Preliminary Examination Paper

Submitted to Committee Member

Under Supervision of: Prof. K. Triantis

Efficiency Evaluation in Railway System Operations: A Focus on Punctuality of Train Paths

Yas Mashayekhy

PhD Student, Industrial and Systems Engineering

Virginia Tech

Abstract

This study addresses the operational efficiency of train paths within a railway system by applying Data Envelopment Analysis (DEA) to train paths, treating them as decision-making units (DMUs). The primary focus is on punctuality as a crucial aspect of service performance, with the goal of enhancing economic infrastructure through timetable scheduling and management. This research innovatively contributes to the field by integrating systems thinking with DEA, thus offering a non-parametric optimization-based method for performance evaluation in complex socio-technical systems (STS) such as railway transportation systems.

By considering train paths as DMUs, this approach enables a detailed performance analysis at both aggregate and disaggregate levels, allowing for a more nuanced understanding of the network's operational dynamics. The study considers the Belgium's railway system (INFRABEL) as a case study and leverages a unique dataset comprised of real-world operational data, which includes various indicators of performance such as punctuality, frequency, and cancellations.

The findings highlight significant potential for improving the technical efficiency of train paths, which can lead to better utilization of existing resources and enhanced passenger satisfaction without necessitating extensive capital investment. The study's implications extend to railway operations management, suggesting that strategic adjustments to scheduling and the management of train paths can substantially mitigate operational inefficiencies and enhance overall system performance.

Keywords: Railway Transportation System, Socio-Technical Systems, Efficiency, Performance Evaluation, Data Envelopment Analysis (DEA), System Thinking, Train Path, Delay

Spring 2025

1. Introduction (Motivation and Background)

The transportation sector plays a pivotal role in the economic development of a country, significantly contributing to the Gross Domestic Product (GDP) (World Bank, n.d.). This sector demands special attention due to its expansive scope and impact, which often requires governmental intervention and substantial investments at a macroeconomic level (The Geography of Transport Systems: Rodrigue, n.d.). Small enterprises alone cannot effectuate meaningful changes within this sector, particularly in national rail systems that are integral to sustainable transportation, especially post-electrification.

Railway systems, characterized by their extensive infrastructure requirements including rail tracks, platforms, and catenary systems, necessitate efficient operation to maximize capacity utilization. The primary objective of service providers, typically governmental bodies, is to deliver effective services to the populace, who are not only the primary beneficiaries but also the principal funders as taxpayers. These providers are accountable to various stakeholders, including the general public and unions, making any infrastructural modifications a matter of careful consideration due to the associated high costs and extended timelines (Makosvek et al, 2015; Lindel et al., 2021; The Geography of Transport Systems: Rodrigue, n.d.).

Railway systems place a high priority on delivering satisfactory service to passengers, with punctuality recognized as a critical performance indicator (Larumbe, 2021; Zakeri & Olsson, 2018). Timely service significantly influences both the technical performance of the railway system and passenger satisfaction, thereby playing a pivotal role in the economic success of the organization. Delays in train operations not only undermine passenger confidence but also lead to economic loss, impacting the overall operational viability of the transport service (Ianonne, 2012).

On the other hand, railway systems are complex socio-technical systems (STS) characterized by intricate interactions between human operators, technological components, and organizational structures (Topcu et al., 2019). The complexity arises from the non-linear relationships and dependencies that exist within the system, making any adjustments or modifications particularly challenging (Carayon et al., 2015; Siegel & Schraagen, 2014). Besides, improving infrastructure is expensive and can be implemented primarily in the long run. Any upgrade in the infrastructure requires that railway service providers spend a lot of the organization’s resources, including time, energy, capital, and labor. Therefore, making small but effective improvements that don't require a lot of resources is crucial to achieve improved performance in railway operations before making decisions about changes in the infrastructure (Songhori et al., 2020).

Timetables and train schedules constitute a promising area for intervention, as they offer the potential to significantly improve railway operations without the need for infrastructure changes and with minimal resource investment at the company level. This approach allows for testing and implementation of different policies and decision-making priorities with minimal side effects, as it does not necessitate changes to the existing infrastructure. Salido et al. (2012) highlight the importance of a robust railway timetable, suggesting that effective scheduling can optimize the use of existing infrastructure while minimizing disruptions (Salido et al., 2012). This perspective aligns with the findings of Oh, who presents a resource-oriented approach to augmenting train timetables, demonstrating that modifications to schedules can be made within the constraints of current resources and infrastructure (Oh, 2023). Learning about the train delays and their complex impacts on railways system performance is highly important topic for the transportation sector (Sobrie et al., 2023).

1.1 **Research Objective**

The objective of this paper is to analyze train path efficiency, since by doing this the railway system managers can identify areas for improvement, optimize scheduling, and strengthen the economic infrastructure through enhanced railway operations. This research employs a novel approach using train paths or itineraries (or routes) as the decision-making unit (DMU or unit of analysis), a critical entity when measuring railway system performance. Train paths, which dictate organized routes from point A to point B, are underexplored in the railway performance measurement literature. The train path as an entity/component is a system that can lead to the representation of overall transportation network performance by considering both the disaggregate and aggregate levels. This can lead to a holistic approach of evaluating the system while considering entities/components that comprise the transportation service and provide a better understanding of the system in terms of its operations.

1.2 **Research Contribution**

This research contributes to the existing body of knowledge by introducing a novel application of systems thinking (Cabrera et al., 2008) and Data Envelopment Analysis (DEA) (Charnes, et al., 1978) as a non-parametric optimization-based method when there is lack of standards exists to evaluate the performance of a complex railway system through its components (the train paths) within the network. While prior studies have examined efficiency in railways through broader operational or infrastructural lenses, this study focuses on the underexplored component of train paths as decision-making units (DMUs). By doing so, it bridges a methodological gap between performance evaluation and system-level operational decision-making. Additionally, it integrates socio-technical perspectives with quantitative efficiency modeling, providing a multi-perspective approach that supports data-driven, safety-conscious optimization strategies for railway scheduling. Besides, it provides some managerial insights for the railway systems staff, helping them for better decisions, by providing the efficiency scores and rankings for the train paths, making a distinction between them and putting priorities for traffic management. Furthermore, it can be used for redefining and modifying the current KPIs used in the system (specifically for punctuality). For this research, a unique dataset from publicly available data and event data from INFRABEL is created. This data set will be used to understand operational realities and test specific operational interventions both in this research and future research.

1.3 **System - Operations and Performance**

The fundamental premise of this research is that the evaluation of different train runs or services can be done through train paths or itineraries. Each train run (service run) is part of one of the train paths, based on its origin and destination (and hence, direction).

Train services are run, managed, and controlled by traffic control centers (TCCs) in the railway systems. Each TCC is operated by zone-based, meaning that it is responsible for covering a specific part of the network. TCCs are in contact with each other as well as with train operators for managing the traffic using the controlling system (the switches and signals). Upon entering a train (service) into their zone, operators in TCCs are making decisions (either real-time or determined from before) to navigate the trains on the tracks and platforms in the stations.

Usually, these controlling operations and actions are done automatically, using the automated system, and humans are in charge of monitoring. However, when the system is operating densely at its maximum possible capacity (for example, in rush hours when many trains are managed within a short time interval), it is more difficult to manage the flow smoothly, and this is where the importance of the real-time operational decision making in the TCCs is pivotal. Making decisions about assigning which trains to which tracks with priority (for both inbound and outbound assignments) as well as decisions about cancelling some train runs based on the disruptions or delays in the system network, are other examples of TCC operator decisions. Each decision for each train run can have a noticeable impact on the train services, and the relations between all the railway system network components are not very clear and easy to track. These are reasons why this system is considered a complex socio-technical system (Richards et al, 2024) and is one of the core assumptions of this research.

The remainder of this paper is organized as follows: Section 2 presents an integrative literature review, outlining key research in system representation, transportation efficiency, and performance evaluation methodologies. Section 3 introduces the analytical framework, detailing the Data Envelopment Analysis (DEA) models employed in this study. Section 4 describes the case study, focusing on the INFRABEL railway network, and discusses the findings related to the technical efficiency of train paths. Finally, Section 5 concludes the paper with key insights, practical implications, and recommendations for future research.

2. Literature Review

An integrative literature review (Torraco, 2020) is conducted in this research to synthesize existing studies across multiple domains, including systems engineering, transportation operations, DEA, safety science, and optimization. The review critically examines and combines insights from this diverse body of work to highlight conceptual and methodological gaps, such as the lack of multi-perspective optimization models for safety-critical socio-technical systems, with the aim of informing and guiding future research. Major scientific publisher databases were consulted, including ScienceDirect, Springer, Taylor & Francis, MDPI, IEEE, INFORMS, the Transportation Research Board (TRB), and the American Society of Civil Engineers (ASCE). Additional relevant publications were identified through reference tracing. In total, 60 peer-reviewed papers were reviewed and included for this study.

The keyword combinations used for the literature search included the terms “transportation performance measurement”, “railway systems efficiency,” “train route”, “train path”, “railway systems optimization,” “socio-technical systems,” “data envelopment analysis (DEA),” “punctuality in transportation,” and “system representation.”

The following sub-sections focus on the literature on functional system representations, efficiency performance measurement in service sector, specifically on the railway systems through metrics and KPIs, optimization models, and more narrowly DEA. Next, some main publications on the safety science literature are explained, followed by the literature gaps and finally, the research question of this work.

2.1 Functional System Representation

Functional system representation is a key aspect of systems engineering that focuses on defining the performance of a system. In the context of efficiency analysis, this approach involves detailing the inputs, outputs, outcomes, and environmental variables that influence system operations. It also requires specifying the transformation function, which describes how inputs are converted into outputs (Helms et al., 2010). This detailed mapping allows for the creation of models designed to compute the efficiency performance of the system (Summers et al., 2017).

2.2 Efficiency Performance Measurement

Efficiency performance in the service sector, specifically transportation systems, can be conceptually and computationally linked to critical operational metrics such as punctuality, cost performance, and resilience. This connection underscores the importance of efficiency in enhancing the overall effectiveness of operations. By applying analytical techniques that focus on these aspects, organizations can directly see how improvements in efficiency lead to better punctuality, reduced operational costs, and greater system resilience. This holistic approach ensures that efficiency improvements are not only theoretical but have tangible impacts on the system's operational success. Nevertheless, the initial focus of this work is on the notion of efficiency and its connection to punctuality, leaving the consideration of other performance dimensions such as cost and resilience for future research.

2.2.1 **Efficiency of Railway Systems**

The transportation literature consists of a considerable amount of research on performance improvement and efficiency analysis that robustly addresses the measurement of transportation systems technical performance, with a notable focus on railway systems (Holvad, 2020), (Victorino and Pena, 2023), (Merkert et al., 2009), (Sameni et al., 2016), (Wang et al., 2022), (Azadeh et al., 2008), (Fadaei Foroutan & Bamdad, 2023). Wang et al. (2019) use a structural and topological analysis with the trains flow in the network to identify key stations in railway networks to enhance overall efficiency of the railway system.

2.2.2. Metrics

The performance of railway systems is frequently assessed using a broad set of metrics and Key Performance Indicators (KPIs), which may include metrics such as service frequency, punctuality, on-time performance, customer wait times, train speeds, employee productivity, and safety-related factors like the number of accidents. Recent studies demonstrate diverse methodological approaches to performance evaluation, often tailored to specific objectives or regional characteristics (Bešinović et al., 2019; 2020; Chen and Wang, 2018; Litherland et al., 2021; Stoilova et al. 2020).

Bešinović et al. (2019) focus on resilience assessment. In their study, they construct KPIs related to passenger inconvenience, train delays, and network-wide disruptions. Their case study in the Netherlands illustrates how performance measurement can inform strategies for managing and mitigating service interruptions. Complementing this, Bešinović (2020) further emphasizes the importance of evaluating resilience in passenger rail systems by integrating operational, infrastructural, and demand-side variables. This analysis highlights the complicated and multi-dimensional nature of railway system performance.

Railway system performance is also measured across countries. Stoilova et al. (2020) adopt a multi-criteria decision-making and clustering approach to classify railway network performance across Eastern European countries. Their framework incorporates macro-level indicators such as railway line length, GDP per capita, and freight usage intensity, revealing how structural and economic factors influence system-wide effectiveness. This approach underscores the need to view railway systems as embedded components of broader economic and logistics networks (Stoilova et al. 2020).

Bešinović (2020) discusses resilience in railway transport systems, highlighting the importance of performance evaluation in passenger railway networks, particularly concerning train service adaptations and passenger comfort. To accurately evaluate the resilience of such systems, methodological approaches should consider the specific characteristics of railways, including operational aspects (train routes, stopping patterns, and timetables), transport demand (passenger and/or freight), and the infrastructure network topology (Bešinović, 2020).

A more technical perspective is offered by Litherland et al. (2021). This study utilizes a RAMS (Reliability, Availability, Maintainability, and Safety) framework, enriched through data analytics, to evaluate asset performance in the UK railway network. Their work reflects a shift toward more granular performance assessment, which can provide tailored insights for managing railway systems.

Additionally, environmental variables also play a critical role in performance outcomes. Chen and Wang (2018) investigated the impact of severe weather events on high-speed rail and aviation systems in China, using a dataset of 350,000 detailed on-time performance (OTP) records, using regression analysis. The study employs OTP as the primary metric for evaluating punctuality, defined as the proportion of services that depart or arrive within a specified time frame, less than 30 minutes late for aviation and less than 5 minutes for high-speed rail systems. In addition to OTP, the study considers other metrics such as the number of delayed and canceled services and the average delay time, offering a comprehensive view of how weather conditions affect the punctuality and reliability of both transport modes. It demonstrates how external disruptions translate into measurable declines in on-time performance, service cancellations, and delays, reinforcing the need to integrate external risk factors into performance modeling. (Chen and Wang, 2018).

Together, these studies illustrate a growing recognition that railway performance evaluation must account for a variety of dimensions: operational, infrastructural, economic, and environmental. While traditional KPIs remain essential, recent research moves toward system-based, context-aware, and resilience-oriented approaches. This ongoing movement paves the way for more integrated and dynamic models of railway system performance.

2.3. Optimization Models

Several research publications have applied operations research methods in transportation predominantly target scheduling, routing optimization, and capacity management and the utilization of these systems, especially for railroads using different types of optimization models (Lai et al. 2010; Xu et al., 2022; Zhou et al., 2020; Carey & Carville, 2003; Dang et al., 2010; Li et al., 2008; Li et al., 2013; D'Ariano et al., 2008; Garisi & Cervelló-Pastor, 2020; Lu et al., 2022; Xu et al., 2017).

The Dutch Railways Decision Support System (DSS) offers an example of train routing and capacity allocation analysis. The DSS used mathematical models to optimize train routing (assignment of trains to inbound and outbound routes) and scheduling (optimizing arrival and departure times for trains) at the station level (Kroon et al., 1997).

Various modeling and algorithmic approaches have been utilized to evaluate railway performance measurement and efficiency. For instance, using a matching algorithm, a study by Levine et al. (2013) introduces a performance measurement framework applied to New York City Transit to measure subway service efficiency through wait assessments. It compares scheduled versus actual headways as one of the variables that captures customers' perspective to support better operations decisions for management (Levine et al., 2013).

Some research, for example (Mo et al., 2021), apply the Simulation-Based Optimization framework to evaluate the efficiency of railway systems. While simulation modeling is useful for capturing system complexity under varying demand conditions, it requires iterative testing and relies on probabilistic assumptions, making it computationally intensive and time-consuming. In contrast, optimization-based algorithmic approaches, such as Data Envelopment Analysis (DEA) introduced by Charnes et. al. (1978) are different, as they generate results based on observed data rather than on probabilistic scenarios.

2.4 Data Envelopment Analysis (DEA)

DEA is a widely recognized non-parametric methodology used to assess efficiency when the production function is unknown (Bray et al., 2015). It applies linear programming techniques to evaluate the relative efficiency of DMUs, accommodating multi-input and multi-output relationships (Bray et al., 2015). Unlike traditional regression-based efficiency evaluations, which rely on average values and central tendencies, DEA determines efficiency based on deviation from the production frontier, allowing for a more nuanced analysis of operational performance (Cooper et al., 2006).

One of the key strengths of DEA is that it does not require assumptions about the shape of the production frontier or the internal structure of the DMUs being analyzed (Bray et al., 2015). This flexibility makes DEA particularly useful for evaluating complex transportation networks where multiple factors contribute to efficiency. Additionally, it facilitates collaboration between analysts and decision-makers by allowing them to define relevant inputs and outputs for measurement (Cooper et al., 2006). Another advantage of DEA is that it can handle multi-dimensional efficiency evaluations, enabling a more comprehensive assessment of railway performance without limiting the selection of input-output variables (Cooper et al., 2006). However, despite its strengths, DEA is highly sensitive to data quality. Small variations in input or output measurements can significantly impact efficiency frontiers, making accurate data collection essential for meaningful analysis (Bray et al., 2015).

DEA has been extensively applied in the transportation sector, particularly in benchmarking performance across different modes of transport. In the aviation industry, DEA has been used to analyze airport efficiency, with studies assessing on-time gate arrivals as a performance criterion (Bray et al., 2015; Diana, 2006). Similarly, seaports have utilized DEA to evaluate cargo-handling efficiency and resource utilization (Bray et al., 2015). The methodology has also gained traction in public transit systems, including U.S. bus transit agencies, where it has been used to compare operational performance between transit providers (Chu et al., 1992).

Performance efficiency evaluation of a bus service is done by a study in 2006 (Sheth et al, 2006) through a Network DEA model. The study defines bus routes as the decision-making units that can create the whole service and hence provide all the stakeholders' perspectives (Sheth et al, 2006).

In the railway sector, DEA has been instrumental in measuring efficiency at train stations. A study in Great Britain used DEA to examine the technical efficiency of 96 major train stations, assessing their ability to handle train stops relative to their existing infrastructure (Khadem et al., 2016). The study further extended DEA analysis to evaluate service effectiveness, measuring how well train stops were converted into passenger flows by considering external factors such as catchment area population and job opportunities (Khadem et al., 2016). These applications demonstrate DEA's versatility in assessing railway performance from multiple perspectives, including infrastructure utilization and service delivery.

2.5. Safety Science

Another related field is safety science. Each organization, such as a transportation system as a safety-critical system (Bozzano & Villafiorita, 2010), spends its resources on different goals, including operational and safety concerns. This is also critical in railway systems because any error can cost people's lives. Sometimes, organizations skip focusing on safety matters and spend more resources on operational tasks. This could happen gradually over time, with no deliberate intention. Hence, these safety-critical STSs (organizations) need to keep a very good track of the tradeoff between safety and efficiency (punctuality).

This leads to production pressure. Hashemian and Triantis explain production pressure and its relation to safety by reviewing 180 publications in the literature (Hashemian and Triantis, 2023). Any changes in the timetable could increase the chance of near misses and errors in the system, specifically for the operators in traffic control centers, as it can increase their workload pressure (Rasmussen, 1997). Hence, when concentrating on any changes in the service schedule, safety matters should be reviewed carefully. All might lead to organizations' loss in the long run. These considerations link to Rasmussen's theory on balancing the workload, safety, and economic boundaries (Rasmussen, 1997).

Rasmussen's framework delineates three main boundaries within an organization viewed as a STS: economic, safety, and workload. The workload boundary defines the system's work capacity, the safety boundary represents the threshold of safety risks beyond which functional failure is likely, and the economic boundary outlines the optimal economic performance necessary to sustain viability. These boundaries are interdependent, meaning the system remains stable as long as none are exceeded. However, given the limited resources available to any STS, organizations may prioritize one boundary over others, potentially destabilizing the system if any boundary is crossed, leading to challenges (Rasmussen, 1997).

2.6 Gaps in the Literature

Despite these contributions, there is a distinct gap in the application of optimization methods for STSs when considering the tradeoffs among workload, safety, and economics. Specifically, the literature does not address the way of implementing such methods, particularly when considering various perspectives for measuring and improving the technical performance of transportation systems, such as railway systems, especially with a concentration on the timetable.

A critical dilemma arises when systems require updates or modifications. For example, which components of the transportation service schedule/timetable (such as paths or itineraries) should be prioritized and changed to meet emerging needs? The decision-making process for applying any modifications becomes even more complex when it incorporates the diverse perspectives of passengers and service providers within an integrated measurement framework. Thus, there is a pressing need for research that bridges this gap by providing a clear starting point for taking the multi-perspective environment into account, ensuring a better service quality (Medina-Borja et al., 2007; Cabrera et al., 2008; Herrera-Restrepo et al., 2016).

This research aims to assist decision-makers of railway transportation systems at the organizational level to address performance efficiency issues across the network. The study focuses specifically on challenges related to punctuality and how it is impacted by the scheduled timetable and the actual performance based on it. As railway systems are complex safety-critical STSs, system thinking approaches seem beneficial for having a better understanding of the system, by providing different points of view and the variables that impact system performance. Hence, measuring system performance to help technical efficiency realization, improvements, and decisions for actions (Cabrera et al., 2008).

Efficiency in railway systems is typically measured as an input-output ratio. This study focuses on logistical efficiency and primarily considers how train paths affect performance. Analyzing train paths could provide valuable insights into network bottlenecks and station congestion, which are critical elements in railway scheduling and capacity planning.

For this research, technical efficiency is defined to provide insights into which train paths or itineraries require modifications. This analysis considers the allocation of resources to each path and the results achieved (Mahmoudi et al., 2020; Wang et al., 2022).

DEA is recognized as a robust framework for modeling efficiency, particularly suited to environments like railway systems where multiple inputs and outputs need evaluation without a predefined functional relationship. In assessing its applicability to safety-critical transportation STSs, it is crucial to consider the production function theory axioms. These axioms generally hold under the assumption of homogeneity among the DMUs, in this case, train paths within a single railway system. However, the axiom of non-increasing returns to scale is particularly relevant, suggesting that beyond a certain point, adding more resources (e.g., trains) to a fixed infrastructure (e.g., tracks) may yield diminishing returns, crucial for planning under capacity constraints.

Technical efficiency in this research refers to the effectiveness with which a train path or itinerary utilizes its allocated resources to maximize its operational objectives, primarily punctuality and safety, under the constraints of railway infrastructure and traffic. It involves minimizing the use of resources such as energy, staff time, and capital while maximizing outputs such as on-time performance, passenger satisfaction, and safety metrics. This definition will guide the application of DEA by setting the framework for what constitutes efficient and inefficient practices within the railway system. It helps quantify the extent to which various segments of the railway network adhere to or deviate from optimal performance levels, providing a basis for comparative analysis across different train paths or operational settings.

2.7 Evolving Research Question

Based on the literature review, the evolving research question of this research can be stated as follows: How to improve the technical efficiency of a railway system in terms of punctuality by analyzing the technical efficiency performance of train paths/itineraries? This research question is being explored using INFRABEL as a case study.

3. Methodology/Analytical Framework

A systematic literature review will be used to identify existing gaps and explore potential solutions. To ensure rigor and reproducibility in this review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines are followed. This method involves a structured process for identifying, evaluating, and synthesizing all relevant research available on a particular topic, conducted across multiple databases to capture a wide range of publications (Page et al., 2021).

In addressing practical problems, particularly those involving real-world organizations, ethnographic research is invaluable for its holistic approach to studying an entity in its entirety. Ethnographic research plays a significant role in enhancing technical efficiency within organizations by providing deep insights into the operational dynamics and cultural contexts that influence performance. It can further enrich this understanding by examining how these factors interact with the daily routines of train operators and passengers. Jensen et al. discuss the personal dimensions of ethnographic work and how these experiences can lead to a deeper understanding of organizational communication and processes (Jensen et al., 2019). By observing and engaging with these stakeholders, researchers can identify practical solutions that are grounded in the realities of everyday operations, leading to improved punctuality outcomes. Additionally, database analysis is commonly employed in applied research to manage and analyze available data (Figure 1).

Figure 1 - Research Methodology

Based on these three methodologies, this research develops a conceptual systems-thinking model and tests it with a real-world case study. It also involves formulating a Data Envelopment Analysis (DEA) model as a linear programming optimization model. Decision-making units are defined along with appropriate variables that address the research needs and adequately respond to the posed research question. For this stage of the research, only technical efficiency and hence, the operational aspects from the supply side are taken into account. Other aspects, including the demand side and the safety factors, could be addressed in later stages of the research, as a future recommendation.

3.1. Overview of DEA Model Specification

The train paths, representing organized sequences of train services across specific segments of the network, are considered as the DMUs in this analysis.

The advantage of using DEA is that it would account for all of the important factors concurrently, and it provides complementary information to the decision makers considering the other performance measures that are unidimensional (i.e., delays; punctuality). With a suitable DEA formulation (based on the choice of the variables, the model orientation and return to scale), it is possible to add the impact/effects of the contextual variables in the analysis. By successfully pursuing this line of research, this could potentially lead to a practical management tool for any railway system management.

Given the infrastructure constraints and varying service intensities across paths, an output-oriented DEA model with Variable Returns to Scale (VRS), also known as the BCC model (Banker et al., 1984), is applied. This orientation is appropriate for railway systems, where the inputs (e.g., infrastructure use, number of stops) are often fixed in the short term, and the objective is to maximize outputs, particularly punctuality.

The BCC model is mathematically formulated as follows:

$$\begin{aligned} \max \quad & \beta \\ s.t. \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \leq x_{io} \qquad (i, \dots, m) \\ & \sum_{j=1}^{n} y_{rj}\lambda_j \geq \beta y_{ro} \qquad (r, \dots, s) \\ & \lambda_j \geq 0 \qquad (j, \dots, n) \end{aligned}$$

Here, x_{ij} and y_{rj} are the inputs and outputs for DMU j, and β represents the radial expansion factor for outputs. A DMU is efficient if β=1.

3.2. Addressing Undesirable Outputs in DEA

In the context of train path performance, undesirable outputs (e.g., delays or number of cancelled trains) pose a modeling challenge, as traditional DEA assumes that outputs are desirable (this means that the more output is the better). There are different ways to deal with undesirable variables in DEA models and settings. This study considers two selected approaches to address this issue: the Negative Output Transformation under BCC (BCC-n) (Seiford and Zhu, 2002) and the Directional Distance Function (DDF) (Chambers et al., 1998).

3.2.1. Negative Output Transformation (BCC-n)

Seiford and Zhu (2002) propose a method where undesirable outputs are transformed by multiplying them by -1 and then adding a constant to ensure all values remain positive. This transformation allows the standard BCC model to handle undesirable outputs while maintaining linearity and convexity. The authors demonstrate that this approach preserves classification invariance, meaning the efficiency classification of decision-making units (DMUs) remains consistent under the transformation (Seiford and Zhu, 2002).

The BCC-negative (BCC-n) approach is a radial model, as it applies a proportional adjustment to all outputs, including undesirable ones that are transformed through monotonic decreasing functions. While this allows for the inclusion of undesirable outputs, it assumes uniform (proportional) expansion across all output dimensions, which may limit flexibility in addressing inefficiencies specific to individual outputs (Seiford and Zhu, 2002). As undesirable outputs need to be minimized, it aligns well with this research on evaluating train path efficiency using DEA.

3.2.2. Directional Distance Function (DDF)

The Directional Distance Function (DDF), introduced by Chambers, Chung, and Färe (1996), provides a more nuanced alternative by allowing simultaneous expansion of desirable outputs and reduction of undesirable ones along a specified direction vector. In this model, a DMU's inefficiency is measured by the extent to which it can move in the defined direction toward the production frontier without violating feasibility (Chambers et al., 1996). The directional distance function provides a flexible, additive framework for measuring inefficiency, allowing non-proportional adjustments in inputs and outputs, in contrast to traditional radial models that assume uniform scaling across all dimensions (Chambers et al., 1998).

For an output-oriented DDF model, let g = (g^y, g^b) where g^y points toward increasing desirable outputs and g^b points toward decreasing undesirable outputs. The DDF model can be expressed as:

$$\sup_{\beta} \{y + \beta g^y, b - \beta g^b \in P(x)\}$$

where P(x) denotes the production possibility set for input x, and β reflects the maximum possible directional improvement. When the value of β or the objective function is 0, it means that the DMU is on the frontier of production and cannot improve further. Thus, the maximum value of the objective function shows the most inefficient DMU.

$$Max\ \beta$$

$$\sum_{j=1}^{n} \lambda_j y_j \le y_i + \beta g_y$$
$$\sum_{j=1}^{n} \lambda_j b_j \le b_i - \beta g_b$$
$$\lambda_j \ge 0\ \forall j$$

The advantage of DDF lies in its ability to explicitly model multi-dimensional efficiency adjustments—increasing the desirable output variables while reducing undesirable (for example, delay-related) variables' inefficiencies—without distorting output values through transformation.

3.3. Justification of Model Choice

The BCC model is suitable for this context because transportation systems typically operate under variable returns to scale (Sheth et al, 2006), especially when infrastructure and train schedules are managed with limited flexibility. Furthermore, DDF is incorporated to accurately handle undesirable outputs, such as delays, which cannot be simply excluded or naively transformed without affecting interpretation.

Together, these models provide a robust framework for analyzing the technical efficiency of train paths under both standard and complex conditions (e.g., disruptions, peak-hour congestion), supporting targeted improvements in scheduling and operations (Table1).

Table 1 - Different DEA models

DEA Model Type	Optimization Type	Objective	Core Method
CCR / BCC	Linear Programming (LP)	Maximize output / minimize input	Linear programming, Simplex, or interior-point solvers
BCC-negative (BCC-n)	Linear Programming (LP) with transformed	Same as BCC (but with transformed	LP

	data	undesirable outputs)	
DDF	Linear Programming (LP)	Maximize directional improvement (β\betaβ)	LP with directional constraints

3.4. Suggested Variables

3.4.1 Inputs, Outputs, and Environmental Variables

For identifying the variables for the performance evaluation analysis, this research starts with the resources available to a railway system organization, as well as the responsibilities of the company for providing the service as the outcome towards the contracting partners and the residents as final customers.

Resources available to the system with the classifications (C: capital, L: labor, E: energy) are including as follows: Budget (C), Network infrastructure including tracks, platforms, stations, signals, etc. (C), Traffic control systems (The centers and the software) (C), Staff (L), Catenary system (E, C) and maintenance (equipment and knowledge) (C).

Responsibilities of the system include: providing the infrastructure for trains to run in the network and control the traffic, providing proper service level in terms of the number of stations and municipalities and regions under the network coverage, punctuality, minimum cancelation, safety (no accidents), and maintenance while not exceeding the budget.

With the limited time and capacity available to the railway system company (based on the expectations from the rolling stock companies, the government, the passengers, etc.), it is possible to map the resources such as infrastructure (network tracks, stations and platforms) into variables such as number of trains running, the duration of the journey, the occupancy of the tracks and the frequency of the service (or the headway which is the time between two consecutive service). On the other hand, by focusing on the production function of the system, the responsibilities of the railway system can be mapped into the number of passengers served, the number of trains operating in the network, the number of incidents, the minutes of delay, the number of cancelled trains, etc.

Considering the train path or itinerary as a decision-making unit (DMU) and using the methods explained in the beginning of this chapter, several variables associated with the service performance were identified. Having the isotonicity assumption[1] on mind (Dyson et al, 2001), these variables are classified into input and output, conceptually. Moreover, there are some environmental variables with considerable impacts on the performance of the train paths, but since they are out of any of the stakeholders' control, they are identified as contextual variables. Table 2 shows the suggested variables with the classifications at this stage of the project. (Undesirable variables are identified with the (-) symbol.). Implementation of these variables is subject to the availability of sufficient data to support the analysis.

[1] isotonicity assumption requires that an increase in any input should not lead to a decrease in any output, under the assumption that other inputs are held constant. In simpler terms, the relationship between inputs and outputs should be non-decreasing. This means that more resources (inputs) should not result in less output.

Table 2 - Possible variables associated with the problem, with their features

Type (input/output/ outcome/ contextual)	***Variable***	***Units of measurement***	***Controllable***
input	Headway	Minute	Yes
input	Duration	Minute	Yes
Input / environmental	Line/track occupancy (score)	Numerical score	No
Input / environmental	Track capacity availability (single/double)	Numerical score	No
output	Dwell time/waiting time (-)	Minute	Yes/No
output	#stops	Count	Yes
output	# train runs/service	Count	Yes
output	# cancelled trains (-) / non-cancelled trains	Count	Yes
output	Delay (-)	Minute	Yes
output	Punctuality metric score	Numerical score	Yes
output	# incidents/accidents (-)	Count	Yes
output	# passengers served	Person	Yes
Input/output / environmental	Demand	Person	Yes/No
environmental	Population density factor (score)	Numerical score	No
environmental	Accessibility factor (score)	Numerical score	No
environmental	Connectivity factor (score)	Numerical score	No

Other than the data availability to support each one of the variables, another factor that is incorporated in considering any of them into the final DEA model is the representation of the combination of the variables for answering the research question, with the focus on the production technology and the transformation function. This needs to be narrowed down after more investigation into the company's performance goals and objectives.

3.5. Model Conceptualization

To have a clear understanding of the system, using system thinking literature, the next sub-section explains the representation of the system.

3.5.1. System and System Thinking Representations

Considering the complexity of these STSs, it is not straightforward how to track the changes within the system. Ropohl's framework for understanding systems through functional, structural, and hierarchical representations could be practically relevant for the operational analysis in safety-critical systems, such as railway systems (Ropohl, 1999).

The functional representation with special attention to the resources, inputs and outputs and outcomes within a black box approach (Cabrera et al, 2008) focuses on the intended operations and behaviors of the system, ensuring that safety measures align with operational goals. This is critical in safety-critical systems where failures can lead to catastrophic outcomes. Focusing on the input/output representation, which can simplify the complexity of the underlying processes (Figure 2), the functional representation of the system does not need to provide the exact mathematical operational definition of the transformation process (Helms et al., 2010). By concentrating on functional analysis, it is possible to build a comprehensive understanding of a system's operation, which is crucial for any attempt to optimize efficiency performance. This approach is used in this research to identify the variables for evaluating the system performance.

Figure 2 - Functional System Representation

3.5.2. A Graphical System Representation

Considering an infrastructure provider railway system (i.e. INFRABEL: the Belgian Railways infrastructure organization company as the case study), a graphical representation of the system is presented in Figure 3. There is enough effort spent on this graphical representation to capture the main elements of a railway system based on the operations and performance, in this problem score. Therefore, this can be applied to any railway system with some modifications, based on the organizational chart and managerial methods. The key players in this representation are the railway infrastructure system, the train company or the rolling stock company system, and the government. The sub-systems working within the bigger systems are the trains traffic controlling centers, where running the traffic of the train services, the passengers (paX), and the government. Modifications for applying the model representation to other train companies systems could be done specifically on the system boundaries.

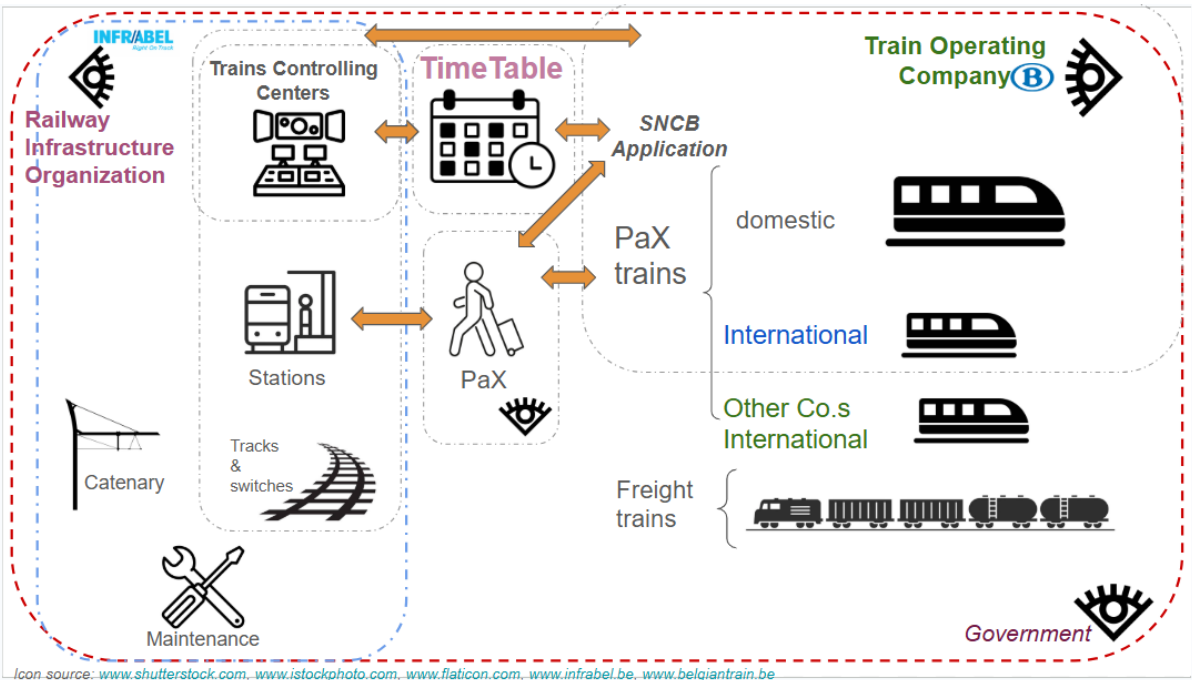


Figure 3 – Graphical representation of the railways system

3.5.3. A Systems Thinking Approach: The DSRP Framework

Understanding the operations of train paths in the INFRABEL railway system can be analyzed through the DSRP (Distinguish, Systems, Relationships and Perspectives) framework (Cabrera et al, 2008), which helps define key elements, their interactions, and the perspectives involved in railway operations. Distinctions within this system begin with train paths themselves—how trains move from Point A to Point B. Additionally, distinctions exist between internal and external stakeholders, such as INFRABEL'S traffic control teams, train operators like SNCB, and European regulatory agencies.

The railway network is composed of multiple systems and subsystems working together to maintain efficiency (e.g., punctuality). The train paths are tied to specific routing mechanisms involving track switches, station stops, and pre-defined schedules/timetables that dictate the start and end times of each journey. The infrastructure supporting train movement includes tracks, stations, and switching mechanisms. The controlling center is another system that includes real-time monitoring technology, safety controllers, signaling systems, and traffic management teams who oversee train dispatching and routing. Together, these subsystems play a critical role in ensuring train movements are coordinated to prevent bottlenecks and enhance punctuality. External entities, such as international rail companies, passengers, and freight operators, interact with INFRABLE's system, contributing to overall network demand and efficiency.

Key relationships exist between these systems, influencing how train paths are managed. Train paths directly impact train scheduling, as modifications to routing or prioritization affect overall timetable efficiency. Traffic controllers, safety controllers, and train drivers work together to maintain smooth operations, ensuring both punctuality and safety in the face of unpredictable disruptions like weather conditions or technical failures. The Belgian government and INFRABEL's financial strategy influence investments in rail infrastructure and technology, affecting the long-term sustainability of the system. Additionally, demand for train services fluctuates based on passenger volume, requiring dynamic scheduling to balance efficiency with customer expectations. External factors like weather conditions can disrupt train paths, leading to delays and necessitating contingency measures within the system (Chen and Wang, 2019).

Various perspectives influence train path management. INFRABEL managers focus on optimizing operations while ensuring cost-effective infrastructure usage. Their controlling centers' staff, responsible for real-time decision-making, deal with train dispatching, signaling, and network disruptions. The Belgian government plays a financial role in railway development, overseeing subsidies and long-term investment plans. Train companies, primarily SNCB, rely on INFRABEL's network to run passenger services efficiently while aligning with international railway operators. Passengers prioritize reliability, punctuality, and convenience, expecting well-managed schedules and minimal delays. Although maintenance teams are crucial for system upkeep, their engagement in daily scheduling decisions may be limited. Finally, European regulatory groups enforce safety and operational standards, ensuring Infrabel complies with broader transportation policies (INFRABEL, n.d.).

The primary goal of this research is to enhance the efficiency of train operations across the railway system managed by decision makers (infrastructure provider with the collaboration of train companies), focusing on improving punctuality and reducing delays without significant infrastructure changes, while keeping safety as a key priority (Hashemian and Triantis, 2023). This research contributes to this goal by assessing train timetables and scheduling, as an integral sub-system of railway systems, employing systems thinking and Data Envelopment Analysis (DEA) (Charnes et al, 1978) methodologies to better capture the dynamic and complex nature of the sub-system.

Some effective environmental, contextual, and managerial factors can have a big impact on the system, which are not controllable by the stakeholders and decision makers within the system. Some of the examples of these factors are as follows:

- Environmental Factors: Weather conditions that impact train runs punctuality and safety; geographical challenges affecting train routes and operations; animals and sometimes people passing through the tracks and tracks area causing safety issues.
- Managerial Factors: Decision-making processes at INFRABEL; strategies for resource allocation; investments in technology; and policy management (evaluation of the demand, ticket pricing, future expansions, etc.) by both INFRABEL and SNCB, considering the observance of the government.
- Contextual Factors: Regulatory requirements from European and national bodies; economic constraints impacting budget allocations; public expectations for reliability and punctuality; and the evolving landscape of public transportation demand influenced by urban development and population growth. Furthermore, density of the population varies in different areas, which has an impact on train service performance in terms of punctuality and safety. Besides, the accessibility of the train services varies in different areas.

This paper tries to get deeper into the topic by these contributions: First, by defining the train paths as the DMU. This choice facilitates a more targeted and granular means for efficiency assessments in railways. Furthermore, this conceptual shift deepens the understanding of how transportation networks function. Second, methodologically, this work contributes by linking systems thinking and performance evaluation methods. This study captures local inefficiencies and links these to their broader system-level implications. Third, empirically/practically, this research contributes by providing insights for railway system managers and policymakers to identify bottlenecks and underperforming routes. The practical application of this train path-based evaluation has the potential to enhance scheduling and improve capacity utilization, etc.; all of which contribute quality of the railway services. Fourth, the dataset created for this research is unique, which shows the operational realities that can test specific operational interventions both in this research and future research.

3.5.4 Defining the Decision-Making Unit (DMU): Train Paths

In the context of this study, train paths—defined as a combination of “Relation” (origin-destination) and “direction” (e.g., inbound or outbound)—have been selected as the Decision-Making Units (DMUs) in our Data Envelopment Analysis (DEA) model. A DMU, in DEA terminology, represents an entity that uses inputs to produce outputs. For INFRABEL, each train path is treated as a unique operational unit within the rail network. This approach enables a granular, performance-based evaluation of each segment of the network, providing insight into how a railway system transforms allocated infrastructure and time resources into productive rail service outcomes.

3.5.4.1 Why Train Paths as DMUs?

Train paths are a foundational unit of railway operation—they represent the scheduled capacity of the network. Each path is associated with a defined slot in time and space across the rail infrastructure. Their efficient management directly affects operational performance, customer satisfaction, and network throughput. Bottlenecks in station areas, delays, or inconsistent service patterns can often be traced back to inefficient use of train paths (Sameni et al., 2016).

Moreover, train paths are one of the primary units used for capacity planning and performance regulation, linking infrastructure availability (tracks, stations, switches) to strategic objectives like punctuality, safety, and network utilization (Harrod, 2012). From a systems thinking perspective, train paths also encapsulate the interactions between subsystems—e.g., signaling, traffic control, timetables—and the actors involved, including SNCB, INFRABEL control centers, and passengers.

The choice to model train paths as DMUs arises from their strategic relevance to railway operations. Unlike train numbers, which are inconsistent and subject to change depending on timetabling and day-to-day operations, train paths defined by their relation and direction offer a stable, repeatable unit of analysis. This approach helps to assess path-level performance, independent of fluctuating service IDs or temporary train identifiers. However, future research could use individual train numbers to explore service-level efficiency or delay propagation at a micro level.

Importantly, defining the DMU at the train path level allows the model to assess the operational efficiency of specific corridors. It answers questions such as: *Is this path being under- or over-utilized relative to its capacity? Are we moving enough trains within reasonable time and headway constraints?* These insights directly support railways organizations’ goals to optimize rail capacity and reduce congestion.

3.5.6. DEA Variables: Selected Inputs and Outputs

DEA models require input and output variables that reflect the resource use (inputs) and service production (outputs) of each DMU. For this research, the following variables were selected based on the literature review, their relevance to operational efficiency and data availability from INFRABEL:

Inputs:

- Duration (minutes): Represents the time a train takes to complete its path. This is a proxy for infrastructure use and scheduling complexity. Longer durations often signal congestion, inefficiencies in routing, or slower speeds.
- Headway (minutes): The temporal spacing between consecutive trains on the same path. High headways suggest underutilization or overly conservative scheduling, whereas lower headways indicate better scheduling and infrastructure utilization.

Outputs:

- Number of Trains: This captures the throughput of the train path, or how many trains successfully used the route in the evaluation period. Important for evaluating train path productivity. In the context of INFRABEL's operational goals and infrastructure responsibilities, this variable provides direct insight into how effectively a train path is being utilized. From a DEA perspective, higher output values imply greater efficiency if they are achieved without proportionally increasing input use (e.g., time or spacing). A path that can accommodate more trains within the same infrastructure and scheduling constraints is considered more efficient.
- # stops: Adds important context to the complexity of a train path. Paths with more stops typically have more scheduling constraints, increased dwell times, and more passenger interactions. Including this as an output allows the DEA model to account for differences in **s**ervice intensity.
- Delay (minutes): Delay helps assess whether throughput is delivered on time. Provides a direct link to **s**ervice quality and reliability. INFRABEL's mission includes improving punctuality, and delay data allows the DEA model to capture inefficiencies arising from congestion, routing conflicts, or infrastructure bottlenecks.

- Number of non-cancelled Trains: This variable includes the number of trains that took the journey completely. Sometimes, some of the trains are operated partially in the network or are completely cancelled. This is an important variable as sometimes in case of disruptions in the system (any incidents or a large amount of delay), operators in the TCCs decide to cancel the train service or run to maintain the traffic flow smoothly. Thus, there could be a trade-off between the number of canceled trains and the amount of delay. The number of cancelled trains is an undesirable variable. However, it could be substituted with the desirable variable "number of non-canceled trains".

4. Case Study and Discussion

4.1. Case Study

To evaluate the proposed method, a case study on the Belgian railway system is conducted to implement the research approach (Figure 3) using real-world operational data.

4.1.1. Context and Organizational Background

INFRABEL (Infrastructure Belgium) is a governmental organization charged with managing and operating Belgium's railway infrastructure. Established in 2005 following its separation from the National Railway Company of Belgium (SNCB/NMBS), INFRABEL builds, owns, maintains, and upgrades the Belgian railway network. It also allocates network capacity to railway operators and oversees train traffic control (INFRABEL, n.d.).

Punctuality is a paramount objective for INFRABEL, particularly concerning passenger trains. The organization adheres to a specific punctuality metric defined by its legal agreements with the government (INFRABEL, n.d.). Recent reports, including a report by The Brussels Times, indicate that train punctuality in Belgium has declined to its lowest levels since records began (The Brussels Times, 2023). This significant drop in performance highlights a critical need for INFRABEL to enhance its punctuality measures, specifically those that the company has more influence on, as some of these factors are environmental factors and require special attention by applying appropriate methods (Topcu et al., 2019).

For this purpose, INFRABEL meticulously records all incidents of delays, capturing both the occurrences and their causes, as part of its commitment to transparency and improvement (INFRABEL, n.d.). The organization implements various initiatives aimed at enhancing punctuality from multiple angles—from systemic changes to passenger awareness. One such initiative is the "Not Spot" project, which educates the public about potential obstructions on tracks and platforms caused by locals (INFRABEL, n.d.).

Additionally, INFRABEL and SNCB focus on the annual train service timetable, adjusting it to achieve punctuality targets. These timetables are developed collaboratively by the two organizations, taking into account requests from the government and politicians as well as performance data from previous years concerning punctuality and delays. Interviews with INFRABLE's experts reveal a particular emphasis on modifying train paths or itineraries to enhance service efficiency.

INFRABEL represents a suitable case due to its centralized infrastructure control, public accountability for punctuality, and extensive historical data collection, making it ideal for performance analysis using DEA when evaluating the relative performance of train paths. The goal of this case study is to apply the developed DEA framework to assess the technical efficiency of train paths within INFRABEL's network and explore the operational factors affecting punctuality.

Figure 3 in section 3 (the DSRP representation of the Belgian railway system), shows the interdependencies between train path configurations, control center operations, and stakeholder influences.

For this study, comprehensive ethnographic activities have been conducted, including interviews with INFRABEL employees and an on-site visit, alongside reviews of documentation and the company's website. An integrative literature review covering transportation efficiency performance measurement, system engineering, and safety science has also been completed (section 2). Thanks to a non-disclosure agreement between INFRABEL and VT, there has been access to INFRABEL's extensive database, including both public and restricted data shared between Virginia Tech and INFRABEL.

The author has arrived at a preliminary definition of the input and output variables that are provided in the previous sections (more details can be found in Appendix). Based on the defined variables and investigating the current datasets (both on the VT ISE server[2] and publicly available[3]), some different DEA models are formed, and results are achieved. Some insights from the preliminary results are captured. The study can be furthered by consulting the results with academics working in the field, as well as practitioners from any railway systems organizations, specifically INFRABEL, and requesting more datasets. Although these

---

[2] Timetable data of September 2023

[3] INFRABEL OpenData: https://opendata.infrabel.be/pages/home/

activities are iterative and ongoing, an initial understanding and result with insights has been achieved to understand the system better and define and address the research problem.

4.1.2. Data

In this research, a train path or itinerary (as defined by INFRABEL) for passenger services (paX) is assumed as the decision-making unit (DMU) for evaluating the railway system. Each DMU corresponds to a train path or itinerary as defined by its origin-destination and scheduled pattern. Each path or itinerary traverses specific train lines and tracks. The transformation of resources, represented as inputs, into services with specified outputs, is evaluated. By developing and applying an appropriate DEA model, efficiency scores can be generated, providing decision-makers (including managers and staff and operators in the traffic controlling centers) with deeper insights into the system's performance.

4.1.2.1 Extraction of DMUs

DMUs as train paths here are extracted from the raw data of the September 2023 timetable (first week from 1 to 7 of September; due to the inconsistency in the dataset shared, only the first week is selected). Each train path is a combination of the values in the column “Relation” and the “Direction”. This research only considers domestic passenger trains, as these trains cover the majority of the service within the network. These trains are identified with Relation starting with IC and L, which represent fast Intercity trains and Local trains, respectively. As these train paths are working in two directions, for each of these relations and directions, we considered one DMU (each unique value of IC and L Relations has two DMUs for the two directions). Overall, there are 204 DMUs.

Due to the homogeneity assumption, all the train runs in the timetable are classified based on their operating time into four categories: The first three categories from the weekdays (divided based on their first departure’s time) include (i) morning peak hour (from 6 am to 9 am), (ii) evening peak hour (from 4 pm to 7 pm) and (iii) off-peak hour (the rest) and (iv) weekend trains. Variables related to each trajectory would have different values in each of these categories and will be compared to each other within the category. It allows the managers to compare the different times of the day performance with each other (i.e., morning peak to evening peak), also makes it possible to run a meta frontier analysis among all DMUs in all categories (O’Donnell et al, 2007), which can bring insights about the performance of all train paths at once.

Meta-frontier approach allows comparisons across different technologies. This approach, known as meta-frontier analysis, enables comparisons across different technologies. A meta-frontier serves as an overarching boundary that encompasses all group-specific frontiers, offering a common benchmark for heterogeneous groups. It represents the best-practice technology across all groups (Yu and Chen, 2019).

Table 3 shows the information on the train runs within a week, to have an understanding of the operation scale in the Belgian railway network in a week. On average, a weekday has about 3800 and a weekend has about 2300 train runs.

Table 3 - Insights from the raw data

# of unique	***Week 1***	***4 Sep*** *(Mon)*	***2 Sep*** *(Sat)*

Train run (associatie_ID_DAT on the Timetable[4])	22976	4457	2723
RELATION	122	119	83

4.1.2.2 Variable Definitions

Some of the variables associated with this research are extracted from the planned data. This means that their nature is determined from the timetable; For example, the duration of the journey for a train service is the planned duration of the journey, based on what was decided for that service when the timetable was designed. It might not be accomplished completely by the actual service when it is in operation. Other variables represent the actual performance of the trains (actual operation reality when the train service follows the designated plan). In this research, there is an undesirable variable as a result of the actual performance, the delay. Table 4 of the variables with their features is shown below:

Table 4 - The variables with their nature

Type	**Input**	**Output**	**Type**
planned	Headway	#trains operated	planned
planned	journey duration	#stops/stations	planned
		delay (-1)	actual
		# non-cancelled trains	actual

Undesirable variables: (-1)
Type: planned - actual

In order to calculate the values associated with these variables, first, the set of unique train run IDs ("associatie_ID_DAT") associated with each DMU is extracted. Then the value for each variable is calculated based on the set of train runs associated with each ID. More details about variable calculations are as follows:

Inputs:

- Duration (minutes): Firstly, the duration of each train run is extracted from the planned columns (for each unique train run ID, the difference between the value of column "TRAIN_THR_DEP" in the first row and "TRAIN_THR_ARR" of the last row is considered as duration of the journey). Then, the median (as a robust measure of central tendency) over the train runs is calculated.
- Headway (minutes): For each set of train IDs sorted in order of departure, the difference between the values of "TRAIN_THR_DEP" column in the first rows of the two consecutive train IDs is calculated. Then, the median over the train runs is

[4] A sample of the timetable with the columns is shown in a separate Appendix file.

calculated. (For the weekday data frames, it could be problematic, as the time interval for morning and evening peaks is three hours, and sometimes a DMU has a headway of more than three hours. Besides, this problem could apply to the off-peak trains set with two big gaps in the headways. To avoid this issue, the headway is calculated for the whole day for these data frames.)

Outputs:

- # trains operated (count): For each DMU, the count of the set of unique IDs represents the number of trains operated or train runs.
- # stops (count): For each of the IDs in the list (set of unique IDs for each DMU), through all the records, if the value of “TRAIN_THR_DEP” and “TRAIN_THR_ARR” is not the same, then it means that the row is a stop. The number of all stops for each train run is calculated this way and the median over the train runs is calculated.
- Delay (minutes): Delay is an undesirable variable. In this research, we consider the positive delays upon arrivals at the stations (due to the fact that this type of delay could be disruptive and raise concerns for all the stakeholders). The train can get delayed at each station, but it could be able to compensate for it over time until its arrival at the last station. However, this would not make the performance acceptable, specifically considering the star-shape of the network, any delay in the middle stations seems to be more disruptive than the first and final stations. Therefore, to better capture the performance of the train in terms of the delay, for each train run, the difference between the planned and actual arrival time at each stop/station is considered. Then the average (mean) delay for each train run is calculated. Finally, the median over the train runs is calculated.
- # non-cancelled trains: This variable is calculated based on the train run actual performance; whether it has taken the journey completely or not (partially or got canceled completely), based on the comparison of the first and last stations that the trains have been taking into account the origin and the destination (the indicators of the train path).

Tables 5 to 8 below show the descriptive statistics for the model variables in this case study, after the data preparation and cleaning (DMUs with the empty values in each timeframe are cleaned, as there was no train runs for that specific path in that time data frame. For example IC 05 is not operating on weekends).

Table 5 - Descriptive statistics for the model variables in the morning peak data frame

MORNING (190 DMUs)		min	max	avg	median	stdev
input	duration	13	180	73.06842105	64.5	38.22788191
	headway	26.5	210	*60.11842105*	*60*	15.3240285
	# trains operated (count)	1	8	2.989473684	3	0.902609232

output	#non_cancelled trains	0	5	1.968421053	2	0.969946615
	delay (-)	0	5.320833333	1.095623684	0.902857143	0.833238863
	#stops	2	21	8.539473684	8	4.27274737

Table 6 - Descriptive statistics for the model variables in the evening peak data frame

EVENING (192 DMUs)		min	max	avg	median	stdev
input	duration	17	189.5	83.32261285	74	44.70539175
	headway	20.675	210	60.00091146	60	15.43830916
	# trains operated (count)	1	6	3.1875	3	0.756299893
output	# non_cancelled trains	0	4	2.666666667	3	0.781549899
	delay (-)	0	13.10406504	1.635539715	1.277339181	1.400596283
	#stops	2	21	9.421875	9	4.647990122

Table 7 - Descriptive statistics for the model variables in the off-peak data frame

OFFPEAK (191 DMUs)		min	max	avg	median	stdev
input	duration	17	180	81.31082024	75	43.4318552
	headway	26.5	210	60.20157068	60	15.21218167

	# trains operated (count)	1	27	13.51832461	13	4.001108967
output	# non_cancelled trains	0	13	9.696335079	10	2.583742119
	delay (-)	0.01547619	4.344444444	1.098317036	0.937671233	0.742807023
	#stops	2	21	9.17539267	9	4.479705315

Table 8 - Descriptive statistics for the model variables in the weekend data frame

WEEKEND (141 DMUs)		min	max	avg	median	stdev
input	duration	7	219.5	69.40248227	57	46.96764438
	headway	18	120	72.05656028	60	24.12538972
	# trains operated (count)	2	30	16.0035461	17	4.983722237
output	# non_cancelled trains	0	18	6.620567376	6.5	6.661161608
	delay (-)	0.022740546	5.145833333	0.915589846	0.712366071	6.661161608
	#stops	2	24	7.546099291	7	4.082038492

4.2. Discussion

This section interprets the technical efficiency results obtained from the DEA models introduced in Section 3, applied to the Belgian train path data described in the case study. The analysis is conducted under the assumption of variable returns to scale, as transportation systems tend to operate this way (Sheth et al, 2006) and encompasses both group-specific models—representing different times of the day—and a consolidated Meta-model that benchmarks all DMUs against a unified reference technology. The following discussion explores patterns in efficiency distributions, relationships between operational variables, and the influence of contextual factors such as route topology and service timing. Emphasis is placed on identifying both consistent trends and noteworthy deviations, offering insight into the operational and structural dimensions shaping train path performance.

4.2.1. Efficiency Score Ranges and Ranking Distribution (BCC-n and DDF Models)

The efficiency score distributions from the BCC-n (radial) and DDF (non-radial) models reveal consistent patterns across all time-of-day classes (Figure 4). In the BCC-n model, most DMUs demonstrate efficiency scores above 0.5, indicating a reasonable level of operational effectiveness in converting inputs into outputs. A particularly dense cluster appears in the [0.9, 1] interval, representing units positioned close to the efficiency frontier.

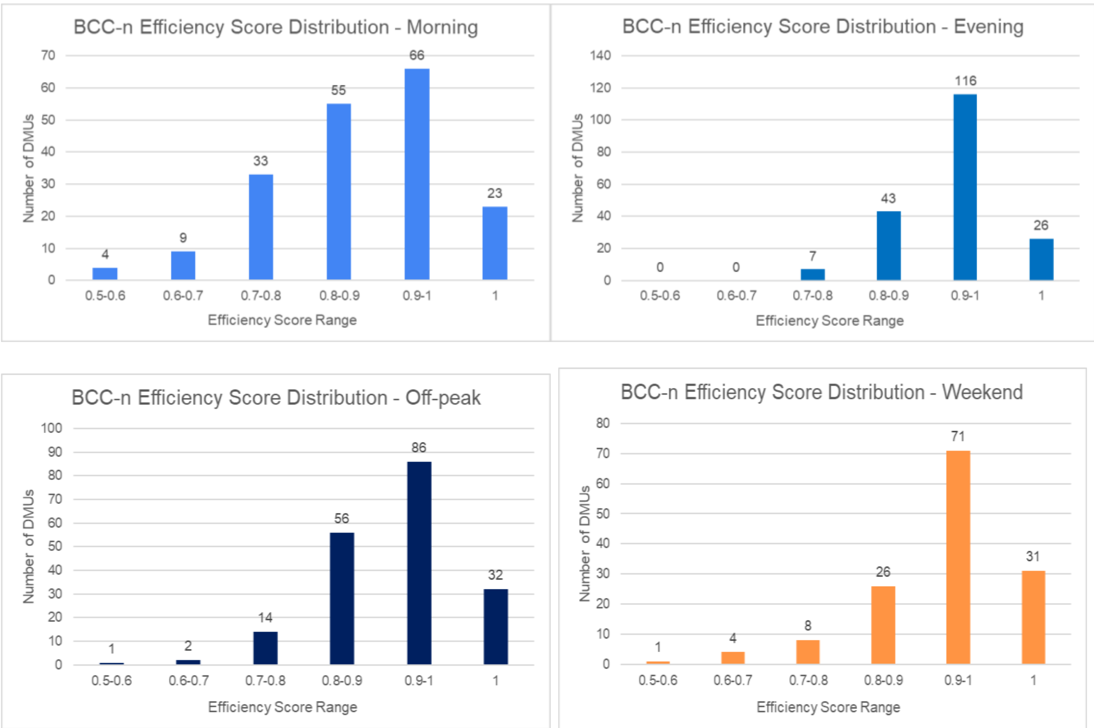


Figure 4 - BCC-n efficiency score distribution in 4 classes

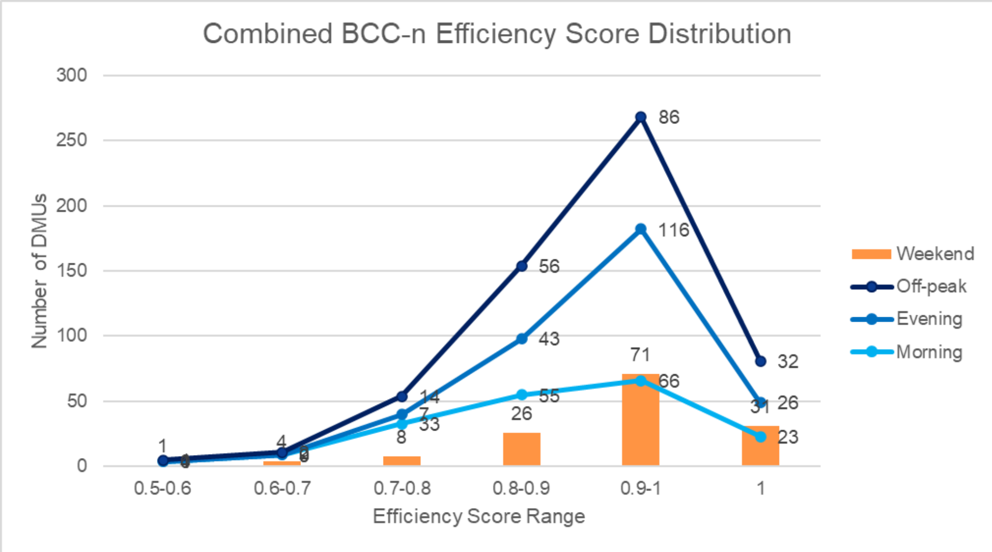


Figure 5 – Combined BCC-n efficiency score distribution of 4 classes

The combined BCC-n (Figure 5) efficiency score distribution chart visually reinforces the findings discussed above. Most DMUs demonstrate efficiency scores well above 0.5, which is consistent with expectations, as all units are involved in operational activity, translating inputs into outputs. The most populated group falls within the [0.90, 1] interval, indicating that a substantial number of DMUs operate near the efficiency frontier. This pattern is particularly evident in the evening and morning datasets, which show the highest frequency of near-efficient units (116 and 71 DMUs, respectively). A smaller share of DMUs occupies the lower efficiency intervals, with off-peak and weekend services exhibiting a more dispersed distribution. This collective visualization affirms the general operational soundness of the Belgian train path system, while also pointing to specific time periods where performance enhancements may be most beneficial.

The DDF model yields objective function values that are an alternative metric for assessing relative performance, which shows the room for operational improvement (same information obtained by the BCC output-oriented DEA models). It serves as an indicator of how far a given DMU is from the directional frontier under a specified vector. The distribution of these values spans a wider range, particularly for the off-peak and weekend datasets, where operational variability tends to be greater (Figure 6). Several DMUs show values close to zero, indicating proximity to the efficient boundary and reflecting strong technical performance in the direction measured. Conversely, elevated objective function values—more frequently observed during evening and weekend periods—suggest increased inefficiency, with room for directional improvement across multiple outputs. These differences highlight the DDF model's ability to capture performance deviations that may be understated in the radial assessments.

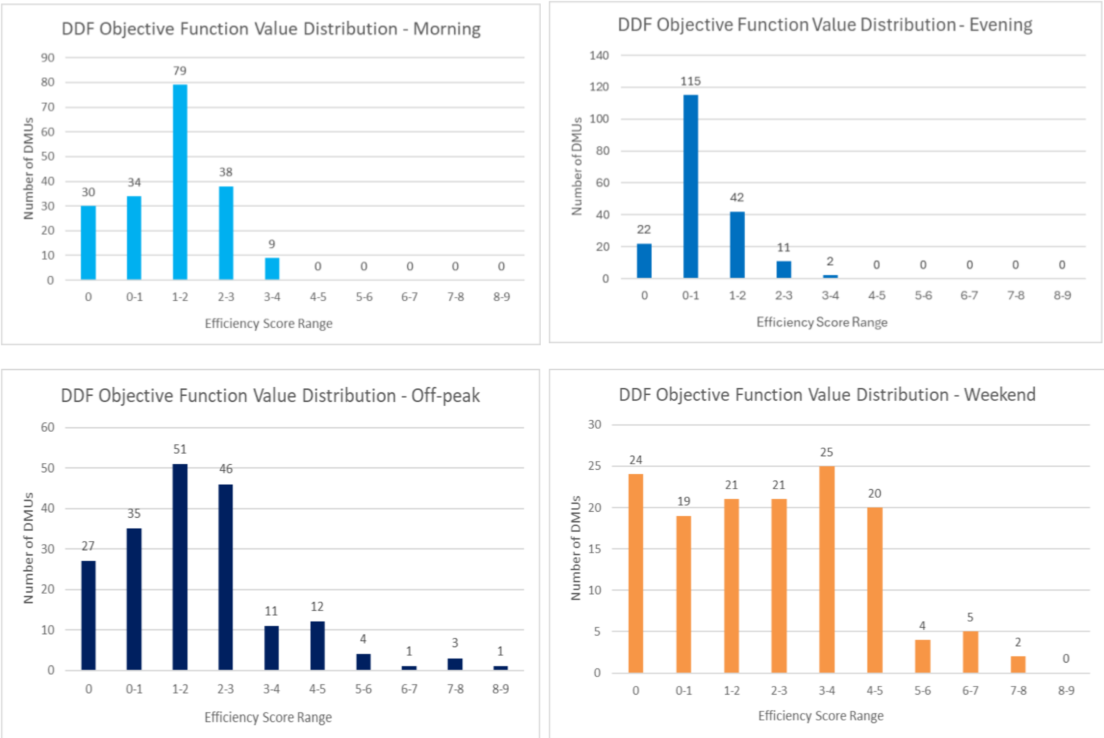


Figure 6 - DDF objective function values distribution in 4 classes

The combined distribution of DDF objective function values (Figure 7) reveals that most DMUs cluster within the [0–2] range, with peak density occurring in the [0–1] interval for morning and evening operations. This indicates that a large portion of train paths operate near the directional frontier. In contrast, off-peak and weekend services display a more even spread across higher value intervals, reflecting relatively greater inefficiencies. These patterns underscore the consistent technical performance during core weekday periods and highlight specific operational contexts, such as reduced frequency or routing constraints, that may impact efficiency during non-peak times.

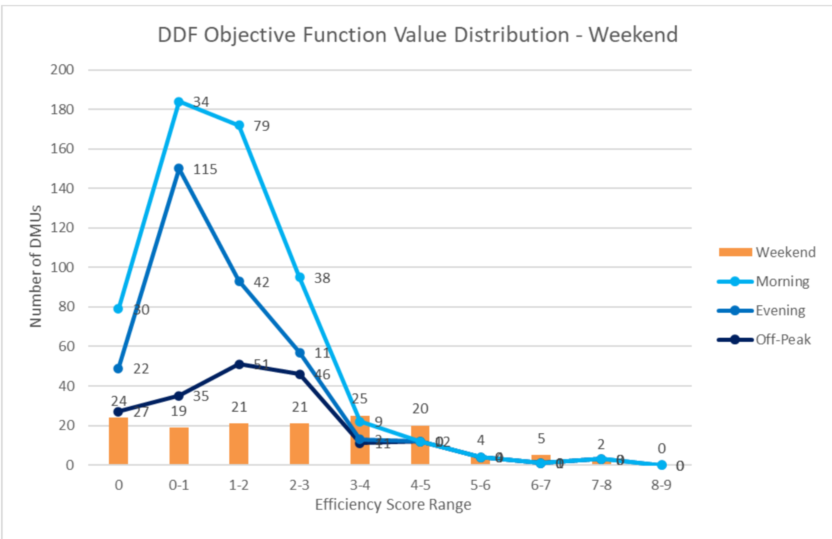


Figure 7 - Combined DDF objective function values distribution of 4 classes

The overall pattern shown in Figure 7 is aligned when comparing with the BCC-n results with respect to the number of existing DMUs in each class in each efficiency range. Thus, for the next set of comparisons, BCC-n is selected.

4.2.2. Delay-Based Efficiency Comparison

An analysis of delay-based groupings (Figure 8) reveals a clear inverse relationship between average delay levels and technical efficiency scores. DMUs with higher delay counts tend to exhibit lower efficiency under the BCC-n model, highlighting delay as a key operational weakness influencing overall performance. While delay is an undesirable output, its impact is evident in the profile of efficient DMUs, which generally report delay values below the dataset mean or the average—and often below the mean plus one standard deviation. This pattern is consistent across all time-of-day classes, further reinforcing delay as a dominant operational discriminator in technical performance. The persistence of this trend among both efficient and non-efficient units suggests that minimizing delay remains one of the most effective levers for improving train path performance. Nonetheless, a few DMUs with relatively high delays still achieve comparatively high efficiency scores, indicating that strong performance in other variables can occasionally offset the impact of delay.

Figure 8 – Delay with the number of DMUs and the average of the efficiency score in 4 classes

4.2.3. Duration-Based Efficiency Comparison

Efficiency patterns analyzed by journey duration reveal that both short and long train paths can achieve high efficiency scores. Efficient DMUs span a wide range of durations, with no consistent preference toward either end of the scale. This distribution implies that duration, in isolation, is not a limiting factor in operational efficiency under the current model configuration. Instead, efficiency appears to depend more on how well the journey characteristics, such as frequency, stops, and delay control, are managed within the time frame allocated. The presence of highly efficient long-distance services, as well as short-distance routes with lower scores (Figure 9), underscores the importance of coordination and operational quality rather than trip length alone.

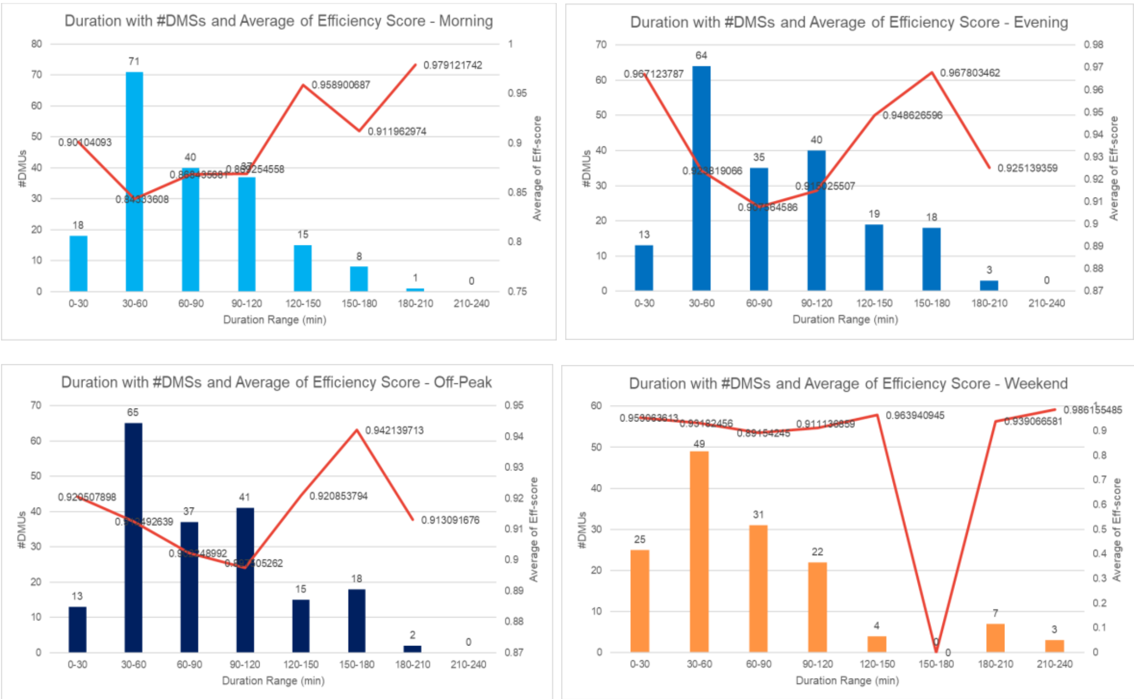
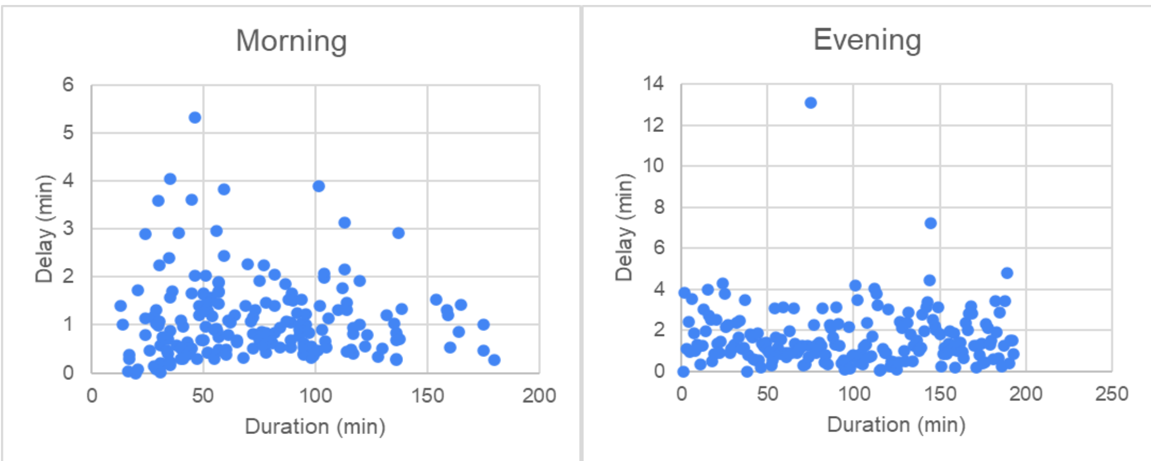


Figure 9 – Duration with number of DMUs and the average of efficiency score in 4 classes

4.2.4. Frontier Visualization: Duration–Delay and Stops–Delay

Two-dimensional frontier plots provide additional insights into the relationships between key output variables and their influence on technical efficiency. The duration–delay (X–Y) plot (Figure 10) illustrates that longer train paths tend to accumulate more delay, but efficient DMUs manage to maintain low delay even across extended durations. This suggests that efficient scheduling and route management can mitigate the expected operational burdens of longer trips.



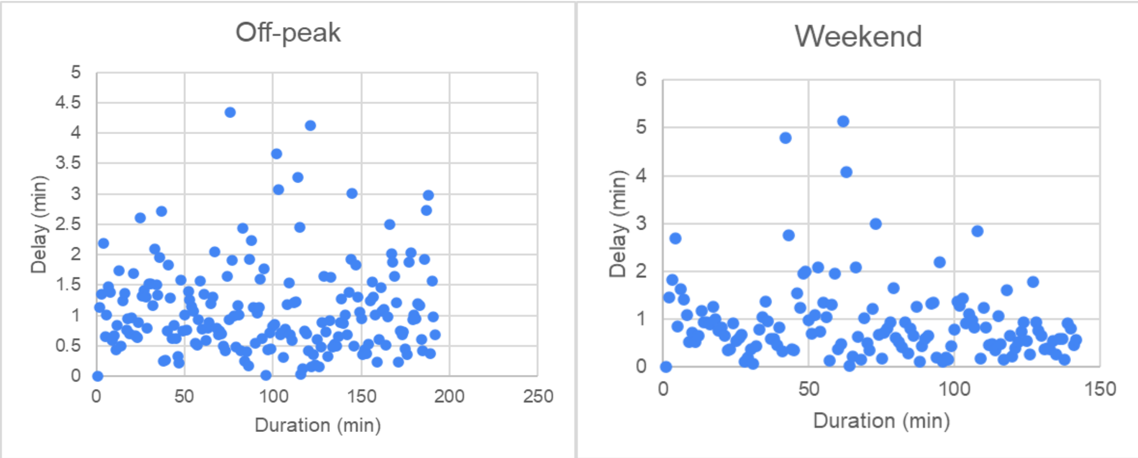


Figure 10 – Scatter plot of delay and duration for 4 classes

In the stops–delay (Y–Y) plot (Figure 11), a mild positive association is observed, indicating that higher numbers of stops often correspond with increased delay. However, several efficient DMUs defy this pattern, achieving low delays despite frequent stops. These observations underline that operational strategies—rather than structural features alone—play a critical role in determining efficiency outcomes. Together, these plots visually reinforce that high-performing train paths are not solely defined by minimal scope but by the system's ability to control performance under varying conditions.

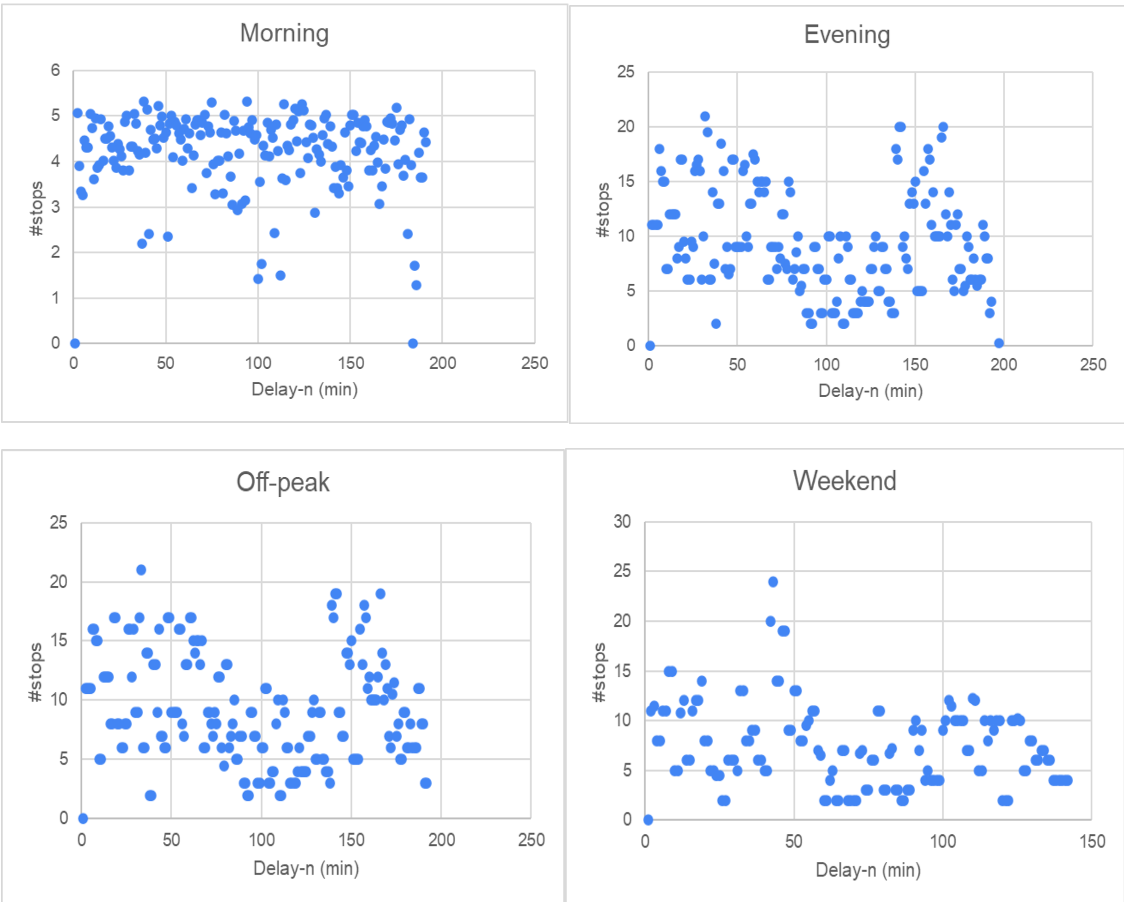


Figure 11 - Scatter plot of number of stops and delay for 4 classes

Two-dimensional frontier plots illustrate the relationship between duration and delay (Figure 10) and stops and delay (Figure 11) across all DMUs. Efficient units appear clustered toward lower delay values regardless of duration or number of stops, highlighting the role of operational control in achieving technical efficiency.

4.2.5. Efficient DMUs Variables Deviations

A descriptive analysis of efficient DMUs under both BCC-n and DDF models reveals notable variability across input and output values, challenging the notion of a consistent efficiency profile (Tables 9-17). Within each time-of-day group, efficient units span a wide range of durations, headways, and delays—indicating that efficiency is not confined to narrowly defined operational characteristics. For instance, several efficient DMUs exhibit above-average delays or extended durations, yet still qualify as efficient due to their relative performance across all variables. This suggests that efficiency emerges from balanced trade-offs among inputs and outputs rather than optimization of a single factor. Additionally, standard deviations within the efficient group remain considerable, further emphasizing the diversity of performance pathways that can lead to high efficiency. These findings imply that operational flexibility and context-specific strengths, rather than uniformity, characterize the most efficient train paths. Statistical tests can be run to test whether the means (avg) and/or medians of the different classes are significantly different for specific operational variables and differences in efficiency scores

Table 9 – Morning – BCC-n

Morning – BCC-n	duration	headway	count	stop	delay	non-canceled
Min	13	26.5	1	2	0	0
Max	175	210	8	21	3.127672956	5
Mean	72.80434783	59.10869565	3.695652174	10.30434783	0.811645776	2.043478261
Median	77	60	3	8	0.541481481	2
St. Dev.	51.05583822	34.8558579	1.520609406	6.783568266	0.864513362	1.223937791

Table 10 – Morning - DDF

Morning – DDF	duration	headway	count	stop	delay	non-canceled
Min	13	26.5	1	2	0	0
Max	175	210	8	21	5.320833333	5

Mean	64.4	65.08333333	3.266666667	8.733333333	1.635123848	1.9
Median	53.5	60	3	7	1.136666667	2
St. Dev.	41.9503812	33.09471344	1.638614497	5.811304877	1.365784421	1.124952106

Table 11 – Evening – BCC-n

Evening – BCC-n	duration	headway	count	stop	delay	non-canceled
Min	17	20.675	1	2	0	1
Max	189.5	210	6	21	3.479577465	4
Mean	92.4400641	56.00673077	3.884615385	11.57692308	1.29190992	2.807692308
Median	95.25	59.75	4	9.75	0.949488994	3
St. Dev.	59.9134409	33.95562071	1.210848399	6.41356735	1.116644755	0.693929724

Table 12 – Evening - DDF

Evening – DDF	duration	headway	count	stop	delay	non-canceled
Min	17	20.675	1	2	0	0
Max	173	210	6	21	13.10406504	4
Mean	80.7365942	61.52934783	3.52173913	9.760869565	1.961553566	2.434782609
Median	75	60	3	7.5	1.151833333	3
St. Dev.	51.22162893	37.6191506	1.533550996	5.787662358	2.713107035	1.036869719

Table 13 - Off-Peak – BCC-n

Off-peak – BCC-n	duration	headway	count	stop	delay	non-canceled
Min	17	26.5	3	2	0.01547619	0
Max	180	119	27	21	2.718627451	13
Mean	81.38235294	60	13.5	8	0.729962366	10.5
Median	71.5	55.88235294	14.79411765	9.485294118	0.912065219	2.85546875
St. Dev.	52.85658968	15.20718472	5.735479224	5.627967627	0.720161555	3.84329868

Table 14 - Off-peak - DDF

Off-peak – DDF	duration	headway	count	stop	delay	non-canceled
Min	20	26.5	1	2	0.01547619	0
Max	177	60	27	19	4.344444444	12
Mean	93.125	60	15	8.5	0.656430841	10
Median	85	51.34375	14.5	10.4375	1.157594152	3.03125
St. Dev.	55.8460682	11.67435758	6.87992248	5.163574343	1.337752878	3.91524797

Table 15 – Weekend – BCC-n

Weekend – BCC-n	duration	headway	count	stop	delay	non-canceled
Min	7	18	2	2	0.0325	0
Max	219.5	120	30	24	5.145833333	18
Mean	67.19354839	59.66129032	18.5	8.564516129	0.786381636	8.403225806

Median	53	60	19	8	0.529166667	9
St. Dev.	55.84139406	14.19236127	4.887057738	5.157505753	0.980710727	8.176612333

Table 16 – Weekend - DDF

Weekend – DDF	duration	headway	count	stop	delay	non-canceled
Min	7	18	2	2	0.2125	0
Max	219.5	60	30	24	5.145833333	18
Mean	76.20833333	57.0625	19.20833333	10.125	1.337088401	10.16666667
Median	62.5	60	19	10	0.926990093	15
St. Dev.	57.02629383	9.866844458	5.012845817	5.337521001	1.31042791	7.913481434

A combination of all the results presented above, as summarized in the table (Table 17) below for the morning class, highlights the range of input and output variables among efficient DMUs across both BCC-n and DDF models. While most variable ranges remain consistent between the two methods, the DDF model allows for slightly greater variation in delay, suggesting a higher tolerance for irregular performance when efficiency is evaluated directionally. Efficient DMUs are observed with both long and short durations, low and high headways, and a broad range of stop counts, underscoring the diversity in operational configurations that can lead to high performance. This supports the view that different models emphasize distinct dimensions of efficiency, and that efficiency is not tied to a uniform structure. Rather, strong performance can emerge from a variety of operational profiles, reinforcing the notion that there is no singular profile for technical efficiency within the train path system.

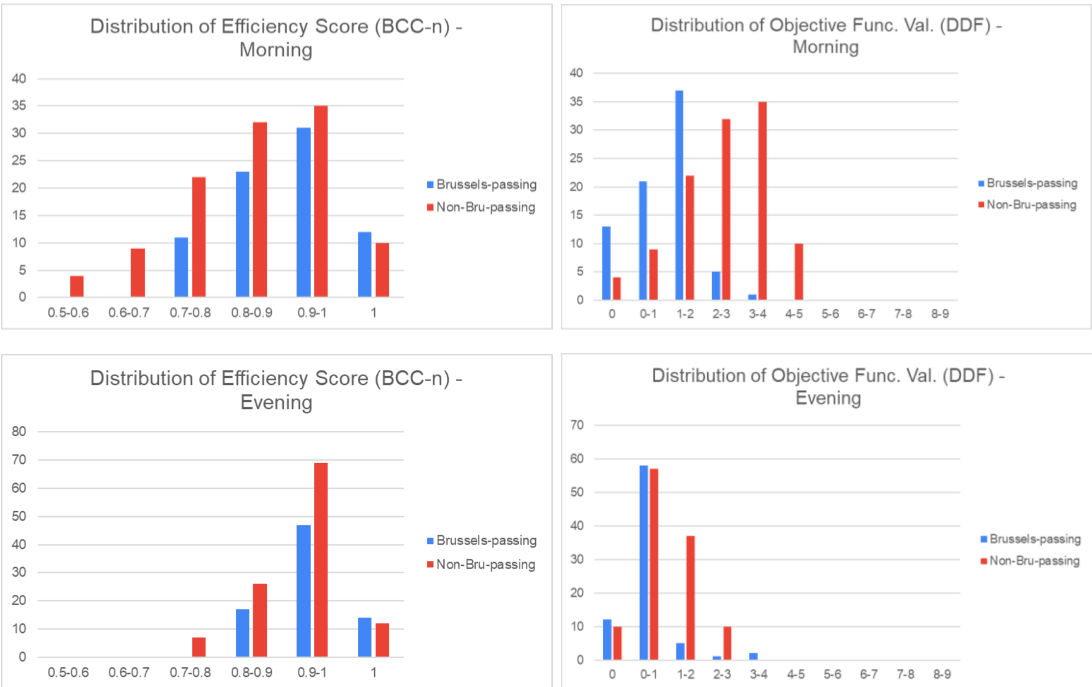
Table 17 – Summary of efficient DMUs variables in morning class

Variable	BCC-n (Min–Max)	DDF (Min–Max)	Observation
Duration (min)	13 – 175	13 – 175	Broad overlap
Headway	26.5 – 210	26.5 – 210	High variance
Count (#)	1 – 8	1 – 8	Similar spread
Stops	2 – 21	2 – 21	Consistent

Delay (min)	0 – 3.13	0 – 5.32	Wider in DDF
Non-Cancelled	0 – 5	0 – 5	Identical

4.2.6. Efficiency Patterns Based on Network Centrality (A Classification of DMUs)

To explore the influence of network topology on performance, DMUs were classified based on whether their routes pass through Brussels, the central hub of the Belgian railway system. Among the 204 DMUs analyzed, 80 (approximately 40%) traverse Brussels stations (Nord, Central, or Midi/Zuid). (Figure 12). This classification reveals that efficient DMUs are present in both categories, suggesting that central network access is not a sole determinant of technical efficiency. However, the presence of efficient train paths that do not rely on the Brussels core indicates that operational quality can be maintained even along peripheral or less centralized routes. This finding underscores the adaptability of the network and suggests that efficiency is achievable both within and outside high-density, high-demand corridors. This classification provides a valuable foundation for future research aimed at integrating contextual or environmental variables—such as network shape, service density, and train occupancy—into the efficiency model. If notable differences emerge between the two groups, such insights could support the development of more refined models that account for structural and geographical factors influencing performance.



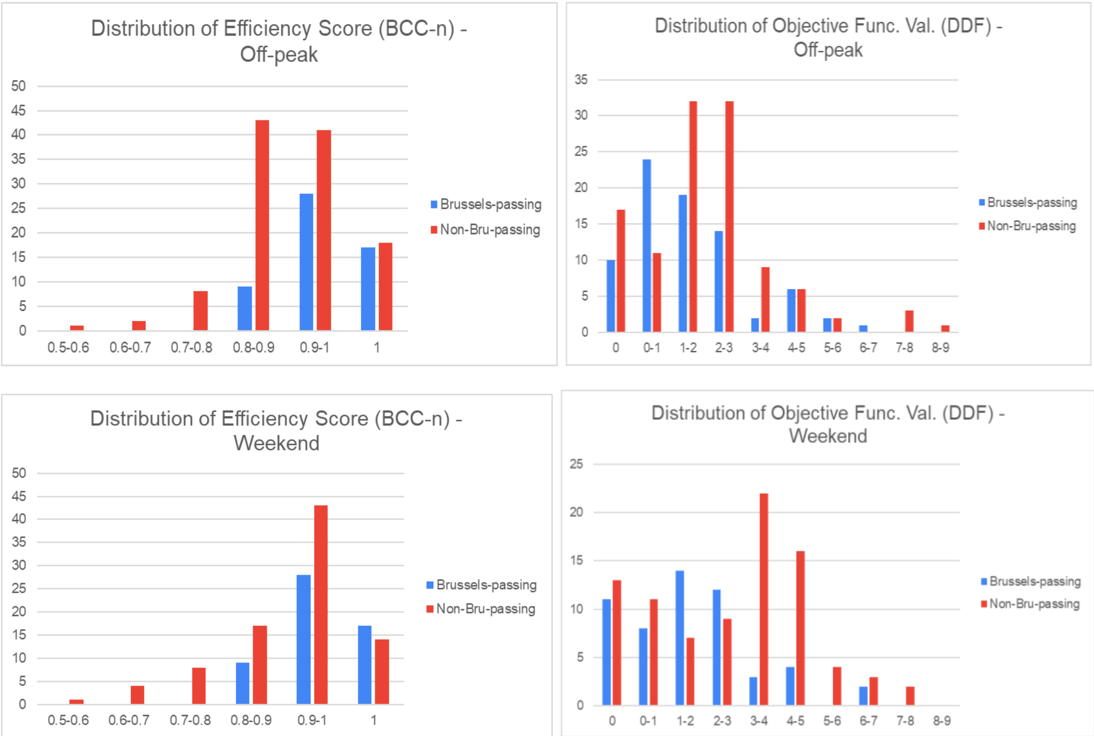


Figure 12 – Distribution of efficiency score for train paths passing through Brussels and not

4.2.7. Cross-Group Efficiency and Variable Deviations (Meta-model analysis)

A cross-group analysis of variable deviations across the combined dataset provides a broader view of efficiency dynamics beyond individual time-of-day segments (Table 18). When aggregating across all DMUs, the range of input and output variables remains wide, reaffirming that technical efficiency can be achieved under highly diverse operational conditions. Variables such as duration and count show particularly high standard deviations, reflecting heterogeneity in train path structures and service frequency. Interestingly, the delay variable spans from zero up to over 13 minutes, yet efficient DMUs are present across nearly the entire range, suggesting that delay alone does not define inefficiency, but rather its interaction with other performance metrics.

Additionally, a comparison of peer units across groups shows that some DMUs consistently serve as references under different operational contexts, indicating their robust performance characteristics. This insight may guide scheduling or planning practices by identifying consistently high-performing paths. The diversity of conditions under which efficiency is achieved supports the notion that flexibility in operations, rather than strict adherence to a single service template, is central to achieving high technical performance in a complex and demand-sensitive railway network.

Table 18 - Deviation of the variables - Meta-frontier analysis

META	duration	headway	count	stop	delay	non-canceled
Min	7	0.895934959	1	2	0	0
Max	219.5	14	30	24	13.10406504	18

Mean	77.30680439	12.79277073	8.429271709	8.75070028	1.207229267	5.142156863
Median	68	13.05196635	4	8	0.948033654	3
St. Dev.	43.47132115	1.028311258	6.563496979	4.440111008	1.028311258	4.607272299

The Meta-model BCC-n distribution is sharply right-skewed (Figure 13), with the majority of DMUs clustered in the [0.9–1] efficiency range. Over 500 DMUs fall in this near-efficient category, confirming the robustness of technical performance across the dataset. A small group of fully efficient units (score = 1) stands apart, representing benchmark paths in the Meta context.

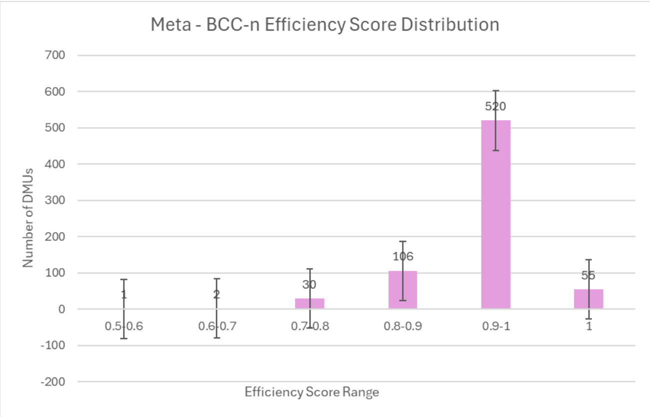


Figure 13 – Meta-frontier analysis – Distribution of BCC-n model efficiency score

The Meta DDF objective function values exhibit a broad distribution, with the largest concentration of DMUs falling between 2 and 4 (Figure 14). While fewer units appear near zero, indicating fewer instances of full directional efficiency, the overall spread suggests variation in performance across outputs when evaluated under a unified directional vector. However, as the appearance of the distribution shape is sensitive to binning choices, comparisons with BCC-n should be made cautiously and with consistent scale considerations. Nonetheless, the DDF results highlight the potential for improvement in multiple dimensions, particularly among mid-performing units.

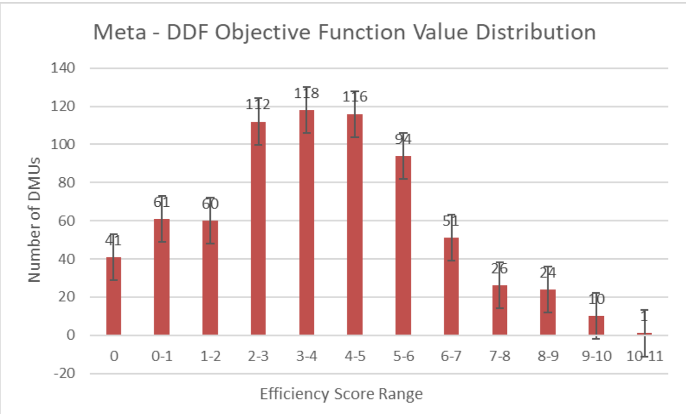


Figure 14 - Meta-frontier analysis – Distribution of DDF model objective value

4.2.7.1. Peers in meta-frontier analysis - based on the time of the day classes:

The count of efficient DMUs by time-of-day reveals an interesting temporal pattern. The weekend group accounts for the highest number of efficient units (24), followed by off-peak (14), while evening operations report the fewest (6) (Figure 15). This may reflect relative slack in network load or operational design during less congested periods, offering an opportunity to identify transferable practices from off-peak and weekend services.

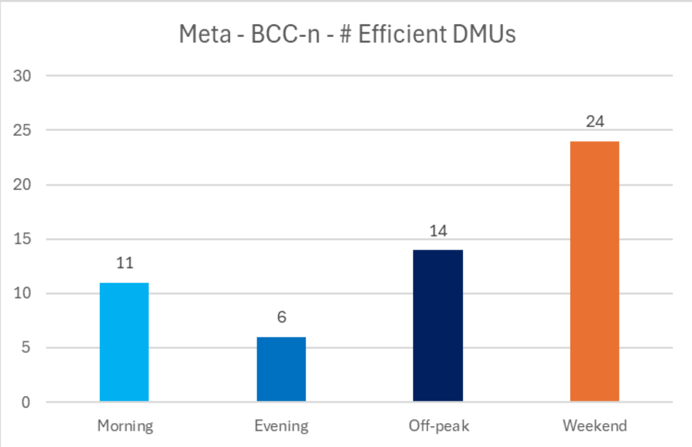


Figure 15 - Efficient DMUs in Meta-frontier analysis by Time of Day

The peer distribution chart below (Figure 16) illustrates which DMUs serve as frequent reference points in the Meta BCC-n model. Several units are repeatedly selected as peers across time-of-day groups, reflecting structural robustness and consistent performance. Notably, many of the most frequently referenced peers belong to the morning and off-peak segments, suggesting that these units offer adaptable efficiency patterns across different operational contexts. In contrast, the evening group contributes fewer efficient DMUs, and those identified as peers tend to support only a limited number of other units. This may indicate narrower applicability or less stable performance during evening operations, potentially due to higher variability or operational constraints specific to that time frame.

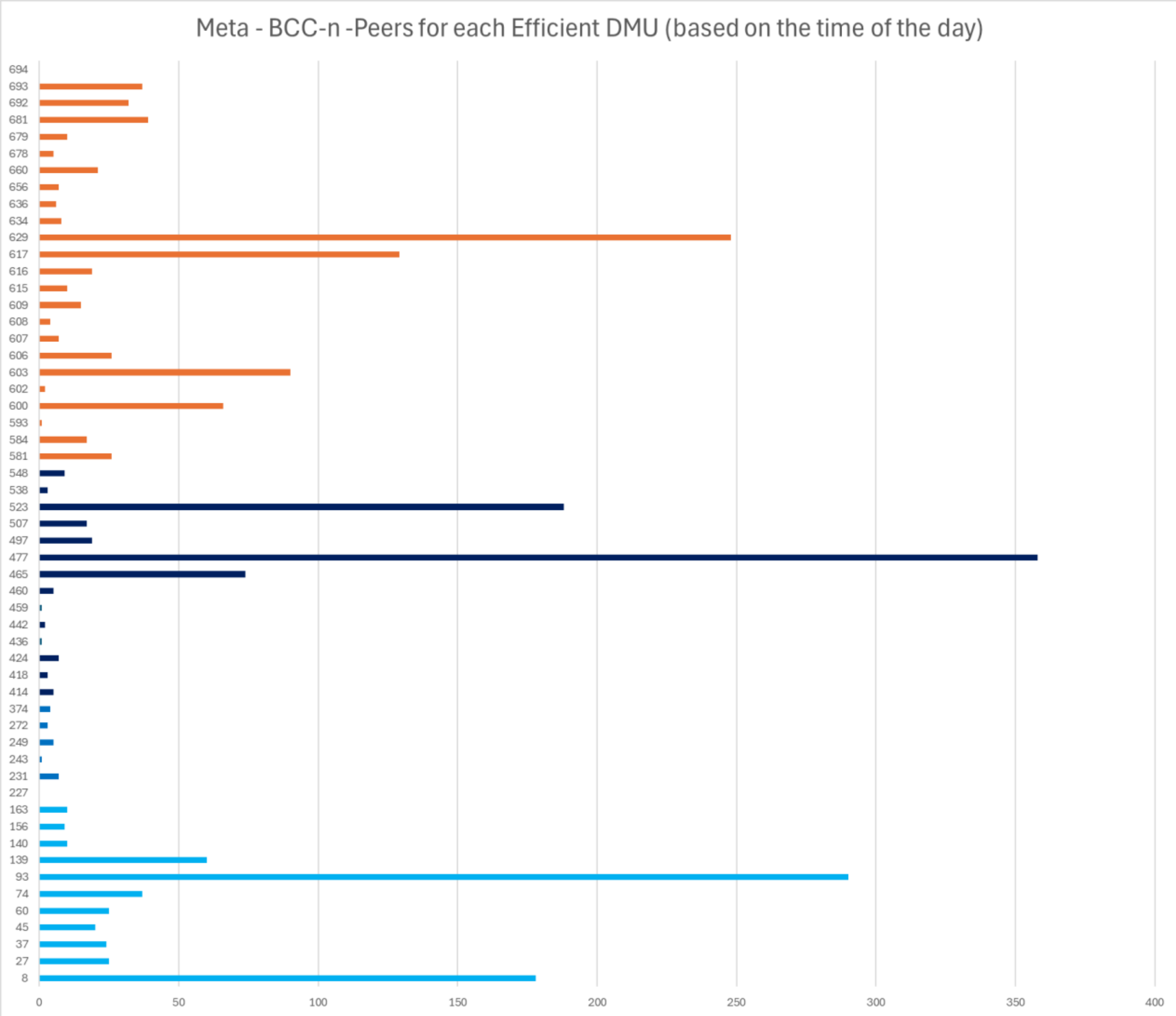


Figure 16 - Peer DMU Distribution per Efficient Unit in Meta

The findings presented in this discussion offer a comprehensive view of technical efficiency in the Belgian railway system, highlighting not only consistent trends across models and time-of-day categories but also the structural and operational diversity within the set of efficient DMUs. The comparison between BCC-n and DDF models reveals the complementary strengths of radial and non-radial approaches, while patterns in delay, duration, and stop frequency underscore the importance of managing variability within the network. Moreover, the presence of efficient train paths across different route topologies, service types, and operational configurations reinforces the notion that efficiency is multifaceted and context-sensitive. These insights contribute to a deeper understanding of performance dynamics and provide a foundation for practical interventions aimed at enhancing network effectiveness. At the same time, they point to numerous avenues for further exploration—aligned with the iterative nature of the methodology—including consultation with domain experts and practitioners (i.e. within INFRABEL), deeper investigation of additional contextual data, and continued engagement with the evolving body of literature to refine the efficiency framework.

5. Conclusions and Future Research

While various approaches to performance measurement have been widely explored in academic research, their practical application by decision-makers within transportation systems remains limited (Sheth et al., 2007). Despite the growing body of literature on efficiency analysis, many operational decisions continue to rely on conventional metrics rather than integrated performance frameworks. This gap underscores the need to critically assess transportation processes using established frontier-based methods, such as those proposed in the literature (Fried et al., 1993), in order to support more informed decision-making and drive meaningful improvements in transit system performance.

The modeling framework developed in this research is intended to support a practical application by transportation planners, agency managers, and traffic engineers. Ideally, these decision-makers would be actively involved in the modeling process to ensure that the formulation and outcomes of the model align with real-world practices—a process often referred to as face validation. In this study, some level of face validation was achieved, particularly through the definition of variables and the interpretation of results. However, fully implementing the framework within an operational setting would require a substantial commitment of resources from transportation agencies. As highlighted in previous studies, successful adoption of DEA-based methods depends not only on technological capacity but also on organizational willingness to embrace new performance assessment paradigms (Borja et al., 2007).

This paper investigates the technical efficiency of the train paths within railway systems through a DEA analysis for 4 classes based on the time of the day and a consolidated Meta-analysis model, focusing on punctual performance of the about 200 train paths from a case study in Belgium.

The findings reveal consistent efficiency patterns across different DEA models (BCC negative and DDF), with most DMUs demonstrating relatively high technical efficiency scores and a significant share falling within the near-efficient range, particularly in the rush hour operations with maximum load in the system (the morning and evening).

Technical efficiency patterns are generally aligned across the BCC-n and DDF models for this case. The exception occurs in a small number of DMUs, such as IC 35-i and L 26-ii, which exhibit substantial differences in efficiency classification, likely due to the contrasting assumptions and measurement approaches in the radial versus non-radial models.

The results indicate that the mix of cases (classified by time of day) significantly affects technical efficiency, with the evening group showing the lowest number of efficient DMUs and contributing fewer peer references. The Meta-model results further reveal that evening-class DMUs are generally positioned farther from the frontier compared to other time categories, which is likely attributable to higher delays and greater variability in operational conditions during evening hours.

The result of this research can be used at different levels in applications:

The efficiency scores (or objective function values) can be used to define an internal priority score within INFRABEL. The more inefficient a train is, the higher priority it should have to be assigned to the tracks and platforms, rather than other trains. This could be reliable because it results from the operational performance and characteristics of that specific path. This score can be used by the operators in the TCCs. Thus, they can make better decisions in real time to decide about assigning trains to the tracks (rather than the first-come, first-served policy that is usually used in INFRABEL or other railway systems).

Furthermore, the result of this work can help change the approach of INFRABEL for redefining the metrics used in the system to evaluate the performance of the company at an aggregate level, such as the punctuality metric (percentage of the trains with less than 6 minutes of the delay in the final destination or the Brussels stations corridor). Comparing the efficiency scores with the punctuality metric for each DMU reveals that this is not a reliable metric, and it does not adequately show the operational and performance reality. (See the Appendix)

In the long run, once the railway system company tests different approaches to decision-making in real-time and if some are still not working well, they can modify the timetable based on the bottlenecks, and even redesign some of the paths to get a better performance out of the existing infrastructure system.

Moreover, even later by running the model for a longer period of time (the data), the managers can identify the bottlenecks of the system in order to decide about the expansions and future infrastructural plans and investments (which would increase the capacity, etc.), along with other research and investigations.

This conclusion is consistent across the trains passing through the center of the star-shaped network and those that are not; there are efficient DMUs in both groups, revealing that passing through the most crowded and dense corridor of the network does not result in an inefficient performance of a path.

5.1. Future Research Recommendations

This research is limited by a problem in the timetable data regarding the completely cancelled trains. Based on the interviews with INFRABEL experts, the data shared for this project lacks these records, because of the recording process. Once a train is cancelled completely, all the records related to that train run would be deleted (as a modification of the actual timetable) in this dataset. This could cause problems with the reliability of the related calculated variables. Besides, DEA models consider a static and linear representation of the system, which makes it limited in capturing some of the operational facts and the causal effects of an incident (for example, delay) on other components of the system. Moreover, verification of the prepared data to feed the DEA model and then the result of it can be extended further to secure more reliability of the insights. There may be other limitations, such as not explicitly accounting for the stochasticity involved in the running of the train paths; An initial attempt to account for that could be Stochastic frontier analysis and fuzzy methods.

A potential direction for future research involves defining a framework that can be applied to other types of railway systems with different organizational structures and operational characteristics, and incorporating additional variables relevant to those systems. This research can be generalized beyond the case of INFRABEL by identifying the system specifications of the target region, whether within a single country, a specific sub-region, or across multiple countries, and then extracting the train paths (defined by origin, destination, and direction) within that region's railway network. For each train path, the relevant variables can be measured, such as journey duration, service headway, number of stops, number of trains operating on the path, total delay in minutes, and the number of trains completing the journey. These variables can be extracted and calculated across various types of railway systems (including train and metro networks), depending on their operational characteristics. It can be argued that any railway system offering regular services inherently contains these features, making the calculation of such variables feasible. However, to further enhance generalizability, this research can be expanded by identifying differing railway management structures—such as systems where infrastructure and train operations are integrated (e.g., DB in Germany), or where systems operate more privately and independently of government funding, unlike INFRABEL which receives its entire budget from the government. A comprehensive framework could then be developed that incorporates variables reflecting these diverse organizational and operational models.

This research presents a model focused on the supply side of the service (i.e., the provider) but does not currently incorporate variables representing the demand side (i.e., the passengers). Some of these variables were suggested in Section 3. However, the model could be expanded for railway systems by drawing on other studies as benchmarks, such as Sheth et al. (2007). Future work may incorporate these suggested variables, especially those related to passenger demand. This could be done using a Network DEA model (such as the one developed for bus routes by Sheth et al., 2006) or through alternative model specifications. Each modeling approach offers different perspectives and can provide valuable insights for decision-makers.

Future research should also focus on incorporating environmental and contextual variables to gain a deeper understanding of factors that are not directly controllable but still impact system performance (Topcu, et al. 2019). These variables require special attention, as their presence can influence operational outcomes in ways not captured by internal metrics alone. Once these variables and their impacts are identified, they may call for different methodological approaches, such as alternative strategies in timetable design or the implementation of specific policies tailored to contextual conditions.

Future research could also focus on evaluating the proposed suggestions, such as applying the inner-priority score based on the efficiency scores results from DEA models, within the Traffic Control Centers (TCCs) and then rerunning the model to observe any resulting changes in the efficiency of the DMUs. More broadly, other policies or operational changes can be implemented within the system and tested post-implementation. In this way, the model could serve as a valuable tool for internal policy evaluation and performance monitoring within the organization.

Another direction for future research involves using system dynamics models within this “black box” framework to better track changes occurring throughout the complex railway system at various levels. This approach could provide deeper insights into system performance and operations, enabling managers to make more informed and effective decisions. Additionally, it allows for a better understanding of potential impacts before policy or operational changes are implemented (Vaneman, 2002).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT in order to improve writing. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of this report.

References

Publications and books:

Azadeh, A., et al. (2008). Integration of DEA and AHP with computer simulation for railway system improvement and optimization. Applied Mathematical Modelling, 32(10), 2747-2759. https://doi.org/10.1016/j.apm.2007.10.005

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. https://doi.org/10.1287/mnsc.30.9.1078

Bešinović, N. (2020). Resilience in railway transport systems: a literature review and research agenda. Transport Reviews, 40(4), 1–22. https://doi.org/10.1080/01441647.2020.1728419

Besinovic, N., Wang, Y., Zhu, S., Quaglietta, E., Tang, T., & Rob M.P. Goverde. (2019). Integrated train and passenger disruption management for urban railway lines. https://doi.org/10.1109/itsc.2019.8917379

Bozzano, M., & Villafiorita, A. (2010). Design and Safety Assessment of Critical Systems. https://doi.org/10.1201/b10094

Bray, S., Caggiani, L., & Ottomanelli, M. (2015). Measuring Transport Systems Efficiency Under Uncertainty by Fuzzy Sets Theory Based Data Envelopment Analysis: Theoretical and Practical Comparison with Traditional DEA Model. Transportation Research Procedia (Online), 5, 186–200. https://doi.org/10.1016/j.trpro.2015.01.005

Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. Evaluation and Program Planning, 31(3), 299–310. https://doi.org/10.1016/j.evalprogplan.2007.12.001

Carayon, P., Hancock, P., Leveson, N., Noy, I., Sznelwar, L., & van Hootegem, G. (2015). Advancing a sociotechnical systems approach to workplace safety – developing the conceptual framework. Ergonomics, 58(4), 548–564. https://doi.org/10.1080/00140139.2015.1015623

Carey, M., & Carville, S. (2003). Scheduling and platforming trains at busy complex stations. Transportation Research Part A: Policy and Practice, 37(3), 195-224. https://doi.org/10.1016/S0965-8564(02)00012-5

Chambers, R.G., Chung, Y. and Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory, 70(2), pp.407-419.

Chambers, R.G., Chung, Y., & Färe, R. (1998). Profit, Directional Distance Functions and Nerlovian Efficiency. Journal of Optimization Theory and Applications, 98(2), 351–364.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. European Journal of Operational Research, 2(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Chen, Z., & Wang, Y. (2019). Impacts of severe weather events on high-speed rail and aviation delays. Transportation Research Part D: Transport and Environment, 69, 168–183. https://doi.org/10.1016/j.trd.2019.01.030

Chu, X., Fielding, G. J., & Lamar, B. W. (1992). Measuring transit performance using data envelopment analysis. Transportation Research Part A: Policy and Practice, 26(3), 223–230. https://doi.org/10.1016/0965-8564(92)90033-4

Cooper, W. W., Seiford, L. M., & Tone, K. (2019). Introduction to Data Envelopment Analysis and Its Uses. SpringerLink. https://doi.org/10.1007-0-387-29122-9

D'Ariano, A., et al. (2008). Reordering and local rerouting strategies to manage train traffic in real-time. Transportation Science, 42(4), 405-419. https://doi.org/10.1287/trsc.1080.0247

Dang, Q. V., et al. (2010). Low-level modeling for routing and scheduling trains through busy railway stations with expandable coupling/decoupling mechanism. International Journal of Modeling and Optimization, 2(6), 741-745. Available at: http://www.ijmo.org/index.php?m=content&c=index&a=show&catid=104&id=962

Diana, T. (2006). Benchmarking Airport Efficiency: An Application of Data Envelopment Analysis. Air Traffic Control Quarterly, 14(3), 183–202. https://doi.org/10.2514/atcq.14.3.183

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

Fadaei Foroutan, Z., & Bamdad, S. (2023). Efficiency measurement of railway passenger stations through network data envelopment analysis. Case Studies on Transport Policy, 11(1), 1-10. https://doi.org/10.1016/j.cstp.2021.11.011

Fried, H., Lovell, C.A.K., Schmidt, S. (Eds.), 1993. The Measurement of Productive Efficiency. Oxford University Press.

Garisi, G., & Cervelló-Pastor, C. (2020). Train-scheduling optimization model for railway networks with multiplatform stations. Sustainability, 12(1), Article 257. https://doi.org/10.3390/su12010257

Hashemian, S. M., & Triantis, K. (2023). Production pressure and its relationship to safety: A systematic review and future directions. Safety Science, 159, 106045. https://doi.org/10.1016/j.ssci.2022.106045

Harrod, S. S. (2012). A tutorial on fundamental model structures for railway timetable optimization. Surveys in Operations Research and Management Science, 17(2), 85–96. https://doi.org/10.1016/j.sorms.2012.08.002

Helms, M., Swaroop Vattam, & Goel, A. K. (2010). The Effect of Functional Modeling on Understanding Complex Biological Systems. https://doi.org/10.1115/detc2010-28939

Herrera-Restrepo, O., Triantis, K., Trainor, J., Murray-Tuite, P., & Edara, P. (2016). A multi-perspective dynamic network performance efficiency measurement of an evacuation: A dynamic network-DEA approach. Omega, 60, 45–59. https://doi.org/10.1016/j.omega.2015.04.019

Holvad, T. (2020). Efficiency analyses for the railway sector: An overview of key issues. Transportation Research Part E: Logistics and Transportation Review, 142, Article 105975. https://doi.org/10.1016/j.retrec.2020.100877

Iannone, F. (2012). The private and social cost efficiency of port hinterland container distribution through a regional logistics system. Transportation Research Part A: Policy and Practice, 46(9), 1424-1448. https://doi.org/10.1016/j.tra.2012.05.019

Jensen, P. R., Cruz, J., Eger, E. K., Hanchey, J. N., Gist-Mackey, A. N., Ruiz-Mesa, K., & Villamil, A. (2019). Pushing Beyond Positionalities and Through “Failures” in Qualitative Organizational Communication: Experiences and Lessons on Identities in Ethnographic Praxis. Management Communication Quarterly, 34(1), 121–151. https://doi.org/10.1177/0893318919885654

Khadem Sameni, M., Preston, J., & Khadem Sameni, M. (2016). Evaluating efficiency of passenger railway stations: A DEA approach. Research in Transportation Business & Management, 20, 33–38. https://doi.org/10.1016/j.rtbm.2016.06.001

Kroon, L. G., Edwin Romeijn, H., & Zwaneveld, P. J. (1997). Routing trains through railway stations: complexity issues. European Journal of Operational Research, 98(3), 485–498. https://doi.org/10.1016/S0377-2217(95)00342-8

Lai, Y., et al. (2010). Optimizing train network routing with heterogeneous traffic. Transportation Research Record, 2159(1), 9-16. https://doi.org/10.3141/2159-09

Larumbe, J. (2021). Evaluating Sustainability of Mass Rapid Transit Stations in Dubai. Sustainability, 13(23), 12965. https://doi.org/10.3390/su132312965

Levine, B., Lu, A., & Reddy, A. V. (2013). Measurement of Subway Service Performance at New York City Transit: Case Study with Automated Train Supervision Track-Occupancy Data. Transportation Research Record, 2353(1), 57-68. https://doi-org.ezproxy.lib.vt.edu/10.3141/2353-06

Li, X., et al. (2008). Efficient scheduling of railway traffic based on global information of train control. Transportation Research Part B: Methodological, 42(10), 1008-1030.

Li, X., Wang, Y., et al. (2013). Train routing model and algorithm combined with train scheduling. Journal of Transportation Engineering, 139(7), 717-726. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000573

Lindner, T., Pachl, J., & Nießen, N. (2021). Quality in rail timetabling – A detailed review of stakeholders' expectations. Case Studies on Transport Policy, 9(2), 936–946. https://doi.org/10.1016/j.jrtpm.2021.100282

Litherland, J., Calvert, G., Andrews, J., Modhara, S., & Kirwan, A. (2021). An Alternative Approach to Railway RAMS: Application of an Extended RAMS Framework to a UK Railway Corridor. Infrastructure Asset Management, 1–12. https://doi.org/10.1680/jinam.21.00003

Lu, J., et al. (2022). Train platforming and rescheduling with flexible interlocking mechanisms: An aggregate approach. Transportation Research Part E: Logistics and Transportation Review, 159, Article 102596. https://doi.org/10.1016/j.tre.2022.102596

Mahmoudi, R., Emrouznejad, A., Shetab-Boushehri, S.-N., & Hejazi, S. R. (2018). The origins, development and future directions of data envelopment analysis approach in transportation systems. Socio-Economic Planning Sciences, 100672. https://doi.org/10.1016/j.seps.2018.11.009

Makovsek, D., Bente, H., & Beck, A. (2015). Efficiency in railway operations and infrastructure management (OECD/ITF Discussion Paper No. 2015/09). International Transport Forum, OECD Publishing.

Medina-Borja, A., Pasupathy, K. S., & Triantis, K. (2007). Large-scale data envelopment analysis (DEA) implementation: a strategic performance management approach. Journal of the Operational Research Society, 58(8), 1084–1098. https://doi.org/10.1057/palgrave.jors.2602200

Merkert, R., et al. (2009). Benchmarking of train operating firms: A transaction cost efficiency analysis. Transport Reviews, 30(3), 369-390. https://doi.org/10.1080/03081060903429330

O'Donnell, C. J., Rao, D. S. P., & Battese, G. E. (2007). Meta-frontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical Economics, 34(2), 231–255.
https://doi.org/10.1007/s00181-007-0119-4

Oh, S., Chung, K., & Choi, J. (2023). Resource-Oriented Augmentation of a Train Timetable. IEEE Access, 11, 114283–114290. https://doi.org/10.1109/access.2023.3323590

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Journal of Clinical Epidemiology, 134, 178-189. https://doi.org/10.1016/j.jclinepi.2021.03.001

Rasmussen, J. (1997). Risk management in a dynamic society: a modeling problem. Safety Science, 27(2-3), 183–213. https://doi.org/10.1016/s0925-7535(97)00052-0

Richards, J., Golightly, D., & Palacin, R. (2024). The railway as a socio-technical system: Human factors at the heart of successful rail engineering. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. https://doi.org/10.1243/09544097JRRT78

Ropohl, G. (1999). Philosophy of Socio-Technical Systems. Techné: Research in Philosophy and Technology, 4(3), 186–194. https://doi.org/10.5840/techne19994311

Salido, M. A., Barber, F., & Ingolotti, L. (2012). Robustness for a single railway line: Analytical and simulation methods. Expert Systems with Applications, 39(18), 13305–13327. https://doi.org/10.1016/j.eswa.2012.05.071

Sameni, M., et al. (2016). Evaluating efficiency of passenger railway stations: A DEA approach. Case Studies on Transport Policy, 4(2), 132-144. https://doi.org/10.1016/j.cstp.2016.05.002

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research, 142(1), 16–20.

Siegel, A. W., & Jan Maarten Schraagen. (2014). Measuring Workload Weak Resilience Signals at a Rail Control Post. https://doi.org/10.1080/21577323.2014.958632

Songhori, M. J., A.M, L., & Rajabalinejad, M. (2020). A Multi-domain Approach Toward Adaptations of Socio-technical Systems: The Dutch Railway Case-Part 2. Data Archiving and Networked Services (DANS). https://doi.org/10.1109/sose50414.2020.9130531

Sheth, C., Triantis, K., & Teodorovic, D. (2006). Performance evaluation of bus routes: A provider and passenger perspective. Transportation Research Part E. https://doi.org/10.1016/j.tre.2005.09.010

Stoilova, S., Munier, N., Kendra, M., & Skrúcaný, T. (2020). Multi-Criteria Evaluation of Railway Network Performance in Countries of the TEN-T Orient–East Med Corridor. Sustainability, 12(4), 1482. https://doi.org/10.3390/su12041482

Summers, J. D., Eckert, C., & Goel, A. K. (2017). Function in engineering: Benchmarking representations and models. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 31(4), 401–412. https://doi.org/10.1017/s0890060417000476

Topcu, T. G., Triantis, K., & Roets, B. (2019). Estimation of the workload boundary in socio-technical infrastructure management systems: The case of Belgian railroads. European Journal of Operational Research, 278(1), 314–329. https://doi.org/10.1016/j.ejor.2019.04.009

Torraco, R. J. (2020). The integrative literature review as a research method. Human Resource Development Review, 19(3), 356–367. https://doi.org/10.1016/j.ssci.2021.105279

Vaneman, W. K. (2002). Evaluating System Performance in a Complex and Dynamic Environment. Virginia Tech.

Victorino, L., & Pena, J. (2023). The development of efficiency analysis in transportation systems: A bibliometric and systematic review. Sustainability, 15(13), Article 10300. https://doi.org/10.3390/su151310300

Wang, Y., et al. (2022). Train routing and track allocation optimization model of multi-station high-speed railway hub. Sustainability, 14(12), Article 7292. https://doi.org/10.3390/su14127292

Wang, L., An, M., Jia, L., & Qin, Y. (2019). Application of Complex Network Principles to Key Station Identification in Railway Network Efficiency Analysis. Journal of Advanced Transportation, 2019, 1–13. https://doi.org/10.1155/2019/1574136

Xu, X., et al. (2017). Train routing and timetabling problem for heterogeneous train traffic with switchable scheduling rules. Transportation Research Part B: Methodological, 106, 102-126. https://doi.org/10.1016/j.trb.2017.09.001

Xu, L., et al. (2022). Train operations organization in high-speed railway station considering variable configuration. Sustainability, 14(4), Article 2365. https://doi.org/10.3390/su14042365

Yu, M.-M., & Chen, L.-H. (2019). A meta-frontier network data envelopment analysis approach for the measurement of technological bias with network production structure. Annals of Operations Research, 287(1), 495–514. https://doi.org/10.1007/s10479-019-03436-3

Zakeri, G., & Olsson, N. O. E. (2018). Investigating the effect of weather on punctuality of Norwegian railways: a case study of the Nordland Line. Journal of Modern Transportation, 26(4), 255–267. https://doi.org/10.1007/s40534-018-0169-7

Zhou, L., et al. (2020). A MINLP method for simultaneous multi-periodic train timetabling and routing on a high-speed rail network. Sustainability, 12(3), Article 1131. https://doi.org/10.3390/su12031131

Websites:

Infrabel. (2023). Home page. Retrieved November 7, 2023, from https://infrabel.be/en

Infrabel. (2023). Punctuality. Retrieved November 7, 2023, from https://infrabel.be/en/punctuality

Infrabel. (2023). Not Spot. Retrieved November 7, 2023, from https://infrabel.be/fr/not-spot

NSF Award Search: Award # 2051685 - LEAP-HI: Safety and Learning from Errors and Near Misses in the Human-Automation Interaction of Socio-Technical Infrastructure Systems. (2024). Nsf.gov. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2051685&HistoricalAwards=false

The Brussels Times. (2023). Paying the price: Belgian train punctuality falls to record low. Retrieved November 7, 2023, from https://www.brusselstimes.com/830305/paying-the-price-belgian-train-punctuality-falls-to-record-low

World Bank: World Bank. (n.d.). Transport. Retrieved November 7, 2023, from https://www.worldbank.org/en/topic/transport

The Geography of Transport Systems: Rodrigue, J.-P. (n.d.). Economic impacts of transportation infrastructure. The Geography of Transport Systems. Retrieved November 7, 2023, from https://transportgeography.org/contents/chapter3/transportation-and-economic-development/economic-opportunities-transport-impacts/

Appendices

I. Double counting: 3-output and 4output models consistency

II. Peers and Lambdas of sample of DMUs in 4 classes (BCC-n and DDF)

Appendix I. Double counting: 3-output and 4output models consistency

(Answer to a concern about possibility of double-counting effect on these sets of models)

First of all, the nature of the #trains operated (count) variable (planned, coming from the timetable design and scheduling phase) and #non-canceled (actual, coming from the performance and operations reality) variables is different. Both of them are important to capture the operational realities, at the same time. For instance, Table i shows an example of three different train paths with various values. Considering only one of them would cause in loss of operational performance information.

Table i – Example of three DMUs with different values in two of the output variables

Variables	#count	#non-cancelled trains
DMU X	10	3
DMU Y	10	10
DMU Z	20	10

Furthermore, to evaluate whether overlapping output variables may introduce double counting, correlation coefficients between the 3-output and 4-output model results were calculated (Table ii).

Table ii - Correlations between two output variables and the efficiency scores of the 4-output with 3-the 3-output model results

Time of Day	**Correlation (Count vs. Non-Cancelled)**	**Correlation (BCC-n Efficiency 4-output vs. 3-output)**
Morning	0.1084	0.9848
Evening	0.1084	0.9376
Off-Peak	0.4353	0.7389
Weekend	0.0227	0.9264

High correlations were observed between efficiency scores in both BCC-n and DDF models (e.g., 0.9848 for Morning and 0.9376 for Evening), indicating that the inclusion of an additional output variable does not significantly distort the efficiency ranking. Furthermore, low correlation between the 'count' and 'non-cancelled' variables (e.g., 0.1084 for Morning and Evening) supports the notion that they capture distinct operational dimensions. These findings suggest minimal risk of redundancy or inflation in model outputs.

Appendix II – Peers and Lambdas of sample of DMUs in 4 classes (BCC-n and DDF)

EVENING:

BCC-n-4out → The most inefficient DMU: #100: L 15-i

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#100 = L 15-i		***104***	***60***	***3***	***10***	***2***	***4.16***
#17 (IC 07-i)	***0.280509136***	141	60	4	17	3	0.51
#45 (IC 19-2-ii)	***0.063785696***	30.5	59	3	7	4	0.17
#60 (IC 27-i)	***0.421514345***	95.5	60	4	15	3	0.68
#158 (L B5-1-i)	***0.234190823***	95	60	5	11	3	0.20

Lambdas and peers in DDF for #100 = L 15-i:

DDF	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#100 = L 15-i		***104***	***60***	***3***	***10***	***2***	***4.16***
#36 (IC 16-1-ii)	***0.053091975***	135	31	6	7.5	3	3.47
#44 (IC 19-2-i)	**0.440445585**	28	60	4	6.5	4	0.54
#74 (IC 35-i)	0.086100402	169	60	3	12	3	13.10
#139 (L B1-1-ii)	0.**420362038**	115	49	5	17	3	2.79

The most inefficient DMU in BCC-n-3output: #74 = IC 35-i

Here are the peers and lambdas from the 3-output model:

BCC-n-3out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#74 = IC 35-i		***169***	***60***	***3***	***12***	***3***	***13.10***

#31 (IC 14-i)	***0.25***	173	60	4	21	3	2.37
#139 (L B1-1-ii)	***0.75***	115	49	5	17	3	2.79

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#74 = IC 35-i		***169***	***60***	***3***	***12***	***3***	***13.10***
#31 (IC 14-i)	***0.5***	173	60	4	21	3	2.37
#45 (IC 19-2-ii)	***0.5***	30.5	59	4	7	3	0.17

delay ⇒ avg: 1.635539715 / stdev: 1.400596283

DMU #29: IC 13-i → the most inefficient DMU in DDF (4-out)

DDF	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#29 = IC 13-i		***75***	***60***	***2***	***6***	***0***	***1.15***
#36 (IC 16-1-ii)	***0.164215427***	135	31	6	7.5	3	3.47
#45 (IC 19-2-ii)	***0.260713152***	30.5	59	3	7	4	0.17
#139 (L B1-1-ii)	***0.217860544***	115	49	5	17	3	2.79
#175 (L C2-1-ii)	**0.357210877**	55.5	35	6	7	3	0.43

BCC-n eff_score = 0.9172 (in the lower half)

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#29 = IC 13-i ***(BCC-n Eff: 0.9172)***		***75***	***60***	***2***	***6***	***0***	***1.15***
#94 (L 12-i)	***0.88415518***	31	60	3	7	3	0.08

#114 (L 31-i)	***0.11584482***	21	60	3	3	3	0.06

(These two peers #94 and #114 are efficient only in BCC-n and are not efficient in DDF)

MORNING:

DMU #88 =L 10-ii – The most inefficient in BCC-n-4out only

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#88 = L 10-ii		***34.5***	***119***	***2***	***3***	***2***	***2.39***
#45 (IC 19-2-ii)	***0.304142012***	30	59	5	7	5	0.094
#93 (L 12-i)	***0.695857988***	31	60	3	7	3	0.007

DDF	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#88 = L 10-ii		***34.5***	***119***	***2***	***3***	***2***	***2.39***
#45 (IC 19-2-ii)	***1***	30	59	5	7	5	0.094

DMU #108 = L 26-ii – Most inefficient in BCC-n (4out) - BUT efficient in DDF: (this probably shows the impact of the radial vs. non-radial method!)

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#108 = L 26-ii		***24***	***60***	***3***	***2***	***1***	***2.9***
#45 (IC 19-2-ii)	***0.515463918***	30	59	5	7	5	0.094
#81 (L 04-i)	***0.053264605***	55	26.5	8	6.5	0	0.29
#182 (L L1-2-ii)	***0.431271478***	13	39.5	5	2	2	1.4

BCC-n-4out	***Eff score in 3-out***	duration	headway	#count	#stops	#non-cancelled	delay

#40 (IC 18-i)	*0.9932*	*137*	*43*	*5*	*15*	*2*	*2.9*
#75 (L 01-i)	*0.8435*	82	32	4	8	3	1.38
#113 (L 31-i)	*0.9914*	21	60	3	3	3	0.069
#137 (L B1-1-i)	*0.9517*	117	60	3	18	3	0.93

Comparing the eff scores of efficient DMUs in DDF only (but not in the BCC-n):

- It seems like DMUs, even with quite big delays (not very small delays), are being efficient in DDF. Which could be interesting.

Sample of some DMUs that are efficient in DDF, but not in BCC-n: #29: IC 13-i and #87: L10-i

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#29 = IC 13-i ***(BCC-n Eff: 0.753)***		***94***	***60***	***1***	***10***	***1***	***1.52***
#8 (IC 04-1-ii)	***0.311901089***	136	60	3	15	3	0.28
#60 (IC 27-i)	***0.472662275***	95	60	3	15	2	0.398
#93 (L 12-i)	***0.215436635***	31	60	3	7	3	0.007

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#87 = L 10-i ***(BCC-n Eff: 0.879)***		***43***	***120***	***1***	***3***	***1***	***0.879***
#37 (IC 16-2-ii)	***0.4***	20	210	1	2	1	0
#93 (L 12-i)	***0.6***	31	60	3	7	3	0.007

Examples of a DMUs that is eff in BCC-n, not in DDF: #6 = IC 03-ii and #75 = L 01-i

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#6 = IC 03-ii ***(BCC-n Eff: 1)***		***14***	***60***	***5***	***2***	***1***	***1.01***
#81 (L 04-i)	***0.023809524***	55	26.5	8	6.5	0	0.29
#182 (L L1-2-ii)	***0.976190476***	13	39.5	5	2	2	1.4

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#75 = L 01-i ***(BCC-n Eff: 1)***		***82***	***32***	***4***	***8***	***3***	***1.38***
#40 (IC 18-i)	***0.261744966***	*137*	*43*	*5*	*15*	*2*	*2.9*
#45 (IC 19-2-ii)	***0.711409396***	30	59	5	7	5	0.094
#81 (L 04-i)	***0.026845638***	55	26.5	8	6.5	0	0.29

<u>OFF-PEAK</u>:

More variety in the data range –> more differences between BCC-n 3-out and 4-out. The BCC-n-4out got 18 more eff DMUs compared to BCC-n-3out! All eff DMUs in 3-out are in 4-out too!

The most inefficient DMUs are the same in 3-out and 4-out models, both BCC-n and DDF!, with eff_score being the same in BCC and BCC-n and the same obj function value in DDF!

DDF 4-output model is similar to 3-output model. All the DMUs, the max obj function, and the most inefficient DMUs are the same.

The most in-eff DMU in both BCC-n: #87= L 09-i

BCC-n-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#87 = L 09-i ***(BCC-n Eff: 0.5649)***		***78***	***60***	***12***	***7***	***0***	***2.22***
#70 (IC 31-ii)	***0.276241199***	57	53	21	9	12	0.702

#78 (L 01-ii)	***0.182146298***	37	35.5	23	4.5	11	0.479
#125 (L 43-i)	***0.100416283***	30	60	11	7	11	0.159
#141 (L B1-2-ii)	***0.441196219***	119	53.5	23	19	12	0.68

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#87 = L 09-i	***(eff: 0.5649046)***	***78***	***60***	***12***	***7***	***0***	***2.22***
#156 (L B4-i)	***0.033333333***	107	60	13	18	10	1.30
# 166 (L B8-4-i)	***0.966666667***	77	60	15	14	12	2.01

#74: IC 33-i == the most in-eff DMU in both DDFs

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#74 = IC 33-i	***(eff: 0.81845)***	***96***	***60***	***12***	***8***	***0***	***0.92***
#95 (L 12-i)	**0.354784013**	31	60	10	7	10	0.01
#141 (L B1-2-ii)	***0.024202664***	119	53.5	23	19	12	0.68
#158 (L B5-1-i)	***0.621013322***	95	60	17	11	11	0.23

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#74 = IC 33-i		***96***	***60***	***12***	***8***	***0***	***0.92***
#156 (L B4-i)	***0.633333333***	107	60	13	18	10	1.30
# 166 (L B8-4-i)	***0.366666667***	77	60	15	14	12	2.01

The eff DMUs in BCC-n, but non-eff in DDF (a sample: IC 07-ii and L 12-i)

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#18 = IC 07-ii (BCC-n Eff: 1)		***140***	***51***	***20***	***17***	***12***	***0.69***
#2 (IC 01-ii)	***0.222222222***	180	60	14	11	13	1.35
#141 (L B1-2-ii)	***0.777777778***	119	53.5	23	19	12	0.68

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#95 = L 12-i (BCC-n Eff: 1)		***31***	***60***	***10***	***7***	***10***	***0.01***
#125 (L 43-i)	***0.978723404***	30	60	11	7	11	0.159
# 166 (L B8-4-i)	***0.021276596***	77	60	15	14	12	2.01

The eff DMUs in DDF, but non-eff in BCC-n (a sample: #75: IC 35-i, #126: L 43-ii)

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#75 = IC 35-i (BCC-n Eff: 0.951219)		***169***	***60***	***18***	***12***	***12***	***4.34***
#2 (IC 01-ii)	**0.128205128**	180	60	14	11	13	1.35
#19 (IC 08-i)	***0.487179487***	93	60	17	8	13	0.96
#141 (L B1-2-ii)	***0.384615385***	119	53.5	23	19	12	0.68

BCC-n-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#126 = L 43-ii (BCC-n Eff: 0.7629)		***93***	***60***	***17***	***8***	***13***	***0.96***
#125 (L 43-i)	**0.978723404**	30	60	11	7	11	0.159

#166 (L B8-4-i)	**0.021276596**	77	60	15	14	12	2.01

WEEKEND:

Raw data (the var range): the difference between this data frame and other times of the day data frames is noticeable. There are different sets of DMUs, relatively fewer than the other three. Reduced service (fewer frequency, fewer trains operated is visible. Less delay. Shorter train journeys.

All the 3-out results can be found in 4-out. 4-out got more:

- All the eff DMUs in 3 out are in 4out
- Same "the most inefficient DMU" in both.
- Same value for max (worst) objective function (DDF)
- Quite close eff scores in BCC-n models
- More efficient DMUs in 4-out.

Quick impression from comparing OFFPEAK and WEEKEND:

- The similarities between the 4-out and 3-out models: It shows that with more data (raw data to support), the result would be more converged/emerge.

The most in-eff DMU in both BCC-n: #72 = L 11-ii (which is also inefficient the most inefficient in DDF)

BCC-n-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#72 = L 11-ii		***109***	***120***	***9***	***7***	***6.5***	***3.00***
#8 (IC 04-1-ii)	***0.373181809***	139	60	17.5	15	16	0.88
#42 (IC 25-ii)	***0.11213934***	219.5	57.5	23.5	24	0	2.75
#43 (IC 27-i)	***0.127057954***	82	60	20.5	14	0	0.37
#106 (L B10-1-ii)	***0.387620896***	57	60	18	10	18	0.83

DDF-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#72 = L 11-ii		***109***	***120***	***9***	***7***	***6.5***	***3.00***
#8 (IC 04-1-ii)	***0.190211907***	139	60	17.5	15	16	0.88
#42 (IC 25-ii)	***0.224016145***	219.5	57.5	23.5	24	0	2.75

#106 (L B10-1-ii)	***0.585771948***	57	60	18	10	18	0.83

The most in-eff DMU in both DDFs: #71= L 11-i

BCC-n-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#71 = L 11-i		***108.25***	***120***	***8***	***6.75***	***5.5***	***1.21***
#30 (IC 19-2-ii)	***0.266592147***	28	60	15	5	15	0.07
#56 (IC 35-ii)	***0.345782309***	213	60	19.5	11	15.5	0.14
#116 (L B5-1-ii)	***0.387625544***	70	60	16.5	9	0	0.15

DDF-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#71 = L 11-i		***108.25***	***120***	***8***	***6.75***	***5.5***	***1.21***
#8 (IC 04-1-ii)	***0.092078708***	139	60	17.5	15	16	0.88
#42 (IC 25-ii)	***0.268920283***	219.5	57.5	23.5	24	0	2.75
#106 (L B10-1-ii)	***0.639001009***	57	60	18	10	18	0.83

The eff DMUs in BCC-n, but inefficient in DDF (a sample: #27: IC 17-i, #83: L 23-i)

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#27 = IC 17-I (BCC-n Eff: 1)		***45***	***60***	***20.5***	***6***	***0***	***0.11***
#34 (IC 21-ii)	***0.348984772***	29	60	17	8	0	1.36
#44 (IC 27-ii)	***0.209446038***	85	60	23.5	14	0	0.35
#61 (L 04-i)	***0.317369234***	31.5	34	30	4	0	5.14
#105 (L B10-1-i)	***0.124199956***	57	60	19	10	18	0.98

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#83 = L 23-i ***(BCC-n Eff: 0.99417)***		***13***	***60***	***18***	***3***	***0***	***0.29***
#34 (IC 21-ii)	***0.26142132***	29	60	17	8	0	1.36
#61 (L 04-i)	***0.010152284***	31.5	34	30	4	0	5.14 *
#120 (L B8-2-ii)	***0.728426396***	7	60	19	2	17.5	0.39

* Max delay in this weekend's data frame.

#108 = L B10-2-ii (Eff in BCC-n, but not in DDF) - it's other direction is exactly opposite. (#107 is eff in DDF, but not in BCC-n) = 107 is not the peer of 108! Even though they are so similar! (another difference of DDF radial methods!)

#108 = L B10-2-ii

DDF-4out	***lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#108 = L B10-2-ii (BCC-n Eff: 1)		***48***	***60***	***16***	***7***	***16***	***0.18***
#29 (IC 19-2-i)	***0.104477612***	28	60	15	6	15	0.36
#105 (L B10-1-i)	***0.776119403***	57	60	19	10	18	0.98
#120 (L B8-2-ii)	***0.119402985***	7	60	19	2	17.5	0.39

The eff DMUs in DDF, but inefficient in BCC-n (a sample: #7 = IC 04-1-i, #107 = L B10-2-i)

BCC-n-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#107 = L B10-2-I (BCC-n Eff: 0.89219)		***46***	***60***	***16***	***7***	***16***	***2.84***
#35 (IC 22-i)	***0.866666667***	52	60	18.5	9	18	0.95
#120 (L B8-2-ii)	***0.133333333***	7	60	19	2	17.5	0.39

BCC-n-4out	***Lambdas***	duration	headway	#count	#stops	#non-cancelled	delay
#7 = IC 04-1-i ***(BCC-n Eff: 0.99417)***		***136***	***60***	***18***	***15***	***15***	***1.09***
#8 (IC 04-1-ii)	***0.821536588***	139	60	17.5	15	16	0.88
#16 (IC 08-ii)	***0.0096083***	107	60	22	12	14.5	1.26
#42 (IC 25-ii)	***0.068637445***	219.5	57.5	23.5	24	0	2.75
#105 (L B10-1-i)	***0.100217667***	57	60	19	10	18	0.98

Recommendations that to investigate further:

- Investigating the physical/managerial/contextual/etc. interpretation of the directional vector in DDF.
- Good to ask for comparison from the railway system staff perspective, such as the operators, to see which result is more aligned based on their experience.