each carrier, where each colour represents a carrier, the allocation of trains for each y-path (each direct connection between two yards). Each stacked bar represents all the y-paths that share the same infrastructure, and the total height of the bar is for the maximum number of trains that can be allocated on the shared infrastructure. The different colours on a bar show that the corresponding infrastructure is of shared use among the carriers represented by their colour, and the height of each colour is for the maximum number of trains the carrier can allocate on this y-path. For instance, the first bar from the left hand side corresponds to the direct y-path from Barcelona to Perpignan, used by all the carriers except Carrier 3, as can be checked in Figure 7.6. The allocation for each carrier is as follows: Carrier 1 and Carrier 2 can run, each one of them, at most 200 trains, while Carrier 4 can run at most 250 trains and Carrier 5 at most 150 trains.

Additionally, Figure 7.9 shows how many y-paths each carrier operates exclusively and how many they share the infrastructure with one or more carriers. As before, each colour corresponds to one carrier. The height of each bar is for the number of y-paths. Starting from the left, the first block of attached bars represents the number of y-paths each carrier operates exclusively. The second block shows how many y-paths are of shared use between two carriers for each carrier, the third block is for the y-paths shared among three carriers, and the fourth block for the y-paths shared among four or five carriers. For instance, Carrier 1 manages two y-paths exclusively: Mannheim to

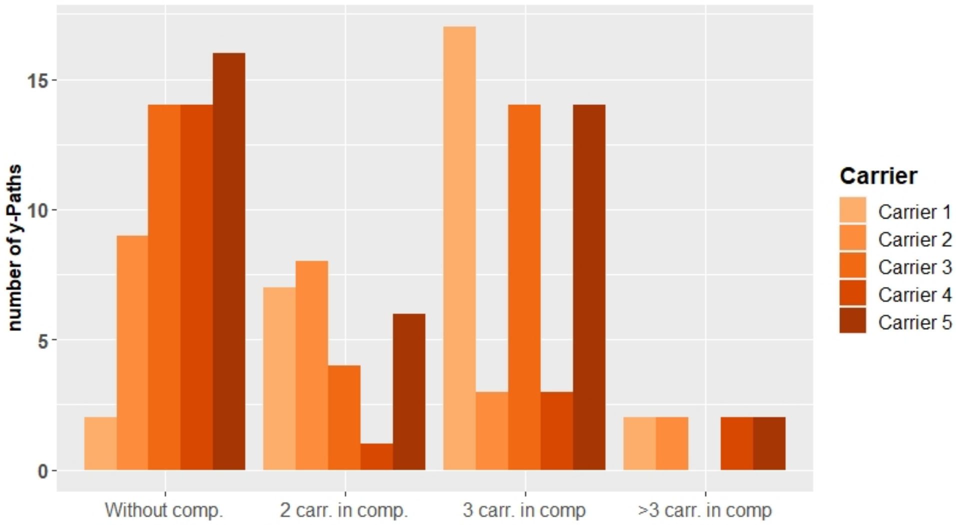


Figure 7.9: Level of competition: y-paths exclusively operated and of shared-use. Source: author

Rotterdam and Rotterdam to Mannheim, as shown in Figure 7.6. Seven of their y-paths share tracks with another carrier, seventeen of their y-paths with two other carriers and two of their y-paths with more than two carriers.

Notice that a terminal that is served exclusively by one carrier means that all the demand from or to this terminal will be transported by that carrier. For instance, in Figure 7.6 Milan, Marseilles or Lyon terminals are only operated by Carrier 5. Then all t-paths with origin or destination these terminals will be operated by Carrier 5. The same occurs with Martorell, linked to Carrier 3, or Warsaw linked to Carrier 4.

7.2.5. Utility criteria

Data preparation for the tests has followed two stages:

a) selection of factors that determine disutilities,

b) estimation of significant parameters of a logit based RUM model, and

The first step is to define the explanatory variables for the systematic term of the generalized cost functions (6.14). The criteria for choosing them must rely on their usage in common data surveys and be easy to calculate for the corresponding transportation mode. Four variables were chosen: distance, travel time, price, GHG-emissions. Also, a dummy variable for each type of product transported was added.

- **Distance**. Information was obtained from Ecotransit (EcoTransIT World), a web tool for calculating transportation environmental impacts.
- **Travel time**. For train, data is calculated from TPNOVA Rail & Logistics Services S.L. (2020), a rail operator agency's web page with calculator for train freight transportation. For road, a calculation was applied based on the average speed for trucks for international freight transport and taking into account mandatory time for resting during the trip.
- **Price**. For train, again TPNOVA gives an approximation, complemented with information from other sources, as DB Cargo AG (2019), Martínez et al. (2015) or Pérez (2015). For road, the data used is the average price per kilometre for international transport by road calculated from the works by Martínez et al. (2015), Pérez (2015), and taking into account that prices of road transport remained steady since 2014 (División de Estudios y Tecnología del Transporte de la Secretaría General de Transporte (2020)).
- **GHG-emissions**. Data obtained from Ecotransit (EcoTransIT World). It was parameterized for transporting two TEUs of 10 t/TEU, from origin yard to destination yard, for both truck and train, using the standard parameters:

Truck: a diesel vehicle of 26-40 t, with a load factor of 95.77% and empty trip factor of 20%.

Train: an electrified container train of 1,000 t, with a load factor of 49.8% and empty trip factor (ETF) of: 20%

- **A dummy variable for each product**. A 0-1 value, trying to catch special characteristics for each product when transported.

7.2.6. Estimation of the utility function parameters

A logit approximation was applied to estimate the values for the β and $\tilde{\beta}$-parameters (the parameters for the disutilities u_r^ω and $\tilde{u}^\omega$ respectively, in the **MINLP**-model), by using the R-package *mlogit* (Croissant (2020)). A common approximation for all OD-pairs was made.

The first difficulty arose with the linear dependency among the explanatory variables, especially between **GHG-emissions** with **distance** and **price**, for train data, and **distance** with **travel time** and **price**, for road data. Different combinations of the

Table 7.1: Disutility function parameters for the tests.

utility component		mode	unit	Estimate	confidence interval 2.5%	97.5%	β^-	β^+	
β_0	(Intercept)	train	-	4.6549	4.3410	4.9688	4.6549	0.3139	
$\tilde{\beta}_d$	distance	truck	km	0.0051	0.0036	0.0065	0.0051	0.0015	$\star$
β_d	distance	train	km	0.0051	0.0036	0.0065	0.0051	0.0015	$\star$
$\tilde{\beta}_{co}$	GHG-emissions	truck	kg	0.0006	0.0006	0.0007	0.0006	0.0001	$\star$
β_{co}	GHG-emissions	train	kg	0.0006	0.0006	0.0007	0.0006	0.0001	
β_p	price diference	train	€	0.0009	0.0003	0.0015	0.0009	0.0006	$\star$
β_{ac}	automot.compl	train	0/1	-0.5919	-1.0091	-0.1747	-0.5919	0.4172	$\star$
β_{gr}	grain	train	0/1	-0.8612	-1.8832	0.1608	-0.9416	0.9416	$\star$
β_{fv}	fruits& veget	train	0/1	1.3259	0.9342	1.7176	1.3259	0.3917	$\star$
β_{ch}	chemical prod	train	0/1	-0.3365	-0.5487	-0.1244	-0.3365	0.2121	$\star$
β_{ve}	vehicles	train	0/1	-4.7270	-4.9468	-4.5072	-4.7270	0.2198	
β_{od}	OD decrease	train	0/1	-2.0661	-2.2039	-1.9283	-2.0661	0.1378	
β_{OD}	OD increase	train	0/1	2.9888	2.6566	3.3209	2.9888	0.3321	

McFadden R^2: 0.37632

variables were tested, with no satisfactory results: either some parameter was not significant or had the wrong sign, or the adjustment was poor. A combination of **distance**, **GHG-emissions** and a new variable defined as the train price minus road price (named as **price difference**), jointly with the dummy variables per product, seemed to fit the data. Nevertheless, new adjustments were necessary: first, the **price difference** was not appropriated for some products, as "vehicle" or "chemical products". In these cases, **price difference** was not applied. Also, because of the special characteristics of train transport, a few combinations of OD-pairs and product have very different behaviour in terms of train share. For these particular OD-pairs-product combinations, two new dummy variables were defined: one, named **OD decrease**, which helps to reduce train costs (and to increase train share); other, named **OD increase**, which helps to raise train costs (and to diminish train share).

The results after calibration can be seen in Table 7.1, which shows the mean estimate, the limits for the 95%-confidence interval and the values for β^-, β^+ parameters. All of the β^-, β^+ parameters were calculated from the confidence interval. The symbol $*$ marks those parameters to which the robustness criteria is applied. The dummy variable for "steel & iron" products was the reference variable. Also, the road was the reference mode. An increasing value in variables **distance**, **GHG-emissions** and **price difference** results in an increasing cost, both for train and road. Some products are more suitable to transport by train than others, as the difference (both in sign and value) among the dummy variables for product shows. Furthermore, parameter **OD decrease**, with its negative sign, implies a reduction in train costs, while parameter **OD increase**, with its positive sign, raises train costs. All parameters are correct in sign, and all have a good level of significance except the dummy parameter for “grain”. Some products are more suitable to transport by train than others, as the difference (both in sign and value) among the dummy variables for product shows. A value of 0.38 for the R^2 of McFadden shows a good adjustment for the model.

Different tests were performed to validate the quality of the estimated coefficients: the marginal effects of the continuous parameters for each product and the capability to reproduce the observed flows, with a high value of the coefficient of determination R^2, equal to 0.99.

Chapter 8

Computational Tests

This chapter presents a summary of the computational tests of the models, based on the scenario described in Chapter 7. The first section describes the **DCECC**-model tests, while the second section explains the **MINLP**-model tests. Tests were carried out on a R5500 workstation using Intel® Xeon® CPU 5645 with 2.40 GHz and 48 Gb RAM.

8.1. DCECC-model computational tests

The maximum train length allowed on tracks is one of the factors that have a significant impact on capacity and interoperability. Increment of the maximum train length on tracks should be done by upgrading the passing loops and refuge sidings, and also the placement of signals, as it was pointed out in Section 2.1. The yards connecting tracks must be adapted to admit trains of the desired length. These actions are costly, but indeed they are less expensive than other actions that increment capacity, as can be doubling the tracks.

8.1.1. Description of the tests

The **DCECC**-tests analyse the impact of upgrading the infrastructure to admit longer trains, focused on the Spanish region of the general scenario. For simplicity, the remainder of the European rail network should permit trains of 750m. Figure 8.1 shows the different maximum train lengths that can run in the Spanish area and the upgrade suggested. As can be seen, there is a wide variety of maximum length of trains allowed, although the most general is 575m. The only section currently adapted to 750m is the direct connection between Castellbisbal and Cerbere. Each colour represents a

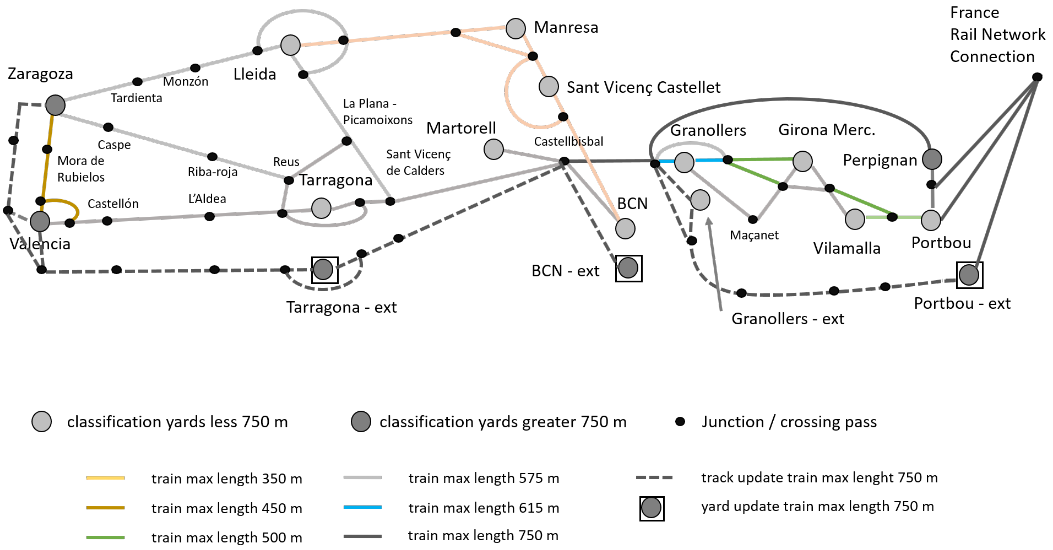


Figure 8.1: Maximum train length allowed on current tracks on the Spanish section of the scenario and the suggested upgrade. Source: author.

maximum allowed train length. Also, most of the yards in this area admit trains with lengths lower than 750m, except Zaragoza and Valencia. The dotted lines represent the tracks adapted to admit trains of 750m. They are pictured as extra-edges so that both options are possible, although they are incompatible.

Table 8.1 displays the dimensions of the rail network. The first block shows the dimensions of the tracks in kilometres: the total kilometres of tracks, the kilometres for new tracks, basically for converting single tracks to double, the total kilometres of tracks that may be upgraded (for instance, by converting a track with Iberian gauge to a track of mixed-use, or electrifying a non-electrified track) and the total tracks that may be upgraded by acting on the block signal system. The second block shows the number of yards: the total number of yards, the number of yards that may be built and the number of yards that may be upgraded (for instance, to admit longer trains or increase capacity).

Table 8.1: Infrastructure characteristics of the rail network for the **DCECC**-tests.

Tracks (km)			Block Signal	Yards (number)		
Total	New	Upgrade	Upgrade (km)	Total	New	Upgrade
16,663	2,740	1,688	4,387	34	2	13

Table 8.2 shows the main characteristics that concern the rail carriers' operation: the period analysed, the tracks occupation percentage that reduces the maximum theoretical

Table 8.2: Operation characteristics of the rail network for the **DCECC**-tests.

Period	% occup.	max.train length	max.train weight
220 days	65%	350m / 750m	650 ton / 1,100 ton

tracks occupation, the interval values for the maximum train length allowed to run on tracks, in metres, and the maximum train weight in tons, for fast/regular freight trains, according to their average speed and slopes on tracks.

Table 8.3 shows the size of the optimisation problem that results from applying the model to this rail network. Here, the sets that conform the problem are detailed, and how they are translated to variables and constraints: total number of nodes ($|N|$), arcs ($|A|$), OD pairs ($|W|$), yard-paths ($|\Gamma|$), number of t-paths ($|R|$), number of carriers ($|C|$), number of railcar types ($|\mathcal{V}|$), number of types of products ($|P|$), number of types of freight trains ($|K_M|$), total of variables of the optimisation problem, how many of them are binary and integer, total number of constraints, and total number of non-zeros. The model was implemented using Python 3, and CPLEX V12.7.

Table 8.3: Problem size for the **DCECC**-tests.

N	A	W	Γ	R	C	$\mathcal{V}$	P	K_M	Var.	Binary	Integer	Rows	Non zeros
95	312	65	192	400	5	2	6	4	14,704	4,699	5,122	5,633	34,830

8.1.2. Results of the tests

Two different groups of tests were executed to check the model and its efficiency. The first deals with the current demand, while the second increases a 15% the demand while maintaining the same origins and destinations and type of products transported. Both consider the possibility of upgrading the rail network to admit trains up to 750m long, and acting on block signals on certain tracks to increase the capacity of the rail network. In each of these groups, the different runs resulted from applying different values for γ, the Pareto coefficient. Previously, we ran an initial test for each group by fixing the signalling distance (d_a variables). The results from these first runs allowed us to obtain an initial MIP-start that helps improve the running time of the original problem. From that, each test solution constituted the MIP-start for the next test. Tables 8.4, 8.5 and 8.6 summarize the results for the experiments. All costs are expressed in thousands of euros, and they are annualised, in a yearly-based approach. Each row in the tables corresponds to one test.

The meaning of the columns for Table 8.4 is as follows. First, the case identification

Table 8.4: Performance and summary of costs in thousands of euros of the best solutions for the **DCECC**-tests.

Case	γ	GAP (%)	CPU (sec)	Total (k€)	Infr (k€)	Ope (k€)
C10	0.1	5.13	22,003	47,028	38,654	8,375
C30	0.3	1.70	22,050	46,801	37,600	9,201
C50	0.5	1.77	22,010	39,867	22,956	16,912
C70	0.7	0.91	18,200	39,831	21,444	18,687
C90	0.9	0.90	17,034	41,052	21,141	19,912
F10	0.1	6.52	22,132	64,751	41,944	22,807
F30	0.3	5.29	22,002	63,385	40,444	22,941
F50	0.5	4.46	22,053	51,408	24,342	27,067
F70	0.7	4.93	22,344	51,443	22,839	28,604
F90	0.9	2.85	22,126	51,859	21,384	30,475

and the γ value (the Pareto coefficient) appear on the left. The first letter of the test case corresponds to the test group: "C" for the current demand group of tests, while "F" for the tests with a 15% increase in demand (for "future"). The following three columns are the GAP percentage for the best feasible solution found, the total CPU time required in seconds, and the total life-cycle costs for the best feasible solution. The next two columns show this total cost divided into infrastructure and operation costs. The majority of tests stopped after 22,000 seconds of CPU consumption (approximately 6 hours). The time limit was imposed on all tests, obtaining good accuracy in the current demand tests but less precision in the increasing demand tests.

Figure 8.2 graphically represents the relationship between infrastructure and operating costs for both test groups. The vertical axis shows the infrastructure costs, while the horizontal axis displays the operational costs. Observe that the problem is of the MIP nature. Only a limited number of points has been obtained, and this does not allow detecting probable unsupported points or discontinuities in the frontier. This situation implies that the Pareto frontier may be discontinuous and even with points corresponding to entire intervals for γ (as shown in several examples in Antunes et al. (2016)). However, some convexity-like shape can be easily guessed.

Notice that the greater the value of γ, the lower the cost for infrastructure and the higher the cost for operations. These results show two tendencies in both groups of tests: increasing infrastructure costs help reduce operating costs while reducing infrastructure investments increases operation costs dramatically. Figure 8.2 depicts clearly this be-

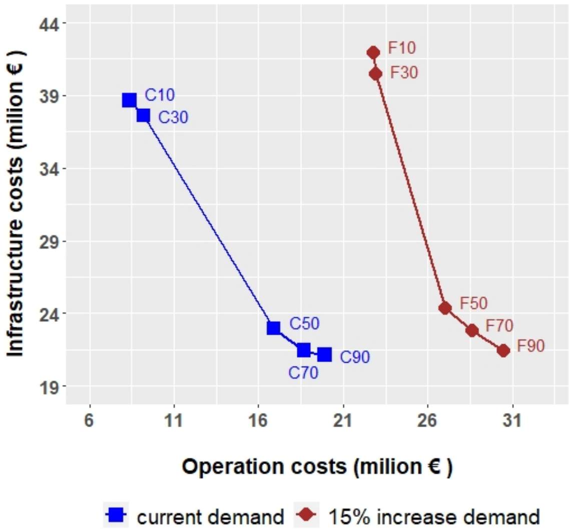


Figure 8.2: Approximate Pareto frontier for the **DCECC**-tests. Source: author.

haviour. For the current demand, with nearly half of test C10 on infrastructure costs, operation costs in test C90 raise more than twofold. These are the most extreme cases, but it looks as if no intermediate point is possible in both test groups. Also, the results in both the table and graph show that operation costs are considerably greater for the "F" tests, with a 15% increase in demand. In contrast, infrastructure costs are quite similar in both groups of tests, and they mainly depend on the weight they have on the objective function.

Table 8.5 decomposes infrastructure and operations costs into the different components of the objective function. The first column identifies the test case. Then, the next three columns correspond to the infrastructure cost investments and maintenance (yards - **Infr. Y**, tracks - **Infr.T.** and signals - **Infr.S.**). The next four columns are for the operation costs: preparation/reclassification costs in the terminals (**Mn. T.**); railroad costs for goods transport on the tracks (**Run.T.**); rolling stock investment and maintenance (**Wag**); and the penalty cost for transporting priority products using non-fast trains (**Prior**).

To complement costs information, Table 8.6 details some of the indicators that result from the best solutions of the tests. As before, the first column is the case identification. The following column shows which of the best solutions for the tests require upgrading the network to admit trains up to 750m in the Spanish region. Column **T(n)** shows the percentage of kilometres of new tracks built of the total kilometres of candidate tracks to be built. Column **T(u)** shows the percentage of tracks to be upgraded due to the

Table 8.5: Costs decomposition in thousands of euros for the best solutions for the **DCECC**-tests.

Case	Infr.Y.	Infr.T.	Infr S	Mn.T.	Run.T.	Wag.	Prior
C10	5,145	33,265	243	1,657	1,970	4,770	-
C30	5,145	32,215	240	2,068	2,241	4,832	61
C50	4,590	18,075	290	2,177	5,414	9,321	-
C70	4,590	16,460	394	2,635	7,021	9,031	-
C90	4,590	16,275	275	2,677	7,914	9,321	-
F10	5,145	36,472	327	5,269	6,882	10,656	-
F30	5,145	34,972	327	5,564	6,706	10,671	-
F50	4,545	19,515	281	3,945	12,595	10,527	-
F70	4,545	18,015	279	4,724	13,334	10,546	-
F90	4,590	16,515	278	5,598	14,334	10,544	-

requirements for admitting trains up to 750m. of the total kilometres of candidate tracks to be upgraded. Column **T(s)** shows the percentage of tracks to be upgraded due to changes in the block signal system of the total kilometres of candidate tracks. Column **C** shows the percentage of tracks of the total number of tracks built and with usage greater than 55%. Columns **Y(n)** and **Y(u)** displays the total new yards built and yards upgraded, respectively. Columns **W(c)** and **W(v)** represent the total number of railcars for transporting containers and vehicles, respectively, needed to cover all transportation.

Table 8.6: Rail network impact of the best solutions for the **DCECC**-tests.

Case	750m	T(n)	T(u)	T(s)	C	Y(n) /Y(u)	W c/v
C10	Y	-	61.5%	18.6%	7.2%	2/4	122/231
C30	Y	-	61.5%	18.6%	9.4%	2/4	124/235
C50	-	-	-	18.6%	6.9%	2/-	242/447
C70	-	-	-	18.6%	7.3%	2/-	239/438
C90	-	-	-	18.6%	6.9%	2/-	242/447
F10	Y	8.2%	94.8%	18.6%	10.7%	2/4	274/530
F30	Y	8.2%	94.8%	18.6%	10.0%	2/4	275/530
F50	-	-	-	18.6%	9.4%	2/-	276/503
F70	-	-	-	18.6%	9.4%	2/-	276/504
F90	-	-	-	18.6%	9.2%	2/-	277/503

Notice that investments and maintenance costs on yards are very similar in all cases. The differences seem to be related to whether or not terminals are to be upgraded to admit trains up to 750m. The investment in tracks does not appear to be associated with improving tracks capacity, although the fact that some solutions do not have enough accuracy may hide this relationship. When operation costs are penalised, tracks and terminals are upgraded, reducing the number of railcars required.

8.2. MINLP-model computational tests

The model was implemented using Python 3, and the solvers *optimize* from Scipy and CPLEX V12.7. The analysis of results was conducted in R. The rail network has $17,406$ km with a maximum train length from 350 m to 750 m. The maximum weight that trains can transport varies from 550 to $1,100$ tons.

8.2.1. Description of the tests

Different groups of tests were executed to check the model utility and its efficiency. A combination of three criteria was used to define these groups. The first criterion considers whether carrier competition exists or not in rail freight transport. A third of the tests simulate the situation where only one carrier operates the rail freight network, acting like a monopoly. In contrast, the reminder two-thirds of the tests correspond to a "Carrier Competition" analysis, where different carriers operate under competition. The second criterion is based on the robustness level applied. The 0-level, where no robustness is applied, is equivalent to a deterministic case. In this case, the β-parameters are based on their average value. The level number indicates the maximum number of β-parameters that are allowed to divert from its average value, taking a value from the interval $[\beta^- - \beta^+, \beta^- + \beta^+]$. The third criterion affects only the "Carrier Competition" cases, defining two types of tests: a) those with a previously fixed allocation of slots for each carrier and b) tests where the number of slots is left variable and determined by the model, thus making possible to get a modal split closer to the a priori utilities of the RUM model.

Two groups of acronyms characterize the tests: first, **CF**, **CD** and **M** apply for the carrier competition cases with fixed allocation (**CF**), carrier competition cases with variable allocation of slots (**CD**), and the monopoly situation (**M**). Second, **Det**, **RL 1**, **RL 2** and so on, label tests depending on the level of robustness applied. **Det** corresponds to the Deterministic version of the model. In order to not saturate with

Table 8.7: Problem size for the **MINLP**-tests with several carriers.

N	T	Y	A	W	Γ	R	$\mathcal{V}$	K	O	Variables	Int.	Bin.	Rows
95	20	34	266	115	140	240	3	3	5	4,013/8,997	3,516	5	4,258/8,750

an excess of very similar data, only some cases are reported. Table 8.7 shows the size of the sets and the problem size for *CF* and *CD* tests. The first value for **Variables** and **Rows** columns corresponds to robustness level 0 or **Det** tests, while the second is for the robust version of the tests. The number of integer variables (**Int.** column) and binary variables (**Bin.** column) is the same for all the tests.

8.2.2. Modal split compliance

In Section 6.1.2 the modal choice properties of the model are detailed. The modal split behaviour, as appears in equation (6.11),

$$\frac{\tilde{h}^{\omega}}{\chi^{\omega}} = \{1 + \exp(\tilde{u}^{\omega} - \vartheta^{\omega})\}^{-1} \ , \tag{6.11}$$

depends on two parameters: $\tilde{u}^{\omega}$ and ϑ^{ω}. The first one corresponds to the road utility function, while the second one does not exactly match the train utility: the Lagrange multipliers associated to the constraints that condition the way products are transported by train exert some influence on the modal split. Whether this influence is strong or weak, it mainly depends on how the right hand terms tighten the different constraints, once calculated the solution to the optimisation problem. In our model, the constraints that mainly put pressure are those which are related with infrastructure capacity (4.29), (4.30) and (4.31), rolling stock capacity (4.11) and (4.12), and cost-effectiveness for carriers (4.13) and (4.14). Remember that $u^{\omega}, \tilde{u}^{\omega}$ are the utilities used in the experiments, and for each $\omega \in W$, $\tilde{h}^{\omega}/\chi^{\omega}$ corresponds to the road share derived from the solution obtained. From (6.11), ϑ^{ω} can be expressed as a function of road utilities and road demand:

$$\vartheta^{\omega} = \tilde{u}^{\omega} - \log\left(\frac{\chi^{\omega}}{\tilde{h}^{\omega} -} 1\right) \ . \tag{8.1}$$

On the other hand, ϑ^{ω} is also a function of train utilities and some extra "penalties" that come from the Lagrange multipliers mentioned above, as appears in (6.10) :

$$\vartheta^{\omega} = \min_{r \in R(\omega)} \{u_r^{\omega} + \gamma_r^{\omega}\}, \ \ \omega \in W. \tag{6.10}$$

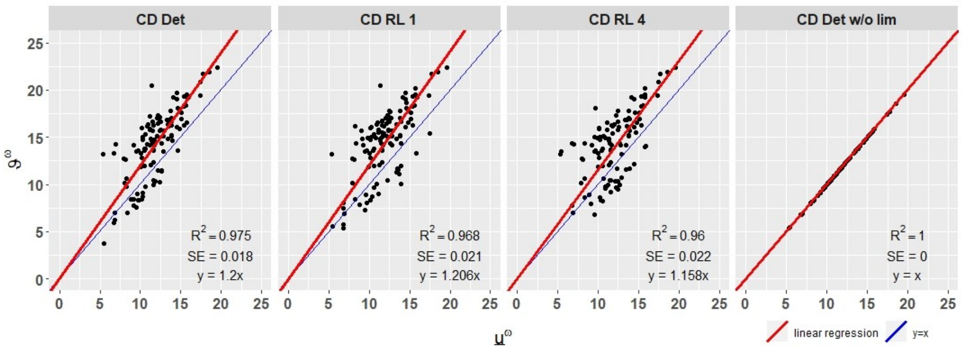


Figure 8.3: Comparison between $\underline{u}^{\omega}$ and ϑ^{ω} values. Source: author.

Then, let $\underline{u}^{\omega}$ be an approximation of ϑ^{ω}:

$$\underline{u}^{\omega} = \min_{r \in R(\omega)} u_r^{\omega}. \tag{8.2}$$

Notice that the fewer constraints are active, the closer $\underline{u}^{\omega}$ will be to the value of ϑ^{ω}. Following this reasoning, a new test-case named **CD Det w/o lim** is defined. This experiment is based on **CD Det**-test, although with constraints (4.29), (4.30) (4.31), (4.11), (4.12) , (4.13) and (4.14) relaxed. The goal is to analyse the behaviour of the Deterministic model when no capacity limits nor cost-effectiveness conditions are imposed to the model, in contrast with the more realistic scenarios.

Figure 8.3 plots the relationship of $\underline{u}^{\omega}$ versus ϑ^{ω} for different examples based on the **CD**-cases. The graph on the left corresponds to the Deterministic (**Det**) test, while the next two graphs show the **RL 1** and **RL 4** tests. The graph on right depicts the results for the **CD Det w/o lim** test. The main linear regression indicators are displayed on each graph (R^2, the standard error SE and the x-coefficient). As can be seen, $\underline{u}^{\omega}$ and ϑ^{ω} values are perfectly correlated for the **CD Det w/o lim** solution, following the line $y = x$, but for the **CD Det**, **CD RL 1** and **CD RL 4** solutions, the constraints that are active alter the expected value for ϑ^{ω}. Although the three graphs are very similar, these examples show that dispersion slightly increases with the robustness level.

8.2.3. Results of the tests

Figure 8.4 illustrates the goodness of fit of the model by comparing the road share that outcomes from the solution, and the observed road share, for the **CD Det**, **CD RL 1** and **CD RL 4** tests, and also, for the **CD w/o lim** example. All graphs compare both values directly by plotting one versus the other. Note that the vertical axis values

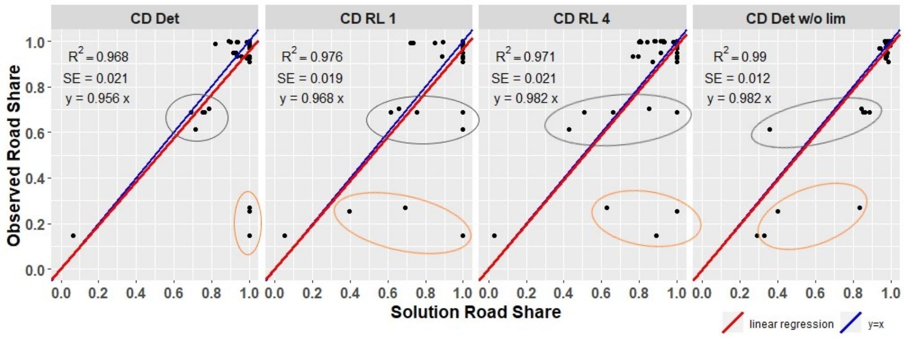


Figure 8.4: Comparison between the solution road share and the observed road share. Source: author.

(the observed values) are the same for all the graphs, while the horizontal axis values correspond to each test's solution. The grouped points, highlighted with two different colors, correspond to the same origin-destination-products triples for all tests. For the yellow group, a better prediction regarding the **Det** test is obtained when applying some level of robustness. In contrast, for the gray group, the **Det**-test solution provides a more accurate prediction. The fact that the **CD Det w/o lim**-test solution has outliers (and the **Det** test gives a better prediction for some of them) illustrates the influence that railway conditions have on modal split when road and train compete.

Table 8.8 summarizes the algorithm performance. The first two columns identify the test. Column **rel. error** shows the relative error when the algorithm stops. Column **number it tot** shows the total number of iterations for solving the full problem **MINLP**, while column **number it NLPP** shows the average number of iterations used to solve the primal subproblem **NLPP**. The next four columns correspond to CPU consumption in seconds. Column **cpu MINLP** shows the total CPU used to solve **MINLP**. Column **cpu it** corresponds to the average CPU-consumption per iteration, and it is decomposed on the next two columns: CPU-use required by the primal subproblem **NLPP** (column **NLPP**) and the CPU consumption for solving the master problem **MLMP** (column **MLMP**). As can be seen, the algorithm is fast and efficient, allowing to solve the different experiments in less than five minutes. A relevant CPU increase in robustness cases appears when different carriers compete against each other, mainly related to the CPU-time required to solve **NLPP**. Figure 8.5 shows the upper bound, the lower bound and the relative error (in logarithmic scale) evolution for the algorithm to solve **MINLP**. The y-axis on the left corresponds to the upper bound and lower bound values (expressed in millions), while the y-axis on the right, in blue, shows the relative

Table 8.8: Algorithm performance for **MINLP**-tests: relative error, iterations, cpu consumption.

case	rob level	rel error	number iter. MINLP	number iter. NLPP	cpu (sec) MINLP	cpu (sec) it	cpu (sec) NLPP	cpu (sec) MLMP
CF	Det	1.9e-5	36	24	87	2.32	1.44	0.88
	RL 1	2.7e-5	34	23	262	7.58	6.66	0.92
	RL 4	1.3e-5	36	22	243	6.65	6.04	0.60
CD	Det	2.9e-5	35	21	200	5.59	1.13	4.46
	RL 1	1.1e-5	35	25	285	8.02	7.21	0.81
	RL 4	1.8e-5	41	20	279	6.70	5.13	1.56
M	Det	0.8e-5	42	5	232	5.45	0.33	5.11
	RL 1	0.7e-5	41	5	196	4.68	0.99	3.69
	RL 4	1.1e-5	36	5	62	1.61	0.92	0.68

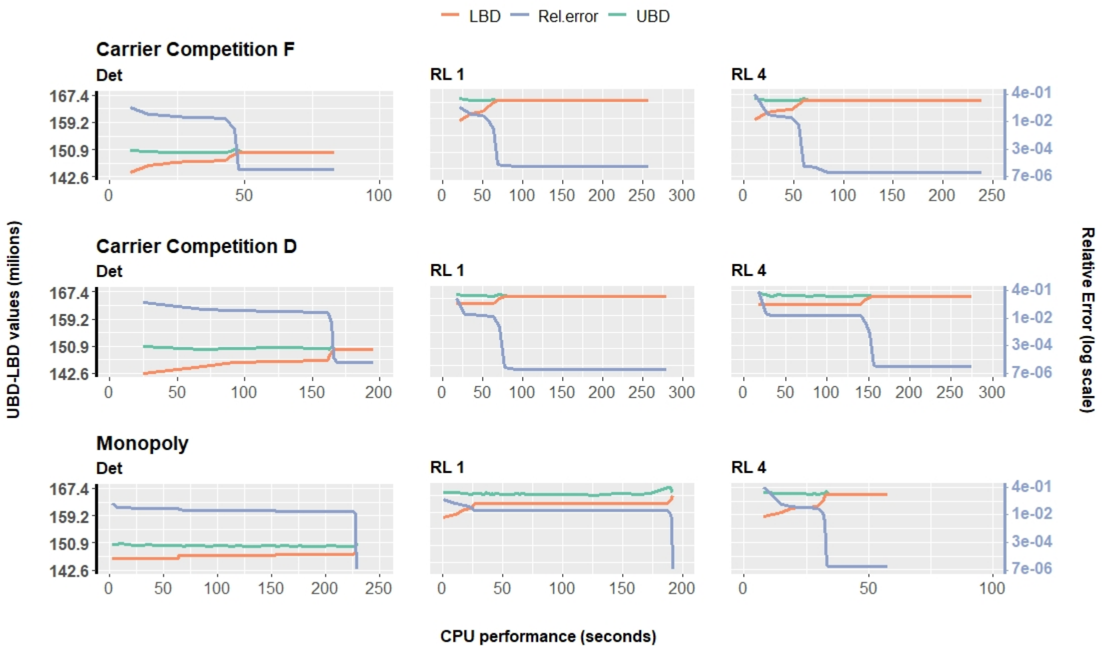


Figure 8.5: UBD vs LBD, and algorithm relative error evolution for **MINLP**-tests. Source: author.

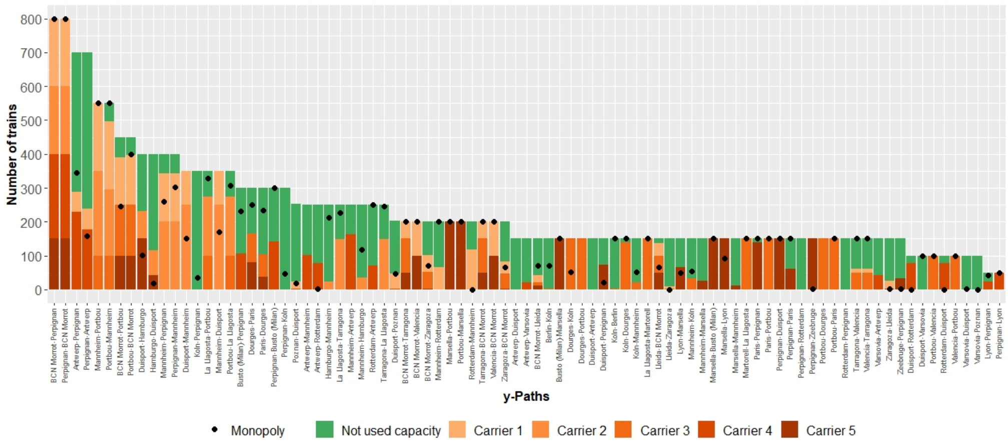


Figure 8.6: y-Paths use, for the **MINLP**-Deterministic case. Source: author.

error values. All the graphics share the same scale on both y-axis. The x-axis represents the CPU consumption in seconds. The first row of graphics corresponds to **CF**-tests, the second row shows the algorithm evolution for **CD**-tests, and the third row for the **M**-tests.

Next, a few examples of the information about the carrier's operation that the model can provide. Figure 8.6 shows a representation of the rail capacity consumption on the different y-paths, compared with the maximum train allocation available for each carrier that Figure 7.8 represents. Figure 8.6 displays the y-paths capacity used for each carrier in the Carriers Competition Deterministic test and also for the Monopoly Deterministic test. Paths are shown following the same order as in Figure 7.8. Each column corresponds to one y-path, and each colour in the orange palettes represents one carrier. Each colour-bar height depicts the number of trains the corresponding carrier operates following the solution from the **CF Det** test. The green colour bars correspond to the free capacity for the **CF Det** test, calculated from the maximum capacity allocated on the path and subtracting the carriers' trains consumption. The black points correspond to the number of trains the Monopoly company operates on the path, also following the solution from the **M Det** test. This graph quickly detects which paths are the rail network bottleneck or are of little use.

For instance, the first column on the left corresponds to the y-path from Barcelona to Perpignan. All carriers use the maximum number of trains they can run: 200 for Carrier 1, 200 for Carrier 2, 250 for Carrier 4 and 150 for Carrier 5 (compare with Figure 7.8). Also, the Monopoly company requires the maximum number of trains they can run (the black point is on the top of the bar). There is no free capacity available in

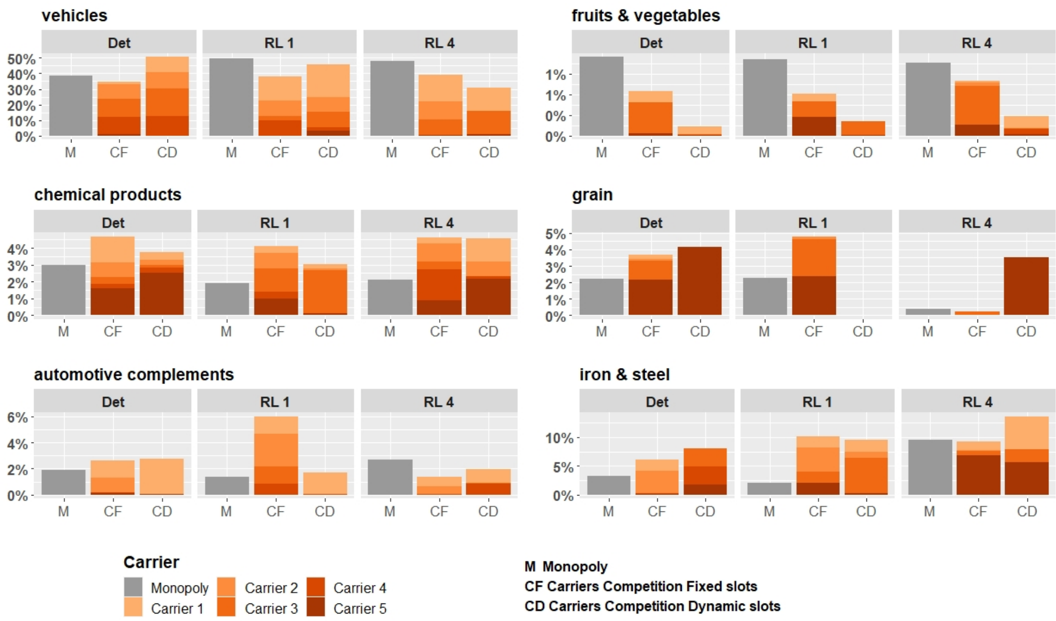


Figure 8.7: MINLP tests. Volume modal share. By products, different robustness levels. Source: author

both tests. In contrast, the third column from the left corresponding to the y-path from Antwerp to Perpignan shows that Carrier 1 requires roughly 50 trains and Carrier 4 just over 200. Still, there is capacity available for just over 400 trains to run. In this case, the Monopoly company requires roughly 350 trains (following the black point position), which is more than the sum of the trains of Carrier 1 and Carrier 4.

Figure 8.7 shows the carrier share and the monopoly share for each product and each test. The range of oranges represents the carriers, while the grey colour corresponds to the monopolistic situation. In most cases, when comparing the **CF** and **CD** results, there is a better distribution among carriers in the **CF** experiments than in the **CD** experiments. This behaviour is due mainly to the initial slots allocation. While the **CD** situation allows one or few carriers to gather the whole train transport easily, the **CF** situation limits more clearly the demand each carrier can transport.

Concerning the modal share, per volume and import, there is barely any difference among the tests, being relatively small for rail freight, as Figure 8.8 shows. The graph on the left corresponds to the volume share, while the chart on the right is for the import share. Experiments give a train share between 4% and 6% in volume and only a 3% in import. Both percentages are coherent with the observed data.

Figure 8.9 represents GHG emissions for the whole transport of the study. The GHG

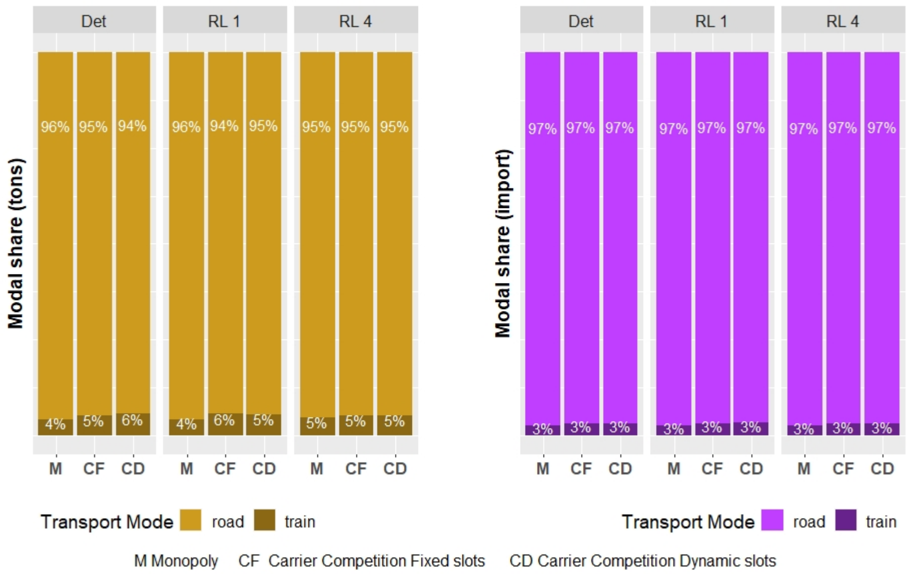


Figure 8.8: MINLP tests. Modal share train vs road. Volume and Import, different robustness levels. Source: author.

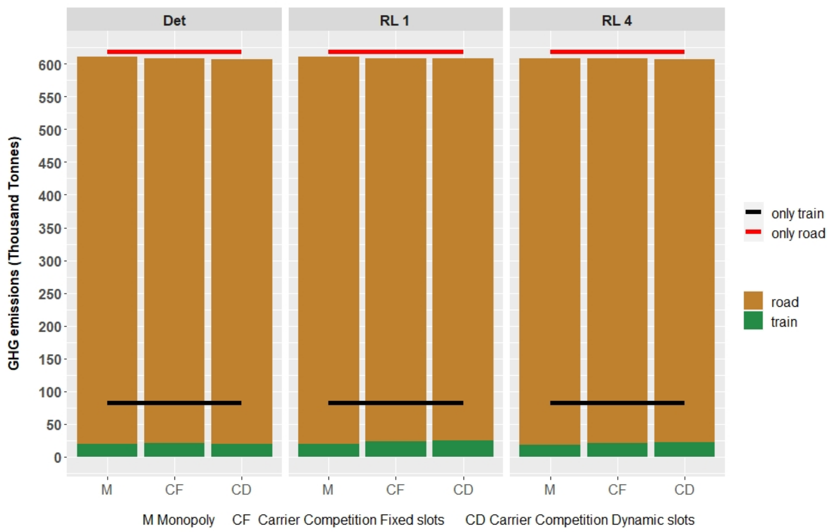


Figure 8.9: MINLP tests. GHG emissions by mode. Different robustness levels. Source: author.

emissions data, both for rail and road, have been obtained from the results provided by the tool **EcoTransIT World** (see Section 2.7). The Well-to-Wheel total consumption for road and train has been assigned to each path connecting the OD pairs after applying the parameters detailed in Section 7.2.5. Brown colour corresponds to road emissions, while green is for rail emissions. The red line would mean GHG emissions if the entire transport was done by truck, while the black line represents GHG emissions if the total transport was done by rail. Comparing both lines and the size of bars, which represent GHG emissions according to the tests executed, it is clear that there is still much work to do to reduce GHG emissions in freight transportation.

Chapter 9

Conclusions

One goal being pursued by the governments of many countries – especially in Europe – is to increase the use of rail networks for freight transport in order to reduce road congestion and pollution. This objective may come up against the complicated trade-off between limiting costly infrastructure investments and minimising operating costs to make rail transport attractive for businesses. In addition, the growth of rail freight transport has to be done above all at the expense of road transport. Then, it is necessary to make it more appealing for those who operate it and those who use it, carriers and shippers. In this thesis, two independent and complementary rail freight transport models have been developed and tested motivated by the interest in helping to boost the use of rail for the transport of goods.

The first, named **DCECC**-model, is a mathematical programming-based design model to evaluate the impact of rail infrastructure improvements and capacity expansion, specifically when applied to a mixed rail network. The problem presented is a multi-objective minimisation problem, where each of the objectives corresponds to: a) cost related to infrastructure investments and maintenance, and b) operating costs. The main decision variables of the model are related to the construction of new tracks or upgrading the existent ones (double/single), the parameters of the blocking system on the tracks, and the capacity of terminals/yards. The problem relies on three blocks of constraints. First, balance equations for products and rolling stock define one set of constraints. Second, a set of hard capacity constraints are defined, given that a rail network is usually of shared use by both passengers and freight. In this case, they are based on blocking distance criteria, following the recommendations of UIC (2013) and the approach by Rotoli et al. (2016). The third block evaluates the necessary rolling stock, i.e., how railcars (full and empty) run throughout the network and how many of them are needed to satisfy the transportation demand. An approximate Pareto efficiency analysis was applied to determine a trade-off solution between infrastructure costs and operating costs.

The second, named **MINLP**-model, is a combined modal-split/traffic assignment model for rail and road freight transport, with detailed modelling of the various railway traffic flows (e.g., railcars either full or empty, volumes of newly formed trains at yards...) for multi-operator scenarios where the modal split has to be applied. The model is formulated as a non-linear integer optimisation problem following the classical relative entropy function maximisation. The model accounts for the large variability that the utility coefficients may have for reasons such as difficulties in the data collection and the predominant role of the road mode of transport. To this end, a robust counterpart of the model is formulated to consider more conservative modal splits under a limited worst-case standpoint. An algorithm based on the outer approximation method is developed to provide accurate solutions in a reasonable computational time for both the robust and non-robust models.

The **DCECC**-model can be a helpful tool for analysing the impact infrastructure investments may have on operating costs, where capacity limitations in the scenarios to be evaluated may necessarily be taken into account. Thus, this model may help assessing rail transport policies to improve rail infrastructure under a criterion of utility for those who will use it. At the same time, the **MINLP**-model may complement the **DCECC**-model by evaluating under the proposed scenarios the possible response from shippers to the different services offered by railway carriers competing with each other and the road.

Examples centred on a section of the Trans-European Transport Network, the **TEN-T** Core network corridors, are reported to test the applicability of the models. Aggregate data from the official statistics web page for Spanish Foreign trade is the base for defining the demand of products. At the same time, railway carriers' services are built from public information obtained from different companies which operate in Europe. A logit approximation is applied to estimate the components of the utility function for both rail and road transportation.

Results prove the applicability of both models. The approximate Pareto frontier graphically shows the relationship between infrastructure investments and operating cost after running two groups of tests for validating the **DCECC**-model, one based on current demand information and the other after applying a 15% increase in demand. While infrastructure investments are quite similar in both groups of tests, operating costs are clearly influenced by the rise in demand. Test cases reveal a clear tendency: increasing infrastructure investments help reduce operating costs, but lessening infrastructure investments increases operation costs dramatically.

Results from the **MINLP**-model show the different behaviour on the railway carrier share when competence is fully open, and none of the carriers has received a previous allocation, or on the contrary, when infrastructure managers have previously fixed the slots to each carrier. In the first case, one or few carriers likely gather the whole train transport easily, while in the second case, limits imposed by the allocation cause the train transport to be better distributed. Also, after evaluating in the **MINLP**-tests the pollution that the transportation of goods causes, it is clear there is still much work to be done to reduce GHG emissions in freight transport.

For the **MINLP**-model, we have analysed the modal split compliance and the influence that some constraints exert on the way that railway carriers transport goods. In our particular examples, these constraints are related to infrastructure capacity, rolling stock limits and cost-effectiveness for carriers. It has been seen that when none of these constraints was imposed on the model, the resultant modal share (the modal share calculated from the solution) exactly fits with the theoretical modal share (the modal share that comes from applying the previously evaluated utilities u_r^ω and $\tilde{u}^\omega$). However, including any of those constraints causes a difference between the resultant and the theoretical modal share. We interpret here this difference as a correction for the utilities u_r^ω because of the effect that these constraints associated with flows exert on shipper decisions could not be appropriately taken into account when estimating the random utility model.

9.1. Future Research

The high cost of expanding the rail network in order to separate rail passenger traffic from rail freight traffic forces both types of traffic to share the rail network partially or even totally. However, the impact of different train speeds on rail network capacity is well known. In the models presented in this thesis, the effects on the passenger network are only considered by limiting some parameter values in the constraints, e.g. capacities or number of slots. A better approximation of passenger flows and how freight and passenger trains share the rail network may be a potential source of future contributions to the design model.

On the other hand, the difficulties in solving some problem instances of the **DCECC**-model show the necessity of defining an algorithm to solve the problem with less computational resources and eventually provide more accurate solutions.

The combined modal split model can be the subject to future extensions, such as including more modes (e.g., inland waterways transport) in a multimodal framework. This aspect is linked to the discrete choice models that are suitable to reflect the decisions of shippers in freight transport systems and on which there are practically no results in the scientific literature. Different formulations and algorithmic solutions could result depending on the type of choice model (e.g. multinomial logit, nested logit,...). Also, the traffic assignment component of the model can be extended by including factors that gradually consider the congestion of the systems and not simply by using capacity limits. In this case, the resulting model would originate some non-linearities that should be treated using non-linear programming methods or methods for variational inequalities.

Bibliography

References

Abate, M., Vierth, I., Karlsson, R., de Jong, G., & Baak, J. (2019). A disaggregate stochastic freight transport model for Sweden. *Transportation*, *46*(3), 671–696. https://doi.org/10.1007/s11116-018-9856-9

Abdulaal, M., & Leblanc, L. J. (1979). Methods for Combining Modal Split and Equilibrium Assignment Models. *Transportation Science*, *13*(4), 292–314. https://doi.org/10.1287/trsc.13.4.292

Abril, M., Barber, F., Ingolotti, L. P., Salido, M. A., Tormos, P., & Lova, A. (2008). An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, *44*(5), 774–806. https://doi.org/10.1016/j.tre.2007.04.001

Antunes, C. H., Alves, M. J., & Clímaco, J. (2016). *Multiobjective Linear and Integer Programming*. Berlin:Springer.

Arencibia, A. I., Feo-Valero, M., García-Menéndez, L., & Román, C. (2015). Modelling mode choice for freight transport using advanced choice experiments. *Transportation Research Part A: Policy and Practice*, *75*, 252–267. https://doi.org/10.1016/j.tra.2015.03.027

Assad, A. A. (1980). Models for rail transportation. *Transportation Research Part A: General*, *14*, 205–220. https://doi.org/10.1016/0191-2607(80)90017-5

Attina, M., Basilico, A., Botta, M., Brancatello, I., Gargani, F., Gori, V., Wilhelm, F., Menting, M., Odoardi, P., Piperno, A., & Ranier, M. (2018). *Assessment of unit costs (standard prices) of rail projects (CAPital EXpenditure)* (tech. rep. No. 1). European Comission. Directorate-General for Regional and Urban Policy. Retrieved December 1, 2021, from https://ec.europa.eu/regional_policy/en/

information/publications/reports/2018/assessment-of-unit-costs-standard-prices-of-rail-projects-capital-expenditure

Bar-Gera, H. (2002). Origin-based algorithm for the traffic assignment problem. *Transportation Science*, *36*(4), 398–417. https://doi.org/10.1287/trsc.36.4.398.549

Baumgartner, J. P. (2001). *Prices and costs in the railway sector* (tech. rep. January). Laboratoire d'Intermodalité des Transports Et de Planification. Laussane. Retrieved December 1, 2021, from http://archiveweb.epfl.ch/litep.epfl.ch/files/content/sites/litep/files/shared/Liens/Downloads/Divers/Baumgartner_Couts_chf_2001_e.pdf

Bešinović, N., & Goverde, R. M. P. (2018). Capacity Assessment in Railway Networks. In R. Borndörfer, T. Klug, L. Lamorgese, C. Mannino, M. Reuther, & T. Schlechte (Eds.), *Handbook of optimization in the railway industry. international series in operations research & management science, vol 268.* (pp. 25–45). Springer, Cham. https://doi.org/10.1007/978-3-319-72153-8_2

Boyce, D., Zhang, Y.-F., & Lupa, M. R. (1994). Introducing "feedback" into four-step travel forecasting procedure versus equilibrium solution of combined model. *Transportation Research Record*, *1443*, 65.

Boysen, N., Fliedner, M., Jaehn, F., & Pesch, E. (2012). Shunting yard operations: Theoretical aspects and applications. *European Journal of Operational Research*, *220*(1), 1–14. https://doi.org/10.1016/j.ejor.2012.01.043

Burdett, R. L. (2015). Multi-objective models and techniques for analysing the absolute capacity of railway networks. *European Journal of Operational Research*, *245*(2), 489–505. https://doi.org/10.1016/j.ejor.2015.03.020

Burdett, R. L. (2016). Optimisation models for expanding a railway's theoretical capacity. *European Journal of Operational Research*, *251*(3), 783–797. https://doi.org/10.1016/j.ejor.2015.12.033

Cacchiani, V., Caprara, A., & Toth, P. (2010). Scheduling extra freight trains on railway networks. *Transportation Research Part B: Methodological*, *44*(2), 215–231. https://doi.org/10.1016/j.trb.2009.07.007

CEN-European Committee for Standardization. (2012). *EN 16258:2012. Methodology for Calculation and Declaration of Energy Consumption and GHG Emissions of Transport Services (Freight and Passengers)*. European Comission.

Crainic, T., Florian, M., & Léal, J.-E. (1990). A Model for the Strategic Planning of National Freight Transportation by Rail. *Transportation Science*, *24*(1), 1–24. https://doi.org/10.1287/trsc.24.1.1

Crainic, T., & Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 409–438. https://doi.org/10.1016/S0377-2217(96)00298-6

Crisalli, U., Comi, A., & Rosati, L. (2013). A Methodology for the Assessment of Rail-road Freight Transport Policies. *Procedia - Social and Behavioral Sciences*, *87*, 292–305. https://doi.org/10.1016/j.sbspro.2013.10.611

Croissant, Y. (2020). *Mlogit: Multinomial logit models* [R package version 1.1-0]. https://CRAN.R-project.org/package=mlogit

Dafermos, S. C. (1976). Integrated Equilibrium Flow Models for Transportation Planning. In M. A. Florian (Ed.), *Traffic equilibrium methods* (pp. 106–118). Springer Berlin Heidelberg.

DB Cargo AG. (2019). Standard Rates and other Provisions of DB Cargo AG. Retrieved November 6, 2019, from https://www.dbcargo.com/rail-deutschland-en/our-offer/prices-services-1695170

de Dios Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*. John wiley & sons.

de Jong, G., Tavasszy, L., Bates, J., Grønland, S. E., Huber, S., Kleven, O., Lange, P., Ottemöller, O., & Schmorak, N. (2015). The issues in modelling freight transport at the national level. *Case Studies on Transport Policy*, *4*(1), 13–21. https://doi.org/10.1016/j.cstp.2015.08.002

de Jong, G., Vierth, I., Tavasszy, L., & Ben-Akiva, M. (2012). Recent developments in national and international freight transport models within Europe. *Transportation*, *40*(2), 347–371. https://doi.org/10.1007/s11116-012-9422-9

Deutsche Bahn AG. (2020). Network statement. Retrieved March 1, 2020, from https://fahrweg.dbnetze.com/fahrweg-en/customers/network_statement

División de Estudios y Tecnología del Transporte de la Secretaría General de Transporte. (2020). *Observatorio del Transporte y la Logística en España. Informe Anual. 2019* (tech. rep.). Ministerio de Transportes, Movilidad y Agenda Urbana. Retrieved February 1, 2020, from https://observatoriotransporte.mitma.es/

Duran, M. A., & Grossmann, I. E. (1986). An Outer-Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs. *Marhematical Programming*, *36*, 307–339. https://doi.org/10.1007/BF02592064

European Commission. (2011). *White paper on transport : Roadmap to a single european transport area : Towards a competitive and resource efficient transport system*. Publications Office. https://doi.org/10.2832/30955

European Commission. (2019). *Handbook on the External Costs of Transport*. Retrieved December 1, 2021, from https://www.cedelft.eu/en/publications/2311/handbook-on-the-external-costs-of-transport-version-2019

European Commission. (2021a). *EU Transports in Figures*. European Union. https://doi.org/10.2832/27610

European Court of Auditors. (2016). *Rail freight transport in the EU: still not on the right track* (tech. rep. No. 08). European Court of Auditors. https://doi.org/10.2865/53961

Evans, S. P. (1976). Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research*, *10*, 37–57.

Feo-Valero, M., García-Menéndez, L., & del Saz-Salazar, S. (2016). Rail freight transport and demand requirements: An analysis of attribute cut-offs through a stated preference experiment. *Transportation*, *43*, 101–122. https://doi.org/10.1007/s11116-014-9566-x

Fernández L., J. E., de Cea, J., & Giesen E., R. (2004). A strategic model of freight operations for rail transportation systems. *Transportation Planning and Technology*, *27*(4), 231–260. https://doi.org/10.1080/0308106042000228743

Florian, M. (1977). Traffic Equilibrium Model of Travel By Car and Public Transit Modes. *Transportation Science*, *11*(2), 166–179. https://doi.org/10.1287/trsc.11.2.166

Florian, M., & Nguyen, S. (1978). A combined trip distribution modal split and trip assignment model. *Transportation Research*, *12*(4), 241–246. https://doi.org/10.1016/0041-1647(78)90065-5

Floudas, C. A. (1995). Mixed-Integer Nonlinear Optimization. In *Nonlinear and mixed-integer optimization: Fundamentals and applications* (p. 480). Oxford University Press, Inc. https://doi.org/10.1093/oso/9780195100563.001.0001

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval research logistics quarterly*, *3*(1-2), 95–110. https://doi.org/10.1002/nav.3800030109

Friesz, T. L., & Kwon, C. (2007). Strategic Freight Network Planning Models and Dynamic Oligopolistic Urban Freight Networks. In H. D.A. & K. Button (Eds.), *Handbook of transport modelling (vol 1)* (pp. 611–631). Emerald Group Publishing Limited. https://doi.org/10.1108/9780857245670-032

Gorman, M. F. (2015). Empty Railcar Distribution. In Bruce W. Patty (Ed.), *Handbook of operations research applications at railroads* (pp. 177–189). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7571-3_7

Guelat, J., Florian, M., & Crainic, T. (1990). A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Transportation Science*, *24*(1), 25–39. https://doi.org/10.1287/trsc.24.1.25

Guinot, V. R. (2008). *Costes del transporte de mercancías por ferrocarril* (tech. rep.). Fundación de los Ferrocarriles Españoles. Retrieved September 15, 2018, from https://www.vialibre-ffe.com/monografias_tecnicas.asp

Harrod, S. (2009). Capacity factors of a mixed speed railway network. *Transportation Research Part E: Logistics and Transportation Review*, *45*(5), 830–841. https://doi.org/10.1016/j.tre.2009.03.004

Hearn, D. W., Lawphongpanich, S., & Ventura, J. (1987). Restricted simplicial decomposition: Computation and extensions. In K. L. Hoffman, R. H. F. Jackson, & J. Telgen (Eds.), *Computation mathematical programming* (pp. 99–118). Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0121181

Hou, B., Zhao, S., & Liu, H. (2020). A Combined Modal Split and Traffic Assignment Model with Capacity Constraints for Siting Remote Park-and-Ride Facilities. *IEEE Access*, *8*, 80502–80517. https://doi.org/10.1109/ACCESS.2020.2991165

Institut Cerdà. (2019). *Observatori de costos del transport de mercaderies per ferrocarril - 1a Edició* (tech. rep.). Institut Cerdà. Retrieved March 18, 2020, from https://cimalsa.cat/observatori-ferrocaril.php

ITF. (2021). *ITF Transport Outlook 2021*. OECD Publishing. https://doi.org/10.1787/16826a30-en

IVE, INFRAS, ifeu, & Fraunhofer. (2021). *EcoTransIT World - Ecological Transport Information Tool for Worldwide Transports. Methodology and Data Update* (tech. rep.). Commisioned by EcoTransIT World Initiative (EWI). Berne - Dortmund - Hannover - Heidelberg.

Jamili, A. (2018). Computation of practical capacity in single-track railway lines based on computing the minimum buffer times. *Journal of Rail Transport Planning and Management*, *8*(2), 91–102. https://doi.org/10.1016/j.jrtpm.2018.03.002

Jensen, A. F., Thorhauge, M., de Jong, G., Rich, J., Dekker, T., Johnson, D., Cabral, M. O., Bates, J., & Nielsen, O. A. (2019). A disaggregate freight transport chain choice model for Europe. *Transportation Research Part E: Logistics and Transportation Review*, *121*, 43–62. https://doi.org/10.1016/j.tre.2018.10.004

Joborn, M., Crainic, T. G., Gendreau, M., Holmberg, K., & Lundgren, J. T. (2004). Economies of scale in empty freight car distribution in scheduled railways. *Transportation Science*, *38*, 121–134. https://doi.org/10.1287/trsc.1030.0061

Jourquin, B. (2006). A Multi-flow Multi-modal Assignment Procedure on Large Freight Transportation Networks. *Studies in Regional Science*, *35*, 929–945. https://doi.org/10.2457/srs.35.929

Jourquin, B. (2016). Calibration and Validation of Strategic Freight Transportation Planning Models with Limited Information. *Journal of Transportation Technologies*, *06*(05), 239–256. https://doi.org/10.4236/jtts.2016.65023

Koster, A. M., Kutschka, M., & Raack, C. (2013). Robust Network Design: Formulations, Valid Inequalities, and Computations. *Networks*, *61*(2), 128–149. https://doi.org/10.1002/net

Larsson, T., & Patriksson, M. (1994). Equilibrium characterizations of solutions to side constrained asymmetric traffic assignment models. *Le Matematiche*, *49*, 249–280. https://lematematiche.dmi.unict.it/index.php/lematematiche/article/view/527

Li, S., Deng, W., & Lv, Y. (2009). Combined modal split and assignment model for the multimodal transportation network of the economic circle in China. *Transport*, *24*(3), 241–248. https://doi.org/10.3846/1648-4142.2009.24.241-248

Little, J. D. (1961). A Proof for the Queuing Formula: L = λW. *Operations Research*, *9*(3), 383–387. https://doi.org/10.1287/opre.9.3.383

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

Maia, L. C., & do Couto, A. F. (2012). An innovative freight traffic assignment model for multimodal networks. *Computers in Industry*, *64*(2), 121–127. https://doi.org/10.1016/j.compind.2012.10.011

Maia, L. C., & do Couto, A. F. (2014). A Rail Network Optimization Model Designed To Model Freight. In J. F. de Sousa & R. Rossi (Eds.), *Computer-based modelling and optimization in transportation* (pp. 209–222). Springer International Publishing Switzerland 2014 209. https://doi.org/10.1007/978-3-319-04630-3_16

Martínez, E., Ampuero, V., & Pascual, C. (2015). *Análisis de oportunidades del transporte ferroviario de productos cárnicos entre Cataluña y Europa* (tech. rep.). from Project IFreightMed-DC. Retrieved January 26, 2018, from http://www.csmmed.eu/wp-content/uploads/2019/04/00952_Informe_Ed2b.pdf

Meersman, H., Ehrler, V. C., Bruckmann, D., Chen, M., Francke, J., Hill, P., Jackson, C., Klauenberg, J., Kurowski, M., Seidel, S., & Vierth, I. (2016). Challenges and future research needs towards international freight transport modelling. *Case Studies on Transport Policy*, *4*(1), 3–8. https://doi.org/10.1016/j.cstp.2015.12.002

Meirich, C., & Nießen, N. (2016). Calculating the maximal number of additional freight trains in a railway network. *Journal of Rail Transport Planning and Management*, *6*(3), 200–217. https://doi.org/10.1016/j.jrtpm.2016.06.005

Ministerio de Fomento. (2019). *Observatorio del Ferrocarril en España. Informe 2018* (tech. rep.). Retrieved March 18, 2020, from https://www.mitma.gob.es/ferrocarriles/observatorios/observatorio-del-ferrocarril-en-espana

Mitradjieva, M., & Lindberg, P. O. (2013). The stiff is moving—conjugate direction frank-wolfe methods with applications to traffic assignment. *Transportation Science*, *47*(2), 280–293. https://doi.org/10.1287/trsc.1120.0409

Mussone, L., & Wolfler Calvo, R. (2013). An analytical approach to calculate the capacity of a railway system. *European Journal of Operational Research*, *228*(1), 11–23. https://doi.org/10.1016/j.ejor.2012.12.027

Nuzzolo, A., Coppola, P., & Comi, A. (2013). Freight transport modeling: Review and future challenges. *International Journal of Transport Economics / Rivista Internazionale di economia dei transporti*, *40*, 151–181.

Pachl, J. (2015). *Railway operation and control* (4th). VTD Rail Publ.

Pérez, P. (2015). *Pla de negoci d'un servei multimodal orientat a la indústria de l'automoció entre Catalunya i el centre d'Europa* (tech. rep.). from Project IFreightMed-DC. Retrieved January 26, 2018, from http://www.csmmed.eu/wp-content/uploads/2019/04/ANNEXES-BP-Automocio.pdf

Petersen, E. R. (1977a). Railyard Modeling: Part I. Prediction of Put-Through Time. *Transportation Science*, *11*, 37–49.

Petersen, E. R. (1977b). Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion. *Transportation Science*, *11*, 50–59.

Rail Freight Forward. (2019). *Rail Freight strategy to boost modal shift* (tech. rep.). Retrieved November 7, 2019, from https://www.railfreightforward.eu/

RailTech. (2021). Track acces charges in Europe. Retrieved December 1, 2021, from https://events.railtech.com/track-access-charges-summit-2022/infographic/

Railway Group Standard. (2015). *GK/RT0075 Line side Signal Spacing and Speed Signage* (tech. rep.). Retrieved November 6, 2018, from https://www.rssb.co.uk/standards-catalogue/catalogueitem/gkrt0075%20iss%203

Rosell, F., & Codina, E. (2020). A model that assesses proposals for infrastructure improvement and capacity expansion on a mixed railway network. *Transportation Research Procedia*, *47*, 441–448. https://doi.org/10.1016/j.trpro.2020.03.119

Rosell, F., Codina, E., & Montero, L. (2022). A combined and robust modal-split/traffic assignment model for rail and road freight transport. *European Journal of Operational Research*, *303*(2), 688–698. https://doi.org/10.1016/j.ejor.2022.03.008

Rosell, F., Ubalde, L., & Saurí, S. (2018). *Potenciació del transport multimodal al corredor del mediterrani des de la implementació de models multiclient i multiproducte eficients* (D. d'Empresa i Coneixement, Ed.; Primera Ed). Generalitat de Catalunya - Departament d'Empresa i Coneixement.

Rotoli, F., Malavasi, G., & Ricci, S. (2016). Complex railway systems: capacity and utilisation of interconnected networks. *European Transport Research Review*, *8*(4). https://doi.org/10.1007/s12544-016-0216-6

Schroten, A., van Wijngaarden, L., Brambilla, M., Gatto, M., Maffii, S., Trosky, F., Kramer, H., Reinhard, M., Bertschmann, D., Killer, M., Lambla, V., El Beyrouty, K., & Amaral, S. (2019). *Overview of transport infrastructure expenditures and costs* (European Commission, Ed.). https://doi.org/10.2832/853267

Sheffi, Y. (1985). *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. Prentice-Hall.

Shi, T., & Zhou, X. (2015). A mixed integer programming model for optimizing multi-level operations process in railroad yards. *Transportation Research Part B: Methodological*, *80*(Supplement C), 19–39. https://doi.org/10.1016/j.trb.2015.06.007

Trabo, I., Landex, A., Otto, A. N., & Schneider-Tilli, 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.*

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Turnquist, M. A., & Daskin, M. S. (1982). Queuing models of classification and connection delay in railyards. *Transportation Science*, *16*, 207–230. https://www.jstor.org/stable/25768049

UIC. (2008). Influence of ETCS on line capacity. Generic Study.

UIC. (2013). Leaflet 406. Capacity, 2nd. Edition.

Web pages

ADIF. (2020). *Declaración sobre la Red*. Retrieved May 15, 2020, from http://www.adif.es/es_ES/conoceradif/declaracion_de_la_red.shtml

Community of European Railway and Infrastructure Companies (CER). (2018). *CER. The Voice of European Railways*. Retrieved November 3, 2021, from http://www.cer.be/

EcoTransIT World. (2021). Retrieved September 5, 2020, from https://www.ecotransit.org

European Commission. (2021b). *Transport sector economic analysis*. Retrieved December 14, 2021, from https://ec.europa.eu/jrc/en/research-topic/transport-sector-economic-analysis

Eurostat. (2021). *Freight transport statistics - modal split*. Retrieved December 13, 2021, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split#Inland_freight_transport_performance_-_the_need_to_adjust_road_transport

FERRMED. (2020). Retrieved December 17, 2020, from http://www.ferrmed.com/

FR8RAIL-IV. (2020). *Use-centric rail freight innovation for Single European Railway Area: Research project granted by the Shift2Rail Joint Undertaking of the European Commission*. Retrieved November 3, 2021, from https://cordis.europa.eu/project/id/101004051

iFreightMED-DC. (2014). *Comitès Desenvolupament Serveis Transport Intermodal: Research project granted by the 2007-2013 Programme MED*. Retrieved October 25, 2021, from https://www.up2europe.eu/european/projects/intermodal-freight-services-development-committees_138833.html

MARATHON. (2019). *Make Rail The Hope for protecting Nature (MARATHON): Research project granted by the 7th Framework Programme - Transport*. Retrieved November 3, 2021, from https://cordis.europa.eu/project/id/265647

Ministerio de Industria, Comercio y Turismo. (2020). *Datacomex*. Retrieved June 1, 2020, from https://comercio.serviciosmin.gob.es/Datacomex/

MOVINGRAIL. (2019). *MOVINGRAIL: Moving Block and Virtual Coupling Next Generations of Rail Signalling: Research project granted by the Shift2Rail Joint Undertaking of the European Commission*. Retrieved November 5, 2021, from https://projects.shift2rail.org/s2r_ip2_n.aspx?p=MOVINGRAIL

ProMedia Group. (2021). *RailFreight.com*. Retrieved November 8, 2021, from https://www.railfreight.com/

RailNetEurope. (n.d.). *Rail Facilities Portal*. Retrieved October 25, 2021, from https://railfacilitiesportal.eu/

RailNetEurope. (2020). Retrieved October 25, 2021, from https://rne.eu/

RailNetEurope. (2021). *Path Coordination System*. Retrieved October 25, 2021, from http://pcs.rne.eu/

Shift2Rail. (2021). Retrieved October 25, 2021, from https://shift2rail.org/

SMART-RAIL. (2021). *SMART-RAIL: smart supply chain oriented rail freight services: Research project granted by the Shift2Rail Joint Undertaking of the European Commission*. Retrieved November 3, 2021, from https://smartrail-project.eu/

SNCF Réseau. (2020). *National rail network statement*. Retrieved March 1, 2020, from https://www.sncf-reseau.com/en/national-rail-network-statement

TPNOVA Rail & Logistics Services S.L. (2020). *TPNOVA* [Accessed 3 june 2020]. https://www.tpnova.com/

TRAILS. (2019). *Transnational Intermodal Links toward Sustainability: Research project granted by the 2014-2020 Programa INTERREG V-A España-Francia-Andorra*. Retrieved October 25, 2021, from https://www.trails-project.com/

Unife. (2021). *ERTMS*. Retrieved October 26, 2021, from https://www.ertms.net/

United Nations Economic Commission for Europe (UNECE). (n.d.). *Transports Statistics*. Retrieved June 3, 2020, from http://www.unece.org/trans/main/wp6/wp6.html

University of Birmingham and Network Rail. (2011). *The Railway Technical Website. Railway Lexicon*. Retrieved February 14, 2022, from http://www.railway-technical.com/glossary/rail-lexicon-mk24.pdf

U.S. Department of Transportation, Bureau of Transportation Statistics. (2021). *Freight Transportation & the Economy*. Retrieved December 14, 2021, from https://data.bts.gov/stories/s/Freight-Transportation-the-Economy/6ix2-c8dn

segment placeholder

Appendix A

Sets, parameters and variables summary

Table A.1 shows the sets and subsets required for **DCECC**-model and **MINLP**-model. Table A.2 shows the list of the parameters that appear in the constraints and the objective functions. Table A.3 summarizes the list of variables used in the models.

Table A.1: List of sets and subsets

A	directed arcs	K_M^H	faster freight locomotive types
K_M	freight locomotive types	$P^H(v)$	priority products compatible with v-type railcars
N	nodes	$R(\rho)$	t-paths which contains line ρ
$\mathcal{O}$	carriers	$R(\omega)$	t-paths from origin $\mathfrak{o}(\omega)$ to destination $\mathfrak{d}(\omega)$
P	products	$R(\omega, o)$	carrier $o \in \mathcal{O}$ t-paths for OD-pair in ω
T	terminals	$R(\omega, \rho)$	t-paths from origin $\mathfrak{o}(\omega)$ to destination $\mathfrak{d}(\omega)$
$\mathcal{V}$	railcars		containing line ρ
W	OD-pairs and product	$\mathcal{V}(\rho)$	railcars compatibles with ρ-line
Y	yards	$Y(o)$	yards where carrier $o \in \mathcal{O}$ operates
Γ	paths between yards	$\Gamma(o)$	lines of carrier $o \in \mathcal{O}$
A^-	$\{a \in A : a = (i,j), i < j\}$	$\Gamma(a, o)$	lines of carrier $o \in \mathcal{O}$ which contain the arc a
A_s	subset of single-tracks	$\Gamma(v, o)$	lines of carrier $o \in \mathcal{O}$ compatible with v-type railcars
	$\Gamma_i^+(o)/\Gamma_i^-(o)(\Gamma_i^+(v,o)$ / $\Gamma_i^-(v,o))$		lines of carrier $o \in \mathcal{O}$ outgoing from/incident to i-yard (also compatible with v-type railcars)

Table A.2: List of parameters. Model 1 = **DCECC**-model, Model 2 = **MINLP**-model

parameter	model	description
$\bar{B}_i$	1	- yard daily-average opening hours (for instance, 12 hours)
$\hat{B}_i^j$	1	- levels of yard congestion (in number of trains handled)
$C_i^{k,o}/\tilde{C}_i^{k,o}$	2	- cost for train formation/decomposition at yard i.
$\hat{C}_\rho^k$	2	- travel cost of k-train when runs on path ρ.
$\tilde{D}_\rho$	1	- penalty for priority products on non-priority trains on y-path $\rho \in \Gamma$
$D^{v,o}$	2	- o-carrier cost for renting and/or maintenance of v-railcars.
$L^{v,o}, \tilde{L}^{k,o}$	1,2	- maximum number of railcars of type v and locomotive of type k that carrier o may dispose, respectively
ℓ_a	1	- length of arc $a \in A$
ℓ_a^{max}	1	- maximum train length allowed on arc $a \in A$
ℓ_a^{min}	1	- lower bound for the block signal distance d_a on arc $a \in A$
ℓ^v	1,2	- length of v-type railcar.
$\bar{\ell}_\rho$	1,2	- maximum train length allowed on line ρ.
$N_\rho, N_a, \tilde{N}_i$	2	- maximum train capacity per line ρ, arc a and yard i, respectively.
$\dot{S}_a/\ddot{S}_a$	1	- build or maintenance costs for tracks (fixed/extra if double) on arc $a \in A$
$\hat{S}_a$	1	- maintenance cost per blocking signal on link $a \in A$.
$S_i/\bar{S}_i$	1	- build or maintenance costs for yard $i \in Y$ (fixed/per track)
$\mathcal{T}$	1,2	- period of study (for instance, a year)
$\bar{\mathcal{T}}$	1,2	- effective time rolling stock runs during the period of study $\mathcal{T}$ (for instance, total working days for a year)
t_ρ	1,2	- average run time for train in line ρ
U_r^ω	2	- price per $\mathfrak{p}(\omega)$-unit paid by shipper when transported by train from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$.
u_r^ω	2	- generalized cost for OD-pair and unit of product ω for train and t-path r, $u_r^\omega = \beta_0^\omega + \sum_{j=i}^m \beta_j^\omega u_{r,j}^\omega$
$\tilde{u}^\omega$	2	- generalized cost for OD-pair and unit of product ω for truck, $\tilde{u}^\omega = \tilde{\beta}_0^\omega + \sum_{j=i}^{\tilde{m}} \tilde{\beta}_j^\omega \tilde{u}_j^\omega$.
$\alpha^{v,\mathfrak{p}(\omega)}$	1,2	- average weight per unit of product $\mathfrak{p}(\omega)$ on railcar of type v
α^v	1,2	- tare of v-type railcar.
$\bar{\alpha}_\rho^k$	1,2	- maximum weight allowed for k-trains on line ρ.
σ^k	1	- average speed for k-type trains
χ^ω	1,2	- total demand for $\mathfrak{p}(\omega)$-product and origin-destination pair $(\mathfrak{o}(\omega), \mathfrak{d}(\omega))$.
γ	1	- efficient points for the Pareto efficiency analysis

Table A.3: List of variables. All of them are non-negative. B: binary - I: integer - C: continuous. Model 1 = **DCECC**-model, Model 2 = **MINLP**-model

variable	type	model	description
b_a	I	1	- total number of block signals on arc $a \in A$.
$f_\rho^{v,\omega}$	I	1,2	- total number of railcars of type v that transport product $\mathfrak{p}(\omega)$ from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$ using $\rho \in \Gamma$.
$f_\rho^{v,\emptyset}$	I	1,2	- total number of empty railcars of type $v \in \mathcal{V}$ running on $\rho \in \Gamma$.
F_ρ^v	I	1,2	- total number of railcars empty and loaded of type $v \in \mathcal{V}$ running on $\rho \in \Gamma$. $F_\rho^v = f_\rho^{v,} + \sum_{\omega \in W} f_\rho^{v,\omega}$
g_i	B	1	- yard $i \in Y$ is built or not.
$\bar{g}_i$	I	1	- size of yard $i \in Y$ (number of tracks).
h_r^ω	C	1,2	- total tons of product $\mathfrak{p}(\omega)$ transported from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$ by train on t-path r.
$\tilde{h}^\omega$	C	2	- total tons of product $\mathfrak{p}(\omega)$ transported from $\mathfrak{o}(\omega)$ to $\mathfrak{d}(\omega)$ by road.
h_ρ^L	C	1	- total tons of priority products transported on $\rho \in \Gamma$ by non-priority trains.
m_ρ^k	I	1,2	- total number of locomotives of type $k \in K_M$ that run on $\rho \in \Gamma$.
n_a^k	I	1	- total number of trains of type $k \in K_M$ that run on track $a \in A$. $n_a^k = \sum_{\rho \in \Gamma(a)} m_\rho^k$.
r_a	B	1	- double track (1) or single track (0) on link $a \in A$.
$\hat{y}_o$	B	2	- binary variable to avoid infeasibility due to relationship between price and cost
z_a	B	1	- track $a \in A$ is built or not.
$\theta_{i,j}^{k,o}$	I	1,2	- total number of locomotives of type k, owned by carrier o, that run from yard i to yard j.
$\lambda^{v,o}$	I	1,2	- minimum number of railcars of type v needed for carrier o to provide the service.
$\phi_{a,i}^k, x_i^j$	B	1	- auxiliary binary variables for the linearisation of the **DCECC**-model
$\zeta_{a,i}^k, \bar{x}_{i,\rho}^{j,v}, \hat{x}_{i,\rho}^{j,k}$	C	1	- auxiliary continuous variables for the linearisation of the **DCECC**-model
$\pi_r^\omega, p_{r,j}^\omega, \tilde{\pi}^\omega, \tilde{p}_j^\omega$	C	2	- auxiliary variables for the robust version of the **MINLP**-model.

Appendix B

Explicit formulation of the models

This appendix shows in full detail the exhaustive and explicit formulation of the different versions of the models DCECC, DCEM and MINLP.

B.1. DCECC-model with non-linear constraints

$$
\min \quad \frac{\gamma}{\hat{f}_0 - \underline{f}_0} \left[\sum_{i \in N} \left(S_i \cdot g_i + \bar{S}_i \cdot \bar{g}_i \right) + \sum_{a \in A^{\cdot}} \left(\dot{S}_a \cdot z_a + \ddot{S}_a \cdot r_a \right) + \sum_{a \in A} \hat{S}_a \cdot b_a \cdot \right] +
$$

$$
\frac{1-\gamma}{\hat{f}_1 - \underline{f}_1} \left[\sum_{\substack{k \in K_M, \\ o \in O}} \left(\sum_{\substack{i,j \in Y(o) \\ i \neq j}} (C_i^{k,o} + \tilde{C}_j^{k,o}) \theta_{i,j}^{k,o} + \sum_{\rho \in \Gamma(o)} \hat{C}_\rho^k m_\rho^k \right) + \sum_{\substack{v \in \mathcal{V}, \\ o \in O}} D^{v,o} \lambda^{v,o} + \sum_{\rho \in \Gamma} \tilde{D}_\rho \cdot h_\rho^L \right]
$$

$$
\chi^\omega = \sum_{r \in R(\omega)} h_r^\omega \qquad \forall \omega \in W
$$

$$
\sum_{r \in R(\rho)} h_r^\omega \leq \sum_{v \in \mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega} \qquad \forall \omega \in W,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}
$$

$$
F_\rho^v = f_\rho^{v,\emptyset} + \sum_{\omega \in W} f_\rho^{v,\omega} \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall o \in \mathcal{O}
$$

$$
F_\rho^v = 0 \qquad v \text{ and } \rho \text{ incompatible}
$$

$$
\sum_{\rho \in \Gamma_i^-(v,o)} F_\rho^v = \sum_{\rho \in \Gamma_i^+(v,o)} F_\rho^v \qquad \forall i \in N,\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}
$$

$$\sum_{v\in\mathcal{V}(\rho)} \ell^v F^v_\rho \le \bar{\ell}_\rho \sum_{k\in K_M} m^k_\rho \qquad \forall \rho\in\Gamma(o),\ \forall o\in\mathcal{O}$$

$$\sum_{v\in\mathcal{V}(\rho)} (\alpha^v F^v_\rho + \sum_{\omega} \alpha^{v,\mathfrak{p}(\omega)} f^{v,\omega}_\rho) \le \sum_{k\in K_M} \bar{\alpha}^k_\rho m^k_\rho \qquad \forall \rho\in\Gamma(o),\ \forall o\in O$$

$$\sum_{v\in\mathcal{V}(\rho)} \sum_{\mathfrak{p}(\omega)\in P^H(v)} \alpha^{v,\omega} f^{v,\omega}_\rho - \sum_{k\in K^H_M} \bar{\alpha}^k_\rho m^k_\rho \le h^L_\rho \qquad \forall \rho\in\Gamma$$

$$m^k_\rho = \sum_{i\in Y(o)} m^k_{\rho,i} \qquad \forall k\in K_M,\ \forall \rho\in\Gamma(o),\ \forall o\in\mathcal{O}$$

$$\sum_{\rho\in\Gamma^+_{i'}(o)} m^k_{\rho,i} - \sum_{\rho\in\Gamma^-_{i'}(o)} m^k_{\rho,i} = \theta^{k,o}_{i',i} \qquad \forall k,\ \forall (i',i)\in Y(o)\times Y(o),\ \forall o\in\mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot F^v_\rho \le \lambda^{v,o} \le L^{v,o} \qquad \forall v\in\mathcal{V},\ \forall o\in\mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot m^k_\rho \le \hat{L}^{k,o} \qquad \forall k\in K_M,\ \forall o\in\mathcal{O}$$

$$b_a \cdot d_a = \ell_a \qquad \forall a\in A$$

$$b_a = 1 \qquad \forall a\in A_s$$

$$\Delta^k_a = \frac{2d_a + \ell^{max}_a}{\sigma^k} + 2t_s + t_r \qquad \forall a\in A, \forall k\in K$$

$$\sum_k n^k_a \Delta^k_a + \sum_k n^k_a (t_{ru} + t_{zu}) \le \mathcal{T}_a \qquad \forall a\in A$$

$$\sum_{o\in\mathcal{O}} \sum_{j\in Y(o)/j\ne i} \sum_{k\in K_M} \theta^{k,o}_{i,j} \le M^0_i \cdot \bar{g}_i \qquad \forall i\in Y$$

$$\sum_{o\in\mathcal{O}} \sum_{j\in Y(o)/j\ne i} \sum_{k\in K_M} \theta^{k,o}_{j,i} \le M^0_i \cdot \bar{g}_i \qquad \forall i\in Y$$

$$z_a = z_{-a} \qquad \forall a\in A^-$$

$$r_a = r_{-a} \qquad \forall a\in A^-$$

$$r_a \le z_a \qquad \forall a\in A$$

$$z_a \le g_i,\ z_a \le g_j \qquad \forall a=(i,j)\in A$$

$$z_a = 0 \implies \sum_{k\in K} \sum_{\rho\in\Gamma} \epsilon_{a,\rho} \cdot m^k_\rho = 0 \qquad \forall a\in A$$

$$h^\omega_r \in \mathbb{R}^+$$

$$f^{v,\omega}_\rho, f^{v,\emptyset}_\rho, m^k_{\rho,j}, \lambda^{v,o}, \theta^{k,o}_{i',i} \in \mathbb{Z}^+$$

$g_i, z_a, r_a \in \{0, 1\}$

$\bar{g}_i \in \mathbb{R}^+$

$b_a \geq 1, \in \mathbb{N},\ d_a \in \mathbb{R}^+,\ d_a \geq \ell_a^{min}$

where $0 \leq \gamma \leq 1$ and $f_0, \underline{f_0}, f_1, \underline{f_1}$ are as detailed in Section 5.1.3.

B.2. DCECC-model: replacing non-linearities

$$\min \quad \frac{\gamma}{\hat{f}_0 - \underline{f}_0} \left[\sum_{i \in N} \left(S_i \cdot g_i + \bar{S}_i \cdot \bar{g}_i \right) + \sum_{a \in A^{\text{-}}} \left(\dot{S}_a \cdot z_a + \ddot{S}_a \cdot r_a \right) + \sum_{a \in A} \hat{S}_a \cdot b_a . \right] +$$

$$\frac{1-\gamma}{\hat{f}_1 - \underline{f}_1} \left[\sum_{\substack{k \in K_M, \\ o \in O}} \left(\sum_{\substack{i,j \in Y(o) \\ i \neq j}} (C_i^{k,o} + \tilde{C}_j^{k,o}) \theta_{i,j}^{k,o} + \sum_{\rho \in \Gamma(o)} \hat{C}_\rho^k m_\rho^k \right) + \sum_{\substack{v \in \mathcal{V}, \\ o \in O}} D^{v,o} \lambda^{v,o} + \sum_{\rho \in \Gamma} \tilde{D}_\rho \cdot h_\rho^L \right]$$

$$\chi^\omega = \sum_{r \in R(\omega)} h_r^\omega \qquad \forall \omega \in W$$

$$\sum_{r \in R(\rho)} h_r^\omega \leq \sum_{v \in \mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega} \qquad \forall \omega \in W,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$F_\rho^v = f_\rho^{v,\emptyset} + \sum_{\omega \in W} f_\rho^{v,\omega} \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall o \in \mathcal{O}$$

$$F_\rho^v = 0 \qquad v \text{ and } \rho \text{ incompatible}$$

$$\sum_{\rho \in \Gamma_i^-(v,o)} F_\rho^v = \sum_{\rho \in \Gamma_i^+(v,o)} F_\rho^v \qquad \forall i \in N,\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}$$

$$\sum_{v \in \mathcal{V}(\rho)} \ell^v F_\rho^v \leq \bar{\ell}_\rho \sum_{k \in K_M} m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$\sum_{v \in \mathcal{V}(\rho)} (\alpha^v F_\rho^v + \sum_\omega \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega}) \leq \sum_{k \in K_M} \bar{\alpha}_\rho^k m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in O$$

$$\sum_{v \in \mathcal{V}(\rho)} \sum_{\mathfrak{p}(\omega) \in P^H(v)} \alpha^{v,\omega} f_\rho^{v,\omega} - \sum_{k \in K_M^H} \bar{\alpha}_\rho^k m_\rho^k \leq h_\rho^L \qquad \forall \rho \in \Gamma$$

$$m_\rho^k = \sum_{i \in Y(o)} m_{\rho,i}^k \qquad \forall k \in K_M,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$\sum_{\rho \in \Gamma_{i'}^+(o)} m_{\rho,i}^k - \sum_{\rho \in \Gamma_{i'}^-(o)} m_{\rho,i}^k = \theta_{i',i}^{k,o} \qquad \forall k,\ \forall (i',i) \in Y(o) \times Y(o),\ \forall o \in \mathcal{O}$$

$$\sum_{o \in \mathcal{O}} \sum_{j \in Y(o)/j \neq i} \sum_{k \in K_M} \theta_{i,j}^{k,o} \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y$$

$$\sum_{o \in \mathcal{O}} \sum_{j \in Y(o)/j \neq i} \sum_{k \in K_M} \theta_{j,i}^{k,o} \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y$$

$$\sum_{\rho\in\Gamma(a)} m_\rho^k = \sum_{i=0}^{c_a^k} 2^i \cdot \phi_{a,i}^k. \qquad \forall k \in K,\ \forall a \in A$$

$$\sum_k \sum_{i=0}^{c_a^k} 2^i \zeta_{a,i}^k + \sum_k n_a^k (t_{ru} + t_{zu}) \leq \mathcal{T}_a \qquad \forall a \in A$$

$$\phi_{a,i}^k = 0 \implies \zeta_{a,i}^k = 0 \qquad \forall i = 0,\dots,c_a^k,\ \forall k \in K,\ \forall a \in A$$

$$\phi_{a,i}^k = 1 \implies \zeta_{a,i}^k = \Delta_a^k \qquad \forall i = 0,\dots,c_a^k,\ \forall k \in K_M,\ \forall a \in A$$

$$\hat{\tau}_i = \sum_{j=0}^{3} b_i^j x_i^j \qquad \forall i \in Y$$

$$\hat{m}_i = \frac{1}{\bar{B}\cdot\mathcal{T}} \sum_{k\in K_M} \Big(\sum_{\rho\in\Gamma_i^+} m_\rho^k + \sum_{\rho\in\Gamma_i^-} m_\rho^k \Big) \qquad \forall i \in Y$$

$$x_i^0 = 1 \Rightarrow \hat{m}_i \leq \hat{B}_i^0 \bar{g}_i \qquad \forall i \in Y$$

$$x_i^1 = 1 \Rightarrow \hat{B}_i^0 \bar{g}_i \leq \hat{m}_i \leq \hat{B}_i^1 \bar{g}_i \qquad \forall i \in Y$$

$$x_i^2 = 1 \Rightarrow \hat{B}_i^1 \bar{g}_i \leq \hat{m}_i \leq \hat{B}_i^2 \bar{g}_i \qquad \forall i \in Y$$

$$x_i^3 = 1 \Rightarrow \hat{m}_i \geq \hat{B}_i^2 \bar{g}_i \qquad \forall i \in Y$$

$$\sum_{j=0}^{3} x_i^j = 1 \qquad \forall i \in Y$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} \Big(\tau_\rho F_\rho^v + \sum_{j=0}^{3} (\bar{x}_{\mathfrak{o}(\rho),\rho}^{j,v} + \bar{x}_{\mathfrak{d}(\rho),\rho}^{j,v})\Big) \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in O$$

$$x_i^j = 0 \Rightarrow \bar{x}_{i,\rho}^{j,v} = 0 \qquad \forall j = 0,\dots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall v \in \mathcal{V}$$

$$x_i^j = 1 \Rightarrow \bar{x}_{i,\rho}^{j,v} = F_\rho^v \qquad \forall j = 0,\dots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall v \in \mathcal{V}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} \Big(\tau_\rho m_\rho^k + \sum_{j=0}^{3} (\hat{x}_{\mathfrak{o}(\rho),\rho}^{j,k} + \hat{x}_{\mathfrak{d}(\rho),\rho}^{j,k})\Big) \leq \hat{\lambda}^{k,o} \leq L^{v,o} \qquad \forall k \in K_M,\ \forall o \in O$$

$$x_i^j = 0 \Rightarrow \hat{x}_{i,\rho}^{j,k} = 0 \qquad \forall j = 0,\dots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall k \in K_M$$

$$x_i^j = 1 \Rightarrow \hat{x}_{i,\rho}^{j,k} = m_\rho^k \qquad \forall j = 0,\dots,3,\ \forall i \in Y,\ \forall \rho \in \Gamma(o),\ \forall k \in K_M$$

$$z_a = z_{-a} \qquad \forall a \in A^{\text{-}}$$

$$r_a = r_{-a} \qquad \forall a \in A^{\text{-}}$$

$$r_a \leq z_a \qquad \forall a \in A$$

$$z_a \leq g_i,\ z_a \leq g_j \qquad \forall a = (i,j) \in A$$

$$z_a = 0 \implies \sum_{k \in K} \sum_{\rho \in \Gamma} \epsilon_{a,\rho} \cdot m_\rho^k = 0 \qquad \forall a \in A$$

$$h_r^\omega \in \mathbb{R}^+$$
$$f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_{\rho,j}^k, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+$$
$$g_i, z_a, r_a \in \{0,1\}$$
$$\bar{g}_i \in \mathbb{R}^+$$
$$b_a \geq 1, \in \mathbb{N},\ d_a \in \mathbb{R}^+,\ d_a \geq \ell_a^{min}$$
$$\phi_{a,i}^k \in \{0,1\} \quad i = 0, \ldots, c_a^k$$
$$\zeta_{a,i}^k \in \mathbb{R}, \quad \forall i = 0, \ldots, c_a^k$$
$$x_i^j \in \{0,1\},\ j = 0,1,2,3$$
$$\bar{x}_{i,\rho}^{j,v} \in \mathbb{R}^+$$
$$\hat{x}_{i,\rho}^{j,k} \in \mathbb{R}^+$$

where

- $0 \leq \gamma \leq 1$,
- $f_0, \underline{f_0}, f_1, \underline{f_1}$ are as detailed in Section 5.1.3,
- $c_a^k \in \mathbb{N}$ such as $\sum_{\rho \in \Gamma(a)} m_\rho^k \leq 2^{c_a^k}$,
- $\hat{\tau}_i$, the waiting time on yard i is estimated as detailed in Section 5.2, and
- $t_\rho := \tau_\rho + \hat{\tau}_{\mathfrak{o}(\rho)} + \hat{\tau}_{\mathfrak{d}(\rho)}$.

B.3. DCEM-model with non-linear constraints

$$\min \quad \frac{\gamma}{\hat{f}_0 - \underline{f}_0} \left[\sum_{i \in N} \left(S_i \cdot g_i + \bar{S}_i \cdot \bar{g}_i \right) + \sum_{a \in A^{\text{-}}} \left(\dot{S}_a \cdot z_a + \ddot{S}_a \cdot r_a \right) + \sum_{a \in A} \hat{S}_a \cdot b_a . \right] +$$

$$\frac{1-\gamma}{\hat{f}_1 - \underline{f}_1} \left[\sum_{k \in K_M} \left(\sum_{\substack{i,j \in Y(o) \\ i \neq j}} (C_i^k + \tilde{C}_j^k) \theta_{i,j}^k + \sum_{\rho \in \Gamma} \hat{C}_\rho^k m_\rho^k \right) + \sum_{v \in \mathcal{V}} D^v \mu^v + \sum_{\rho \in \Gamma} \tilde{D}_\rho \cdot h_\rho^L \right]$$

$$\chi^\omega = \sum_{r \in R(\omega)} h_r^\omega \qquad \forall \omega \in W$$

$$\sum_{r \in R(\rho)} h_r^\omega \leq \sum_{v \in \mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega} \qquad \forall \omega \in W,\ \forall \rho \in \Gamma(o)$$

$$F_\rho^v = f_\rho^{v,\emptyset} + \sum_{\omega \in W} f_\rho^{v,\omega} \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v)$$

$$F_\rho^v = 0 \qquad v \text{ and } \rho \text{ incompatible}$$

$$\sum_{\rho \in \Gamma_i^-(v)} F_\rho^v = \sum_{\rho \in \Gamma_i^+(v)} F_\rho^v \qquad \forall i \in N,\ \forall v \in \mathcal{V}$$

$$\sum_{v \in \mathcal{V}(\rho)} \ell^v F_\rho^v \leq \bar{\ell}_\rho \sum_{k \in K_M} m_\rho^k \qquad \forall \rho \in \Gamma$$

$$\sum_{v \in \mathcal{V}(\rho)} (\alpha^v F_\rho^v + \sum_\omega \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega}) \leq \sum_{k \in K_M} \bar{\alpha}_\rho^k m_\rho^k \qquad \forall \rho \in \Gamma$$

$$\sum_{v \in \mathcal{V}(\rho)} \sum_{\mathfrak{p}(\omega) \in P^H(v)} \alpha^{v,\omega} f_\rho^{v,\omega} - \sum_{k \in K_M^H} \bar{\alpha}_\rho^k m_\rho^k \leq h_\rho^L \qquad \forall \rho \in \Gamma$$

$$b_a \cdot d_a = \ell_a \qquad \forall a \in A$$

$$\Delta_a^k = \frac{2d_a + \ell_a^{max}}{\sigma^k} + 2t_s + t_r \qquad \forall k \in K,\ \forall a \in A$$

$$\Delta_a^{k,k'} = \frac{d_a}{\sigma^{k'}} + \frac{d_a + \ell_a^{max}}{\sigma^k} + 2t_s + t_r \qquad \forall k, k' \in K, k \to k'\ \forall a \in A$$

$$\sum_{j \in Y / j \neq i} \sum_{k \in K_M} \theta_{i,j}^k \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y$$

$$\sum_{j \in Y / j \neq i} \sum_{k \in K_M} \theta_{j,i}^k \leq M_i^0 \cdot \bar{g}_i \qquad \forall i \in Y$$

$$n_a^\varsigma \le \pi^\varsigma \sum_{\rho\in\Gamma_a} m_\rho^{k(\varsigma)} \qquad \forall\varsigma \in \Theta(K), k(\varsigma)\in K_M, \forall a\in A$$

$$\sum_{\varsigma\in g(k)} n_a^\varsigma = \sum_{\rho\in\Gamma_a} m_\rho^k \qquad \forall k\in K_M, \forall a \in A$$

$$\tilde{n}_a^\varsigma = \max(0, n_a^\varsigma - 1) \qquad \forall\varsigma\in\Theta(K), \forall a\in A$$

$$\tau_a^\varsigma + \delta_a^\varsigma \cdot \theta_a^{k(\varsigma)} + \Delta t_a^\varsigma \cdot \tilde{n}_a^\varsigma \le t_a^\varsigma \qquad \forall\varsigma\in\Theta(K),\ \forall a\in A$$

$$\tau_a^{\varsigma'} \ge \tau_a^\varsigma + \Delta t_a^\varsigma \cdot \tilde{n}_a^\varsigma + \Delta t_a^{\varsigma,\varsigma'} \qquad \forall\varsigma,\varsigma'\in\Theta(K), \varsigma\to\varsigma',\ \forall a\in A$$

$$\tau_a^{\varsigma'} + \delta_a^{\varsigma'} \cdot \theta_a^{\varsigma(\varsigma')} \ge t_a^\varsigma + \Delta t_a^{\varsigma,\varsigma'} \qquad \forall\varsigma,\varsigma'\in\Theta(K), \varsigma\to\varsigma',\ \forall a\in A$$

$$\delta_a^\varsigma = 0 \Rightarrow \sum_{\rho\in\Gamma_a} m_\rho^{k(\varsigma)} = 0 \qquad \forall\varsigma\in\Theta(K), \forall a\in A$$

$$r_a = 0 \Rightarrow t_a^{\varsigma_s} \le \tau_{-a}^{\varsigma_0} \qquad \forall a \in A^{-}$$

n_a^ς has a fixed value, previouly known $\qquad \forall\varsigma\in\Theta(k), k(\varsigma)\in K_P, \forall a\in A$

$$\sum_{q\in P} f_\rho^{v,q} + f_\rho^{v,\emptyset} = \sum_{c\in C(\rho)} \lambda_c^v \qquad \forall\rho\in\Gamma,\ \forall v\in\mathcal{V}$$

$$w_c^v \ge \frac{1}{\mathcal{T}}(\tau_c \cdot \lambda_c^v) \qquad \forall c\in C,\ \forall v\in\mathcal{V}$$

$$\sum_{c\in C} w_c^v = \mu^v \cdot \eta^v \qquad \forall v\in\mathcal{V}$$

$$z_a = z_{-a} \qquad \forall a\in A^{-}$$

$$r_a = r_{-a} \qquad \forall a\in A^{-}$$

$$r_a \le z_a \qquad \forall a\in A$$

$$z_a \le g_i,\ z_a \le g_j \qquad \forall a=(i,j)\in A$$

$$z_a = 0 \implies \sum_{k\in K}\sum_{\rho\in\Gamma} \epsilon_{a,\rho}\cdot m_\rho^k = 0 \qquad \forall a\in A$$

$$h_r^\omega \in \mathbb{R}^+$$

$$f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_{\rho,j}^k, \lambda^v, \theta_{i',i}^k \in \mathbb{Z}^+$$

$$g_i, z_a, r_a, \delta_a^\varsigma \in \{0,1\}$$

$$\bar{g}_i \in \mathbb{R}^+$$

$$b_a \ge 1, \in \mathbb{N},\ d_a\in\mathbb{R}^+,\ d_a \ge \ell_a^{min}$$

$$\tau_a^\varsigma, t_a^\varsigma \in \mathbb{R}^+,\ \tau_a^\varsigma \le \mathcal{T},\ t_a^\varsigma \le \mathcal{T}$$

$$\lambda_c^v, \tau_c, w_c^v, \mu^v \in \mathbb{R}^+$$

where

- $0 \leq \gamma \leq 1$,
- $f_0, \underline{f_0}, f_1, \underline{f_1}$ are as detailed in Section 5.1.3,
- superscript o has been removed from variable $\theta^k_{i,i'}$ and parameters $C^k_i, \tilde{C}^k_i$ due the non-existent of carriers competence in the model,
- variable λ^v in **DCECC**-model here in **DCEM**-model corresponds to variable μ^v, while λ^v_c is for cycles, and represents the total v-class railcars for each cycle c and the total period $\mathcal{T}$.

B.4. MINLP-model: the Deterministic version

$$\min \sum_{\omega\in W}\sum_{r\in R(\omega)} u_r^\omega h_r^\omega + \sum_{\omega\in W} \tilde{u}^\omega \tilde{h}^\omega + \sum_{\omega\in W} \int_0^{\tilde{h}^\omega} \left(\log \frac{x}{\chi^\omega - x}\right) dx$$

$$\chi^\omega = \sum_{r\in R(\omega)} h_r^\omega \qquad \forall \omega \in W$$

$$\sum_{r\in R(\rho)} h_r^\omega \leq \sum_{v\in\mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega} \qquad \forall\omega \in W,\ \forall\rho\in\Gamma(o),\ \forall o\in\mathcal{O}$$

$$F_\rho^v = f_\rho^{v,\emptyset} + \sum_{\omega\in W} f_\rho^{v,\omega} \qquad \forall v\in\mathcal{V},\ \forall\rho\in\Gamma(v,o),\ \forall o\in\mathcal{O}$$

$$F_\rho^v = 0 \qquad v \text{ and } \rho \text{ incompatible}$$

$$\sum_{\rho\in\Gamma_i^-(v,o)} F_\rho^v = \sum_{\rho\in\Gamma_i^+(v,o)} F_\rho^v \qquad \forall i\in N,\ \forall v\in\mathcal{V},\ \forall o\in\mathcal{O}$$

$$\sum_{v\in\mathcal{V}(\rho)} \ell^v F_\rho^v \leq \bar{\ell}_\rho \sum_{k\in K_M} m_\rho^k \qquad \forall\rho\in\Gamma(o),\ \forall o\in\mathcal{O}$$

$$\sum_{v\in\mathcal{V}(\rho)} (\alpha^v F_\rho^v + \sum_{\omega} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega}) \leq \sum_{k\in K_M} \bar{\alpha}_\rho^k m_\rho^k \qquad \forall\rho\in\Gamma(o),\ \forall o\in O$$

$$m_\rho^k = \sum_{i\in Y(o)} m_{\rho,i}^k \qquad \forall k\in K_M,\ \forall\rho\in\Gamma(o),\ \forall o\in\mathcal{O}$$

$$\sum_{\rho\in\Gamma_{i'}^+(o)} m_{\rho,i}^k - \sum_{\rho\in\Gamma_{i'}^-(o)} m_{\rho,i}^k = \theta_{i',i}^{k,o} \qquad \forall k,\ \forall (i',i)\in Y(o)\times Y(o),\ \forall o\in\mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot F_\rho^v \leq \lambda^{v,o} \leq L^{v,o} \qquad \forall v\in\mathcal{V},\ \forall o\in\mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot m_\rho^k \leq \hat{L}^{k,o} \qquad \forall k\in K_M,\ \forall o\in\mathcal{O}$$

$$\sum_{\omega\in W}\sum_{r\in R(\omega,o)} U_r^\omega h_r^\omega \geq \sum_{k\in K_M}\sum_{i\in Y(o)}\sum_{\substack{j\in Y(o)\\ j\neq i}} (C_{i'}^{k,o} + \tilde{C}_i^{k,o})\theta_{i',i}^{k,o} + \sum_{k\in K_M}\sum_{\rho\in\Gamma(o)} \hat{C}_\rho^k m_\rho^k + \sum_{v\in\mathcal{V}} D^{v,o}\lambda^{v,o} - M_o(1-\hat{y}_o) \qquad \forall o\in\mathcal{O}$$

$$\sum_{r\in R(\omega,o)} h_r^\omega \leq \bar{\chi}^\omega + \chi^\omega \cdot \hat{y}_o \qquad \forall\omega\in W,\ \forall o\in\mathcal{O}$$

$$\sum_{k \in K_M} m_\rho^k \leq N_\rho \qquad \forall \rho \in \Gamma(o), \forall o \in \mathcal{O}$$

$$\sum_{o \in \mathcal{O}} \sum_{\rho \in \Gamma(a,o)} \sum_{k \in K_M} m_\rho^k \leq N_a \qquad \forall a \in A$$

$$\sum_{o \in \mathcal{O}} \sum_{i' \in Y(o)} \sum_{k \in K_M} \theta_{i,i'}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y$$

$$\sum_{o \in \mathcal{O}} \sum_{i' \in Y(o)} \sum_{k \in K_M} \theta_{i',i}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y$$

$$h_r^\omega \in \mathbb{R}^+$$
$$\tilde{h}^\omega \in \mathbb{R}^+$$
$$f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_{\rho,j}^k, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+$$
$$\hat{y}_o \in \{0,1\}$$

where:

- $u_r^\omega = \beta_0^\omega + \sum_{j=1}^m \beta_j^\omega u_{r,j}^\omega$ are the generalized costs for transport products by rail,
- $\tilde{u}^\omega = \tilde{\beta}_0^\omega + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^\omega \tilde{u}_j^\omega$ are the generalized costs for transport products by road.

B.5. MINLP-model: the Robust version

$$\min \sum_{\omega\in W}\sum_{r\in R(\omega)} \mathbf{u}_r^\omega h_r^\omega + \sum_{\omega\in W} \tilde{\mathbf{u}}^\omega \tilde{h}^\omega + \sum_{\omega\in W}\int_0^{\tilde{h}^\omega} (\log \frac{x}{\chi^\omega - x})dx$$
$$+ \sum_{\omega\in W}\sum_{r\in R(\omega)} \left(H\pi_r^\omega + \sum_{j=0}^{m} p_{r,j}^\omega \right) + \sum_{\omega\in W} \left(\tilde{H}\tilde{\pi}^\omega + \sum_{j=0}^{n} \tilde{p}_j^\omega \right)$$

$$\chi^\omega = \sum_{r\in R(\omega)} h_r^\omega \qquad \forall \omega \in W$$

$$\sum_{r\in R(\rho)} h_r^\omega \le \sum_{v\in\mathcal{V}(\rho)} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega} \qquad \forall \omega \in W,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$F_\rho^v = f_\rho^{v,\emptyset} + \sum_{\omega\in W} f_\rho^{v,\omega} \qquad \forall v \in \mathcal{V},\ \forall \rho \in \Gamma(v,o),\ \forall o \in \mathcal{O}$$

$$F_\rho^v = 0 \qquad v \text{ and } \rho \text{ incompatible}$$

$$\sum_{\rho\in\Gamma_i^-(v,o)} F_\rho^v = \sum_{\rho\in\Gamma_i^+(v,o)} F_\rho^v \qquad \forall i \in N,\ \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}$$

$$\sum_{v\in\mathcal{V}(\rho)} \ell^v F_\rho^v \le \bar{\ell}_\rho \sum_{k\in K_M} m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$\sum_{v\in\mathcal{V}(\rho)} (\alpha^v F_\rho^v + \sum_{\omega} \alpha^{v,\mathfrak{p}(\omega)} f_\rho^{v,\omega}) \le \sum_{k\in K_M} \bar{\alpha}_\rho^k m_\rho^k \qquad \forall \rho \in \Gamma(o),\ \forall o \in O$$

$$m_\rho^k = \sum_{i\in Y(o)} m_{\rho,i}^k \qquad \forall k \in K_M,\ \forall \rho \in \Gamma(o),\ \forall o \in \mathcal{O}$$

$$\sum_{\rho\in\Gamma_{i'}^+(o)} m_{\rho,i}^k - \sum_{\rho\in\Gamma_{i'}^-(o)} m_{\rho,i}^k = \theta_{i',i}^{k,o} \qquad \forall k,\ \forall (i',i) \in Y(o)\times Y(o),\ \forall o \in \mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot F_\rho^v \le \lambda^{v,o} \le L^{v,o} \qquad \forall v \in \mathcal{V},\ \forall o \in \mathcal{O}$$

$$\frac{1}{\mathcal{T}} \sum_{\rho\in\Gamma(o)} t_\rho \cdot m_\rho^k \le \hat{L}^{k,o} \qquad \forall k \in K_M,\ \forall o \in \mathcal{O}$$

$$\sum_{\omega\in W}\sum_{r\in R(\omega,o)} U_r^\omega h_r^\omega \ge \sum_{k\in K_M}\sum_{i\in Y(o)}\sum_{\substack{j\in Y(o)\\ j\ne i}} (C_{i'}^{k,o} + \tilde{C}_i^{k,o})\theta_{i',i}^{k,o} +$$
$$\sum_{k\in K_M}\sum_{\rho\in\Gamma(o)} \hat{C}_\rho^k m_\rho^k + \sum_{v\in\mathcal{V}} D^{v,o}\lambda^{v,o} - M_o(1-\hat{y}_o) \qquad \forall o \in \mathcal{O}$$

$$\sum_{r\in R(\omega,o)} h_r^\omega \leq \bar{\chi}^\omega + \chi^\omega \cdot \hat{y}_o \qquad \forall \omega \in W,\ \forall o \in \mathcal{O}$$

$$\sum_{k\in K_M} m_\rho^k \leq N_\rho \qquad \forall \rho \in \Gamma(o), \forall o \in \mathcal{O}$$

$$\sum_{o\in\mathcal{O}} \sum_{\rho\in\Gamma(a,o)} \sum_{k\in K_M} m_\rho^k \leq N_a \qquad \forall a \in A$$

$$\sum_{o\in\mathcal{O}} \sum_{i'\in Y(o)} \sum_{k\in K_M} \theta_{i,i'}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y$$

$$\sum_{o\in\mathcal{O}} \sum_{i'\in Y(o)} \sum_{k\in K_M} \theta_{i',i}^{k,o} \leq \tilde{N}_i \qquad \forall i \in Y$$

$$\pi_r^\omega + p_{r,j}^\omega \geq \beta_{r,j}^{+,\omega} u_{r,j}^\omega h_r^\omega \qquad \forall j = 1, \dots, m,\ \forall r \in R(\omega),\ \forall \omega \in W$$

$$\pi_r^\omega + p_{r,0}^\omega \geq \beta_{r,0}^{+,\omega} h_r^\omega, \qquad \forall r \in R(\omega),\ \forall \omega \in W$$

$$\tilde{\pi}^\omega + \tilde{p}_j^\omega \geq \tilde{\beta}_j^{+,\omega} \tilde{u}_j^\omega \tilde{h}^\omega \qquad \forall j = 1, \dots, \tilde{m},\ \forall \omega \in W$$

$$\tilde{\pi}^\omega + \tilde{p}_0^\omega \geq \tilde{\beta}_0^{+,\omega} \tilde{h}^\omega \qquad \forall \omega \in W$$

$$h_r^\omega \ \in \mathbb{R}^+$$

$$\tilde{h}^\omega \ \in \mathbb{R}^+$$

$$f_\rho^{v,\omega}, f_\rho^{v,\emptyset}, m_{\rho,j}^k, \lambda^{v,o}, \theta_{i',i}^{k,o} \in \mathbb{Z}^+$$

$$\hat{y}_o \in \{0,1\}$$

$$\pi_r^\omega,\ p_{r,j}^\omega,\ \tilde{\pi}^\omega,\ \tilde{p}_j^\omega \in \mathbb{R}^+$$

where:

- $u_r^\omega = \beta_0^\omega + \sum_{j=1}^m \beta_j^\omega u_{r,j}^\omega$ are the generalized costs for transport products by rail,
- $\tilde{u}^\omega = \tilde{\beta}_0^\omega + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^\omega \tilde{u}_j^\omega$ are the generalized costs for transport products by road,
- $\mathbf{u}_r^\omega \triangleq \beta_{r,0}^{-,\omega} + \sum_{j=1}^m \beta_{r,j}^{-,\omega} u_{r,j}^\omega$, and
- $\tilde{\mathbf{u}}^\omega \triangleq \tilde{\beta}_0^{-,\omega} + \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j^{-,\omega} \tilde{u}_j^\omega$.