Network-level analysis of the performance of European intermodal freight transport

Yoram Houtsma

Master Thesis Transport & Planning

TU Delft
Panteia
Research to Progress

Network-level analysis of the performance of European intermodal freight transport

Yoram Houtsma

to obtain the degree of Master of Science in

Civil Engineering, Transport & Planning

at the Delft University of Technology.

Student number:	4301153	
Thesis committee:	Prof. dr. L.A. Tavasszy	TU Delft
	Dr. B. Wiegmans	TU Delft
	Dr. J.H.R. van Duin	TU Delft
	Dr. H. Saeedi	Tilburg University
	MSc. J. Kiel	Panteia

TU Delft

Panteia
Research to Progress

Preface

This thesis concludes my time as student at TU Delft and is the last step towards finishing the master Transport & Planning. My previous courses were mostly focused on passenger transport and public transport. However, I have always been interested in broadening my knowledge, which is why I chose the topic of freight transportation for my master thesis. The execution of this project was challenging and insightful. I learned both professionally and personally a lot through doing my thesis. Therefore, I would like to take this opportunity to express my gratitude to everyone that supported me.

Firstly, I would like to thank Panteia for giving me the opportunity to perform this research. Furthermore, I am grateful for the opportunities they gave me to join their projects, from which I learned a lot. I also would like to thank my colleagues who made me feel welcomed from the beginning. In particular, I would like to thank my daily supervisor Jan Kiel. Our conversations and your feedback helped me with keeping an eye on the practical side of this research. Furthermore, thank you for your patience and personal advice.

Secondly, I would like to acknowledge the assistance of my supervisors from TU Delft. Thank you Lóri Tavasszy and Ron van Duin, for being patient with me and giving me feedback that I could directly apply. This helped me steer in the right direction. I want to especially thank my daily supervisor Bart Wiegmans. It was very helpful to discuss the many details of my research and report. Besides that, thank you for helping me develop on a personal level. I am also grateful for the help of Hamid Saeedi, on whose PhD thesis I continued. The introduction in DEA and later our in-depth discussions were very valuable for my research.

Thirdly, I would like to thank all the people that have been there for me outside of working hours. I want to thank my girlfriend Maneka and my friends for giving me the much needed distractions when I needed them. I would like to thank my family, for believing in me throughout all my years at TU Delft.

Dear all, this accomplishment would not have been possible without you. Thank you.

Yoram Houtsma

September 2020

Contents

List of Figures

List of Tables

List of Abbreviations

CEF	Connecting Europe Facility
CRS	Constant Returns-to-Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
EC	European Commission
ETIS	European Transport Policy Information System
EU	European Union
IFT	Intermodal Freight Transport
ITS	Intelligent Transportation System
ITU	Intermodal Transportation Unit
IWW	Inland Waterways
KPI	Key Performance Indicator
LP	Linear Programming
main-to-main	main haulage to main haulage
NDEA	Network Data Envelopment Analysis
NS-M	North Sea - Mediterranean
OD	Origin-Destination
R-A	Rhine - Alpine
SB	Slacks-Based
SFA	Stochastic Frontier Analysis
SSS	Short Sea Shipping
TEN-T	Trans-European Transport Networks
TEU	Twenty feet Equivalent Unit
VIFTS	Value of Intermodal Freight Transport Service
VRS	Variable Returns-to-Scale

Executive summary

Road transport is the most used mode in freight transport because. While road transport gives flexibility, it also comes with large external costs. Therefore, The European Commission stimulates the more sustainable alternative intermodal freight transport (IFT). Despite large European investments in IFT, its modal share is still relatively small.

To increase the modal share of IFT, it is needed to know what the performance is of the current IFT-network, which consists of nine TEN-T corridors (Trans European Transport Network). Most research in the performance of IFT is done on individual segments, because independent decision makers can optimize their own technical efficiency. However, the performance of entire IFT-corridors is interesting on a European level, but to the best of the author's knowledge, there is no research on this yet. To fill this gap, the main objective of this thesis is to assess the performance of an IFT-network. Ti be more specific, the parts of the IFT-network assessed in this thesis are the transshipment segments and the rail haulage segments (see Figure 1). To achieve the objective, there are three sub-objectives. Firstly, an overview of the performance indicators that are currently used is needed. Secondly, a method to measure performance of IFT-corridors is needed, taking into account parallel chains, overlap and different chain structures. Lastly, the developed method can be used to evaluate a part of the European IFT-network.

The main research question is: **How can the performance of an intermodal freight transport chain be assessed?**

There are two requirements to answer the main research question. The first is the parameters with which the corridors are assessed, which are called performance indicators. The second is a model, which uses the aforementioned indicators to evaluate the performance of an IFT-chain.

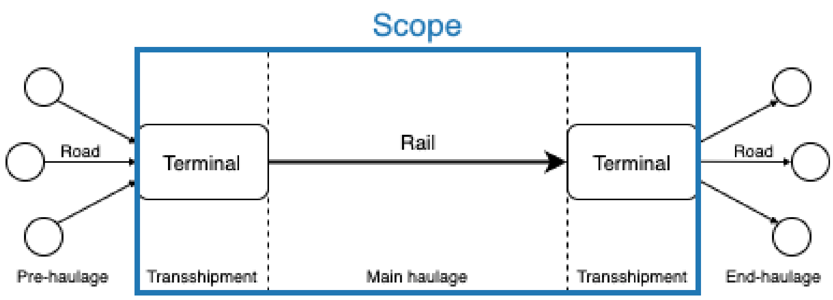


Figure 1: Scope of this research within the IFT-chain.

To see which performance indicators are useful to evaluate an IFT-network, both scientific and practical literature is used. Both use key performance indicators (KPIs) to describe the performance of European intermodal freight transport (IFT).

In scientific literature, we distinguish two domains: chains & transportation and terminals & ports. The most common KPIs for IFT in general are environment & emissions, waiting time and reliability. For each domain, there are also specific KPIs. For terminals & ports, these are transshipment time, safety & security, employee & equipment productivity and volume. The chains & transportation specific KPIs are transport costs and travel time. In practical literature, the most common objectives for both transshipment and haulage segments are price of transportation, travel time and waiting time. Also capacity and capacity utilization are seen as KPIs on the operator level.

Besides the parameters, a model is required to assess the performance of an IFT-network. A well known method to assess performance is Data Envelopment Analysis (DEA). It measures the relative performance of decision making units (DMUs) such as transshipment- or haulage segments in the case of IFT-chains. The difficulty with IFT-networks is the overlap between chains, parallel chains and the different structures of chains. Therefore, a regular Network DEA cannot be applied and Saeedi et al. (2019) developed a Network DEA that takes into account some of these difficulties. However, this model has some issues, mainly that the model gives a division (such as transshipment or haulage segment) later in a chain on average a higher efficiency, unrelated to the actual technical efficiency of the division.

Therefore, a new model is developed, a weighted slacks-based DEA, which takes into account all the difficulties of IFT-chains. It uses the simpler formulation from a traditional DEA and combines this with the weighted objective function of a Network DEA. To see how the adapted DEA (aDEA) performs compared to a traditional DEA and the Network DEA, two comparisons have been made.

- A numerical experiment, which randomly assigns terminals to a position in a chain and then runs a traditional DEA, a Network DEA and the newly developed aDEA. Because this is done in 1000 simulations, many possible configurations are assessed. The Network DEA does indeed have a significant higher efficiency later in the chain, while the developed aDEA does not have this problem (see Figure 2).

- A comparison is run between the Network DEA and the developed DEA with a real data sample of 10 corridors with different structures in Europe. Although the sample is too small to have a significant difference, the Network DEA has again a higher efficiency on later positions. The developed DEA works mostly as expected, although the efficiency of the transshipment and haulage segments are smaller.

Future comparisons between traditional DEA, Network DEA and the developed DEA using real data with more complex chain structures should show if there are actual lower efficiencies for the developed weighted slacks-based DEA.

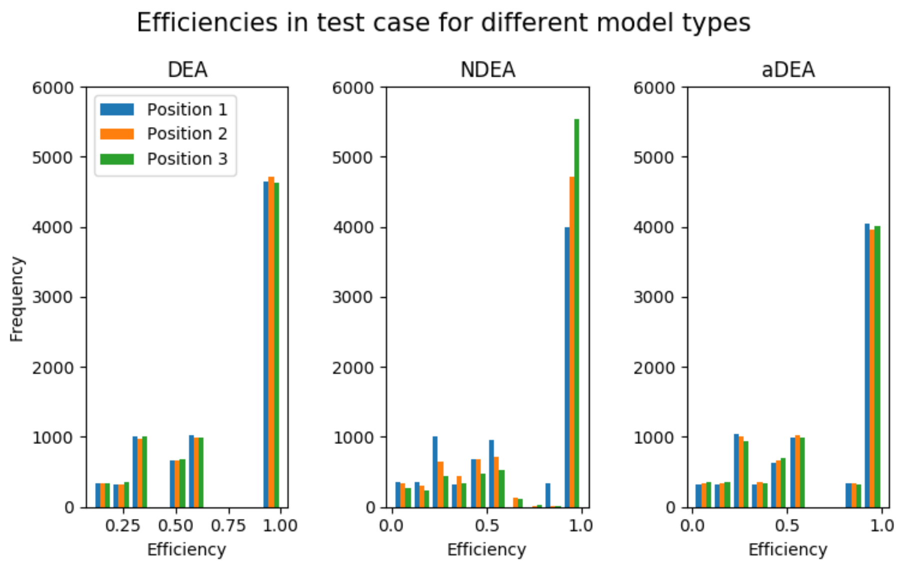


Figure 2: Efficiencies of single divisions on different positions in the chain for numerical experiment with DEA, NDEA and aDEA with 1000 simulations.

Since the model works as expected, it can be applied on a part of the TEN-T corridors. The two corridors are selected on best data availability and are the North Sea - Mediterranean and Rhine - Alpine corridors. However, there are still limitations in the availability and quality of the data. The inputs for transshipment are characteristics of terminals: area, number of tracks, total length of tracks, number of cranes and number of reachstackers. For haulage segments, the number of trains is the proxy attribute for transport cost as input. Volume is the output for both segments.

The application of the model results in an efficiency score for the North Sea - Mediterranean corridor (47%) and the Rhine - Alpine corridor (62%). The latter is mainly more efficient because of the efficient haulage segments between Rotterdam and Milano, with a large number of trains and large volumes. This might indicate some advantages of economies of scale. However, a limitation for the North Sea - Mediterranean corridor seems a missing number of trains, decreasing the efficiency of the other haulage segments. For transshipment, both corridors have similar efficiency scores. There seems to be no advantages of economies of scale, but this could be due to a limitation of the data. There is no data on transshipment between two main haulage segments. Large transshipment segments such as ports, for which this is a large share of the total transshipment, do therefore get a lower efficiency score. These limitations show the importance of data quality, which is an important point of attention for future research.

Overall, there are three main scientific contributions:

- Literature review of KPIs in scientific and practical literature. The literature review updates on previous literature reviews, together with a comparison between scientific and practical literature.
- Show issues with cumulative intermediates for Network DEA. The use of cumulative intermediates has some issues that are shown in this thesis with theoretical explanation and simulations.
- Development of weighted slacks-based DEA. The developed model is a hybrid version of the traditional DEA and a Network DEA. It optimises for the entire chain, but without the use of intermediates. It can be applied on chains with different structures, parallel segments and overlap between chains.

Furthermore, this thesis shows that data quality and availability are essential to the assessment of performance of IFT-networks. Essential data to assess the performance of European IFT are reliable data for terminals, number of trains and volumes. Furthermore, when the EU wants to assess the performance of individual corridors, all data needs to be collected on a corridor level.

A suggestion for future research is to test the effect of using different weights in the weighted slacks-based DEA. These can be used to assign relative importance of divisions, which impacts the efficiency of the entire chain. Including relative importance is interesting for fine-tuning the results in specific applications.

A practical suggestion is to further evaluate the developed weighted slacks-based DEA and see if it is more widely applicable. Using a sensitivity analysis, the impact of aforementioned limitations can be researched. This should give better insights on the working and application of the model.

The weighted slacks-based model has been developed for IFT-network. IFT-corridors do not have fixed structures, have possible overlap and have parallel divisions. However, one could also see other applications for the model. An example that has quite some parallels with IFT is public transport. A network of bus or train lines can be assessed, where transportation and stations are the divisions. Another example could be modern power grids. Both modern power grids and public transport do not have a fixed structure and can have overlapping and parallel divisions, which makes it interesting to apply the weighted slacks-based DEA.

1 Introduction

Road transport is the most used mode in freight transport. While road transport gives flexibility, it is also the mode that comes with most external costs such as air and noise pollution, greenhouse gases, accidents and congestion (Kos, Vukić, & Brčić, 2017). A more sustainable alternative is intermodal freight transport (IFT) (Kos et al., 2017), which has been stimulated by the European Commission (EC). However, the current market share of IFT is relatively low (Pastori, 2015) and the EC has set a goal to shift 30 % of road freight over 300 km distance to other modes like rail and water (European Commission, 2011), which are the main haulage elements of IFT services. To accomplish this goal, investments in infrastructure and services are required (European Commission, 2011).

Between 2007 and 2013, 28 billion euro was invested in rail projects by the EC (European Court of Auditors, 2016). This large funding did not show an improvement in the volume and modal share of rail freight between 2000 and 2013 (European Court of Auditors, 2016). Volume and modal share depend on the performance of the complete IFT-network. Performance of IFT services can be seen as performance of separate IFT-chain divisions, such as transshipment- and haulage segments, but also as the overall performance of IFT-chains (Yang, Wu, Liang, Bi, & Wu, 2011). Most attention has been paid to the separate divisions because independent decision makers can optimize their own technical efficiency (Yang et al., 2011). There is however a gap in research to the overall performance of IFT-chains (Yang et al., 2011), while these are most important for competition with road freight transport and the goals set by the EC.

Efficiency of individual chain divisions can be explained as the ratio of outputs to inputs (Charnes, Cooper, & Rhodes, 1978). A well-known method of assessing efficiency is data envelopment analysis (DEA), which is a linear programming based technique introduced by (Charnes et al., 1978). It measures the relative performance of decision making units (DMUs) such as transshipment- or haulage segments in the case of IFT-chains. DEA is a non-parametric approach to analyse relative efficiency of DMUs and does not require assumptions about the functional form (Charnes, Cooper, Lewin, & Seiford, 1994). Therefore it is typically used when the presence of multiple inputs and outputs makes it difficult to compare alternatives.

Most DEA models are used to analyse a single process that acts as a black box, where aggregated inputs and outputs of the single process are analysed and no attention is paid to the inner structure of the process (Lozano & Gutiérrez, 2014). An IFT-chain could be seen as a single process, but next to inputs and outputs also consists of intermediate flows between chain divisions. Liang, Yang, Cook, and Zhu (2006) were the first to come up with a DEA model that takes into account intermediate flows within a process, which is called a Network DEA (NDEA). Tone (2001) developed a slacks-based DEA model which deals with input excesses and output shortfalls. Tone and Tsutsui (2009) applied this slacks principle on a NDEA, which is suitable for performance measures for non-proportional changes in inputs and outputs.

Regular NDEAs assume a fixed chain structure, which is a fixed number of chain divisions, but in the case of IFT, this is often not the case. To compare IFT-chains with different chain structures, Saeedi et al. (2019) developed a slacks-based NDEA that takes these differences into account. However, this has not been applied to a realistic European IFT-network yet.

The remainder of this chapter consists of the objectives and research questions in Sections 1.1 and 1.2. These are followed by the scope in Section 1.3. Finally, the research approach of this thesis is presented in Section 1.4.

1.1. Objectives

The main objective of this thesis is to assess the performance of an intermodal freight network. This main objective can be divided into three sub-objectives. The first sub-objective is having an insight in the performance indicators that are used in the literature to evaluate IFT-networks. Secondly, this thesis aims to develop a reliable method to measure performance of an IFT-corridor. The third sub-objective is analysing the performance of European IFT-services at a network level.

1. Insight in the performance indicators used in literature for evaluation of IFT-chains. A literature review evaluating frequently used KPIs could give an insight on this.
2. Developing a reliable method to measure performance of an IFT-corridor. There are multiple challenges to this analysis.
   (a) Variable length of IFT-chains. When chains in an IFT-network have different structures, the efficiency should still be assessed.
   (b) Resource sharing in overlapping chain divisions. When different chains have overlapping divisions, the efficiency of these chains should still be assessed.
   (c) Aggregate parallel chains into a corridor. When parallel chains are part of the corridor, the efficiency of the corridors should still be assessed.
3. Analysing the performance of European IFT-services at a network level. The developed model with multiple model innovations should be applied on the European IFT-network. There is one main challenge for this objective.
   (a) Availability and quality of data of the European IFT-network. Data is needed for the different inputs and outputs of all analysed chain segments, which could be difficult to obtain because the firms that offer IFT-services often do not share data for competitive reasons. Furthermore, these data would preferably be relatively recent, since IFT-services are constantly evolving.

1.2. Research questions

In order to achieve the objectives of this thesis, the following main research question has been formulated:

How can the performance of an intermodal freight transport chain be assessed?

The main research question can be divided into the following three sub-questions:

1. How is performance measurement done currently in European intermodal freight transport?
2. How can the performance of an IFT-corridor in a network be assessed reliably?
   (a) How can the efficiency of IFT-chains with different structures be assessed?
   (b) How can the efficiency in overlapping IFT-chain divisions in a network be measured?
   (c) How can the efficiency of a corridor with parallel IFT-chains be measured?
3. How can the developed model be applied on a part of the European IFT-network?

1.3. Scope

The goal to shift from road to IFT has already been mentioned before, but another goal of the EC is to have “a fully functional and EU-wide multimodal TEN-T ‘core network’ by 2030, with a high-quality and capacity network by 2050” (European Commission, 2011, p. 9). The core network of this Trans-European transport network (TEN-T) is split in nine core network corridors (European Commission, n.d.-a). The application of the slacks-based NDEA in this thesis will be on a part of these nine core network corridors.

IFT-chains can be divided in haulage and transshipment segments. The haulage segments are pre-haulage, main haulage and end-haulage, with transshipment segments in between two consecutive haulage segments. The European corridors or chains consist mainly of road, rail and inland waterways (IWW) transport. However, within IFT, road freight is only considered for pre- and end haulage. Main haulage can then be either rail, IWW or a combination of these two.

This thesis focuses on certain segments of the IFT-network, which is containerized rail. Although rail infrastructure is shared between passenger and freight transport, the access to the network is managed by (national) rail infrastructure managers. Furthermore, there are not that many route alternatives within a chain. An overview of the scope is shown in Figure 1.1.

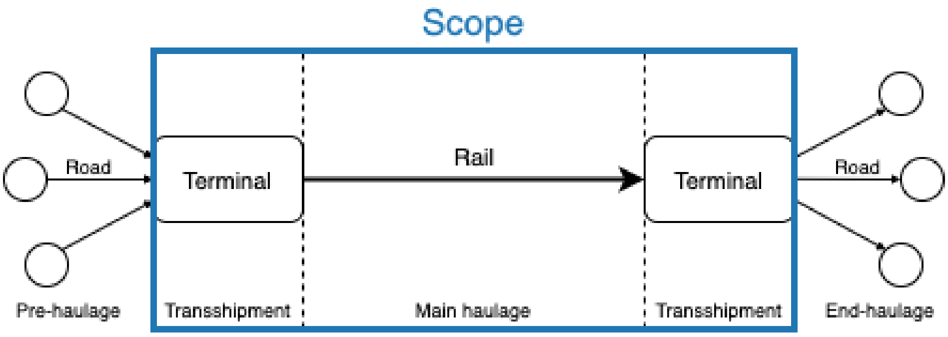


Figure 1.1: Scope of this research within the IFT-chain.

Origins and destinations can be anywhere within an analysed region, where transport to and from terminals is done by pre- and end haulage. Because the analysis of road pre- and end-haulage is too detailed for the scope of this thesis, it is excluded. Therefore, this thesis includes transport between terminals, including transshipment- and main haulage segments.

1.4. Research approach

An outline of the research approach is shown in Figure 1.2. The numbers represent the corresponding sections. Firstly, the background information of IFT is presented in Chapter 2. Chapter 3 follows with a literature review of the usage of KPIs in IFT. An explanation of (Network) DEA from literature and a newly adapted DEA is presented in Chapter 4. This model is then applied in Chapter 5 in a case study with a part of the TEN-T corridors, to evaluate the performance of the European IFT-network. Finally, the conclusions for this thesis, some limitations and recommendations for future research are discussed in Chapter 6.

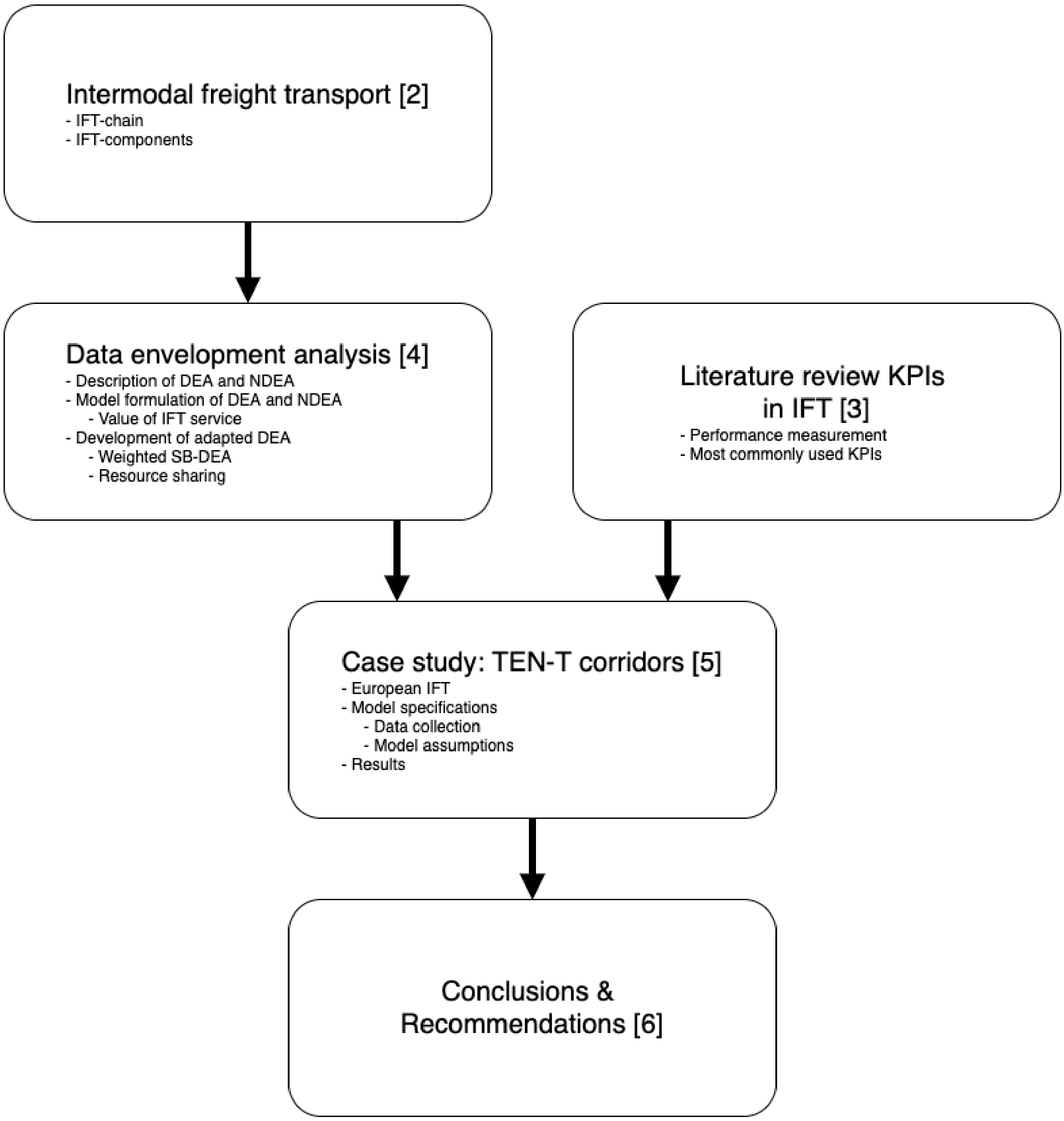


Figure 1.2: Outline research approach with corresponding chapters of this thesis.

2

Intermodal Freight Transport

This section provides the underlying theoretical background of intermodal freight transport (IFT). It gives an overview of definitions and components within IFT, making sure that there is a common basis of understanding. Section 2.1 explains the components of an IFT-chain. Section 2.2 describes the actors that play a role in IFT.

Intermodal transport is defined as the “movement of goods (in one and the same loading unit or a vehicle) by successive modes of transport without handling of the goods themselves when changing modes” (Organisation for Economic Co-operation and Development., 2008). The modes are commonly shortened to road, rail and inland waterways (IWW). In intermodal transport, the goods are transported in a loading unit with standardized dimensions, called an intermodal transport unit (ITU). An ITU is a “container, swap body or semi-trailer/goods road motor vehicle suitable for intermodal transport” (Eurostat, 2003). Containers are the most common in IFT.

2.1. IFT-chain

An IFT chain consists of transport and transshipment segments. The transport segments are pre-haulage, main haulage and end-haulage, with transshipment segments in between two consecutive transport segments. A typical IFT chain is shown in Figure 4.2, but there can be more than one main haulage segment.

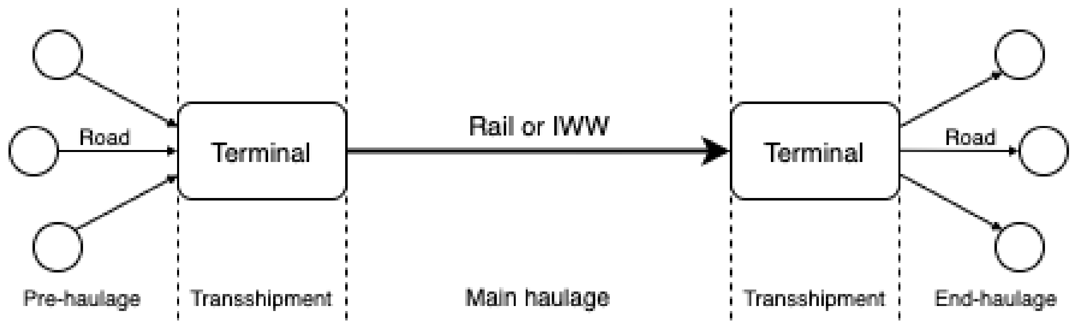


Figure 2.1: The process of intermodal transportation. Adapted from De Vries (2019, p. 8).

Pre- and end-haulage

Pre-haulage and end-haulage are the first and last segments of an IFT-chain. In pre-haulage, ITUs are collected at the consignor and transported by a carrier to the first transshipment division. This is typically done by road transport.

End-haulage is the last division of the IFT-chain. It connects the main haulage via a transshipment node to the consignee. Just like pre-haulage, this is mainly done by truck.

For intercontinental IFT, deep sea shipping is a main haulage division, beginning and ending in a transshipment node. When considering the IFT network of Europe however, deep sea shipping can be seen as preor end-haulage transportation. Containers can enter or exit the European IFT network at seaports, acting as transshipment nodes, which are connected to main haulage segments for European IFT.

Main haulage

Main haulage is the long-distance leg between different transshipment nodes. The first transshipment node bundles the freight from pre-haulage for transport by rail, inland waterway or short sea transport.

Air transport is not considered here as part of IFT, since the transportation units needed for air transportation are different from the standardized units used for IFT.

In the Netherlands, IWW transports by far the largest volume in IFT (*Mobiliteitsbeeld 2019*, 2019). In the rest of Europe however, rail plays a much larger role than IWW (*Freight transport statistics - modal split*, n.d.).

Transshipment

An essential part of the IFT-chain are the transshipment nodes. These are terminals, and their function is to transship freight from one mode to another. Because of that, they are usually connected to at least two modes, of which road is almost always one. Besides that, the terminals are connected to rail, IWW or both. Next to their transshipment function, they can also fulfill a function of storing the freight.

2.2. Actors in IFT

IFT requires the cooperation and involvement of many different actors. Every division of the IFT-chain requires a new actor. This subsection gives an overview of the most important actors of the IFT system.

Consignor

The consignor or shipper is typically the owner or supplier of the goods. The consignor wants to transport its goods to the consignee. The consignor usually has a contract with a transport integrator. In the IFT-chain, the consignor is located as origin, from where pre-haulage departs.

Consignee

The consignee is the actor to whom the goods are shipped. In the IFT-chain, the consignee is located as destination, where the end-haulage finishes.

Forwarder

The forwarder acts as agent for the consignor to transport its goods. The consignor wants to transport his goods from an origin to a destination with a set of quality requirements. The forwarder organises the full process of transportation on behalf of the consignor, including booking at intermodal transport companies, connecting services and possible formalities like customs.

The forwarder acts as intermediary for multiple consignors and usually has long-term agreements with logistic companies. Extra services they can offer are warehousing and providing ITUs.

Road haulier

The (road) haulier performs the road transport and is usually responsible for the pre- and end-haulage of the IFT-chain. He is the owner of the trucks and pays for the use of infrastructure to the road infrastructure manager. The road haulier is almost always a private actor.

Railway company

The railway company or railway undertaker provides rail transport and is usually responsible for the main haulage of the IFT-chain. He is the owner of rail transport equipment and pays for the use of infrastructure to the rail infrastructure manager. In contrast to road hauliers, railway companies can be both private and public operators.

Shipping company/Barge owner

The shipping company performs transportation via IWW. Similarly, a barge owner is an independent person who transports freight via IWW. Most vessels are operated by barge owners, usually for 14 hours per day, while shipping companies operate often 24 hours a day (Posset et al., 2010).

Infrastructure manager
Roads, railways and IWW all have their own infrastructure managers. The infrastructure manager provides the infrastructure for road, rail or IWW. It builds and maintains the infrastructure and also operates its safety and control systems.

The infrastructure managers charge the road hauliers, railway companies and shipping company for use of the infrastructure. The road infrastructure managers in the EU charge hauliers for use of the infrastructure via either via distance- or time-based charges (European Commission, n.d.-b). For IWW, the infrastructure manager does not charge for the use of free flowing waterways, but it does charge for the use of canals and locks (Posset et al., 2010). Most infrastructure managers operate on a national level.

Terminal operator
The terminal operator provides the transshipment of ITUs between two different modes of transport. They often also provide extra services like empty container depots and temporary storage facilities, since direct transshipment between means of transport is often impossible.

Intermodal transport company
The intermodal transport company or intermodal operator combines ITUs from different consignors in a single consignment. It buys a train or shipping service from the railway company, shipping company or barge owner in advance. An intermodal transport company can either provide only main haulage or door-to-door transport, including pre- and end-haulage.

3

Literature review performance measurement in IFT

As explained in Chapter 1, DEA is a method for performance measurement and it needs inputs and outputs. These can also be seen as individual key performance indicators (KPIs). It is interesting to identify the most commonly used KPIs for IFT in the literature, to give a direction for the choice of inputs and outputs in the application of (N)DEA models in the future.

Section 3.1 describes the theory about performance measurement, how to select these indicators and defines the key concepts. Section 3.2 then presents a literature review of KPIs in scientific literature, followed by practical literature in Section 3.3. Finally, the conclusions and some recommendations for future research are presented in Section 3.4.

3.1. Performance measurement

Performance measurement is “the process of quantifying the efficiency and effectiveness of action” (Neely, Gregory, & Platts, 1995). It is needed to understand and improve the performance of a system (Lindberg, Tan, Yan, & Starfelt, 2015). The metrics that are used to quantify performance measurement are performance measures or performance indicators (Neely et al., 1995). The most important performance indicators of a system are called key performance indicators (KPIs). In performance measurement terminology, firms or other entities of which performance can be measured are called Decision Making Units (DMUs) (Charnes et al., 1978). KPIs can be used to track and evaluate an individual DMU, but is also used for benchmarking comparable DMUs.

Performance indicators should be in line with a DMU’s goal and “reflect the degree to which an objective is being achieved” (Posset et al., 2010, p.41). These objectives can be expressed in terms of input, output or outcome Posset et al. (2010). Beamon (1999) gives input, output and flexibility as performance measure types or KPIs.

Inputs can be seen as resources needed for the process of a DMU (Cook & Zhu, 2014; Farrell, 1957), where outputs should capture what the DMU generates (Cook, Green, & Zhu, 2006). Examples in IFT are transportation and handling cost as input (Beamon, 1999) and value of IFT service based on transshipment or haulage volumes and prices (Saeedi et al., 2019). Although the difference between inputs and outputs (resources and production) seem quite clear, there are some cases in which a performance indicator can be seen as both input and output (Beasley, 1990). However, generally speaking, DMUs try to minimize inputs and maximize outputs (Cook et al., 2006).

In multiple-criteria decision analysis, performance indicators are often called attributes, which can be divided in natural, proxy and artificial attributes (Eisenfuhr, Weber, & Langer, 2010). Table 3.1 (Afsharian et al., 2016) gives an explanation of the differences and use of these attributes, together with an example for each type of attribute from IFT.

Neely et al. (1995) and De Toni and Tonchia (2001) both give 4 distinctive performance dimensions, namely cost, time, flexibility and quality. The selection of KPIs depends on the scope and domain of the research. Specifically, the performance measurement should be in line with the DMU’s goal, since "Performance is defined as the fulfillment of goals pursued" (Afsharian et al., 2016).

Table 3.1: Characterization of attributes as performance criteria to measure the achievement of objectives. Adapted from Afsharian et al. (2016, p.1897).

Type of attribute	**Description**	**Example**	**Remarks**
Natural attribute	Indicator directly related to the objective	*Transshipment time* and *haulage time* as natural attributes for the objective *total shipping time*	Prerequisite is that attributes are equivalent to the objective
Proxy attribute	An indicator or a means for the achievement of an objective	*Haulage distance* as proxy attribute for the objective *fuel consumption*	Data availability may be the reason why *fuel consumption* itself is not measured
Artificial attribute	A constructed combination of indicators relevant for the objective	Combination of country-specific *network access charges*, *personnel costs* and *emissions* as artificial attributes for the objective *railway costs*	Artificial attributes can be seen as a combination of several proxy attributes

According to Afsharian et al. (2016), three steps are needed for choosing performance indicators for DEA. We have generalized these for different methodologies with the following three steps:

1. Development of a system of objectives. The goals for the analyzed DMUs should be set and split in lower-level goals;

2. Derivation of suitable performance indicators. Ideally, these indicators would be natural attributes, but when these are not available, proxy or artificial attributes can be used (see Table 3.1 (Afsharian et al., 2016));

3. Construction of estimation functions. When dealing with proxy or artificial attributes, an estimation for the objective should be made.

To clarify the three steps, we use rail haulage as example in IFT. We use a selection of indicators and sub-indicators from Islam, Zunder, and Jorna (2013) as simplification for explanation purposes. The three steps of the example are shown in Table 3.2. In the first step, we could say the main goal is customer satisfaction. This can be split into lower-level goals. Some lower-level goals for customer satisfaction are transport cost, transport time, emissions and reliability. The next step is to find suitable performance indicators. These indicators are can be found in Table 3.2. For transport cost and time, there are natural attributes that are usually readily available for DMUs. Emissions are more difficult however, so haulage distance is used as proxy indicator. Reliability is here an artificial attribute, which uses punctuality and complaints. The third and last step is the construction of estimation functions. With natural attributes this is straightforward, but for emissions and reliability, factors (β ond γ) are needed to account for proxy and artificial attributes.

The objectives from Table 3.2 that should be minimized can be seen as inputs, the objectives to be maximized as outputs. Emissions seems to be an exception of this. Emissions are produced in the IFT process and are therefore an output. However, undesirable outputs should be minimized and in calculating efficiencies this is solved by modelling them as input (Cook & Zhu, 2014). Depending on the goals of a DMU, another possibility is to choose fuel consumption as objective, which is an input to be minimized and is directly related to emissions.

Another observation is that the main goal and lower-level objectives are customer oriented. However, most of the performance indicators are on an operational level. The estimation level is the connection between these two levels. This can be seen parallel to the customer-operator relationship. A customer is mainly interested in an aggregation of main indicators, while the operator brings together many resources on a detailed level to get to this aggregation and is therefore interested in more detailed indicators. The levels of customer, operator and system integrator and their KPIs are described in PROMIT (Davydenko, Jordans, & Krupe, 2007) and will be discussed in Section 3.3.

The example from Table 3.2 showed some of the objectives and indicators used for IFT as shown in (Islam et al., 2013). It is interesting however, to see what other literature says about the use of performance indicators in IFT. There are existing literature reviews that evaluate the use of KPIs in European IFT (Posset et al., 2010; Posset, Pfliegl, & Zich, 2009). To the knowledge of the author, the most recent is from 2010 (Posset et

Table 3.2: Rail haulage example for determining IFT-criteria according to three steps of Afsharian et al. (2016).

Goal level	Fundamental goal: customer satisfaction						
	Objectives to be minimized					Objectives to be maximized	
	x_1: Transport cost		x_2: Transport time		x_3: Emissions	y_1: Reliability	
	↓		↓		↓	↓	
Estimation level	$x_1 = VC + NC$		$x_2 = WT + DT$		$x_3 = \beta_1 HD$	$y_1 = \gamma_1 P + \gamma_2 C$	
	↑	↑	↑	↑	↑	↑	↑
Performance indicator level	VC Vehicle cost	NC Network charge	WT Waiting time	DT Driving time	HD Haulage distance	P Punctuality	C Complaints

al., 2010). The more recent literature review by Saeedi, Behdani, Wiegmans, and Zuidwijk (2018) evaluates the performance indicators used in various modes of transport, it is however not focused on IFT in Europe. Therefore, there is a gap in literature of nearly a decade of literature in European IFT, which this chapter of the thesis aims to fill.

The rest of this chapter contains of a literature review about the most widely used KPIs within IFT in both scientific and practical literature. This can give a direction for the choice of inputs and and outputs in the application of (N)DEA models in the future.

3.2. KPIs in scientific literature

This section presents a literature review of KPIs in scientific literature. First, Section 3.2.1 explains the methodology to select relevant scientific literature. Section 3.2.2 then presents the results.

3.2.1. Methodology scientific literature

For the review of scientific literature, the methodology from Wee and Banister (2016) is used to determine the relevant literature. The literature search is performed using the Scopus database (`www.scopus.com`). The selection methodology is shown in Figure 3.1 and the extensive version in Figure A.1 in Appendix A.

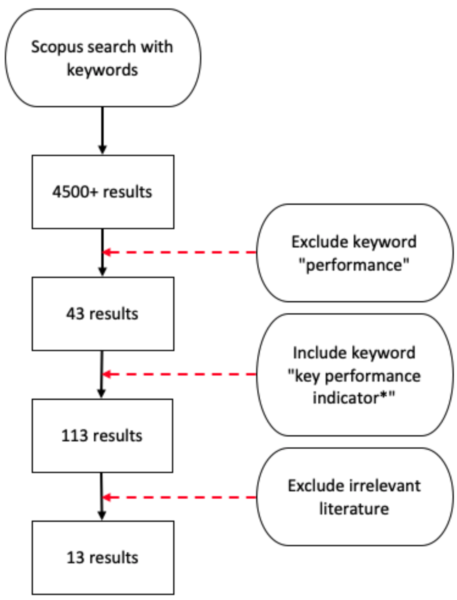


Figure 3.1: Methodology of selecting papers in the literature review.

Firstly, the following keywords were used: “performance” or “KPI”, “intermodal”, “transport*”. This resulted in too many results, where “performance” did not seem to be specific enough and resulted in mostly irrelevant papers. After removing this search term, there were too few results. Adding KPI fully written out, “key performance indicator”, resulted in 113 results. For the application in Chapter 5, we are interested in European IFT. Finally, the 13 results where manually filtered for their relevance, excluding papers not regarding Europe or IFT.

It is notable that the literature found in the last Scopus search with 113 results, are all dated 2006 and later. 95 of the 113 results (84%) are from 2013 onwards. This could either mean that this topic has become much more popular in recent years, or that the Scopus database provides only limited access to less recent literature. This is to be noted but is not a problem for this thesis, since this gives a better overview of up-to-date research in the field of performance in IFT.

Backward snowballing is used to find more background literature about performance and key performance indicators but is not included in the final selection of reviewed papers to have a consistent methodology.

3.2.2. Results KPIs scientific literature

The 13 selected papers are presented in Table 3.3. The papers have been categorised within two main categories. 7 papers belong to terminals & ports and 6 to chains & transportation, both containing 4 domains. Various different KPIs have been used and these will be discussed in the following paragraphs.

Table 3.3: KPIs from papers selected in the literature review, sorted by terminal & ports and chains & transportation.

Paper	**Domain**	**KPIs**
Chains & transportation		
Ivanov and Sokolov (2013)	Supply chains	• Robustness • Flexibility • Costs • Level of service
Fancello et al. (2018)	Supply chains	• Travel time • Waiting time • Handling time • Tariffs
Posset et al. (2009)	Inland waterways	• Employment • Emissions • Maintenance costs per tkm • ITS coverage • Infra availability • Reliability • Capacity ports • Utilization ports • Storage time ports • Safety & security
Islam et al. (2013)	Freight transport chains	• Transport cost • Transport time • Flexibility • Reliability • Quality • Sustainability

Table 3.3: KPIs from papers selected in the literature review, sorted by terminal & ports and chains & transportation.

Paper	**Domain**	**KPIs**
Gianpiero et al. (2015)	IFT	• Emission costs • Congestion costs • Fuel costs • Human resources costs • Road + rail costs
Panagakos and Psaraftis (2017)	IFT	• Cost • Speed • CO2 & SOx
Terminals & ports		
Gogas et al. (2017)	Intermodal terminals	• Transshipment time • Handling cost • Punctuality • Loss and damages • Supply chain visibility • Information availability • Employee & equipment productivity • Fair and equal access companies • Safety and security • Environmental burden • Saturation ratio
Ricci et al. (2016)	Intermodal terminals	• Handling equipment rate • Storage • Energy consumption • Equipment performance • Truck waiting ratge • Terminal occupancy • Reliability • Equipment haul • Maintainability • System utilization rate • Personel distribution rate
Nathnail et al. (2016)	Intermodal terminals	• Handling cost • Punctuality • Employee & equipment productivity • Transshipment time • Loss and damages • Supply chain visibility • Information availability • Saturation ratio • Fair and equal access companies • Safety and security • Environmental burden

Table 3.3: KPIs from papers selected in the literature review, sorted by terminal & ports and chains & transportation.

Paper	**Domain**	**KPIs**
Ballis (2004)	Intermodal terminals	• Waiting time • Reliability • Flexibility • Safety & security • Terminal accessibility
Antognoli et al. (2018)	Rail terminals	• Total transit time • Utilization rate of equipment/storage/system • Energy consumption rate • Capacity of equipment • Truck waiting rate • Terminal occupancy • Reliability
Schipper et al. (2017)	Ports	• Environment • Volume • Quality of handling • Investments • Market share
Di Pierro et al. (2017)	Ports	• Waiting time • Throughput • Customer satisfaction ITS • Vehicle speed • Travel time • Distance • Fuel consumption • CO2 • Weather conditions

It is interesting to see that there are almost 50 different terms for the KPIs used and after combining some comparable KPIs like system utilization rate and utilization of ports, there are still more than 40 somehow different KPIs. A list of all different KPIs and their frequencies can be found in Table A.1 in Appendix A. The most common KPIs that were used are environment & emissions and waiting time, used in 8 respectively 6 of the 13 selected papers. Other common KPIs for both domains are transshipment time, employee productivity, safety & security, energy consumption, reliability and transportation costs, all at least present in 4 of the papers.

Some externalities are included as KPIs, such as safety, congestion costs and environment & emissions as the most common KPI. Including these makes sense from a social optimum perspective since these are costs for the entire society. Most of the selected scientific literature have a societal perspective, which explains the frequent choice of externalities as KPI. Externalities are usually less important from a company's perspective (Bask & Rajahonka, 2017). However, since fuel consumption is closely related to emissions, it is logical that also companies would include some KPI related to externalities (Macharis, Van Hoeck, Pekin, & van Lier, 2010).

Some KPIs are mentioned mostly for terminal & ports. These are transshipment time, reliability, safety & security, employee and equipment productivity and volume. The last two are not mentioned for chains & transportation at all, so they are terminal specific. For equipment productivity this is expected, since it has to do with terminal specific equipment. Economies of scale play an important role in intermodal terminals (B. W. Wiegmans, Masurel, & Nijkamp, 1999), so it is understandable that volume is used as KPI for terminals. It is interesting however, that volume is not used in any of the papers regarding chains & transportation, since there are also benefits with economies of scale for transportation (Crainic, 2003).

Besides the aforementioned environment & emissions, waiting time and reliability, there are two chains & transportation specific KPIs, which are transport costs and travel time. It is interesting that there is not a large difference between the KPIs used for chains & transportation and terminals & ports, showing that the interests are often aligned.

3.3. KPIs in practical literature

Besides the scientific literature, there is also information available about the performance indicators in the so called grey literature. Types of grey literature include reports and white papers from organizations such as governments departments and agencies, private companies and consultants. This literature can add insight, because it provides a practical view of the approach to measure and evaluate performance in IFT.

First, the methodology for selection of practical literature is presented in Section 3.3.1. Secondly, Section 3.3.2 presents the results of the practical literature review.

3.3.1. Methodology practical literature

The methodology is different from methodology for the scientific literature. Search engines Google and Ecosia were used with a combination of the following search terms: 'IFT', 'Intermodal freight transport', 'performance', 'performance indicators', 'KPI'. Furthermore, the websites of some well-known organisations and projects regarding IFT or segments of IFT-chains were visited. From the search results, forward snowballing was used to find more relevant literature (Wee & Banister, 2016). A weakness of this methodology is that it leads to a relatively subjective and it depends heavily on the results from the used search engines. However, the strength of this methodology is that it leads to a varied range of sources.

3.3.2. Results KPIs practical literature

The EU funding project Connecting Europe Facility (CEF) has the following objectives, as stated in article 4(2) of Regulation No 1316/2013 of the European Parliament and of the Council of 11 December 2013 establishing Connecting Europe Facility: removing bottlenecks, ensuring sustainable and efficient transport systems and optimising the integration and interconnection of transport modes. The measurements or indicators for these objectives are on a macro level, such as the length of upgraded rail network, the number of inland ports with alternative fuel supply points and the number of connections of terminals to the railway network.

International Union of Railways (UIC) (2016) gives macro indicators like intermodality, safety and environmental footprint, without mentioning the specific ways to measure these indicators. Infrastructure managers like Austrian ÖBB often also have a macro view on performance indicators, such as market share, trains delayed due to infrastructure (service reliability), service coverage (accessibility) and train km per track km (asset utilization) (ÖBB, n.d.). These macro views give an overview of the relative importance of indicators in IFT for different stakeholders. Most of these indicators are focused on the entire network, but some could also be used to evaluate the performance of IFT-chains or individual transport or transshipment segments.

The Organisation for Economic Co-operation and Development (OECD) (2002) uses the following four indicators that can be used to evaluate individual modes and routes: price of transportation, cost of total travel time, cost of total waiting time and value of container load.

Marco Polo II was a European funding programme from 2007-2013 which had the main goal of shifting 144 billion tonne-kilometres off the road. The indicators to evaluate this goal are tonne-kilometres shifted to rail, IWW and SSS, tonne-kilometres of road traffic avoided and externalities such as reduction of growth in air pollution, noise, congestion and traffic accidents (ECORYS, 2004). It is interesting to see that all are outputs or outcomes.

PROMIT (Davydenko, Jordans, & Krupe, 2007) has KPIs for three actor levels of abstraction. Level 1 is the customer or shipper level, where their main interest is the transport from source to destination. Level 2 consists of the viewpoint for the logistics service providers, where the chain consists of links of black box activities (Davydenko, Zomer, & Krupe, 2007). Level 3 depicts the operations of each of these activities from level 2. These three levels make it possible to include indicators from the different types of actors. The structure and the main KPIs of these three levels can be found in Table B.1 (Posset et al., 2010) in Appendix B.

It is interesting to see that many KPIs apply to all three levels, such as price, lead time and reliability (Davydenko, Jordans, & Krupe, 2007). The operator level (level 3) however, also has more operation specific KPIs. An example of these operation KPIs are storage capacity and train arrival rate for intermodal terminals, driving/waiting time ratio for trains and cargo capacity and fuel consumption for IWW (Davydenko, Jordans, & Krupe, 2007). An overview of all of these KPIs can be found in Table B.2 (Posset et al., 2010) in Appendix B.

Table 3.4: Most important indicators for IFT in general, and the main domains terminals & ports and chains & transportation according to scientific and practical literature.

	Scientific literature	**Practical literature**
IFT	Environment & emissions Reliability Waiting time	Price Reliability Capacity Capacity utilization
Terminals & ports	Transshipment time Safety & security Employee & equipment productivity Volume	Waiting time Cranes
Chains & transportation	Travel time Transport costs	Travel time

3.4. Conclusions & recommendations

This chapter discussed performance measurement in European IFT. Performance measurement should stem from a DMU's goal. KPIs come from the lower-level goals of a DMU and can be divided in inputs and outputs. The abstraction level of KPIs can be on the broader customer level or the more detailed operator level. The perspective determines which KPIs should be used.

To evaluate an IFT-network, there are two requirements. The first is the parameters with which the network is assessed. These are the KPIs as discussed in this chapter. These are then applied in a model, such that efficiency of IFT-chains can be assessed. The development of such a model is discussed in the next chapter. The parameters from this chapter and the model from next chapter will then come together in Chapter 5, where the model will be applied with the performance indicators.

The scientific literature was reviewed for commonly used KPIs in European IFT. Of the 13 selected scientific papers, the three most frequently used KPIs are environment & emissions, waiting time and reliability. Some domain specific KPIs are equipment productivity and volume for terminals & ports, for chains & transportation these are transport costs and travel time. The literature review of practical literature came with many KPIs on a macro view. PROMIT (Davydenko, Jordans, & Krupe, 2007) gave a more detailed overview of KPIs used for different abstraction levels, which are in line with the most common KPIs from scientific literature. Depending on the chosen abstraction level, this gives a good indication of the choice for KPIs for future analyses of European IFT. The most common KPIs for both scientific and practical literature can be found in Table 3.4. The KPIs are needed to assess

There are some limitations regarding the methodology for the scientific literature. There are many more papers available regarding European IFT and KPIs, but these didn't show up in the literature review with the final search terms in Scopus. One way to get a larger number of papers is using backward or forward snowballing (Wee & Banister, 2016). In the case of KPIs for IFT, this results in hundreds of papers. Selecting only some of them would be rather arbitrarily, therefore we have chosen to not follow this approach for this thesis. However, we recommend to use snowballing starting at the current selection of papers for future research to get a broader overview of KPIs for European IFT.

To evaluate an IFT-network, there are two requirements. The first is the parameters with which the network is assessed. These are the KPIs as discussed in this chapter. These are then applied in a model, such that efficiency of IFT-chains can be assessed. The development of such a model is discussed in the next chapter. The parameters from this chapter and the model from next chapter will then come together in Chapter 5, where the model will be applied with the performance indicators.

4
Data Envelopment Analysis for Intermodal Freight Transportation

Chapter 3 resulted in the most frequently used KPIs in literature. We can use some of these in a model to evaluate the performance of an IFT-chain. The two main methods for benchmarking are Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). DEA is a non-parametric method and does not require assumptions about the functional form (Charnes et al., 1994). Because it is often difficult to determine the functional form for efficiency in IFT, DEA is preferred over parametric approaches such as SFA. Therefore, DEA will be used in this thesis to evaluate the performance of IFT-chains.

This chapter describes the method of Data Envelopment Analysis in detail. In Section 4.1, the principles of DEA and Network DEA are discussed. Section 4.2 presents the formulation of these models, followed by a description of Value of IFT Service (VIFTS) and some of its issues in Section 4.3. In Section 4.4, we propose an adapted DEA in Section 4.4. Finally, Section 4.5 gives a conclusion of this Chapter.

4.1. Description of (Network) Data Envelopment Analysis

An IFT-network can be split in separate chains, which consist of transshipment- and haulage segments, represented by nodes and links respectively. The firms that operate on these separate chains are called decision-making units (DMUs) in the literature, which can be used for different scales in traditional DEA or Network DEA. A traditional DEA is often used for an analysis of comparable segments (e.g. rail terminals) and DMUs represent individual segments. For analyses on a network level, a Network DEA is more commonly used. Here, DMUs represent entire IFT-chains and divisions represent individual segments such as transshipment and haulage segments. This section will first describe the traditional DEA, followed by the Network DEA.

As seen in Chapter 3, performance is often measured with KPIs. In a DEA however, performance is seen as efficiency of a segment, where performance indicators can be used as inputs and outputs for the DEA. In the rest of this thesis, efficiencies will be the main measure of to describe performance.

The efficiency of DMUs can be explained as the ratio of outputs to inputs (Charnes et al., 1978) as presented in (4.1), (4.2) and (4.3). Maximizing efficiency can be done by either minimizing inputs for fixed outputs or maximizing outputs for fixed inputs (Ockwell, 2001).

$$\text{efficiency} = \frac{\text{output}}{\text{input}} \tag{4.1}$$

$$\text{efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}} \tag{4.2}$$

$$\text{efficiency} = \frac{\sum_r u_r y_r}{\sum_i v_i x_i} = \frac{u_1 y_1 + u_2 y_2 + \ldots}{v_1 x_1 + v_2 x_2 + \ldots} \tag{4.3}$$

Where: u_r is the weight for output r

y_r is output r

v_i is the weight for input r

x_i is input i

4.1.1. Traditional DEA

Data Envelopment Analysis (DEA) is a method to evaluate the efficiency of a single DMU relative to all other representative DMUs (e.g. one rail operator on a certain link with the other representative rail operators on individual links). DEA is a non-parametric approach which does not require a functional form, meaning that the weights for the input and output variables (v_i and u_r) are not determined beforehand.

A traditional DEA model can be applied in the following three steps (Golany & Roll, 1989):

1. Selection of the DMUs. All DMUs should perform similar activities and use similar technology, having the same input and output variables;

2. Selection of input and output variables. These variables are used to analyse the efficiency of the DMUs. This can be done by using the goal-oriented approach as described in Section 3.1;

3. Selection of DEA method and its application.

An example of these three steps can be seen in an analysis of inland waterway terminals by B. Wiegmans and Witte (2017). The selection of DMUs here are container terminals in Europe, using DEA to compare the efficiency of these terminals. For this case, the inputs are the the number of cranes, quay length and terminal area among others, with the handling capacity and throughput of the terminal as output. Finally, the DEA method needs to be selected, which depends on two aspects: returns-to-scale assumption and model orientation. B. Wiegmans and Witte (2017) used an input oriented DEA with constant returns-to-scale (CRS).

The *returns-to-scale assumption* depends on the used technology and can be constant or variable. A constant returns-to-scale (CRS) model can be used when the ratio of outputs to inputs (efficiency) is constant independent of the scale, i.e. when an increase in inputs always gives the same relative increase in outputs (Charnes et al., 1978). When this assumption does not hold up, the model has variable returns-to-scale (VRS) (Banker, Charnes, & Cooper, 1984). The *model orientation* depends on the aim of the analysis. This can be input-oriented with minimizing costs for a fixed output, or output-oriented with maximizing outputs for a fixed input. Since the outputs in IFT are dependent on external factors such as demand, inputs are usually minimised and therefore an input-oriented model is used in this thesis. Figure 4.1 shows the CRS- and VRS-frontier for a sample of five DMUs, with a fixed output and two input variables per DMU in an input-oriented DEA.

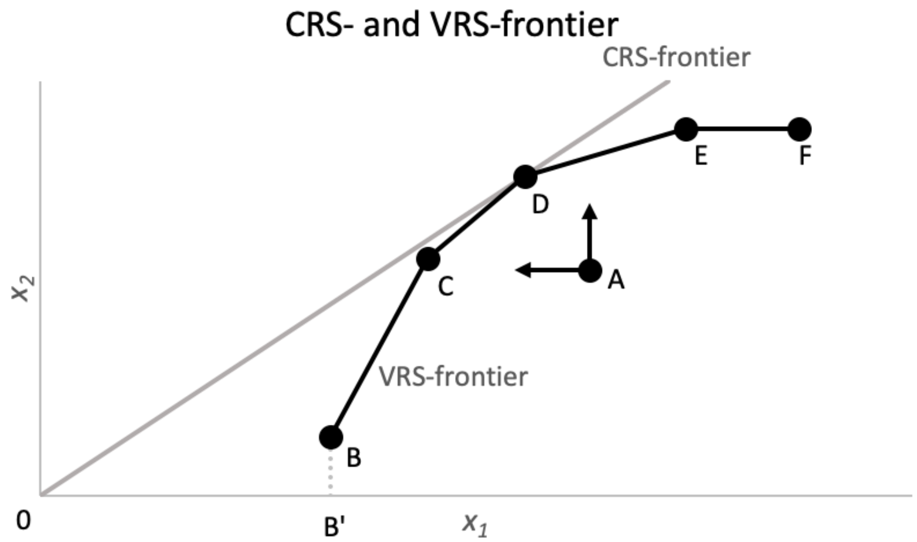


Figure 4.1: CRS- and VRS-frontier for an input-oriented DEA model. x_1 and x_2 are inputs, with a fixed output. DMUs B, C, D, E and F are on the VRS-frontier and are efficient according to VRS, while A is not efficient. According to CRS, only DMU D is efficient.

The traditional DEA is useful for comparing individual firms on either transshipment or haulage segments as DMUs. This can give some useful information about the performance of individual segments of IFT. However, to look further to the performance of entire IFT-chains, we need to use another method.

4.1.2. Network DEA

Network DEA can be used to evaluate the performance of an entire IFT-chain consisting of multiple segments. An IFT-chain could be seen as a single process, but next to inputs and outputs also consists of intermediate flows between chain divisions, as shown in Figure 4.2. Furthermore, each chain is often active in multiple chains, making it difficult to attribute the right amount of resources to each chain.

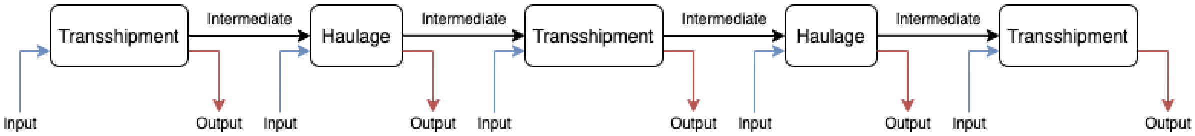


Figure 4.2: Example of IFT-chain with three transshipment and two haulage segments. Besides inputs and outputs for each division, there are intermediates between consecutive divisions.

Most DEA models are used to analyse a single process that acts as a black box, where aggregated inputs and outputs of the single process are analysed and no attention is paid to the inner structure of the process (Lozano & Gutiérrez, 2014). An IFT-chain could be seen as a single process with these inputs and outputs, but also consists of intermediate flows between chain divisions. Liang et al. (2006) came up with a DEA model that takes into account intermediate flows within a process, which is called a Network DEA (NDEA). Tone (2001) developed a slacks-based DEA model which deals with input excesses and output shortfalls (such as DMU F in Figure 4.1). Tone and Tsutsui (2009) applied this slacks principle on a Network DEA, which is suitable for performance measures for non-proportional changes in inputs and outputs.

The aforementioned (slacks-based) NDEAs assume a fixed chain structure, meaning a fixed number of chain divisions, but for IFT, this is often not the case. To compare IFT-chains with different chain structures, so with a variable amount of divisions for each chain, Saeedi et al. (2019) developed a slacks-based NDEA (SB-NDEA) that takes these differences into account, this is further explained in Section C.2 in Appendix C. This has not yet been applied to a realistic and important part of the European IFT network yet.

4.2. Model formulation (Network) Data Envelopment Analysis

This section presents the model formulation for the traditional DEA and the later developed slacks-based NDEA as formulated by Saeedi et al. (2019).

4.2.1. Traditional DEA

Equation (4.3) shows the objective function of the traditional DEA model that needs to be maximized for every DMU_a. Maximizing this objective function subject to some constraints can be formulated as a linear programming (LP) problem. This problem is known as the multiplier problem, with the weights for the input and output variables as unknowns (Santos, Negas, & Cavique, 2013). The dual of the multiplier problem is the envelopment problem, which is the most widely used formulation in DEA literature. It indicates which efficient DMUs the inefficient DMU should be compared with to find the source of the input inefficiencies. Those efficient DMUs are referred to as reference set or peer group (Santos et al., 2013).

First we present the traditional input-oriented envelopment model. We want to calculate the efficiency of DMU_o by comparing it with n DMUs, where each DMU_j ($j \in J, J = 1, 2, \ldots, n$) uses m inputs $x_{i,j}$ ($i \in I, I = 1, 2, \ldots, m$) and s outputs $y_{r,j}$ ($r \in R, R = 1, 2, \ldots, s$). The envelopment model is formulated as follows:

$$\min_{\lambda} Z_o \tag{4.4}$$

s.t.

$$\sum_{j \in J} \lambda_j x_{i,j} \leq Z_o x_{i,o} \qquad \forall i \in I \tag{4.5}$$

$$\sum_{j \in J} \lambda_j y_{r,j} \leq Z_o y_{r,o} \qquad \forall r \in R \tag{4.6}$$

$$\lambda_j \geq 0 \qquad \forall j \in J \tag{4.7}$$

λ_j is called the multiplier and describes the weight of each DMU_j compared to DMU_o. The DMUs for which $\lambda_j > 0$ can be seen as the peer group. The formulation of the DEA model can be interpreted as follows:

"Given DMU_o, find the composite unit which has no smaller outputs than this one and whose inputs are smaller than those of DMU_o scaled down by a factor Z_o as small as possible." (Santos et al., 2013)

The formulation of (4.4) to (4.7) assumes constant returns-to-scale (CRS). Adding the constraint of (4.8) makes this various returns-to-scale (VRS).

$$\sum_{j\in J} \lambda_j = 1 \tag{4.8}$$

Figure 4.1 shows the VRS frontier for a sample of DMUs. DMU E and F are both on the frontier and would both have an efficiency of 1 according to the traditional DEA model. However, DMU F could still reduce its input x_1 to reach DMU E, with this property DMU F is called weakly efficient. This input reduction is called input slack and also output slacks may exist (Tone, 2001).

To incorporate the inefficiency of slacks into the DEA, Tone (2001) developed a slacks-based DEA, formulated as below.

$$\min_{\lambda, s^-_{i,o}} \rho_o = \frac{1 - \frac{1}{m}\sum_{i\in I}\frac{s^-_{i,o}}{x_{i,o}}}{1 - \frac{1}{m}\sum_{r\in R}\frac{s^+_{r,o}}{y_{r,o}}} \tag{4.9}$$

s.t.

$$\sum_{j\in J} \lambda_j x_{i,j} + s^-_{i,o} = x_{i,o} \qquad \forall i \in I \tag{4.10}$$

$$\sum_{j\in J} \lambda_j y_{r,j} + s^+_{r,o} = y_{r,o} \qquad \forall r \in R \tag{4.11}$$

$$\lambda_j, s^-_{i,o}, s^+_{r,o} \geq 0 \qquad \forall j \in J, \forall i \in I, \forall r \in R \tag{4.12}$$

ρ is the efficiency score for DMU_o. $s^-_{i,o}$ and $s^+_{r,o}$ With this slacks-based model, a weakly efficient DMU such as DMU F in Figure 4.1 has an efficiency lower than 1. The slacks-based model assumes CRS, adding the constraint of (4.8) would turn it into Variable Returns-to-Scale (VRS).

4.2.2. Network DEA

To look at the efficiency of a chain instead of only individual segments like transshipment or haulage, the Network DEA can be used. Its formulation is an extension of the (slacks-based) DEA.

Most NDEAs like the NDEA by Lewis and Sexton (2004) and the SB-NDEA by Tone and Tsutsui (2009) have chains with a fixed structure, but Saeedi et al. (2019) modified the latter for chains with a variable number of divisions, as is often the case in a network with different IFT-chains. Traditional NDEAs compare divisions based on their position in a chain, while Saeedi et al. (2019) takes into account that there can be divisions with comparable activities on different positions in the chain (compare e.g. a terminal to all other terminals). This increases the discriminatory power of the model, but with the requirement that compared divisions undertake similar activities and produce comparable services (Dyson et al., 2001). For more information about the advantage of the input-oriented modified NDEA, we refer to Saeedi et al. (2019). This is the model that is described below.

Because the focus is on a network level, chains are here considered as DMUs, each divided in a number of divisions (e.g. transshipment and haulage segments). The model considers n DMUs, where each DMU_o ($o \in O, O = 1,2,,\ldots,n$) consists of a set of divisions K_o. Each division k consists of m_k inputs $x_{i,k}$ and s_k outputs $y_{r,k}$.

$$\min_{\lambda_k, s^-_{i,k}} \rho_o = \sum_{k\in K_o} w_k \left(1 - \frac{1}{m_k}\sum_{i\in I_k}\frac{s^-_{i,k}}{x_{i,k}}\right) \tag{4.13}$$

s.t.

$$\sum_{c \in C_k} \lambda_{k,c} x_{i,c} + s^-_{i,k} = x_{i,k} \qquad \forall k \in K_o, \forall i \in I_k \tag{4.14}$$

$$\sum_{c \in C_k} \lambda_{k,c} y_{r,c} - s^+_{r,k} = y_{r,k} \qquad \forall k \in K_o, \forall r \in R_k \tag{4.15}$$

$$\lambda_{k,c}, s^-_{i,k}, s^+_{r,k} \geq 0 \qquad \forall k \in K_o, \forall c \in C, \forall i \in I_k, \forall r \in R_k \tag{4.16}$$

$$\sum_{c \in C_k} \lambda_{k,c} = 1 \qquad \forall k \in K_o \tag{4.17}$$

$$\sum_{c',c \in C_k} \lambda_{k',c} z_{c'c} = z_{k'k} \qquad \forall (k',k) \in K_o \tag{4.18}$$

$$\sum_{c',c \in C_k} \lambda_{k,c} z_{c'c} = z_{k'k} \qquad \forall (k',k) \in K_o \tag{4.19}$$

The objective ρ_o is the total efficiency of DMU_o and is shown in (4.13). This chain efficiency is the weighted efficiency of divisions k, with w_k the relative weight of division k, and $\sum_{k \in K_o} w_k = 1$, $w_k \geq 0 \forall k \in K_o$. The weight w_k is determined corresponding to the importance of division k, e.g. volume share (Tone & Tsutsui, 2009). Equations (4.14) to (4.17) are comparable to the slacks-based DEA. C_k is the set of all divisions with the same activities as division k. Equations (4.18) and (4.19) describe the linking constraints for the intermediate services within DMU_o. $z_{k'k}$ is the intermediate value from divisions k' to k, where k' and k are consecutive divisions in the same chain. Similarly, $z_{c'c}$, is the intermediate value from consecutive divisions c' and c. The description of these intermediates is given in Section 4.3. More explanation about the model formulation can be found in Appendix C.

Because of the weighted efficiency in the objective function and the use of λs from consecutive divisions in (4.18) and (4.19), the model optimises the efficiency of the entire chain instead of individual divisions.

4.3. Intermediates for NDEA

Saeedi et al. (2019) tested their modified NDEA by analysing an IFT case study. An IFT-chain usually consists of multiple transshipment and haulage segments. An intermediate should connect division k to the previous division k'. This means the intermediate should act as an input and output of both transshipment and haulage segments. A typical output for haulage segments is ton-km or TEU-km (Markovits-Somogyi, 2011). However, these cannot be interpreted as inputs for both transshipment and haulage segments and are therefore not fit as intermediates.

Saeedi et al. (2019) presents the Value of the IFT Service (VIFTS) as intermediate. The VIFTS represents the value that is created by transshipment and haulage in spatial value and time value. A product can have one value at the consignor, but another value at the consignee. By changing the location of these products, spatial value is created (Kilibarda, Andrejić, & Popović, 2013). Time value is created by delivering these products at the consignee at the required time (Kilibarda et al., 2013).

The VIFTS is based on hedonic pricing to show the relation between the price of service (freight charge), distance and time (Massiani, 2008). A freight charge function is defined by Massiani (2008) with characteristics such as distance, time and weight. Saeedi et al. (2019) gives the hedonic formulation for the VIFTS.

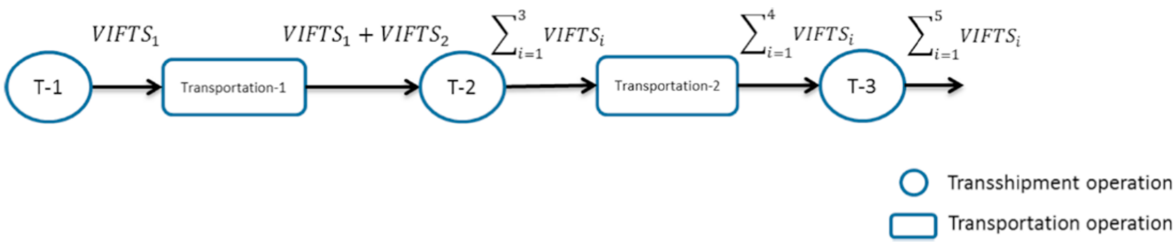


Figure 4.3: Incremental value of the service in an IFT-chain. Reprinted from Saeedi et al. (2019, p. 75).

The VIFTS by Saeedi et al. (2019) is a cumulative value of all divisions' intermediates located before the current division, as can be seen in Figure 4.3. From a theoretical standpoint, this means that the position of a division in the chain influences the value of the intermediate before and after the division. Because of (4.18) and (4.19), this influences the efficiency of a single division.

The dependency of efficiency on position could raise some issues. It means that efficiencies of individual divisions cannot be compared fairly, thus giving less information about the efficiency of individual segments of a chain. Furthermore, the length of a chain could have an impact on the overall efficiency of the chain, regardless of the actual performance of the chain. The numerical experiment as presented in Appendix D shows the differences in efficiency between different positions in an IFT-chain. While there is no difference seen in efficiency among positions for traditional DEA, the NDEA by Saeedi et al. (2019) has frequently a larger efficiency for terminals on later positions than on the first positions and this difference is significant.

It can be concluded that the use of VIFTS seems to be valuable as output for divisions, but the the cumulative character of VIFTS causes some issues in application of the model.

4.4. Adapted DEA for IFT

In this section, we propose adaption of the DEA for the use in IFT. We use several of the advantages from the modified NDEA by Saeedi et al. (2019), such as the coupling of all divisions within a chain and the substitutability of divisions to account for different chain structures. However, instead of the intermediate constraints from (4.18) and (4.19), we propose to use a non-cumulative output, such as volume or VaIFTS (Value added of IFT Service). This results in a combination of the slacks-based DEA and NDEA. The adapted DEA model is formulated in (4.13) to (4.17).

The idea of Saeedi et al. (2019) to optimise for entire chains instead of individual divisions is very relevant for the analysis of an IFT network. Also, using a set of all divisions with comparable activities in a NDEA increases the discriminatory power and accuracy of the model. These two advantages are still present in the adapted DEA model.

4.4.1. Weighted slacks-based DEA

The objective function as seen in (4.13) uses weights corresponding to the relative importance of divisions. In the case of IFT, the importance can be based on various indicators e.g. volume, cost share or a type of value. The choice of the indicator depends on the goal of the analysis and the availability of data.

4.4.2. Overlap and resource sharing

In IFT networks, there can be chains that partly overlap in a transshipment or haulage segment, an example is shown in Figure 4.4. In this case, resources (inputs) of the overlapping segments must be shared between overlapping chains. Using the actual proportions of resources used by each chain gives a good discriminatory power between chains in shared divisions. However, in IFT there can be a lack of available data (Tavasszy & de Jong, 2014), and the detailed level of the use of resources in a segment for different chains is especially difficult.

When it is unknown how the inputs for a division are divided between overlapping chains, other data can be used. This can be either the outputs like VaIFTS, or other data such as freight flows. Using one of these two can then be used as proportion of the resources used for each chain. A disadvantage is that this method gives similar efficiencies to overlapping segments in different chains, losing some of its discriminatory power. However, unless different actors such as the forwarders, railway companies and terminal operators share data on actual resource distribution, a substitute like outputs must be used.

4.4.3. Formulation adapted DEA

We propose a new formulation of a weighted slacks-based DEA (weighted SB-DEA) with resource sharing. The formulation in Section 4.2.2 does not take into account resource sharing, but we can rewrite the formulation to do so. Equations (4.13) to (4.15) use total inputs $x_{i,k}$ and outputs $y_{r,k}$ for division k. However, when sharing resources, they should be replaced by the share of inputs or outputs belonging only to the chain analysed (DMU_o). We call $x_{i,ko}$ the inputs and $y_{r,ko}$ the outputs for division k in DMU_o. The relationship between the total and chain inputs and outputs for a division are $\sum_{a\in O} x_{i,ka} = x_{i,k}$ and $\sum_{a\in O} y_{r,ka} = y_{r,k}$. The adapted model formulation is given in (4.20) to (4.24).

$$\min_{\lambda_k, s^-_{i,ko}} \rho_o = \sum_{k\in K_o} w_{ko}\left(1 - \frac{1}{m_k}\sum_{i\in I_k}\frac{s^-_{i,ko}}{x_{i,ko}}\right) \tag{4.20}$$

s.t.

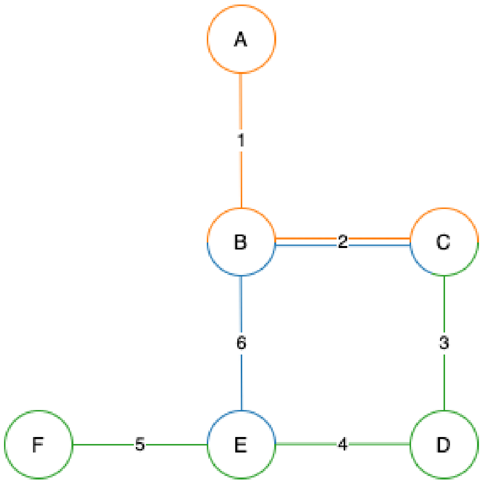


Figure 4.4: Example of IFT network with overlapping chains. The different colours represent different chains. Transshipment divisions B, C and E and haulage division 2 share their resources.

$$\sum_{c\in C_k} \lambda_{k,c} x_{i,ca} + s^-_{i,ko} = x_{i,ko} \qquad \forall k \in K_o, \forall i \in I_k, \forall a \in O \qquad (4.21)$$

$$\sum_{c\in C_k} \lambda_{k,c} y_{r,ca} - s^+_{r,ko} = y_{r,ko} \qquad \forall k \in K_o, \forall r \in R_k, \forall a \in O \qquad (4.22)$$

$$\lambda_{k,c}, s^-_{i,ko}, s^+_{r,ko} \geq 0 \qquad \forall k \in K_o, \forall c \in C, \forall i \in I_k, \forall r \in R_k \qquad (4.23)$$

$$\sum_{c\in C_k} \lambda_{k,c} = 1 \qquad \forall k \in K_o \qquad (4.24)$$

As explained in Section 4.4.2, the share of resources used for each DMU in IFT is often not known. To use a proportion of the resources, we need an estimation. This can be done by using the proportion of either one of the outputs or another data type such as freight flows. α_{ko} is the estimated proportion of resources for division k in DMU_o. Using output $y_{r,ka}$ with output type r or another data type f_{ka}, we can estimate the proportion of resources with either (4.25) or (4.26).

$$\alpha_{ko} = \frac{y_{r,ko}}{\sum_{a\in O} y_{r,ka}} \qquad \forall k \in K_o \qquad (4.25)$$

$$\alpha_{ko} = \frac{f_{ko}}{\sum_{a\in O} f_{ka}} \qquad \forall k \in K_o \qquad (4.26)$$

When the entire IFT network is analysed, $\sum_{a\in O} \alpha_{ka} = 1$. When f_{ko} is the freight flow, α_{ko} is also called the utility ratio. The proportion α_{ko} is then used to estimate the inputs for division k in DMU_o with (4.27).

$$x_{i,ko} = \alpha_{ko} x_{i,k} \qquad \forall k \in K_o \qquad (4.27)$$

This can now be used as inputs in the model as in (4.20) to (4.24). The same numerical experiment as for the DEA and NDEA has been done in Appendix D and the position of the divisions in the chain does not influence the efficiency.

4.4.4. Comparison weighted SB-DEA with NDEA

The developed weighted SB-DEA has not been tested on real data yet and it is unknown what the effect is of adapting a traditional SB-DEA for analysing a chain. This subsection compares the developed DEA with the NDEA by Saeedi et al. (2019), since both are developed to analyse IFT-chains with different structures. The sample case with real data of corridors in Europe from Saeedi et al. (2019) is used for comparison (see Figure 4.5). For an explanation of how the data has been obtained, see Saeedi et al. (2019).

There are two remarks for using these data and results. The first is that the original results from Saeedi et al. (2019) could not be replicated completely, even after making various small adaptations. We use the results that are closest to the original results, where the differences are small for most transshipment and haulage segments. The second remark is that the sample is relatively small. It consists of 10 chains with 15 haulage segments and 20 transshipment segments (see Figure 4.5). This decreases the discriminatory power of the models (Dyson et al., 2001). Therefore, it is too soon to draw conclusions solely based on these results. Further research is needed to better understand the differences and applications for both the NDEA and weighted SB-DEA.

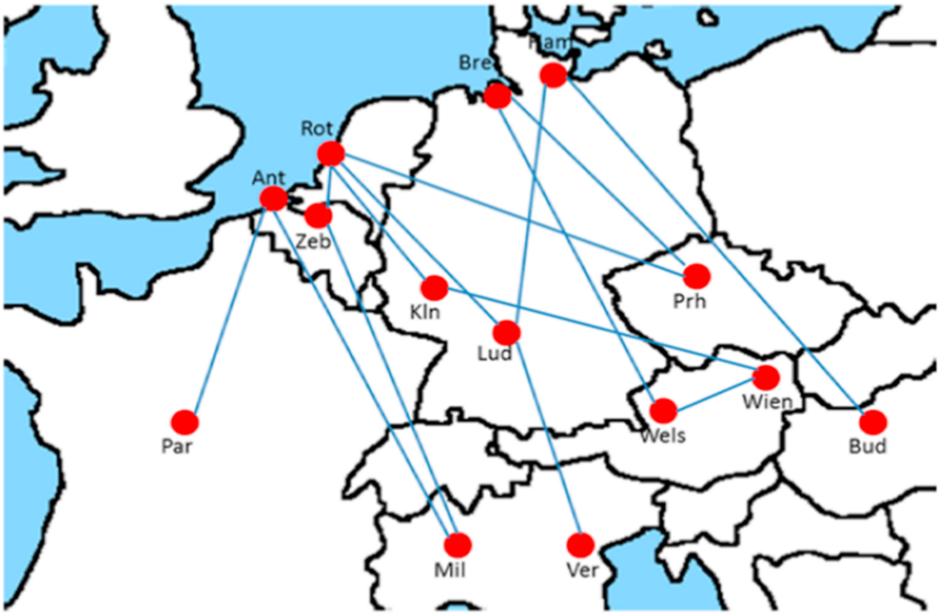


Figure 4.5: Different corridors in the sampled network. Reprinted from Saeedi et al. (2019, p. 76).

The results for both the NDEA with cumulative intermediates and the developed weighted SB-DEA can be found in Tables 4.1 and 4.2. To evaluate if the position of the division in the chain influences its efficiency, the mean for every position is calculated. For the NDEA, it can be seen that the first terminal and haulage segments are smaller than the mean of the total efficiency, while they are larger than the mean later in the chain. This is not the case for the weighted SB-DEA. This is similar to the results from the numerical experiments in Appendix D. However, both the total efficiencies of the chains and divisional scores are significantly smaller for the weighted SB-DEA. Some divisions that are seen by the NDEA as fully efficient (having a score of 1), are scored lower by the weighted SB-DEA (see for example chain 7). This difference can partly be explained by not using the cumulative intermediates. Including intermediates increases the number of unique parameters in the model, making it possible to have more efficient divisions. The number of fully efficient divisions (23 for NDEA, 12 for weighted SB-DEA) points in the same direction. However, there could be other reasons for having smaller values in the weighted SB-DEA on top of the number of parameters. Future research is needed to further evaluate the working of the weighted SB-DEA in comparison to other methods to measure performance of IFT-networks.

4.5. Conclusions DEA for IFT

This chapter discussed the DEA and Network DEA and their formulation. There are issues with using a flexible NDEA with cumulative intermediates. Therefore, we presented an adapted DEA, that extends on the weighted slacks-based DEA. We use the advantages of substitutability to increase discriminatory power and couple the chain for a network effect of DMUs. After comparing the model with a traditional DEA and a Network DEA, the model seems to work as designed without fixed structures and overlapping segments. It is therefore expected that will not be large issues with the model itself for the application on an IFT-network. One remark is that efficiency scores of the model are slightly lower than for a Network DEA. This needs to be taken into account in the next chapter, where this model will evaluate the efficiency of a real IFT-network.

Table 4.1: Reproduced results for the 10 chains from Saeedi et al. (2019) using the NDEA with cumulative intermediates. T are terminals, H are haulage segments.

Chain	DMUs	Total Efficiency	Divisional score T1	H1	T2	H2	T3
1	Beatrix Terminal - HUPAC - Ludwigshafen KTL - CEMAT - Verona Quadrante Europa	0.70	0.42	0.76	0.78	0.78	0.73
2	Container Terminal Altenwerder - IMS - Rail Cargo Terminal BILK	0.61	0.34	0.47	1.00	-	-
3	Combinant (Quay 755) - HUPAC - Busto Arsizio (Gallarate)	0.81	1.00	0.43	1.00	-	-
4	Eurogate C.T. - IMS - Enns Hafen CTE - IMS - Wien Freudenau Hafen CCT	1.00	1.00	1.00	1.00	1.00	1.00
5	Progeco Zeebrugge - Danser - Rail Service Center (RSC) - METRANS - METRANS Praha	0.86	1.00	1.00	1.00	0.83	0.45
6	RCT Rotterdam - Kombiverkehr - DUSS Terminal Duisburg - HUPAC - Wien Freudenau Hafen CCT	1.00	1.00	1.00	1.00	0.98	1.00
7	DUSS Billwerder - Kombiverkehr - Ludwigshafen KTL - CEMAT - Verona Quadrante Europa	0.84	0.18	1.00	1.00	1.00	1.00
8	Van Doorn - Naviland Cargo - Paris Valenton	1.00	1.00	1.00	1.00	-	-
9	Eurogate C.T. - METRANS - METRANS Praha	0.28	0.28	0.33	0.24	-	-
10	Progeco Zeebrugge - HUPAC - Busto Arsizio (Gallarate)	1.00	1.00	1.00	1.00	-	-
Mean		**0.81**	**0.72**	**0.80**	**0.90**	**0.92**	**0.84**

Table 4.2: Results for the 10 chains from Saeedi et al. (2019) using the developed weighted slacks-based DEA. T are terminals, H are haulage segments.

Chain	DMUs	Total Efficiency	Divisional score T1	H1	T2	H2	T3
1	Beatrix Terminal - HUPAC - Ludwigshafen KTL - CEMAT - Verona Quadrante Europa	0.46	0.44	0.49	0.54	0.61	0.21
2	Container Terminal Altenwerder - IMS - Rail Cargo Terminal BILK	0.44	0.34	0.47	0.52	-	-
3	Combinant (Quay 755) - HUPAC - Busto Arsizio (Gallarate)	0.60	1.00	0.36	0.43	-	-
4	Eurogate C.T. - IMS - Enns Hafen CTE - IMS - Wien Freudenau Hafen CCT	0.74	1.00	1.00	1.00	0.18	0.53
5	Progeco Zeebrugge - Danser - Rail Service Center (RSC) - METRANS - METRANS Praha	0.58	1.00	0.21	1.00	0.46	0.23
6	RCT Rotterdam - Kombiverkehr - DUSS Terminal Duisburg - HUPAC - Wien Freudenau Hafen CCT	0.69	1.00	1.00	0.51	0.76	0.19
7	DUSS Billwerder - Kombiverkehr - Ludwigshafen KTL - CEMAT - Verona Quadrante Europa	0.53	0.19	1.00	0.54	0.70	0.20
8	Van Doorn - Naviland Cargo - Paris Valenton	0.63	1.00	0.46	0.44	-	-
9	Eurogate C.T. - METRANS - METRANS Praha	0.32	0.44	0.28	0.24	-	-
10	Progeco Zeebrugge - HUPAC - Busto Arsizio (Gallarate)	0.78	1.00	1.00	0.35	-	-
Mean		**0.58**	**0.74**	**0.63**	**0.56**	**0.54**	**0.27**

5

Efficiency of TEN-T corridors

The previous chapters discussed the theory and mathematics for a model to evaluate the efficiency of an IFT network. This chapter will apply that theory and the model to a case study of the Trans-European Transport Network (TEN-T). This should give a better understanding of the model itself and of the efficiency of European intermodal freight transport. First, the description of the case study of TEN-T corridors is presented. This is followed by the model specifications of the DEA, including data requirements, data collection and model assumptions. Section 5.3 then presents the results for the corridors and its haulage and transshipment segments.

5.1. Description case study TEN-T corridors

The European Union EU aims to improve the transport system in Europe to increase competitiveness and sustainability goals. To do this, a modern well-functioning infrastructure network is required. The Trans-European Transport Network (TEN-T) policy is developed to build this network and involves the implementation of a European network of roads, railway lines, IWW, SSS, ports and terminals (European Commission, 2020c). The core network consists of 9 core network corridors, as depicted in Figure 5.1 (European Commission, 2020c).

The TEN-T corridors are funded by both national governments and multiple European funding programs, such as Connecting Europe Facility (CEF), European Fund for Strategic Investment (EFSI), Horizon 2020 and European Structural and Investment Funds (ESIFs) (European Commission, 2020a). The TEN-T corridors exist of both haulage and transshipment segments for the modalities rail, road and water, including ports and terminals. In this thesis, we look at the rail network as haulage segments and intermodal terminals as transshipment segments. For the reason of data quality and availability, the corridors Rhine - Alpine (R-A) and the continental part of North Sea - Mediterranean (NS-M) are selected for the case study, as can be seen in Figure 5.2 (European Commission, 2020b). They both have a part in the Netherlands, for which we have the best and most data available.

5.2. Model specifications DEA for TEN-T corridors

This section describes the model specifications for the application of the DEA model as described in Section 4.4 on the TEN-T corridor case study from Section 5.1. It consists of the choice for input and output data and the collection of this data, followed by the assumptions made for the model.

5.2.1. Data requirements and data collection

As seen in Section 3.3.2, there are three levels of performance indicators according to Davydenko, Jordans, and Krupe (2007). Level 1 describes the customer level, level 2 the system integrator level and level 3 the operator level. A DEA evaluates the efficiency of a system from an operator level, therefore these indicators will be used. Table B.1 (Davydenko, Jordans, & Krupe, 2007, p. 105) in Appendix A gives generic indicators for all modalities on level 3, such as price (cost), capacity and reliability. Furthermore, it makes a distinction in operation specific indicators between road, IWW, train and intermodal terminals, of which the last two fall within our scope. Preferably, the indicators from Table 5.1 should be used to evaluate the efficiency of an IFT network. However, due to the lack of available data, a selection must be made.

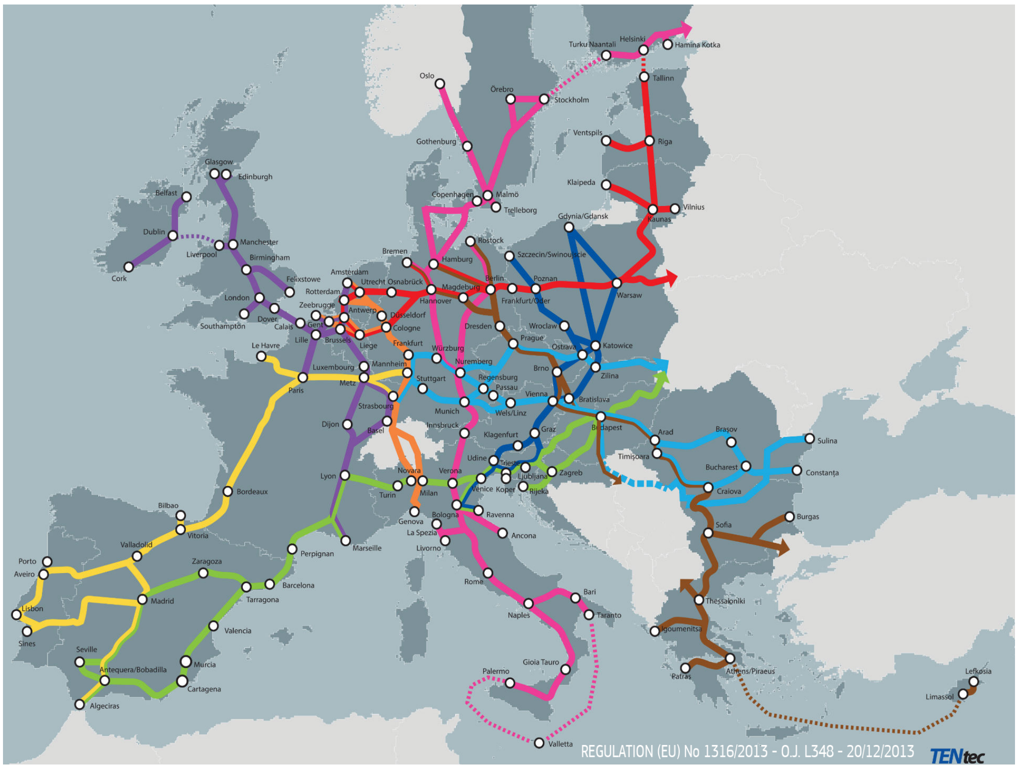


Figure 5.1: The core network corridors of TEN-T (European Commission, 2020c)

Table 5.1: Data requirements for transshipment and haulage segments with perfect data availability.

	Inputs	**Outputs**
Transshipment	• Transshipment cost • Area • Number of tracks • Track length • Number of cranes and reachstackers • Train arrival rate	• Volume • Price
Haulage	• Transport cost • Driving/waiting time ratio • Emissions	• Volume • Price

5.2.1.1. **Transshipment data for TEN-T corridors**

The data required with perfect data availability can be found in Table 5.1. The generic indicators on operator level as found in Table B.1 (Davydenko, Jordans, & Krupe, 2007) in Appendix A are among others capacity, cost and reliability. For terminals specifically, cargo handling equipment, cranes and train arrival rate are specified, of which the first two could be seen as capacity as well. These are comparable to the most frequently found indicators in scientific literature as in Table A.1. However, there is a lack of availability for cost and train arrival rate, but characteristics of terminals in terms of capacity are shared more widely. These are area, number of tracks, track length, number of cranes and number of reachstackers and will be used as inputs.

This input data is available from two main sources. The first is intermodal-terminals.eu/database/ from the AGORA Marco Polo project by the EC (KombiConsult, n.d.), containing 408 terminals as of 2020. The main source of this database are the intermodal terminal owners themselves, which result in some missing data. The website inlandlinks.com by the Port of Rotterdam (n.d.) had audited information of 85 terminals, including the inputs, as well as a yearly volume capacity and CO2 emissions. The audits make it a more reliable source and its data is prioritized over intermodal-terminals.eu. However, because the first has a much larger number of terminals included, it is still the largest source for intermodal terminal inputs.

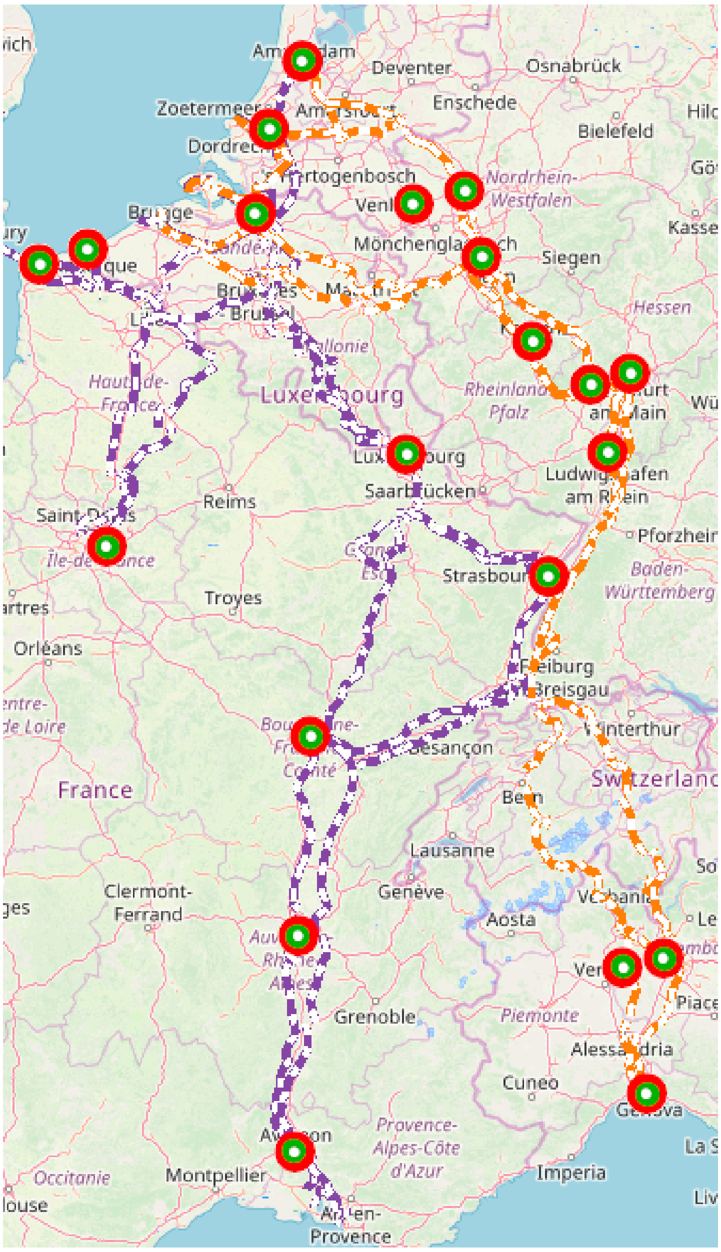


Figure 5.2: The core network corridors of TEN-T Rhine - Alpine (orange) and North Sea - Mediterranean (purple) (European Commission, 2020b)

From these two data sources, there are 83 terminals in the NS-M and R-A corridors, of which 33 have at least one input data point missing. For the application of the DEA, we need to have all input and output values for all terminals, therefore the missing data is estimated by using the data from similar terminals. The estimation is done by calculating the weighted arithmetic mean or weighted average for the missing input, based on similarity of terminals. The description of this method is described in detail in Appendix E. The inputs for terminals in the Rhine - Alpine and North Sea - Mediterranean corridors can be found in Tables F.1 and F.2 in Appendix F, including estimated data in bold. No difference is made in the type and size of a crane or reachstacker, just the number of cranes and reachstackers is taken into account.

Volume is the production of the IFT system and will be used as output for both transshipment and haulage segments. The data comes from the ETISplus database, the follow-up project of European Transport Policy Information System (ETIS) from 2005 (TRIMIS - European Commission, n.d.). The data is available on a NUTS2 level (Eurostat, n.d.). For example, the provinces in the Netherlands are all individual regions. However, countries like Germany (37 regions) and France (25 regions) have larger regions. Because of this avail-

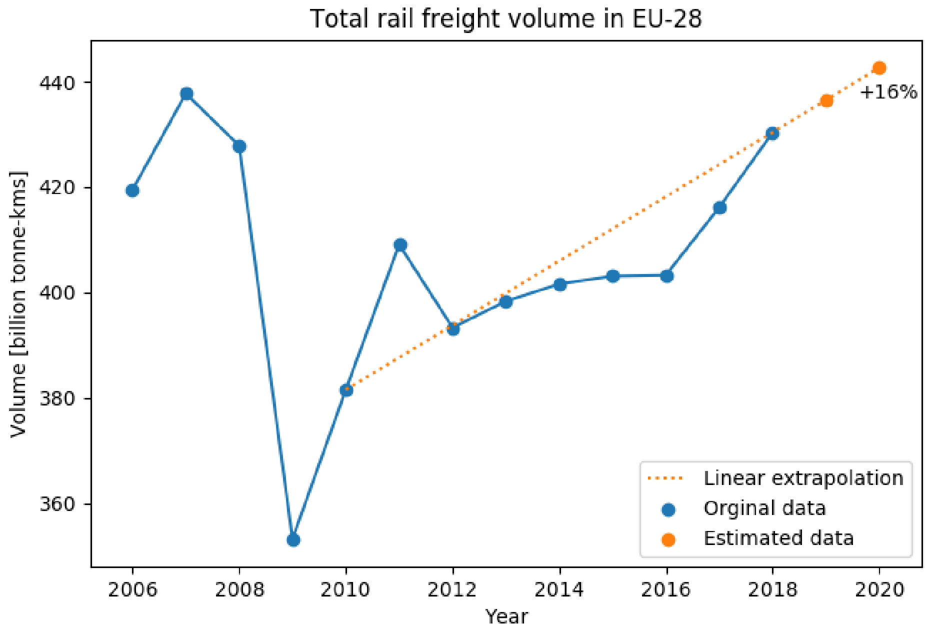


Figure 5.3: Rail freight volumes in all 28 EU countries (EU28) for 2006-2018 (Eurostat, 2020), with linear extrapolation from 2010-2018 to 2020. This results in a 16% higher volume in 2020 compared to 2010.

able detail level, the detail level of transshipment segments are cities and their according regions within a radius of approximately 50 km. The inputs from terminals in the same cities are summed for the total inputs per city.

For the output, transshipped volume per city is needed. However, there is no data available in ETIS about transshipped volumes. Instead, the data from ETIS is available as volumes between origin and destination regions. We therefore assume that the volume with an origin or destination region belonging to a transshipment segment is transshipped in that segment. Because the transshipment segment is the origin or destination, this is mainly transshipment between pre- and main haulage or main and end-haulage. This means that transshipment between two main haulage segments is not included, which is a the first limitation of the data for the model. This will be called main-to-main transshipment in the rest of this thesis. Transshipment segments for which main-to-main transshipment is a large share of their total transshipment, are thus not assigned a sufficiently large volume. Examples of this are ports such Antwerpen and Rotterdam or cities with many rail terminals such as Frankfurt am Main. It is expected that the model will assign these transshipment segments a smaller efficiency than their actual technical efficiency (see Figure 5.4c.

Another limitation regarding transshipment is that the databases do not necessarily include all existing terminals in and around the researched cities, some might be missing. Intermodal-terminals.eu uses several sources for the terminals, but the main source is that terminal owners had to apply themselves (KombiConsult, n.d.). For inlandlinks.com terminal owners had to pay to be audited and included Port of Rotterdam (n.d.), hence the lower amount of terminals. This is the reason that some terminals might be missing from the data set, increasing the calculated efficiency of the cities where the terminals are missing (see Figure 5.4a). This is difficult to prevent, but a sensitivity analysis could be applied to see what the effect would be on the efficiencies when a terminal would be missing.

A third limitation of the data is that no difference can be made between a transshipment segment in multiple different corridors. The aforementioned characteristics of terminals are used as inputs for the transshipment segments. However, no distinction can be made between inputs used for a transshipment segment that is part of multiple corridors. Therefore, the inputs are shared among all corridors for overlapping transshipment segments. Because this is done for inputs, this is also done for the outputs. Therefore, volumes are summed for a transshipment segment as origin for all possible destination regions (and as destination for all possible origin regions), not only within that corridor. This makes a fairer comparison between transshipment segments possible. However, this decreases the discriminatory power.

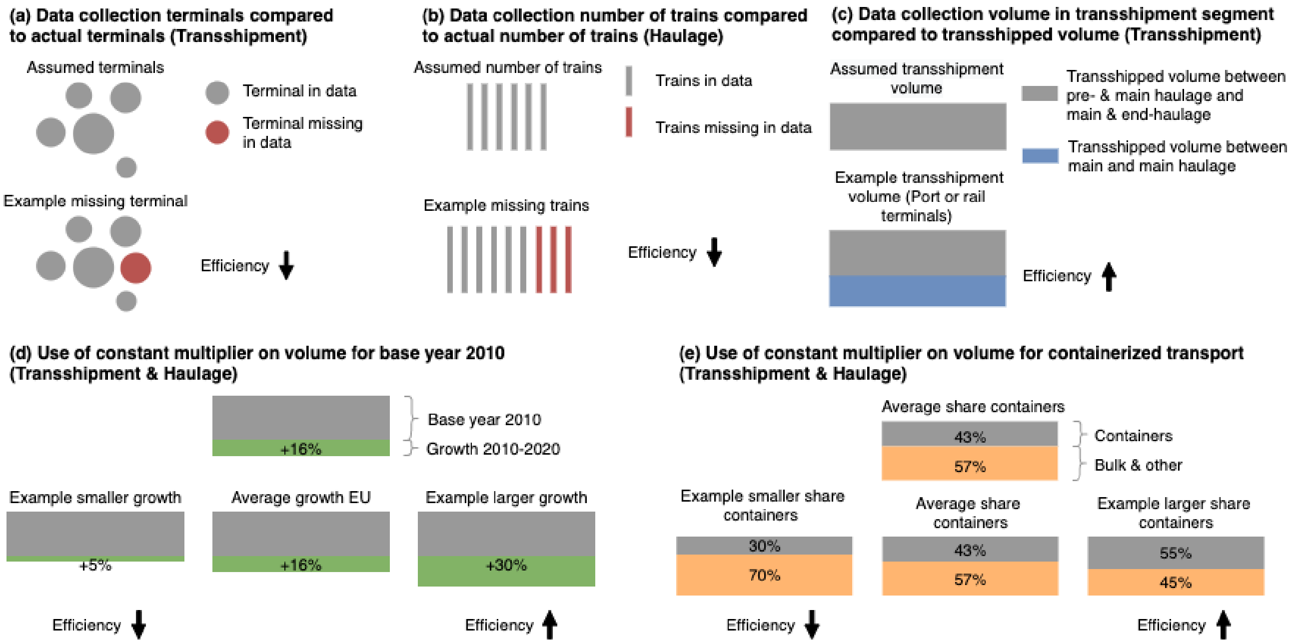


Figure 5.4: The impact from 4 limitations in input and output data on efficiency. Between brackets is the type of segment (transshipment and/or haulage) which the limitation influences. The effect on efficiency is described as what the efficiency would be when the data would be improved.

The volumes from ETIS are from 2010, which is used as data source for the volumes for both transshipment and haulage, since we have no newer data available. However, freight rail volumes have increased in Europe, although there are differences between countries (Eurostat, 2020). Data from Eurostat (2020) go to 2020, but with a linear extrapolation the volumes have increased with 16% in EU28 between 2010 and 2020, see Figure 5.3. To take the increased volumes into account, the volumes from NEAC10 can be multiplied with a factor of 1.16 (+16%). However, a limitation is that the volumes will have increased differently for the different transshipment and haulage segments. Some segments will have increased more than 16%, while others will have increased less or even decreased, which would result in a wider range of efficiencies from the DEA. The effect on the efficiency of segments is explained in Figure 5.4d.

The last limitation for transshipment is that the used volumes from ETIS are not only containerized transport, but also include bulk volume. To get the tonnes for IFT only, we can multiply this total with the share of containerized rail transport. In the Netherlands, this is about 43% (ProRail, 2017). There is no data available for shares for individual segments, therefore the same share is used for all segments. However, transshipment or haulage segments that have a higher share of containerized, will get a smaller efficiency score than their technical efficiency (see Figure 5.4e.

All these limitations are shown in Figure 5.4. Limitations from Figure 5.4d and 5.4e can have an impact on the efficiency scores of individual segments. However, since the percentage differences will not be large for most cases, the effects cause probably a maximum of 10% difference for a terminal. The lack of data on main-to-main transshipment (5.4c) will likely have a larger effect. Main-to-main transshipment is a large share of the total transshipment for some transshipment segments. Missing this part will decrease their efficiency score, which is expected to be seen for ports or cities with many main-to-main transshipment terminals. However, the largest effects will come from the other two limitations as mentioned in Figure 5.4a and b. Missing a single terminal (5.4a) for a transshipment segment with only a small number of terminals can decrease its efficiency by much. The same is true for missing a number of trains for haulage segments (5.4b). A transshipment or haulage segment which misses terminals or trains can result in a very high efficiency score. However, the largest effect is that other segments will be compared to these seemingly efficient segments, resulting in a much lower efficiency score for them. High quality data is needed to prevent have missing terminals or trains.

Summarizing the effects of all these limitations, the net effect of the limitations is a small efficiency score for many segments that are not necessarily technically inefficient. Furthermore, ports and cities with many main-to-main transshipment terminals, will also score lower than their actual technical efficiency.

5.2.1.2. Haulage data for TEN-T corridors

The KPIs used in literature are mainly inputs. The operation specific KPIs from Table B.2 (Davydenko, Jordans, & Krupe, 2007, p. 106) are special cargo storage/handling possibilities and driving/waiting time ratio. The data required for haulage with perfect data availability can be found in Table 5.1. All are known by the railway company, but are not publicly available. Indicators that are most frequently mentioned in literature are transport cost and emissions environment. There is a lack of data for all of the above-mentioned data. However, transport cost can be estimated using the simple formula from Janic (2007) in Equation 5.1.

$$\text{Transport cost} = \text{Frequency} \times \text{Cost per frequency} \tag{5.1}$$

There is no data available for the cost per frequency to differ between the haulage segments, so this is set as constant but decreases the discriminatory power of the model. Because the multipliers in a DEA (see Equation 4.21) are variables, the assumed value for this constant is irrelevant and is set to 1. It follows that the frequency or number of trains is the proxy attribute for transport cost. This is chosen as input parameter for haulage.

The website railway.tools (Deutsche Bahn, n.d.) is used for the number of trains. It has a database of direct intermodal connections between origin and destination terminals within Europe. To find the number of trains on a haulage segment, all possible origin-destination (OD) pairs within a corridor are searched. All connections within the corridor that go through the haulage segment are summed, which gives the total number of trains on the segment belonging to the corridor. Just like for terminals, volume is used as output. The data from ETIS is again used for volumes. However, now only volumes that belong to the researched corridors are included. Similarly to the number of trains, all volumes within the corridor that go through the haulage segment are summed, which gives the total volume on the segment belonging to the regions within the corridor.

A drawback of the collection of volume and number of trains is that haulage segments can be part of multiple corridors, as shown in Figure 5.4. Regarding the R-A and NS-M corridors, there is overlap with the Atlantic corridor in France and Germany, the North Sea - Baltic corridor in the Netherlands and Germany, the Rhine - Danube corridor in Germany and the Mediterranean corridor in France and Italy. Both volumes and number of trains do not specifically belong to that corridor, but belong to the regions within the corridor. Therefore, the inputs and outputs are shared among all corridors for overlapping haulage segments. Therefore, no difference can be made between a haulage segment in two different corridors. This decreases the discriminatory power of the model.

Table 5.2: Data sources for transshipment and haulage segments.

Source	**Data**	**Comment**
inlandlinks.com	• Area • Number of tracks • Track length • Number of cranes • Number of reachstackers	Audited data of 85 terminals
intermodal-terminals.eu	• Area • Number of tracks • Track length • Number of cranes • Number of reachstackers	Data of 408 terminals, estimation of missing data
railway.tools	• Number of trains	Data collection for some trains is incomplete. Can be solved by using other sources for number of trains.
ETIS	• Volume transshipment • Volume haulage	NUTS2 detail level. Volumes for transshipment between two haulage segments is missing.

5.2.2. Modelling assumptions DEA for TEN-T corridors

The previous section discussed the methods, assumptions and limitations regarding the data. This section lists the other assumptions that are made for the data collection and application of the weighted SB-DEA specifically for the TEN-T corridors.

- We are interested in the performance of entire corridors, but also how this performance is established. For this, we need the performance of its transshipment and haulage segments. From an organisational perspective, both the entire corridor and its transshipment and haulage segments could be seen as DMUs. Decisions are made by the EU about the entire corridor at once, but companies make decisions about the operations in single terminals and train services. According to NDEA theory, the DMU is the entire network with divisions as elements. Translating this to an IFT network, the corridors are the DMUs and the divisions are the transshipment and haulage segments. However, it remains interesting to analyse the performance of single divisions and its efficiencies will be calculated post analysis.
- The weights in the objective function, Equation 4.20 are all set to $\frac{1}{|I_k|}$, or in other words all divisions of the DMU have the same weight. The weights can be changed according to the relative importance of divisions. Inputs or outputs could be used to assign this relative importance, for example volume or the length of haulage segments. A requirement for this is reliable data, since extreme results from outlier input or output data could be amplified when weights also depend on these data. This can be done in future research.
- Transshipment segment Zeebrugge only uses cranes. This means that the number of reachstackers is zero for these terminals. However, an input can't be 0 in a DEA, since it would require division by 0 in the objective function (Equation 4.20). When the value of one of the inputs is zero, such as the number of reachstackers in Zeebrugge, this is replaced by a very low value (0.0001).
- For overlap between the R-A and NS-M corridor, a segment is seen as part of both corridors. Since there is no data available about which share of volume belongs to which corridor, it is difficult to determine α_{ko}, the estimated proportion of resources. Therefore, we assumed these segments to be fully part of both corridors.
- A requirement of DEA is that divisions have the same activities. The transshipment segments are an aggregation of terminals, so the terminals are required to be comparable. Although some terminals are connected to IWW and other terminals are not, we assume that all terminals are comparable and can thus be compared fairly for their efficiency. Future research could have a separate category for terminals with and without access to IWW.

5.3. Efficiencies of TEN-T corridors

The weighted SB-DEA has been applied on the two TEN-T corridors R-A and NS-M, using the data as described in Section 5.2.1. All data used for the application of the DEA can be found in Appendix F.

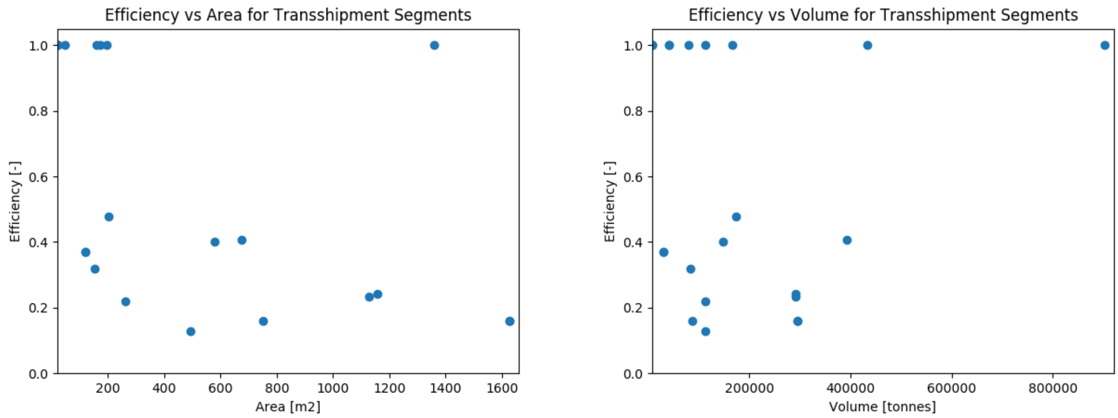
The most direct result is the efficiency of the DMUs, the corridors (see Table 5.3). The Rhine - Alpine corridor has the highest efficiency. Following Equation 4.20 in Section 4.4, the DMU efficiencies come from the weighted average of the efficiencies of the divisions in each DMU. The efficiencies for individual divisions can be found in Tables G.1 and G.2 in Appendix G. From the corridor results, it seems that both chains are quite inefficient. However, there are multiple reasons why the assigned efficiencies for both haulage and transshipment segments could be lower than their actual technical efficiency, this will be discussed in the following two sections.

Table 5.3: Efficiencies for North Sea - Mediterranean and Rhine - Alpine corridors.

Corridor	Efficiency
North Sea - Mediterranean	0.469
Rhine - Alpine	0.622

5.3.1. Efficiencies for transshipment segments in TEN-T corridors

The efficiencies for transshipment segments in the R-A and NS-M corridors can be found in Table G.1 in Appendix G. The efficiencies of individual segments show some remarkable results. First, it is notable that there



(a) Input: Area

(b) Output: Volume

Figure 5.5: Relation between area & efficiency and volume & efficiency for transshipment segments.

are some low efficiency scores, with 5 cities under 30% (Ludwigshafen, Rotterdam, Lille, Paris and Antwerpen), going as low as 13% (Ludwigshafen). For other DEA analyses of rail or IWW terminals, values lower than 30% are occasionally occur (Saeedi et al., 2019; B. Wiegmans & Witte, 2017; B. W. Wiegmans, Rietveld, Pels, & van Woudenberg, 2004), but are more uncommon than in the results from this research. Although it is expected to have a certain range of efficiencies, it is striking that two of largest ports of Europe (Rotterdam and Antwerpen) seem to be quite inefficient for rail transshipment according to the model.

Figure 5.5a shows the efficiency vs area for transshipment segments, Figure G.1 in Appendix G shows similar figures for all inputs. Generally speaking, the segments with low inputs have the highest efficiency. The only exception to this is Ludwigshafen (top right corner of all efficiency vs input Figures G.1), which has some of the highest inputs, but also by far the highest output. That the most efficient terminals have small inputs is somewhat as expected, since larger inputs decrease the efficiency when volumes are kept the same. However, Evers (1994) and B. Wiegmans and Konings (2015) suggest that economies of scale play a role in terminals, such that larger terminals are usually more efficient, resulting in lower unit prices. Although generally speaking, transshipment segments with larger inputs transship a larger volume (see increasing trend in Figures 5.6 and Figure G.2), there is no sign of economies of scale. Economies of scale would show a high efficiency for segments with higher inputs, but this is not the case (see Figures 5.5a and G.1). Economies of scale would mean lower unit prices, meaning there should be an increasing rate of volume growth for larger segments. Figures G.2a and G.2 does however not show this, it seems more like a linear growth rate instead of economies of scale. There are multiple possible explanations for the difference between literature and the results for the TEN-T corridors.

Firstly, it could be that there is actually a large difference between the operational efficiency of transshipment segments. Probably a part of the differences could be explained by this, but as stated above, the differences between transshipment segments are larger than found in for terminals in literature (Saeedi et al., 2019; B. Wiegmans & Witte, 2017; B. W. Wiegmans et al., 2004).

A second possible explanation is the working of the model. The weighted slacks-based DEA has not been applied to other cases according to the knowledge of the author, so there could be some unintended consequences of the model itself. This can be checked by comparing the results of the model to results of a regular slacks-based DEA. The model has been tested on a simple test network in Chapter 4 and on sample data from Saeedi et al. (2019). The latter showed that efficiencies of the weighted SB-DEA can be a bit lower than in the NDEA. However, this is probably not as much as the differences seen for the TEN-T corridors. Still, future research comparing the two models with realistic data in a more complex network should confirm the intended working of the model.

A third possible explanation is the quality of the data. Section 5.2.1.1 already described some of the limitations of the data collection and data sources. The limitations with probably the largest impacts are missing terminals (Figure 5.4a and no main-to-main transshipment (Figure 5.4c). The list of terminals from the two data sources can be incomplete, which influences the sum of inputs for transshipment segments. In this case, the database missing terminals in a city decreases the amount of inputs and thus increases the efficiency of the DEA. Well documented transshipment segments (like port of Rotterdam) will be assigned lower efficien-

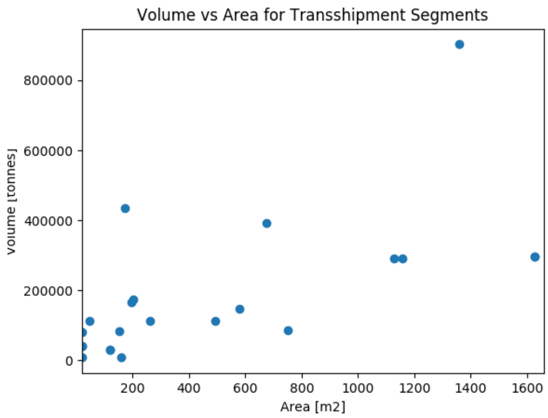


Figure 5.6: Relation between area and volume for transshipment segments.

cies than smaller segments that possibly misses a terminal in the database. Furthermore, when both a smaller and larger transshipment segment are missing one terminal, a single terminal is a relatively larger share for the smallest segment. It therefore would impact the efficiency of a smaller transshipment segment more than a larger segment. This in line with the findings of smaller transshipment segments with the highest efficiency.

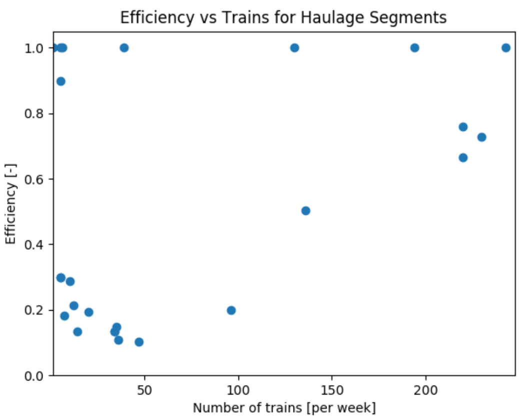
The other important limitation is the lack of data on main-to-main transshipment volumes (see Figure 5.4b). There is only data available which mainly assumes transshipment between pre- & main haulage or main & end-haulage. Transshipment between two main haulage segments is thus not included. However, transshipment segments such as ports have a large share of transshipment between two main haulage segments and thus the available data excludes a large share of their transshipment volumes. This results in a lower efficiency score, which is the case for the ports of Rotterdam and Antwerpen.

It is possible that lack of economies of scale in the results it is a combination of all of the above reasons. Steps to be taken to confirm this, are further testing the weighted SB-DEA and increasing the quality of data for terminals and transshipment volumes.

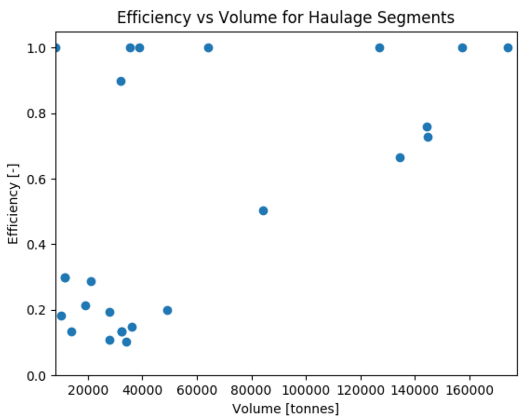
5.3.2. Efficiencies for haulage segments in TEN-T corridors

The efficiencies for haulage segments in the R-A and NS-M corridors can be found in Table G.2 in Appendix G. It stands out that there are even more low efficiency scores than for transshipment, see also Figure 5.7a. In the North Sea - Mediterranean corridor alone, there are 7 segments with an efficiency of lower than 20%. This cannot be explained alone by the technical efficiency of these segments.

The low efficiencies mainly arise for haulage segments with low number of trains and low volumes (see



(a) Input: Number of trains



(b) Output: Volume

Figure 5.7: Relation between number of trains & efficiency and volume & efficiency for haulage segments.

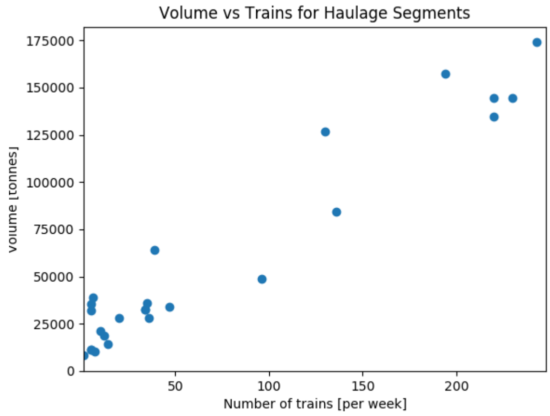


Figure 5.8: Relation between number of trains and volume for haulage segments.

Figure 5.7. Figure 5.8 shows that there is indeed a strong relation between number of trains and the volume. Because there is only one input and one output used for haulage, we can trace the origin and calculation of the efficiencies. Haulage segments between Antwerpen and Paris have the highest volumes of the North Sea - Mediterranean corridor. However, the number of trains on the segments Gent-Lille (6) and Lille-Paris (1) is extremely low. This gives them an efficiency score of 1.00, but also decreases the efficiency scores of other haulage segments with similar volumes, but a larger and more realistic number of trains. If we look at the segment Gent-Lille, there are 6 trains per week for a volume of 3 million tonnes per year. This low number of trains is not realistic, it would mean that 6 trains would move 3 million tonnes per year or 60,000 tonnes per week. This is 10 times more than in normal cases, where a single freight train in Europe would handle around 1,000 tonnes (Saeedi et al., 2019). This means that the number of trains for some segments is unrealistically low, which decreases the efficiency of many other segments, as also shown in Figure 5.4b.

The Rhine-Alpine corridor has more haulage segments with a large number of trains and large volume. This results in a higher efficiency score for the R-A corridor, indicating that there could be some advantages of economies of scale for haulage. The part of the corridor between Rotterdam and Milano have large number of trains and volumes and their efficiencies are all at least 0.50. Using the calculation from previous paragraph, the number of trains realistically fits with the volumes. However, it seems again that segments with the ports Rotterdam and Antwerpen do have lower score. This might have to do with the limitation as mentioned in Section 5.3.1 (see Figure 5.4c). Although assumed in Section 5.2.2 that all segments are comparable, ports and their haulage segments might thus not be comparable to other rail-terminals. There are two ways to solve this in future research. First, different groups can be made for different types of segments, comparable to the two groups in the current research (transshipment and haulage segments). Different categories can be created for ports and port-haulage segments. A requirement is that there are sufficient ports, such that set of all divisions with the same activities C_k is large enough.

The second possibility for future research is to try to get a more realistic number of trains. This could be done by using multiple IFT planners or collecting data from the national rail infrastructure managers.

5.4. Conclusions results TEN-T corridors

This chapter discussed the application of the weighted SB-DEA on the TEN-T corridors North Sea - Mediterranean and Rhine - Alpine. The corridors are split in transshipment and haulage chains. For transshipment segments, the inputs are characteristics of terminals: area, number of tracks, total length of tracks, number of cranes and number of reachstackers. For haulage segments, the number of trains is the proxy attribute for transport cost as input. Volume is the output for both segments.

The data collection is done from 4 sources. There are five main limitations of the data: possible missing terminals, missing trains, no main-to-main transshipment data, volumes based on year 2010 and no difference in volume between bulk and containerized transport. The last 3 mainly have an effect on individual segments, of which the lack of main-to-main transshipment data can have quite a large effect on a segment such as a port. However, possible missing terminals or trains would have the largest effect on the efficiency scores of multiple segments. With a missing terminal or train, the collected inputs can be much lower than in reality. A transshipment or haulage segment would then seem much more efficient and is used as reference for other segments, that would all get a much lower efficiency score. High quality of data for the existing terminal belonging to each transshipment segment and the number of trains for each haulage segment has thus the highest priority in the data collection.

With these limitations in mind, the weighted SB-DEA is applied on the TEN-T corridors NS-M and R-A. The R-A corridor is the most efficient, with an efficiency score of over 62%. This is mainly because of the haulage segments between Rotterdam and Milano, which score all over 50% efficiency. On these segments, there are a large number of trains and large volumes. This might indicate some advantages of economies of scale. The lower score for NS-M is likely due to one of the earlier mentioned limitations, missing a number of trains. The number of trains for haulage segments Gent-Lille (6) and Lille-Paris (1) probably decrease the efficiency scores of the other segments with comparable volumes, which are mainly part of the NS-M corridor. More data sources for IFT trains are needed to have more reliable inputs for haulage segments.

For transshipment, there seem to be no advantages of economies of scale. Most of the large transshipment segments have a low efficiency score, with ports Antwerpen and Rotterdam as main example. The limitation of lack of main-to-main transshipment is likely involved. Both the NS-M and R-A corridor seem to be equally efficient. Future research could estimate the main-to-main transshipment to see if there is a difference between these two corridors.

6

Conclusions & Recommendations

The final part of this thesis consists of conclusions and recommendations based on this research. Firstly, conclusions are discussed in Section 6.1, where the research questions as defined in Chapter 1 are answered. Based on these conclusions, scientific and practical contributions are discussed in Section 6.2. Section 6.3 consists of the discussion and limitation of this research. This chapter concludes with recommendations for future research and application.

6.1. Conclusions

Section 1.2 defined the research questions for this research where the main research objective was to assess the performance of an intermodal freight network. To assess performance of an IFT-Network, a DEA model was developed. This model continued on the work of Saeedi et al. (2019). This model has then been applied to a part of the European TEN-T corridors. The inputs and outputs for this application are based on scientific and practical literature, taking into account the data availability.

Three sub-questions are posed in Section 1.2. These will be answered below, followed by an answer to the main research question.

1. How is performance measurement done currently in European intermodal freight transport?
Although the performance of an entire IFT-network with both haulage and transshipment segments is rarely measured, haulage and transshipment segments are often assessed individually in literature. Both scientific and practical literature use key performance indicators (KPIs) to describe the performance of European intermodal freight transport (IFT).

In scientific literature, many different KPIs are used. The two domains chains & transportation and terminals & ports have been distinguished. The most common KPIs are environment & emissions and waiting time. Other common KPIs for both domains are transshipment time, employee productivity, safety & security, energy consumption, reliability and transportation costs. For each domain, there are also specific KPIs. For terminals & ports, these are transshipment time, safety & security, employee & equipment productivity and volume. There are two chains & transportation specific KPIs, which are transport costs and travel time. However, it is interesting to see that there is not a large difference between the KPIs used for terminals & ports and chains & transportation.

In practical literature, many objectives are mentioned without their specific ways to measure the objectives. This gives an overview of the relative importance of objectives for stakeholders in IFT, but does not provide direct KPIs. The most common objectives are price of transportation, travel time and waiting time, which is a combination of both transshipment and haulage objectives. Davydenko, Jordans, and Krupe (2007) provides an overview of KPIs on 3 different levels of abstraction, namely the level for customers (level 1), logistics service providers (level 2) and operators (level 3). The operator level is most relevant to evaluate the performance of IFT and its KPIs are price, cost, capacity and capacity utilization. These are overlapping with a part of the KPIs from scientific literature.

However, an important remark is that data availability is often leading for the choice of KPIs. Although it is important to assess beforehand which indicators are most relevant to measure performance, the assessment of performance needs available data. Making more data publicly available could thus be a very important step to better assess performance of IFT.

2. How can the performance of an IFT-corridor in a network be assessed reliably?
The requirements for the assessment of performance for a IFT-network are presented in three sub-sub-questions that accompanied the sub-question.

Data envelopment analysis (DEA) is a commonly used method in literature to assess the performance of individual IFT segments like terminals and haulage. To assess entire IFT-corridors, a Network DEA has been used before (Saeedi et al., 2019). However, in this thesis we adapted a Network DEA to a weighted slacks-based DEA, where the weights represent the relative importance of its divisions.

Section 4.4 presents the mathematical formulation and an elaborate explanation of the weighted slacks-based DEA. The weighted slacks-based DEA takes into account all the requirements from sub-sub-questions 2a to 2c. Because of the general formulation and explanation, it can not only be applied to IFT-chains, but can also be used for other networks for which the requirements as mentioned below are not fulfilled by a Network DEA.

(a) How can the efficiency of IFT-chains with different structures be assessed?
IFT-chains consist of haulage and transshipment segments. Furthermore, there are multiple of each of these segments in a single corridor or chain. In Network DEA, these segments are called divisions, where the Decision Making Unit (DMU) is the chain. Traditional Network DEA's only compare divisions with each other on the same position in the chain. However, we compare all divisions using the same technology with each other (e.g. rail haulage with rail haulage), also within the same chain. This makes it possible to assess IFT-chains with different structures. Furthermore, it increases the discriminatory power of the model, especially when assessing few IFT-chains with a large amount of divisions.

(b) How can the efficiency in overlapping IFT-chain divisions in a network be measured?
When IFT-chains are part of a larger IFT-network, there is often overlap between a number of haulage or transshipment segments. This means such an overlapping segment shares it's input or resources among multiple IFT-chains. If it's known which inputs and outputs for each division belong to which chain, this is not a problem and these values can be used, seeing the overlapping divisions as separate divisions for each chain. Because of lack of data availability however, the inputs belonging to a chain are often unknown, while the outputs for each chain can be easier retrieved. We can use these outputs to estimate the proportion of resources used for each chain. This can then be used as separate divisions of each IFT-chain. Because the proportion of inputs and outputs is the same for such an overlapping division, efficiencies will be similar. This decreases the discriminatory power of the model for these divisions, but it is not possible to differentiate between chains with the same division because of a lack of data for these divisions.

(c) How can the efficiency of a corridor with parallel IFT-chains be measured?
In a traditional Network DEA model, divisions are compared to each other, based on their position in the network. However, we use the method as explained under sub-sub-question 2a and in Section 4.4, which compares divisions with similar activities to each other, independent on the position of that division in the chain. For a Network-DEA, this gives an issue because of intermediate values, which are dependent on the position of divisions. By using a weighted slacks-based DEA without intermediates, position of divisions in the chain are not taken into account, which means parallel divisions can be assessed and compared to each other.

3. How can the developed model be applied on a part of the European IFT-network?
A part of the TEN-T corridors is used to apply the model on the European IFT-network. After the development of the model, the largest difficulty for application is the lack of available data. Both operators and infrastructure managers and owners do not share much of their data publicly. Therefore not all the indicators from sub-question 1 can be used for the evaluation of the European TEN-T corridors. For transshipment segments, capacities such as area, total track length and number of cranes are used as output. For haulage segments, the number of trains are used as input as proxy for transport cost. Volume is used as output for both transshipment and haulage segments.

The results are the efficiencies of the North Sea - Mediterranean (NS-M) and Rhine - Alpine (R-A) corridors and its individual divisions. These divisions are the largest cities (transshipment segments) and tracks between these cities (haulage segments). The efficiency for the NS-M and R-A corridors are 47% and 62%

respectively. The efficiency of the NS-M corridor seems quite low, which is because of the many haulage segments that have an efficiency score below 20%. This is much lower compared to other applications of DEA in the field of IFT, where efficiencies below 20% only occur occasionally. It is unlikely that all these low values are the actual technical efficiencies. A likely issue is the data quality, for which 5 main drawbacks are identified. These will be discussed in Section 6.3. Section 6.4 will provide some recommendations to improve the data quality and availability for future research.

Although the values of the efficiencies cannot be guaranteed to point us in the right directions, there are some clear trends in the efficiencies for both transshipment and haulage segments. For both, efficiency mostly decreases for larger inputs. At first, this seems logical from a theoretical standpoint, because increasing inputs will decrease the efficiency if output remains equal. However, it would be expected that there are some segments that accompany larger inputs with larger outputs, making larger segments more efficient according to the principle of economies of scale. The R-A corridor is mostly more efficient because of the higher efficiencies for larger haulage segments, so this might indicate some advantage of economies of scale. For transshipment however, it seems like the opposite. This is probably because of the limitations, which will be discussed in Section 6.3.

Using the three sub-questions, we can answer the main research question:

How can the performance of an intermodal freight transport chain be assessed?

The performance of an IFT-chains within a network can be determined using a weighted slacks-based DEA. This can be used for IFT-networks with IFT-chains that overlap, have parallel segments and have different structures. KPIs from both scientific and practical literature as inputs and outputs can then be used to assess the performance of corridors in the European IFT-network.

6.2. Contributions

This section describes both the scientific and practical contribution that this thesis adds.

6.2.1. Scientific contributions

This thesis contributed in various way to scientific knowledge about IFT and performance assessment. These contributions are discussed below.

Literature review of KPIs in scientific and practical literature

There were already existing literature reviews that evaluate the use of KPIs in European IFT (Posset et al., 2010, 2009), of which the most recent is from 2010. The more recent literature review by Saeedi et al. (2018) evaluates the performance indicators used in various modes of transport, it is however not focused on IFT in Europe. Therefore, this thesis contributes with an updated literature review of KPIs used in IFT in Europe, toghether with a comparison between scientific and practical literature.

Issues with use of cumulative intermediates for Network DEA

Saeedi et al. (2019) presents a Network DEA for the evaluation of IFT-chains. The method compares divisions with each other, independent of the position in the chain, which is a novel development within DEA literature and is also incorporated in the developed DEA in this thesis. The use of intermediates however, has some issues. Due to the choice of cumulative intermediates, the efficiency is dependent on the position in the chain. This is shown with a mathematical explanation, as well as with a simulation of a fictional network. This contradicts the aforementioned position-independent comparison of divisions. Therefore, this thesis shows that both developments cannot be applied at the same time. The simulation of a fictional network shows that efficiencies of the same division increases when that division is placed later in the chain, unrelated to the actual technical efficiency. It is therefore not recommended to use a cumulative intermediate in a Network DEA, neither for chains with or without a fixed structure.

Weighted slacks-based DEA

There were some limitations for both regular and Network DEAs. Regular DEAs only take into account DMUs with similar activities and cannot be used on entire chains. However, their discriminatory power is relatively large, because they assume all DMUs with the same technology to be separate, even if they are part of a chain. Network DEAs on the other hand, can take into account chains, but need a fixed structure and intermediates between divisions. We developed a weighted slacks-based DEA, which combines the discriminatory power of a regular DEA with the network functionalities of a Network DEA. It is a hybrid version of a regular DEA and Network DEA, since it optimises for an entire chain, but without the use of intermediates. It can be applied on chains without fixed structure, with parallel segments and overlap between chains. Although the model was developed for IFT-chains, the formulation is generic and could also be used for other applications.

6.2.2. Practical contributions

Besides the scientific contributions, this thesis has practical contributions as well.

Model for assessment of IFT-chains in a network

In practice, some DEAs have are already been used to evaluate the performance of IFT. This is however mostly done on either a haulage or transshipment segment level and rarely on a chain or corridor level. Because of overlap, different structures and parallel segments, Network DEAs are rarely used to evaluate IFT-chains. The developed weighted slacks-based DEA can take all these characteristics of IFT-chains into account. This makes it possible to assess the performance of IFT-chains within a network, such as done in this thesis with a part of the European IFT-network. When there is sufficient quality and availability of data, this can give an overview of the relative performance of entire chains and its individual haulage and transshipment segments.

Lack of data quality and availability in European IFT

As mentioned in the previous paragraph, the quality and availability of data is a requirement for the evaluation of IFT-network. From the application of the model in this thesis, it is clear that there are some limitations in the data about European IFT. Although Tavasszy and de Jong (2014) already gives an overview of the available data sources for freight transport modelling and its limitations, this thesis again shows the difficulties of obtaining this data for application. Essential data to assess the performance of IFT in Europe are reliable data for terminals and number of trains. Furthermore, when the EU wants to see the performance of individual corridors, all data needs to be collected on a corridor level.

Indication of efficiency of European IFT

This thesis has applied the weighted SB-DEA on two TEN-T corridors NS-M and R-A. Especially the main part of the R-A corridor has a consistently high efficiency score. Although there limitations to the application, this indicates that this corridor is more efficient than the NS-M corridor. This thesis opens up the possibility to further research the efficiency of the different TEN-T corridors.

6.3. Discussion

This section describes the limitations of this research and consists of three parts. Firstly, the limitations of the data availability is discussed, followed by a discussion about the limitations of the application and especially the data collection. Lastly, the limitation of the model application is discussed.

Data availability

- Afsharian et al. (2016) warns for choosing input and output variables based on availability, instead of their appropriateness. "a sound DEA application requires first to systematically deriving the objectives to be taken into account. This is a prerequisite for the next step of selecting reasonable performance factors" (Afsharian et al., 2016). With the use of the literature reviev, the objective and appropriate indicators were selected, but were not all available. Therefore, some proxy indicators such as number of trains are used, while indicators closer to the goal level are preferred.

- The preferred detail level for the performance of haulage and transshipment segments is on operator level, since even within a region, performance can vary between segments such as terminals. Although the inputs for transshipment are known on a terminal level, the detail level of the volumes is regional (NUTS2). Therefore, the transshipment segments are aggregated in regions around large cities. It is thus not possible to say something about the efficiency of individual terminals and thus decreases the discriminatory power of the model. For later investment and policy recommendations of individual terminals, volume data on a more detailed level for the different corridors is needed.

Data collection and estimation

- Some terminals might be missing from the data. The terminal input data comes from two data sources and seems to include most terminals in the North Sea - Mediterranean and Rhine - Alpine corridors, but it is possible that some terminals are not included in either of the terminal sources. This can have an impact on the total amount of resources/inputs and thus on the efficiency. When most sufficiently large terminals are in the data sources, missing a smaller terminal in a region likely has only a small effect on the total inputs and efficiency. However, when a terminal is missing in a smaller region and thus is a relatively large share of that regions' inputs, the impact can be much larger. That region would be assigned a larger efficiency than its technical efficiency, while it can significantly decrease the assigned efficiency of the other regions.

- The number of trains for some haulage segments are likely larger than in the collected data. To collect the number of trains, all possible OD-pairs within corridors are searched for in an IFT-planner by hand. This should result in all trains belonging to that corridor, but there are still haulage segments with an unrealistically low number of trains. If this share of missing trains is large, it can have a significant impact on the efficiencies of haulage segments. A solution is to analyse the entire IFT-network and include all possible OD-pairs. This should then be automated, because there are too many combinations of origins and destinations to do this manually.

- The transshipped volumes can be much larger than in the data. The data available are volumes between pre- & main haulage and between main & end-haulage. However, data that is transshipped between two main haulage segments (main-to-main transshipment) are not available. Regions with a large share of main-to-main transshipment, will thus be assigned a lower volume, resulting in lower efficiency scores. This is mainly the case for ports, such as Antwerpen and Rotterdam, which indeed both have an efficiency of lower than 25%. Estimation of transshipped volumes based on the inputs is not a solution, since it would give every region the exact same ratio of inputs to outputs and would assign all segments the same efficiency. Actual transshipment volumes per region or terminal are needed to solve this issue.

- The volumes used as output for both haulage and transshipment are from 2010 and are all scaled with the same factor to 2020. However, the relative increase might be different for different segments, creating underestimating volumes in some segments, while overestimating volumes in others. This influences the assigned efficiencies. However, unless there are very large differences for a specific region or segment between the average increase and the actual increase, this is likely not influencing efficiencies in a very large way.

- The volume data does not differentiate between bulk or containerized volumes. The same constant multiplier is used as a share for the volumes on all segments, but could differ per segment. This again can influence the assigned efficiencies and show divisions more or less efficient than they actually are.

- For the terminals, some input data was missing. This data is estimated by calculating the weighted arithmetic mean for the missing input, based on similarity of terminals. This was needed to run the model, but an estimation based on other terminals with which it will be compared in the DEA decreases the discriminatory power.

- The distribution of both inputs and outputs for overlapping haulage and transshipment segments is unknown, because lack of data. This means that there is no difference visible in the model for these segments. The efficiency of these overlapping segments is taken into account for both chains, decreasing the discriminatory power.

Model application

- All divisions in the chain are assigned the same weights in the DEA ($\frac{1}{|I_k|}$). set to 1/(number of segments). Including an extra unneeded division would get the same weight as the other divisions and have an impact on the efficiency of the entire chain. This gives researches the responsibility to only include relevant divisions or vary the weight in the application of the DEA.

- Some of the rail terminals transships to IWW, SSS or maritime shipping, while most do not. All these terminals are compared to each other. A prerequisite of DEA is that its divisions and DMUs use the same technologies (Dyson et al., 2001). However, there could be a difference in efficiencies between terminals with and without IWW, since terminals including IWW could have other inputs that are not taken into account in the DEA.

- The developed weighted slacks-based DEA has not been tested on realistic data set with complex chain structures yet. The weighted SB-DEA has been compared to a traditional DEA and a Network DEA using both a numerical example and a sample of real data, which proved to solve some issues of a Network DEA with cumulative intermediates. However, it has not been tested on realistic data with complex chain structures. It is expected that it will present realistic efficiencies, but this cannot be guaranteed yet and should be tested in the future.

6.4. Recommendations

This section describes the recommendations and next steps that can be taken for both future research and practical application.

6.4.1. Scientific recommendations

- The developed model could be tested more extensively to see its behaviour for various situations. The developed model has been tested using a simple test network and simulations. After comparing it with a traditional DEA and a Network DEA, it performed as expected. Furthermore, the data has been tested on some real data, using the sample data from Saeedi et al. (2019). However, this was a small sample with straight-forward IFT-chains, so it is interesting to see how the model behaves in a more complicated network with multiple chains with different structures. This can give more information about the working of the model. Secondly, a sensitivity analysis can be performed, to see how changes in the data impact the behaviour and results of the model.

- The weights in the model are currently all the same for all divisions. However, it can be tested what the behaviour would be when the weights are adapted based on one of the variables. An example would be to let the weights be dependent on the output, such that divisions with larger outputs are taken more into account. However, changing the weights should be in line with the objective of the DEA. Therefore the method to determine the weights should preferably be defined beforehand. If only adjusted afterwards, there is the risk that the efficiencies are afterward adjusted to the view of the person that runs the DEA, until possible desired efficiencies are produced by the model, similar to p-hacking

6.4.2. Practical recommendations

- The performance of all 9 TEN-T corridors can be assessed by using the developed model. This should give a more complete overview of the performance of IFT in Europe.

- The data collection for this thesis had limitations and could be improved with finding and using more data sources. Especially the number of trains and terminals have a large impact and would benefit from using more and reliable data sources.

- A sensitivity analysis as described in the scientific recommendations is not only useful for the understanding of the model, but also to see how reliable the results are for the North Sea - Mediterranean and Rhine - Alpine corridors. An example could be to add the inputs of an average terminal to all transshipment segments, or change the share of volumes transshipped. After doing this either sequentially or simultaneously and run the DEA for every case, it should be more clear how robust the efficiencies obtained in this thesis are.

- The developed model could also be used for other applications. Cases in which there is a network without a fixed structure and where divisions can occur multiple times, could benefit from the approach as developed in this thesis. An example that has quite some parallels with IFT is public transport. The DMUs could be bus or train lines, where transportation and stops or stations are the divisions. There are many lines with different structures and overlap between lines is common. Another example could be modern power grids. Because power is supplied by many local power suppliers such as power plants, solar panels and wind turbines, the power grid is a complicated network without a fixed structure, with many possible parallel elements. Separate (micro)grids can be seen as DMUs, while transformers and electricity cables can be seen as divisions.

- The European Commission should discuss what goal they have for the TEN-T corridors and how they want to assess the performance of IFT in Euorpoe. The EC should then from these goals decide which performance indicators measure these goals best, using the method as described in Afsharian et al. (2016).
- When the goals for European IFT are clear, the needed data for the TEN-T corridors should be collected and shared publicly in one central place. Currently, it is difficult to assess the TEN-T corridors, because much of the data is owned by the operators and infrastructure managers and only some of this data is shared with researchers. Having one central place where all data is collected, makes it easier to research the performance of European IFT. All indicators the EC decides are important for the goals of IFT as discussed in the previous recommendation, should be available here. Examples could be: volumes for each haulage segment and the actual transshipped volumes for terminals, detailed data of terminals and distribution of number of passenger and freight trains on haulage segments.

References

Afsharian, M., Ahn, H., & Neumann, L. (2016). Generalized DEA: an approach for supporting input/output factor determination in DEA. *Benchmarking*, *23*(7), 1892–1909. doi: 10.1108/BIJ-07-2015-0074

Antognoli, M., Capodilupo, L., Marinacci, C., Ricci, S., Rizzetto, L., & Tombesi, E. (2018). Present and Future Operation of Rail Freight Terminals. In (pp. 233–273). Retrieved from `http://link.springer.com/10.1007/978-3-319-78295-9_6` doi: 10.1007/978-3-319-78295-9{_}6

Ballis, A. (2004, 1). Introducing Level-of-Service Standards for Intermodal Freight Terminals. *Transportation Research Record: Journal of the Transportation Research Board*, *1873*(1), 79–88. Retrieved from `http://journals.sagepub.com/doi/10.3141/1873-10` doi: 10.3141/1873-10

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, *30*(9), 1078–1092. Retrieved from `http://pubsonline.informs.org.https://doi.org/10.1287/mnsc.30.9.1078http://www.informs.org` doi: 10.1287/mnsc.30.9.1078

Bask, A., & Rajahonka, M. (2017). The role of environmental sustainability in the freight transport mode choice: A systematic literature review with focus on the EU. *International Journal of Physical Distribution and Logistics Management*, *47*(7), 560–602. doi: 10.1108/IJPDLM-03-2017-0127

Beamon, B. M. (1999, 3). Measuring supply chain performance. *International Journal of Operations & Production Management*, *19*(3), 275–292. Retrieved from `https://www.emeraldinsight.com/doi/10.1108/01443579910249714` doi: 10.1108/01443579910249714

Beasley, J. E. (1990). Comparing university departments. *Omega*. doi: 10.1016/0305-0483(90)90064-G

Bijma, F., Jonker, M., & van der Vaart, A. (2017). *Introduction to Mathematical Statistics*. Amsterdam University Press.

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology, and Application*. doi: 10.1007/978-94-011-0637-5

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. Retrieved from `https://ac.els-cdn.com/0377221778901388/1-s2.0-0377221778901388-main.pdf?_tid=bcf0a2a2-0959-4f37-8aea-5e7c975964dd&acdnat=1541679408_c9c7aefa03fed5920b88a2e4216372e8` doi: 10.1016/0377-2217(78)90138-8

Cook, W. D., Green, R. H., & Zhu, J. (2006, 2). Dual-role factors in data envelopment analysis. *IIE Transactions*, *38*(2), 105–115. Retrieved from `https://www.tandfonline.com/doi/full/10.1080/07408170500245570` doi: 10.1080/07408170500245570

Cook, W. D., & Zhu, J. (2014). *Data Envelopment Analysis - A Handbook on the Modeling of Internal Structures and Networks* (Tech. Rep.). Retrieved from `http://www.springer.com/series/6161`

Crainic, T. G. (2003, 2). Long-Haul Freight Transportation. In *Handbook of transportation science* (pp. 451–516). Kluwer Academic Publishers. doi: 10.1007/0-306-48058-1{_}13

Davydenko, I., Jordans, M., & Krupe, S. (2007). *Promoting Innovative Intermodal Freight Transport: D.4.1 European Benchmarks in Intermodal Transport* (Tech. Rep.).

Davydenko, I., Zomer, G., & Krupe, S. (2007). *Promoting Innovative Intermodal Freight Transport: D4.3 Recommendations processes and quality improvements in intermodal transport* (Tech. Rep.).

De Toni, A., & Tonchia, S. (2001). Performance measurement systems Models, characteristics and measures. *International Journal of Operations and Production Management*, *21*(1-2), 46–70. Retrieved from `http://www.emerald-library.com/ft` doi: 10.1108/01443570110358459

Deutsche Bahn. (n.d.). *Intermodal Connections railway.tools*. Retrieved from `https://railway.tools/#/en/connections`

De Vries, M. V. I. (2019). *Exploring the effects of carbon pricing on the decarbonization of freight transportation in the Netherlands* (Tech. Rep.).

Di Pierro, B., Iacobellis, G., Turchiano, B., & Ukovich, W. (2017, 9). Performance assessment for intermodal transportation systems: A case study. In *2017 ieee international conference on service operations and logistics, and informatics (soli)* (pp. 236–241). IEEE. Retrieved from `http://ieeexplore.ieee.org/document/8121000/` doi: 10.1109/SOLI.2017.8121000

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001, 7). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259. doi: 10.1016/S0377-2217(00)00149-1

ECORYS. (2004). *Ex ante Evaluation Marco Polo II (2007-2013)* (Tech. Rep. No. June 2004).

Eisenfuhr, F., Weber, M., & Langer, T. (2010). *Rational decision making*. Springer.

European Commission. (n.d.-a). *Corridors*. Retrieved from `https://ec.europa.eu/transport/node/2443`

European Commission. (n.d.-b). *Mobility and Transport*. Retrieved from `https://ec.europa.eu/transport/`

European Commission. (2011). *White Paper on transport: Roadmap to a single European transport area - towards a competitive and resource-efficient transport system* (Tech. Rep.). Retrieved from `http://europa.eu` doi: 10.2832/30955

European Commission. (2020a). *Infrastructure and Investment*. Retrieved from `https://ec.europa.eu/transport/themes/infrastructure_en`

European Commission. (2020b). *TENtec Interactive Map*. Retrieved from `https://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/map/maps.html`

European Commission. (2020c). *Trans-European Transport Network (TEN-T)*. Retrieved from `https://ec.europa.eu/transport/themes/infrastructure/ten-t_en`

European Court of Auditors. (2016). *Special Report Rail freight transport in the EU: still not on the right track* (Tech. Rep.). Retrieved from `http://europa.eu` doi: 10.2865/53961

Eurostat. (n.d.). *NUTS Classification*. Retrieved from `https://ec.europa.eu/eurostat/web/nuts/background`

Eurostat. (2003). *Glossary for Transport Statistics* (Tech. Rep.). Retrieved from `http://www.oecd.org/cem`

Eurostat. (2020). *Railway freight transport statistics*. Retrieved from `https://ec.europa.eu/eurostat/statistics-explained/index.php/Railway_freight_transport_statistics`

Evers, P. (1994). The occurrence of statistical economies of scale in intermodal transportation. *Transportation journal*, *33*(4), 51–63.

Fancello, G., Schintu, A., & Serra, P. (2018, 1). An experimental analysis of Mediterranean supply chains through the use of cost KPIs. *Transportation Research Procedia*, *30*, 137–146. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352146518300887` doi: 10.1016/J.TRPRO.2018.09.016

Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society: Series A (General)*, *120*(3), 253–281.

Freight transport statistics - modal split. (n.d.). Retrieved from `https://ec.europa.eu/eurostat/statistics-explained/index.php/Freight_transport_statistics_-_modal_split#Modal_split_in_the_EU`

Gianpiero, M., Andrea, B., Massimiliano, C., Matteo, S., Mirko, C., Luca, C., & Roberto, K. (2015, 6). GIS-based decision support system for multi criteria analysis of intermodal transport networks. In *2015 ieee international conference on engineering, technology and innovation/ international technology management conference (ice/itmc)* (pp. 1–10). IEEE. Retrieved from `http://ieeexplore.ieee.org/document/7438674/` doi: 10.1109/ICE.2015.7438674

Gogas, M., Adamos, G., & Nathanail, E. (2017, 1). Assessing the performance of intermodal city logistics terminals in Thessaloniki. *Transportation Research Procedia*, *24*, 17–24. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352146517303423` doi: 10.1016/J.TRPRO.2017.05.061

Golany, B., & Roll, Y. (1989, 1). An application procedure for DEA. *Omega*, *17*(3), 237–250. Retrieved from `https://www.sciencedirect.com/science/article/pii/0305048389900297?via%3Dihub` doi: 10.1016/0305-0483(89)90029-7

International Union of Railways (UIC). (2016). *GRI G4 indicators handbook Guidelines for railway companies*.

Islam, D. M. Z., Zunder, T. H., & Jorna, R. (2013, 4). Performance evaluation of an online benchmarking tool for European freight transport chains. *Benchmarking: An International Journal*, *20*(2), 233–250. Retrieved from `http://www.emeraldinsight.com/doi/10.1108/14635771311307696` doi: 10.1108/14635771311307696

Ivanov, D., & Sokolov, B. (2013). Control and system-theoretic identification of the supply chain dynamics domain for planning, analysis and adaptation of performance under uncertainty. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2012.08.021

Janic, M. (2007, 1). Modelling the full costs of an intermodal and road freight transport network. *Transportation Research Part D: Transport and Environment*, *12*(1), 33–44. doi: 10.1016/j.trd.2006.10.004

Kilibarda, M. J., Andrejić, M. M., & Popović, V. J. (2013). *Creating and measuring logistics value* (Tech. Rep.). Retrieved from https://www.researchgate.net/publication/259527793

KombiConsult. (n.d.). *Intermodal Terminals in Europe.* Retrieved from http://www.intermodal-terminals.eu/database

Kos, S., Vukić, L., & Brčić, D. (2017, 4). Comparison of External Costs in Multimodal Container Transport Chain. *PROMET - Traffic&Transportation*, *29*(2), 243. Retrieved from http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/2183 doi: 10.7307/ptt.v29i2.2183

Lewis, H. F., & Sexton, T. R. (2004, 8). Network DEA: efficiency analysis of organizations with complex internal structure. *Computers & Operations Research*, *31*(9), 1365–1410. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305054803000959 doi: 10.1016/S0305-0548(03)00095-9

Liang, L., Yang, F., Cook, W. D., & Zhu, J. (2006). DEA models for supply chain efficiency evaluation. *Ann Oper Res*, *145*, 35–49. Retrieved from https://link.springer.com/content/pdf/10.1007%2Fs10479-006-0026-7.pdf doi: 10.1007/s10479-006-0026-7

Lindberg, C.-F., Tan, S., Yan, J., & Starfelt, F. (2015, 8). Key Performance Indicators Improve Industrial Performance. *Energy Procedia*, *75*, 1785–1790. Retrieved from https://www.sciencedirect.com/science/article/pii/S1876610215012424 doi: 10.1016/J.EGYPRO.2015.07.474

Lozano, S., & Gutiérrez, E. (2014). Transportation Planning and Technology A slacks-based network DEA efficiency analysis of European airlines A slacks-based network DEA efficiency analysis of European airlines. , *Transportation Planning and Technology*, *37*(7), 623–637. Retrieved from http://www.tandfonline.com/action/journalInformation?journalCode=gtpt20 doi: 10.1080/03081060.2014.935569

Macharis, C., Van Hoeck, E., Pekin, E., & van Lier, T. (2010, 8). A decision analysis framework for intermodal transport: Comparing fuel price increases and the internalisation of external costs. *Transportation Research Part A: Policy and Practice*, *44*(7), 550–561. doi: 10.1016/j.tra.2010.04.006

Markovits-Somogyi, R. (2011). Measuring efficiency in transport: The state of the art of applying Data Envelopment Analysis. *Transport*, *26*(1), 11–19. doi: 10.3846/16484142.2011.555500

Massiani, J. (2008). *Can we use hedonic pricing to estimate freight value of time?* Retrieved from www.econstor.eu

Mobiliteitsbeeld 2019 (Tech. Rep.). (2019). Kennisinstituut voor Mobiliteitsbeleid (KiM). Retrieved from https://www.kimnet.nl/publicaties/%0Ahttps://www.kimnet.nl/publicaties/rapporten/2019/11/12/mobiliteitsbeeld-2019-vooral-het-gebruik-van-de-trein-neemt-toe

Nathnail, E., Gogas, M., & Adamos, G. (2016, 1). Urban Freight Terminals: A Sustainability Cross-case Analysis. *Transportation Research Procedia*, *16*, 394–402. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352146516306512 doi: 10.1016/J.TRPRO.2016.11.037

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: a literature review. *International Journal of Operations & Production Management*, *15*(4), 35. Retrieved from https://www.emerald.com/insight/content/doi/10.1108/01443579510083622/full/pdf?title=performance-measurement-system-design-a-literature-review-andresearch-agenda doi: 10.1108/01443570510633639

ÖBB. (n.d.). *Key Performance Indicators in Railways-Who measures what?* (Tech. Rep.).

Ockwell, A. (2001). Benchmarking the performance of intermodal transport. *Report. Division of Transport, Organisation for Economic Co-operation and Development, Paris.*

Organisation for Economic Co-operation and Development. (2008). *OECD glossary of statistical terms.* OECD.

Organisation for Economic Co-operation and Development (OECD). (2002). *Benchmarking Intermodal Freight Transport* (Tech. Rep.). Retrieved from https://www.itf-oecd.org/

Panagakos, G., & Psaraftis, H. N. (2017). Model-based corridor performance analysis – An application to a European case. *European Journal of Transport and Infrastructure Research*, *17*(2), 225–247. Retrieved from https://orbit.dtu.dk/portal/en/publications/id(efc78311-4df4-44e4-b2ca-2550b8b4a124).html

Pastori, E. (2015). *Modal share of freight transport to and from EU ports* (Tech. Rep.). Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2015/540350/IPOL_STU(2015)

540350_EN.pdf

Port of Rotterdam. (n.d.). *InlandLinks*. Retrieved from http://www.inlandlinks.com

Posset, M., Gronalt, M., & Häuslmayer, H. (2010). *COCKPIIT – Clear Operable and Comparable Key Performance Indicators for intermodal transportation* (Tech. Rep.). Retrieved from https://trimis.ec.europa.eu/project/clear-operable-and-comparable-key-performance-indicators-intermodal-transportation

Posset, M., Pfliegl, R., & Zich, A. (2009, 1). An Integrated Set of Indicators for Assessment of Inland Waterway Transportation Performance. *Transportation Research Record: Journal of the Transportation Research Board*, *2100*(1), 86–93. Retrieved from http://journals.sagepub.com/doi/10.3141/2100-10 doi: 10.3141/2100-10

ProRail. (2017). *Goederen over het spoor (in Nederland en daarbuiten) - ProRail.* Retrieved from https://www.prorail.nl/spoorgoederenvervoer/goederen-over-het-spoor-in-nederland-en-daarbuiten

Ricci, S., Capodilupo, L., Mueller, B., Karl, J., & Schneberger, J. (2016). Assessment Methods for Innovative Operational Measures and Technologies for Intermodal Freight Terminals. *Transportation Research Procedia*, *14*, 2840–2849. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S235214651630357X doi: 10.1016/j.trpro.2016.05.351

Saeedi, H., Behdani, B., Wiegmans, B., & Zuidwijk, R. (2019, 6). Assessing the technical efficiency of intermodal freight transport chains using a modified network DEA approach. *Transportation Research Part E: Logistics and Transportation Review*, *126*, 66–86. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1366554518307932 doi: 10.1016/j.tre.2019.04.003

Saeedi, H., Behdani, B., Wiegmans, B., & Zuidwijk, R. A. (2018). Performance Measurement in Freight Transport Systems. *SSRN Electronic Journal.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3122275 doi: 10.2139/ssrn.3122275

Santos, J., Negas, E. R., & Cavique, L. (2013). Introduction to Data Envelopment Analysis. In *Efficiency measures in the agricultural sector* (pp. 37–50). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-94-007-5739-4_3 doi: 10.1007/978-94-007-5739-4{_}3

Schipper, C., Vreugdenhil, H., & de Jong, M. (2017, 12). A sustainability assessment of ports and port-city plans: Comparing ambitions with achievements. *Transportation Research Part D: Transport and Environment*, *57*, 84–111. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S136192091730384X doi: 10.1016/j.trd.2017.08.017

Statistical functions SciPy. (2019). Retrieved from https://docs.scipy.org/doc/scipy/reference/stats.html

Tavasszy, L., & de Jong, G. (2014). Data Availability and Model Form. In *Modelling freight transport* (pp. 229–244). Elsevier Inc. doi: 10.1016/B978-0-12-410400-6.00010-0

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, *130*(3), 498–509. Retrieved from https://www-sciencedirect-com.tudelft.idm.oclc.org/science/article/pii/S0377221799004075 doi: 10.1016/S0377-2217(99)00407-5

Tone, K., & Tsutsui, M. (2009). Dynamic DEA: A slacks-based measure approach. *European Journal of Operational Research*, *197*(1), 243–252. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221708004517 doi: https://doi.org/10.1016/j.ejor.2008.05.027

TRIMIS - European Commission. (n.d.). *European Transport Policy Information System Developement and Implementation of Data Collection Methodology for EU Transport Modelling (ETISplus).* Retrieved from https://trimis.ec.europa.eu/project/european-transport-policy-information-system-developement-and-implementation-data-collection

Wee, B. V., & Banister, D. (2016, 3). How to Write a Literature Review Paper? *Transport Reviews*, *36*(2), 278–288. Retrieved from http://www.tandfonline.com/doi/full/10.1080/01441647.2015.1065456 doi: 10.1080/01441647.2015.1065456

Wiegmans, B., & Konings, R. (2015, 6). Intermodal Inland Waterway Transport: Modelling Conditions Influencing Its Cost Competitiveness. *The Asian Journal of Shipping and Logistics*, *31*(2), 273–294. Retrieved from https://www.sciencedirect.com/science/article/pii/S2092521215000346?via%3Dihub doi: 10.1016/J.AJSL.2015.06.006

Wiegmans, B., & Witte, P. (2017, 12). Efficiency of inland waterway container terminals: Stochastic frontier and data envelopment analysis to analyze the capacity design- and throughput efficiency. *Transporta-*

tion Research Part A: Policy and Practice, *106*, 12–21. doi: 10.1016/j.tra.2017.09.007

Wiegmans, B. W., Masurel, E., & Nijkamp, P. (1999). Intermodal freight terminals: An analysis of the terminal market. *Transportation Planning and Technology*, *23*(2), 105–128. doi: 10.1080/03081069908717643

Wiegmans, B. W., Rietveld, P., Pels, E., & van Woudenberg, S. (2004). Container terminals and utilisation of facilities. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, *31*(3), 313–339.

Yang, F., Wu, D., Liang, L., Bi, G., & Wu, D. D. (2011). Supply chain DEA: production possibility set and performance evaluation model. *Ann Oper Res*, *185*, 195–211. Retrieved from https://link.springer.com/content/pdf/10.1007%2Fs10479-008-0511-2.pdf doi: 10.1007/s10479-008-0511-2

A

Methodology and KPIs scientific literature

Figure A.1 shows the exact methodology of the selection of papers for the literature review of scientific literature. The excluding of irrelevant literature (last step) is explained in Section 3.2.1.

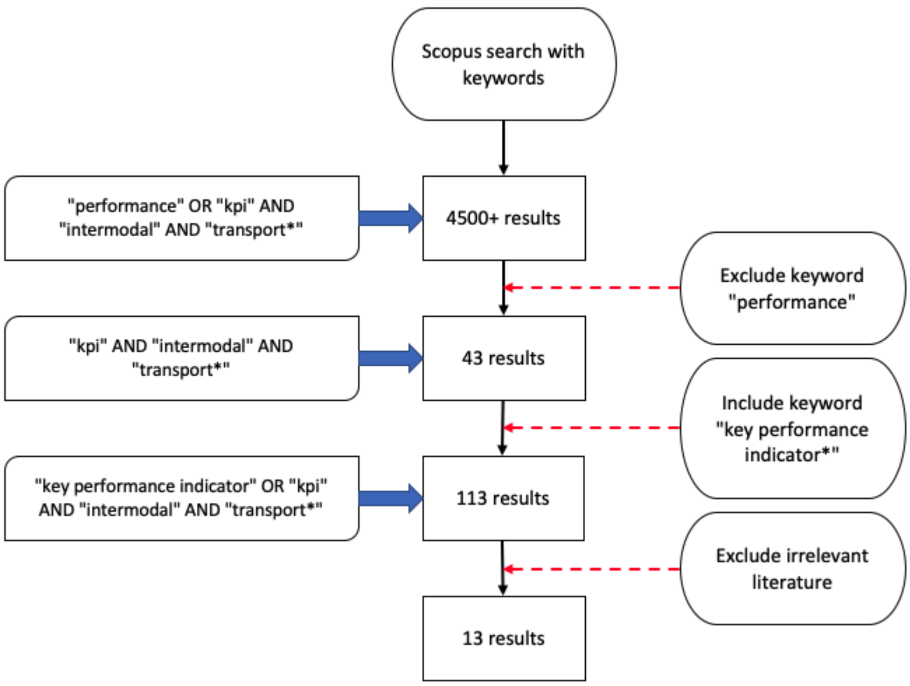


Figure A.1: Extensive methodology of selecting papers in the literature review with scientific literature.

There were a total of almost 50 KPIs in the 13 selected papers. Similar indicators such as system utilization rate and utilization of ports have been combined to the KPIs presented in Table A.1. It could possibly be further simplified, e.g. speed, distance and travel time could be combined in two indicators, but since they were specifically mentioned this way in some papers, this is not done.

Table A.1: All used KPIs in the selected papers of the literature review, sorted by frequency.

KPI	Terminals & ports	Chains & transportation	Total
Environment & emissions	4	4	8
Waiting time	3	3	6
Reliability	3	2	5
Transshipment time	3	1	4
Employee productivity	3	1	4
Safety & security	3	1	4
Energy consumption	3	1	4
Transport costs	0	4	4
Information availability	2	1	3
Equipment productivity	3	0	3
Travel time	1	2	3
Saturation ratio	2	0	2
Handling cost	2	0	2
Punctuality	2	0	2
Loss and damages	2	0	2
Supply chain vissibility	2	0	2
Fair and equal access companies	2	0	2
Terminal occupancy	2	0	2
Maintanability	1	1	2
System utilization rate	2	0	2
Flexibility	1	1	2
Volume	2	0	2
Quality of handling	1	1	2
Vehicle speed	1	1	2
Storage	1	0	1
Handling equipment rate	2	0	2
Terminal accessibility	1	0	1
Utilization rate equipment	1	0	1
Utilization rate storage	1	0	1
Investments	1	0	1
Market share	1	0	1
Customer satisfaction ITS	1	0	1
Distance	1	0	1
Weather conditions	1	0	1
Robustness	0	1	1
Level of service	0	1	1
Tariffs	0	1	1
Infra availability	0	1	1
Capacity ports	0	1	1
Storage time ports	0	1	1
Congestion costs	0	1	1
Human resource c osts	0	1	1

B

KPIs practical literature

Table B.1: PROMIT Performance Indicator Structure. Reprinted from Posset et al. (2010, p. 105).

	Level 1: Customer Level (transport from source to destination)	Level 2: 4PL system integrator Level (transportation as a number of activities: road-terminal-rail/inland waterway-terminal-road)	Level 3: Operator Level (specific operations witihn all activities of Level 2)
Key Performance Indicators	- price (cost) - lead time (transit time) effective 95% reiliability on-time - lead time variability - frequency of service - shipment capability loading unit business model regulations compliance cargo - damages and theft	- price (cost) - lead time (execution time) effective 95% reiliability on-time - lead time variability - shipment capability loading unit business model regulations compliance cargo - damages and theft	- price (cost) - lead time (execution time) - capacity terminal fleet - capactiy utilization - reliability - damages - theft **- operation specific**
Additional Indicators	- flexibility capacity re-routing - information flow general track and trace administration - invoicing accuracy - relationship with shippers behaviour responsibility - market different providers competition - energy consumption ton/km	- flexibility capacity re-routing - information flow general track and trace administration - invoicing accuracy - relationship with provider behaviour responsibility - market different providers competition - energy consumption per ton/km	- schedule convenience - IT (ICT) systems cost savings reliability flexibility information exchange quality management cargo booking process - energy consumption - number of accidents - operation-specific - frequency of service
	Most important performance indicators which are present at all abstraction levels price - lead time (execution time) - lead time variability - frequency of service - shipment capability – damages - theft		

Table B.2: PROMIT Operation Specific Indicators. Reprinted from Posset et al. (2010, p. 106).

	Level 3: Operator Level (specific operations within all activities of Level 2)	

Operation Specific Indicators	**Road**: Source – Terminal and Terminal – Destination
	Special cargo storage/handling possibilities
	Frequency of train service (trains and ships)
	Cost structure
	    Vehicle related
	    Wages and salaries (staff only)
	    Materials and loading fees
	    Net profit
	Fleet indicators
	    AVG distance travelled per vehicle per week
	    Cost of vehicle per kilometer
	    AVG distance between breakdowns per vehicle
	    AVG distance travelled per day
		Intermodal Terminals
		Terminal working hours
		Proximity to market
		Cargo handlng equipment (investment)
		Loading/unloading time per unit handled
		Storage capacity
		Average storage time and variance
		Ship (train) arrival rate
		    Average arrival rate
		    Peak arrival rate
		    Variance of arrival rate
		Truck arrival rate
		    see ship (train)
		Truck loading/unloading turnover time
		    average, 95%, variance
		Cranes
		    AVG movement time/distance between yard and crane
		    AVG number of moves per crane per hours
		    AVG utilization rate per crane per year
		    Max speed of equipment with full load
		    Max lifting weight
		    AVG number of injuries per crane per year
		Tug boats/manoeuvre locomotives necessary
		AVG vessels' (train) waiting time between end of loading and start of voyage
		AVG vessels' (train) waiting time between end of voyage and start of loading/unloading
		AVG length and variance of trucks' queue
		Min time required between truck arrival and ship (train) departure
		Proportion of English-speaking personnel
		
		AVG: average

	Inland Waterway Ship: Terminal - Terminal
	Cost of vessel per day
	Maximum speed of vessel
	Cargo capacity
	AVG speed between ports (terminals)
	AVG time between stops (ports/terminals)
	AVG distance between stops (ports/terminals)
	AVG time at port per roundtrip
	Consumption of fuel
	    Per kilometer travelled
	    Per TEU/km or ton/km
	    Per day	

	Train: Terminal - Terminal Special cargo storage/handling possibilities Driving/waiting time ratio	

C

Extra information data envelopment analysis

This appendix provides a further and more in depth explanation of the model description and formulation from Chapter 4.

C.1. NDEA

The formulas from Chapter 4 are presented here again for readability. It can be difficult to see the difference between Equations 4.18 and 4.19, but they use different λs ($\lambda_{k',c}$ and $\lambda_{k,c}$), which are for consecutive divisions k' and k. These consecutive divisions and its intermediates can be seen the selected in Figure C.1. This helps to let the model optimise the efficiency for the entire chain.

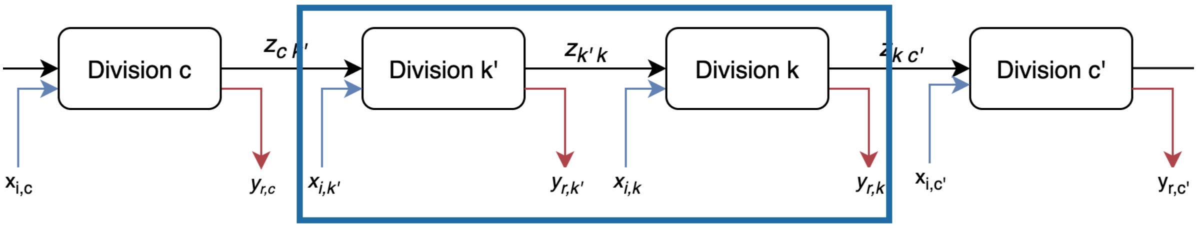


Figure C.1: Two consecutive divisions k and k' in a part of an IFT-chain.

$$\min_{\lambda_k, s^-_{i,k}} \rho_o = \sum_{k \in K_o} w_k \left(1 - \frac{1}{m_k} \sum_{i \in I_k} \frac{s^-_{i,k}}{x_{i,k}}\right) \tag{4.13}$$

s.t.

$$\sum_{c \in C_k} \lambda_{k,c} x_{i,c} + s^-_{i,k} = x_{i,k} \qquad \forall k \in K_o, \forall i \in I_k \tag{4.14}$$

$$\sum_{c \in C_k} \lambda_{k,c} y_{r,c} - s^+_{r,k} = y_{r,k} \qquad \forall k \in K_o, \forall r \in R_k \tag{4.15}$$

$$\lambda_{k,c}, s^-_{i,k}, s^+_{r,k} \geq 0 \qquad \forall k \in K_o, \forall c \in C, \forall i \in I_k, \forall r \in R_k \tag{4.16}$$

$$\sum_{c \in C_k} \lambda_{k,c} = 1 \qquad \forall k \in K_o \tag{4.17}$$

$$\sum_{c',c \in C_k} \lambda_{k',c} z_{c'c} = z_{k'k} \qquad \forall (k', k) \in K_o \tag{4.18}$$

$$\sum_{c',c \in C_k} \lambda_{k,c} z_{c'c} = z_{k'k} \qquad \forall (k', k) \in K_o \tag{4.19}$$

k' is the division in DMU_o on the location just before division k. Since there is no division on the location before the starting division in a chain (e.g. the first transshipment division), the constraints in Equations 4.18 and 4.19 start at second division in a chain o. This is formulated as $\forall(k', k) \in K_o$.

C.2. Different chain structures

For IFT, chains can have different structures, depending on the amount of transshipments and haulages and the type of haulage. For example, a chain consisting of three terminals and two haulage segments is different from a chain consisting of two terminals and a single haulage segment. For the original NDEA, these two chains could not be compared and no information can be cross compared. However, Saeedi et al. (2019) developed a model that can use information of single divisions, to evaluate entire chains. This is presented in Figure C.2.

The reference sets for the original NDEA are very small, which limits the discriminatory power. Since all transshipment segments here are comparable, the NDEA by Saeedi et al. (2019) has a reference set with all transshipment segments. This is done similarly for the haulage segments. An important requirement here is that the technology of the transshipment and haulage segments within a reference should be comparable (Tone & Tsutsui, 2009).

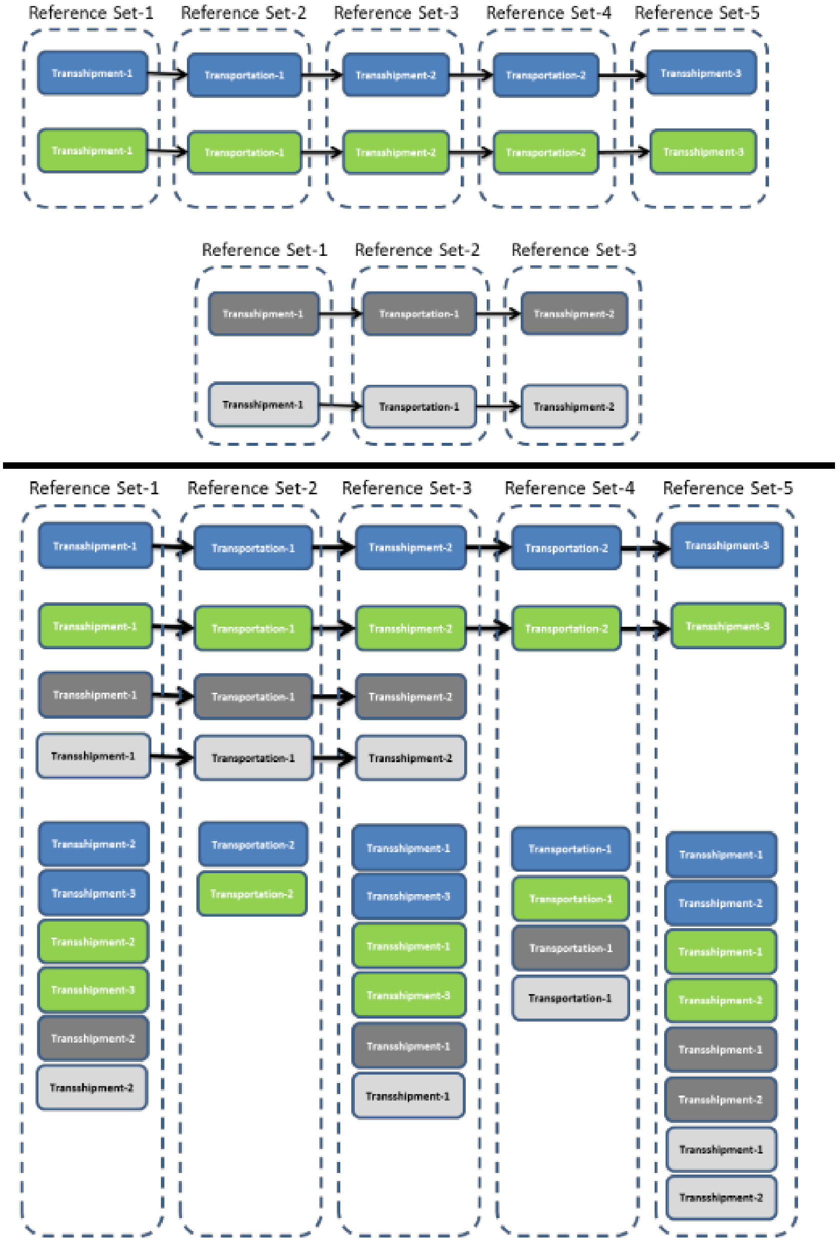


Figure C.2: The reference set for chains with different structures for a original NDEA (above) and for the NDEA by Saeedi et al. (2019) (below).

D

Numerical experiments DEA

As explained in Section 4.3, there are some issues with the use of a cumulative VIFTS for a NDEA. These issues are related to the mathematical formulation of the modified NDEA. To confirm that these issues arise when a NDEA is actually performed, a numerical experiment has been done. Both a regular DEA and a NDEA are used to compare the effect of position of the terminals on their efficiency. The experiment is done with a network of 8 fictive separate IFT chains. It is simplified by using terminals only, haulage segments are excluded. A set of 24 terminals with fictive values for their inputs and outputs is used. These terminals are randomly distributed over 8 non-overlapping chains with 3 terminals each, so each terminal is placed at either 1$^{\text{st}}$, 2$^{\text{nd}}$ or 3$^{\text{rd}}$ position. This is done in 1000 simulations, because calculating efficiencies for all $\frac{24!}{8!} = 1.5 \times 10^{19}$ possible combinations would consume too much computation time. The result of the 1000 simulations for DEA and NDEA can be seen in Figure D.1.

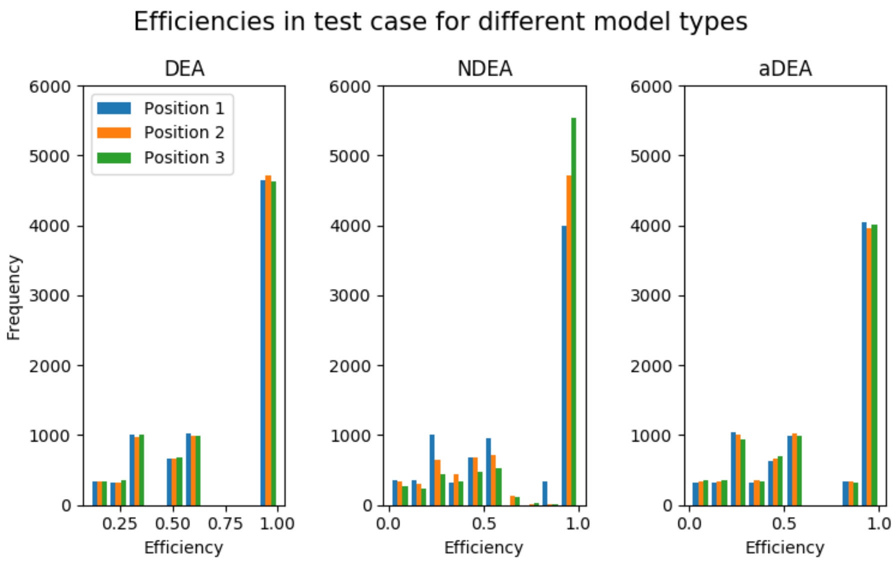


Figure D.1: Efficiencies of single divisions on different positions in the chain for numerical experiment with DEA, NDEA and aDEA with 1000 simulations.

A regular DEA does not take into account the position of a terminal in a chain when calculating the efficiency and therefore there are no large differences between efficiencies on each position in Figure D.1: DEA. For NDEA however, Figure D.1: NDEA shows a higher efficiency is reached more frequently when a division is positioned further in the chain, as was expected from the model formulation. A statistical test is needed to see if the difference between position 1, 2 and 3 is significant.

Both the DEA and NDEA models do not follow the normal distribution, which is clear from the shape of the sample distribution in Figure D.1. Even more so, after trying to fit the sample to all 95 different continuous probability distributions found in Python module Scipy (*Statistical functions SciPy*, 2019), there is no good fit for any of these distributions. Therefore, a non-parametric test is needed. For the NDEA, we are interested to see the difference between the samples of position 1 and position 3. We are comparing two related, matched samples, and thus the Wilcoxon signed-rank test can be used (Bijma, Jonker, & van der Vaart, 2017). This test is used to see if two dependent samples were selected from populations with the same distribution. We use a significance level of $\alpha = 0.01$. The null and alternative hypotheses are presented below.

H_0 : The samples of efficiencies at positions 1 and 3 come from the same population.

H_1 : The samples of efficiencies at positions 1 and 3 do not come from the same population

Applying the Wilcoxon signed-rank test gives a p-value of 3.1×10^{-110}, which is much smaller then α. Therefore, we can reject the null hypothesis and say that there is a significant difference between positions in the chain for a NDEA.

As stated before, this gives the issue of not being able to compare individual DMUs fairly. The issue lies within the choice of an intermediate service for IFT chains when using a NDEA. There is no clear intermediate product or service that is transmitted from transhipment to haulage or vice versa. Saeedi et al. (2019) proposed VIFTS as intermediate, but that raised the model issues because it is a cumulative value. It is also questionable whether there is a physical meaning in the value of the previous divisions influencing the efficiency of divisions later in the chain.

The adapted DEA has also been simulated with the test case and the result can be seen in Figure D.1. Applying again the Wilcoxon signed-rank test gives a p-value of 0.51, which is larger than $\alpha = 0.01$ or even the commonly used $\alpha = 0.05$. Therefore, there is no reason to believe there is a difference in efficiency between the 1^{st} and 3^{rd} positions. Table D.1 shows the p-value for all three models.

Table D.1: P-values from Wilcoxon signed-rank test for numerical experiment for different model types.

	DEA	NDEA	aDEA
p-value	0.80	0.00	0.51
significant	no	yes	no

E

Estimation of missing data for terminals

The input data for terminals that are used in the case study had some missing data. This appendix describes the method for estimating these missing data points based on similar terminals in the data set.

E.1. Data sources

The input data of the terminals is collected from two different sources, namely http://www.inlandlinks.com from the Port of Rotterdam and http://www.intermodal-terminals.eu/database/ from the AGORA project by the EC. The former used data was audited and is used as main source. However, the latter has data from a larger set of terminals and is used to provide data about the other terminals. These combined resulted in a list of 83 terminals for the North Sea - Mediterranean and Rhine - Alpine corridors.

There were 33 terminals of which at least one of the input data was missing. Data was mainly missing for the surface area, number of cranes or number of reachstackers. These data have been estimated by finding terminals that are similar for the input data that is known fir them. For example, when the surface area is missing, the terminal is compared to all other terminals with regard to number of railway tracks, track length, number of cranes and number of reachstackers.

E.2. Estimation of missing data

Several steps have been taken to identify similar terminals and then give an estimation of the missing data points. The method is described below and generalized for division k with missing input values $x_{\not{i},k}$ ($\not{i} \in \not{I}_k$, where $\not{I}_k$ is the set of missing inputs for division k). $x_{i,k}$ are the known input values ($i \in I_k \setminus \not{I}_k$, where I is the set of all inputs, so the set difference $I_k \setminus \not{I}_k$ is the set of all known inputs).

First, the similarity between divisions for each individual input is calculated using a pairwise comparison. Division k is compared to division c ($c \in C_k$, where C_k is the set of all divisions with the same activities). The ratio $f_{i,kc}$ between $x_{i,k}$ and $x_{i,c}$ is calculated, which is always between 0 and 1, as seen in Equation E.1. The product of the ratios for all known inputs are used as weight u_{kc} to calculate a weighted arithmetic mean or weighted average as estimate for the missing inputs. The calculation of the weight and the estimate for missing input is shown in Equations E.2 and E.3.

$$f_{i,kc} = \frac{\min(x_{i,k}, x_{i,c})}{\max(x_{i,k}, x_{i,c})} \tag{E.1}$$

$$u_{kc} = \prod_{i \in I \setminus \not{I}} f_{i,kc} \tag{E.2}$$

$$x_{\not{i},k} = \frac{\sum_{c \in C_k} u_{kc} x_{\not{i},c}}{\sum_{c \in C_k} u_{kc}} \tag{E.3}$$

By taking the product of the ratios, divisions with more similar inputs get a larger weight for the weighted arithmetic mean. A division c with one much larger and one much smaller input than division k, won't be seen as similar.

When the estimate is for example 3.86 reachstackers, this is not rounded to integers. Although decimal reachstackers are not possible in reality, this the best estimate according to this method and rounding would decrease the accuracy of the estimation. It does not give a problem for a DEA model, so we refrained from rounding.

E.3. Limitations of missing data estimation

This method of estimating missing data points has some limitations. Taking the product of the ratios gives the most similar terminals the largest weight for calculating the weighted arithmetic mean. However, it does not take into account if division k is larger or smaller than division c. This could be a problem for a division k that is at the extreme of the division spectrum. As example, let's assume a division k has the largest inputs for $i \in I \setminus \not{i}$. It could be expected that $x_{\not{i},k}$ would also have the largest value of $x_{\not{i},c} \forall c \in C_k$. However, because the use of the weighted arithmetic mean, the estimated value for $x_{\not{i},k}$ will always be lower than the largest known $\not{i}$ from C_k.

Another possible limitation of the estimation of missing data lies in the following application of the DEA. Using average values for the missing data points could give more average efficiencies of the divisions, which decreases the discriminatory power of the model. This is mainly an issue because of missing data itself and using most other methods of estimation would not change this. One method to increase differences in efficiencies between divisions is to assume the worst value each missing input value. Since inputs need to be minimized, this would mean using the largest value from $x_{\not{i},c} \forall c \in C_k$. It would practically eliminate the use of the missing input from the DEA, only using the known data. However, it would result in a lower efficiency score for most divisions with missing data. Therefore, we think the use of estimation is preferred, to get more realistic results at the cost of some discriminatory power.

F
Data for TEN-T corridors

Table F.1: Inputs for terminals in the Rhine - Alpine corridor. The numbers in bold are estimated according to the method as described in Appendix E.

Terminal	**City**	**Area**	**No. of tracks**	**Track length**	**No. of cranes**	**No. of reach-stackers**
Katoen Natie Terminals (Quay 1227)	Antwerpen	563430	2	1000	7	12
Van Doorn Container Depot	Antwerpen	160000	1	300	0	0
Antwerpen ATO	Antwerpen	95000	1	150	2	3
Antwerpen Cirkeldyk	Antwerpen	77000	4	2600	2	4
Antwerpen Combinant	Antwerpen	125000	5	3100	2	**3.85**
Antwerpen HTA Hupac Terminal Antwerp	Antwerpen	53000	5	3100	3	3
Antwerpen Main-Hub	Antwerpen	20000	8	5600	3	4
Contargo Rhein-Waal-Lippe Terminal GmbH	Arnhem	45000	3	550	2	2
Contargo Basel	Basel	23000	6	1220	2	0
Basel - Weil am Rhein (DUSS)	Basel	**92611**	6	3660	3	**3.8**
Basel - Weil am Rhein (Rheinhafen)	Basel	15000	1	300	**1.56**	1
Ottmarsheim	Basel	40000	2	800	3	5
Trimodal Terminal Brussels	Bruxelles	17000	2	700	1	**1.91**
Neuss Trimodal	Dºsseldorf	88000	6	3270	2	8
Duisburg Intermodal Terminal GmbH (DIT)	Duisburg	185000	6	700	3	5
DeCeTe	Duisburg	190000	1	700	4	5
Rhein Ruhr Terminal (Gateway West/Home Terminal)	Duisburg	149000	6	400	4	6
Duisburg Kombiterminal (DKT)	Duisburg	60000	6	2820	2	**3.6**
Duisburg KV-Hub Rhein-Ruhr	Duisburg	**87697**	4	2840	2	**3.55**
Duisburg logport III (Hohenbudberg)	Duisburg	140000	8	5920	2	2
Duisburg Ruhrort Hafen	Duisburg	140000	9	5980	3	4
Duisburg Trimodal Terminal (D3T)	Duisburg	37500	4	1400	1	**2.72**
Emmelsum (Contargo)	Duisburg	34000	2	590	1	**2.15**
Emmelsum (Jerich)	Duisburg	136000	1	280	1	1

Table F.1: Inputs for terminals in the Rhine - Alpine corridor. The numbers in bold are estimated according to the method as described in Appendix E.

Terminal	**City**	**Area**	**No. of tracks**	**Track length**	**No. of cranes**	**No. of reach-stackers**
Contargo Neuss GmbH	Dusseldorf	80000	2	550	3	4
Krefelder Container Terminal	Dusseldorf	32500	1	400	2	3
Contargo Frankfurt Ost	Frankfurt am Main	38000	2	50	2	0
Frankenbach Container Terminal	Frankfurt am Main	80000	4	600	5	1
Arluno	Frankfurt am Main	65000	2	880	**2.59**	3
Frankfurt am Main (West)	Frankfurt am Main	18000	6	1940	2	2
Ghent Container Terminal	Gent	**120437**	3	2250	**2.79**	**3.92**
Contargo Karlsruhe	Karlsruhe	23000	2	400	1	1
Karlsruhe (DUSS)	Karlsruhe	**79727**	4	2000	2	2
Lauterbourg - R3flex	Karlsruhe	48000	2	800	1	2
CSA Container Service Andernach	Koblenz	42000	3	750	2	3
Contargo Koblenz	Koblenz	20000	1	300	2	2
CTS Container Terminal	Koln	132000	9	730	5	4
Dormagen	Koln	62000	2	673.5	1	1
Contargo Mannheim	Ludwigshafen	95000	3	1250	3	2
Contargo Ludwigshafen	Ludwigshafen	81000	2	660	3	1
DP World Germersheim GmbH Co KG	Ludwigshafen	**69203**	3	410	0	2
Ludwigshafen KTL	Ludwigshafen	130000	13	7836	7	2
Mannheim DP World	Ludwigshafen	9000	2	250	1	1
Mannheim-Handelshafen	Ludwigshafen	**87093**	5	3150	2	**3.62**
Worms	Ludwigshafen	22500	2	480	2	1
Trimodales Containerterminal Aschaffenburg GmbH	Luxembourg	20000	2	211	1	1
CIM S.p.A. Centro Interportuale Merci di Novara	Milano	160598	7	600	0	7
Busto Arsizio-Gallarate	Milano	245000	13	8755	12	1
Melzo	Milano	160000	5	1300	**4.09**	10
Terminal Intermodale di Mortara	Milano	110000	3	2100	**2.57**	2
Moerdijk Container Terminals (MCT)	Rotterdam	380000	3	950	7	16
Rotterdam Container Terminal (RCT)	Rotterdam	170000	3	350	2	6
Van Doorn Container Depot	Rotterdam	107000	3	350	2	1
Beatrix Terminal	Rotterdam	262300	3	937	10	7
Rail Service Center Rotterdam B.V.	Rotterdam	240000	8	750	4	5
CTT Rotterdam	Rotterdam	42342	3	200	1	2
ECT Delta Terminal	Rotterdam	100000	4	2800	2	2
Pernis Combi Terminal	Rotterdam	50000	3	1400	**2.32**	2
Rotterdam APM Terminals Maasvlakte II	Rotterdam	**87908**	4	3000	2	**3.6**
Rotterdam Euromax	Rotterdam	**100854**	6	4800	**2.91**	3
Rotterdam RWG	Rotterdam	**87432**	6	4500	2	**3.66**
Progeco Zeebrugge	Zeebrugge	20000	6	600	8	0

Table F.2: Inputs for terminals in the North Sea - Mediterrenean corridor. The numbers in bold are estimated according to the method as described in Appendix E.

Terminal	**City**	**Area**	**No. of tracks**	**Track length**	**No. of cranes**	**No. of reach-stackers**
Katoen Natie Terminals (Quay 1227)	Antwerpen	563430	2	1000	7	12
Van Doorn Container Depot	Antwerpen	160000	1	300	0	0
Antwerpen ATO	Antwerpen	95000	1	150	2	5
Antwerpen Cirkeldyk	Antwerpen	77000	4	2600	2	5
Antwerpen Combinant	Antwerpen	125000	5	3100	2	**3.86**
Antwerpen HTA Hupac Terminal Antwerp	Antwerpen	53000	5	3100	3	3
Antwerpen Main-Hub	Antwerpen	20000	8	5600	3	4
Muizen (Ambrogio)	Antwerpen	35000	2	1500	2	**2.85**
Charleroi Dry Port	Bruxelles	40000	5	2250	2	2
Garocentre Terminal	Bruxelles	60000	3	1650	1	1
Trimodal Terminal Brussels	Bruxelles	17000	2	700	1	**1.91**
Dijon	Dijon	20000	3	1050	1	**2.19**
Ghent Container Terminal	Gent	**120437**	3	2250	**2.79**	**3.92**
Lille Dourges Container Terminal (LDCT)	Lille	600000	7	750	5	4
Contargo North France SAS	Lille	45000	1	**645**	1	2
Delcatransport Rekkem	Lille	5000	3	1350	**2.57**	5
Mouscron Dry Port	Lille	**90167**	2	600	**2.44**	2
Prouvy	Lille	10000	1	450	1	2
Athus Container Terminal	Luxembourg	160000	1	4000	3	6
Lyon-St. Priest	Lyon	45000	5	1580	2	**2.3**
Avignon-Courtine (Novatrans)	Marseille	85296	9	2765	**2.89**	1
Miramas	Marseille	495000	3	2100	**4.56**	4
Bonneuil sur Marne	Paris	42000	2	1700	**2.2**	2
Gennevilliers	Paris	**148671**	4	1100	3	6
Noisy-le-Sec	Paris	70000	10	3490	**2.68**	**3.52**
Moerdijk Container Terminals (MCT)	Rotterdam	380000	3	950	7	16
Rotterdam Container Terminal (RCT)	Rotterdam	170000	3	350	2	6
Van Doorn Container Depot	Rotterdam	107000	3	350	2	1
Beatrix Terminal	Rotterdam	262300	3	937	10	7
Rail Service Center Rotterdam B.V.	Rotterdam	240000	8	750	4	5
CTT Rotterdam	Rotterdam	42342	3	200	1	2
ECT Delta Terminal	Rotterdam	100000	4	2800	2	2
Pernis Combi Terminal	Rotterdam	50000	3	1400	**2.32**	2
Rotterdam APM Terminals Maasvlakte II	Rotterdam	**87908**	4	3000	2	**3.6**
Rotterdam Euromax	Rotterdam	**100854**	6	4800	**2.91**	3
Rotterdam RWG	Rotterdam	**87432**	6	4500	2	**3.66**
ETK Euro Terminal Kehl GmbH	Strasbourg	40000	4	1700	3	3
Strasbourg Terminal Conteneurs Nord	Strasbourg	81800	5	3050	1	6
Strasbourg Terminal Conteneurs Sud	Strasbourg	107500	4	1280	2	6
Progeco Zeebrugge	Zeebrugge	20000	6	600	8	0

G
Results for TEN-T corridors

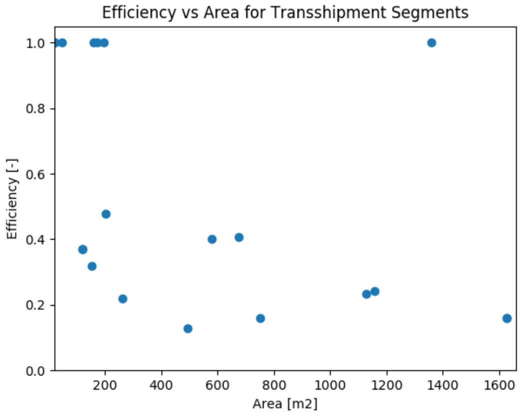
This appendix shows the input data and results for the application of the weighted SB-DEA of the North Sea - Mediterranean and Rhine - Alpine corridors. Tables G.1 and G.2 show the pre-processed data that is used directly in the DEA. The result are the efficiencies, that are both shown as values for the individual transshipment and haulage segments, and in Figures G.1 and G.2.

Table G.1: Efficiency from DEA, inputs and outputs for transshipment for the cities in the North Sea - Mediterranean and Rhine - Alpine corridors. The volume is given in tonnes per year.

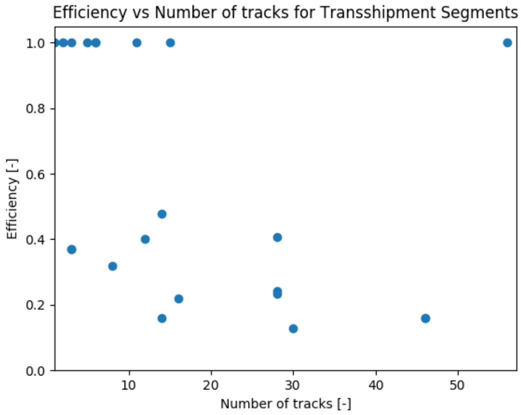
Corridor	**City**	**Efficiency**	**Inputs**					**Output**
			Area [m2]	**No. of tracks**	**Track length [m]**	**No. of cranes**	**No. of reach-stackers**	**Volume [tonnes]**
North Sea - Mediterranean	Marseille	0.40	580296	12	4865	7.45	5	14676233
	Rotterdam	0.16	1627836	46	20037	37.23	51.26	29558161
	Zeebrugge	1.00	20000	6	600	8	0	4037129
	Dijon	1.00	20000	3	1050	1	2.19	8036521
	Antwerpen	0.23	1128430	28	17350	21	35.71	29069039
	Luxembourg	1.00	160000	1	4000	3	6	765275
	Lille	0.16	750167	14	3795	12.01	15	8600564
	Paris	0.22	260671	16	6290	7.88	11.52	11225106
	Lyon	1.00	45000	5	1580	2	2.3	11310342
	Gent	0.37	120437	3	2250	2.79	3.92	2962849
Rhine - Alpine	Gent	0.37	120437	3	2250	2.79	3.92	2962849
	Frankfurt am Main	0.48	201000	14	3470	11.59	6	17313030
	Zeebrugge	1.00	20000	6	600	8	0	4037129
	Antwerpen	0.24	1156903	28	17100	20	30.85	29069039
	Karlsruhe	0.32	150727	8	3200	4	5	8326810
	Basel	1.00	170611	15	5980	9.56	9.8	43380814
	Luxembourg	1.00	20000	2	211	1	1	765275
	Koln	1.00	194000	11	1403.5	6	5	16631956
	Dusseldorf	1.00	1359697	56	25850	30	50.02	90269121
	Ludwigshafen	0.13	493796	30	14036	18	12.62	11291676
	Rotterdam	0.16	1627836	46	20037	37.23	51.26	29558161
	Milano	0.41	675598	28	12755	18.66	20	39231152

Table G.2: Efficiency from DEA, inputs and outputs for the haulage segments in the North Sea - Mediterranean and Rhine - Alpine corridors. Trains are the number of trains per week, the volume is given in tonnes per year.

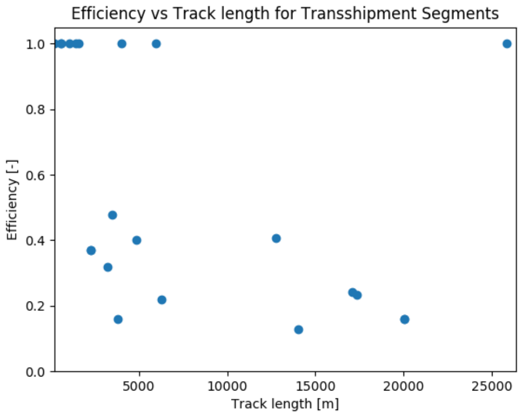
Corridor	**Haulage segment**	**Efficiency**	**Input**	**Output**
			Trains	**Volume**
North Sea - Mediterranean	Zeebrugge-Antwerpen	0.13	34	3233038
	Strasbourg-Dijon	0.13	14	1408286
	Metz-Dijon	0.21	12	1893293
	Luxembourg-Metz	0.29	10	2098283
	Metz-Strasbourg	0.18	7	1006476
	Charleroi-Luxembourg	0.19	20	2794002
	Antwerpen-Charleroi	0.90	5	3213140
	Lille-Paris	1.00	1	813266
	Gent-Lille	1.00	6	3892158
	Antwerpen-Gent	1.00	39	6406587
	Zeebrugge-Gent	0.11	36	2805233
	Dijon-Lyon	0.15	35	3610413
	Lyon-Marseille	0.10	47	3398109
	Rotterdam-Antwerpen	0.30	5	1151214
Rhine - Alpine	Milano-Genova	1.00	5	3557404
	Basel-Milano	1.00	194	15732050
	Karlsruhe-Basel	0.66	220	13448861
	Ludwigshafen-Karlsruhe	0.76	220	14433689
	Frankfurt am Main-Ludwigshafen	0.73	230	14447930
	Koln-Frankfurt am Main	1.00	243	17390456
	Antwerpen-Koln	0.20	96	4895576
	Zeebrugge-Antwerpen	0.13	34	3233038
	Rotterdam-Antwerpen	0.30	5	1151214
	Dusseldorf-Koln	1.00	130	12671303
	Rotterdam-Dusseldorf	0.50	136	8429549



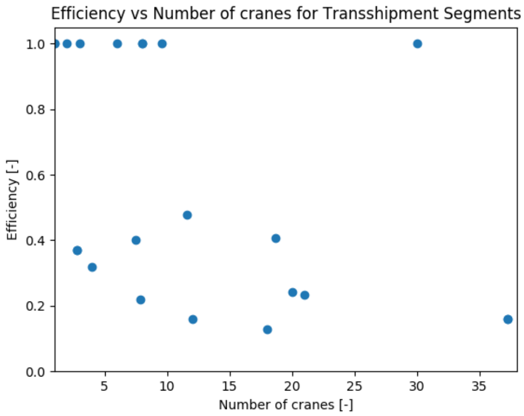
(a) Input: Area



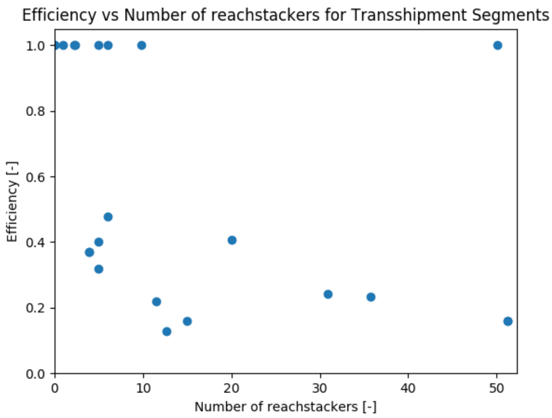
(b) Input: Number of tracks



(c) Input: Track length

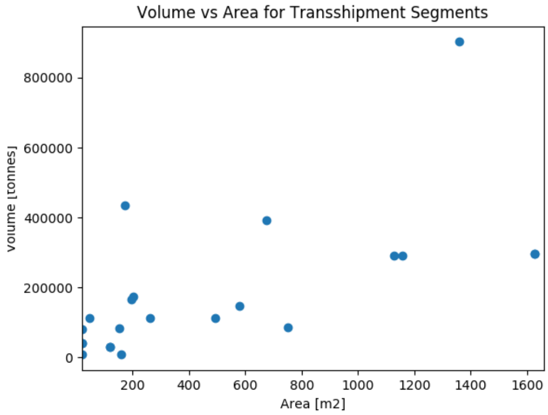


(d) Input: Number of cranes

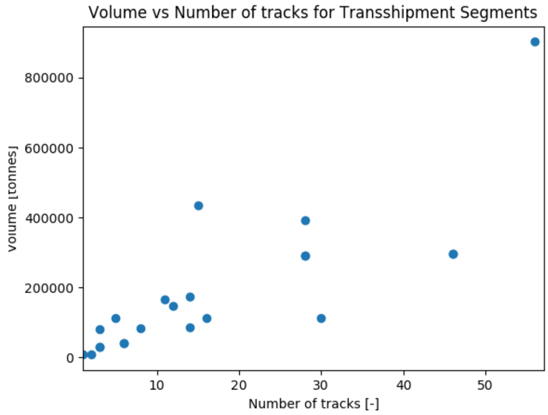


(e) Input: Number of reachstackers

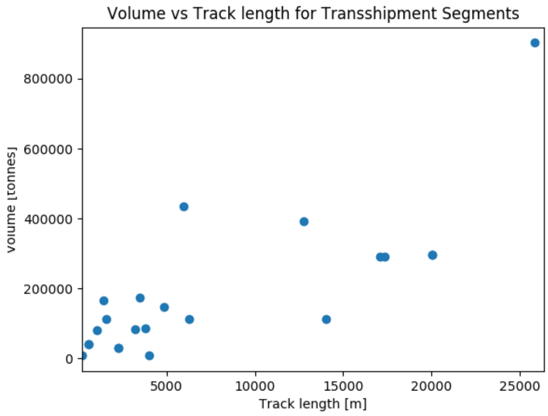
Figure G.1: Relation between inputs and efficiency from weighted slacks-based DEA for transshipment segments.



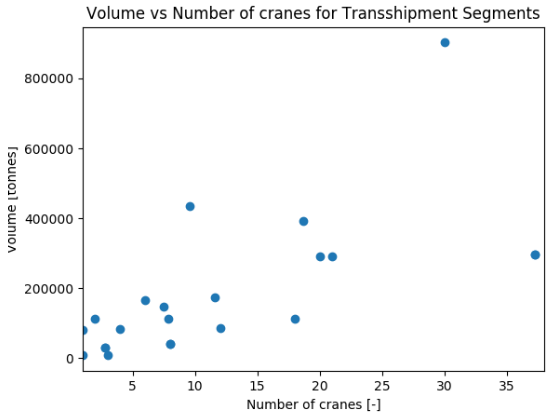
(a) Input: Area



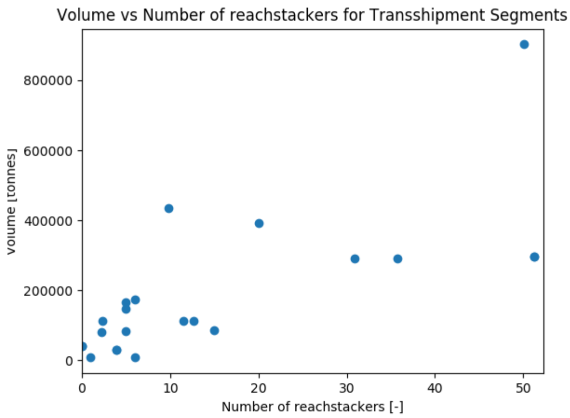
(b) Input: Number of tracks



(c) Input: Track length



(d) Input: Number of cranes



(e) Input: Number of reachstackers

Figure G.2: Relation between inputs and volume for transshipment segments.