Improving transport demand forecasting: ex-post evaluation using smart card data

A case study of the Hoekse Lijn

MSc Thesis

F.B. (Fyodor) Onck

Delft University of Technology

TUDelft

Goudappel
MOBILITY MOVES US

Improving transport demand forecasting: ex-post evaluation using smart card data

A case study of the Hoekse Lijn

by

F.B. (Fyodor) Onck

MSc Transport, Infrastructure & Logistics

Delft University of Technology

Student number:	4888383
Chair TU Delft:	Dr. Ir. N (Niels) van Oort
Supervisor TU Delft:	Dr. (Jan Anne) Annema
Supervisor Goudappel:	Monica van Luven
Supervisor Goudappel:	Erik Oerlemans

Cover: RET Beeldbank, Joan van der Veen

TU Delft

Goudappel

MOBILITY MOVES US

Preface

This report contains my final thesis for obtaining the master's degree in Transport, Infrastructure and Logistics at Delft University of Technology. The research project was conducted in collaboration with Goudappel from February to July 2025. Throughout this project, I explored the potential of smart card data for evaluating and improving transport demand forecasts.

I would like to express my sincere gratitude to my graduation committee for their support and guidance. Jan Anne and Niels, thank you for your clear and constructive feedback, your flexibility in rescheduling meetings, and your continued support throughout the process. Monica, thank you for always responding so quickly to my questions, for your useful and constructive feedback, and for helping me find my way at Goudappel. Erik, thank you for your enthusiasm and the valuable substantive feedback and help you provided during the project.

F.B. (Fyodor) Onck
Delft, September 2025

Summary

Transport infrastructure investments represent some of the largest and most long-lasting public expenses, with the decisions that are made today influencing urban mobility for generations to follow. As cities all over the world face growing populations, climate targets and thereby the requirement for sustainable mobility solutions, accurate transport demand forecasts become increasingly important for effective infrastructure planning, resource allocation and policy evaluation (de Vries & Willigers, 2011). Otherwise this could lead to either insufficient capacity that constraints economic development or over-investment that wastes public resources.

Despite their importance, transport demand forecasts are rarely subjected to systematic ex-post evaluation, making it difficult to learn from past successes and failures, and to improve forecasting methodologies over time (Tempert et al., 2010). This represents a significant shortcoming in transport planning practice, where accountability for forecast accuracy is limited and systematic biases in models may go undetected across multiple projects (Flyvbjerg et al., 2005).

This thesis addresses this gap by using a systematic ex-post evaluation based on smart card data to investigate the accuracy of multimodal transport demand forecasts, with a focus on the Hoekse Lijn metro conversion project in the Rotterdam metropolitan area. The study provides a systematic, smart card data-driven ex-post assessment that compares forecasted and realised public transport demand. Using anonymised smart card transaction data from the *OV-chipkaart* system enables this research to offer a scalable and replicable approach to validating and improving demand forecasting methodologies. This contributes to a more accountable and data-informed planning process of transport projects. The research is guided by the main question:

How can multimodal transport models be improved to enhance the accuracy of future public transport demand forecasts?

A methodological framework was developed to guide the analysis, comprising four sequential steps: (1) demand reconstruction and exploration using smart card data; (2) forecast comparison through temporal and spatial indicators; (3) diagnostic analysis to trace the origins of discrepancies; and (4) synthesis of insights to inform model improvement. This framework was applied to the Hoekse Lijn case study, a former heavy rail line converted to metro operation, supported by detailed transport forecasts from 2011 and 2015. The demand reconstruction was based on already aggregated smart card data, requiring only limited preprocessing. The analysis involved applying correction factors to account for alternative payment methods. Key demand indicators, such as boardings per station, segment occupancies and total passenger kilometres, were derived. The latter two required additional processing using a secondary dataset. Forecast errors were assessed using quantitative metrics, including mean absolute error (MAE) and weighted mean absolute percentage error (WMAPE). To account for the structural impact of the COVID-19 pandemic, a correction factor was applied to adjust observed demand upwards to a pre-pandemic baseline, ensuring comparability with the original forecasts.

The results show that demand forecasts significantly overestimated actual ridership. Relative to the 2015 forecasts, weekday boardings were on average 25–30% lower after applying the COVID-19 correction factor. Eight out of ten stations showed underperformance in total weekday boardings, ranging from -9% to -40%. Westbound boardings were particularly overestimated, with most stations experiencing overestimates of more than 50%. In contrast, some stations performed near or above forecast levels: Maassluis Steendijkpolder came close to the projected figures, while Hoek van Holland Strand exceeded expectations, with an increase of 121% due to unexpectedly high levels of recreational travel.

While total station-level boardings were lower than forecasted, passenger kilometres were up by 18%, indicating that the average trip length was longer than anticipated. This discrepancy is explained by the high share of long-distance, non-commuting trips. For example, beach-related travel to Hoek van Holland Strand generated up to 27 times the normal weekday boardings during peak summer days. This

illustrates that the project's broader objective of improving leisure accessibility was achieved despite shortcomings in commuter demand projections.

Four systematic deviations were identified in the forecast:

- **Spatial asymmetry:** Easternmost stations showed the largest absolute errors in number of boardings.
- **Directional bias:** Westbound boardings were consistently overestimated across nearly all stations.
- **Trip purpose misalignment:** Commuter demand being overestimated and leisure travel being underestimated.
- **Trip length distribution change:** Average trip length were fundamentally underestimated, with a higher proportion of longer-distance trips and integration into the bigger Rotterdam metro network attracting travellers from outside the corridor.

These results point to a mixed outcome. Although the strategic target of a 52% uplift in weekday boardings was not achieved at most stations, the line did outperform on leisure travel and, as a result, total passenger-kilometres. At the same time, percentage errors (especially westbound, where base volumes are low) can exaggerate apparent deviations; for operational interpretation, absolute deviations are more informative. Taken together, the findings underline the difficulty of forecasting for corridors that must serve both regular commuters and irregular, weather-sensitive leisure users.

The systematic deviations also had implications that extended beyond the immediate outcomes of the project. Spatial asymmetry and directional bias resulted in suboptimal resource allocation, with capacity and service assumptions concentrated in areas that ultimately experienced lower-than-expected demand. Meanwhile, trip-purpose misalignment and longer-than-assumed journey times shifted the operational profile away from the anticipated commuter-oriented service towards longer recreational journeys, thereby influencing urban mobility patterns in ways that were originally not planned for.

The observed inaccuracies in the forecast were due to a mix of limitations in the model and external developments:

- **Overestimated service levels:** Forecasts included an additional peak-hour metro line that was never implemented and assumed a dense feeder bus network, both of which inflated perceived accessibility and demand.
- **Misaligned cost assumptions:** Outdated parameters related to value of time and distance costs did not align with actual traveller behaviour, leading to misrepresentations of generalised travel costs in the forecasts.
- **Outdated demographic forecasts:** Forecasts misjudged spatial patterns of population and employment growth, particularly outside Rotterdam.
- **Post-2015 policy shifts:** Parking fee increases and fare reforms influenced mode choice more than anticipated, especially due the rising relative costs of public transport compared to other modes.
- **Pandemic effects:** COVID-19 led to lasting reductions in commuting and a rise in long-distance leisure trips, notably to Hoek van Holland Strand.
- **E-bike adoption:** Growth in e-bike use diverted medium-distance travellers from metro access trips, impacting mode competition.

After applying the −12% structural correction for the effects of the pandemic, systematic attribution analysis (see figure 5.8) shows that approximately 60% of the total residual deviation on an average weekday can be explained by identifiable factors. These include network representation (30%), which is driven by the assumed frequencies of Line A and a feeder bus network that is more extensive than that delivered. Other factors include behavioural and economic parameter misalignment (20%), which is due to time and distance cost factors that are anchored in 2010, despite the fact that public transport costs have risen relative to car travel. There are also socioeconomic inputs (5%), which reflect employment that is concentrated in Rotterdam rather than along the corridor. Finally, there is unmodelled competition from e-bikes (10%) for trips of between 5 and 15 km. These effects are partly offset by stronger-than-forecast leisure demand at Hoek van Holland Strand (-7%). The remaining 40% is best understood as interaction effects (factors reinforcing each other), normal post-opening ramp-up, seasonality and data/measurement limitations, rather than a single missing cause.

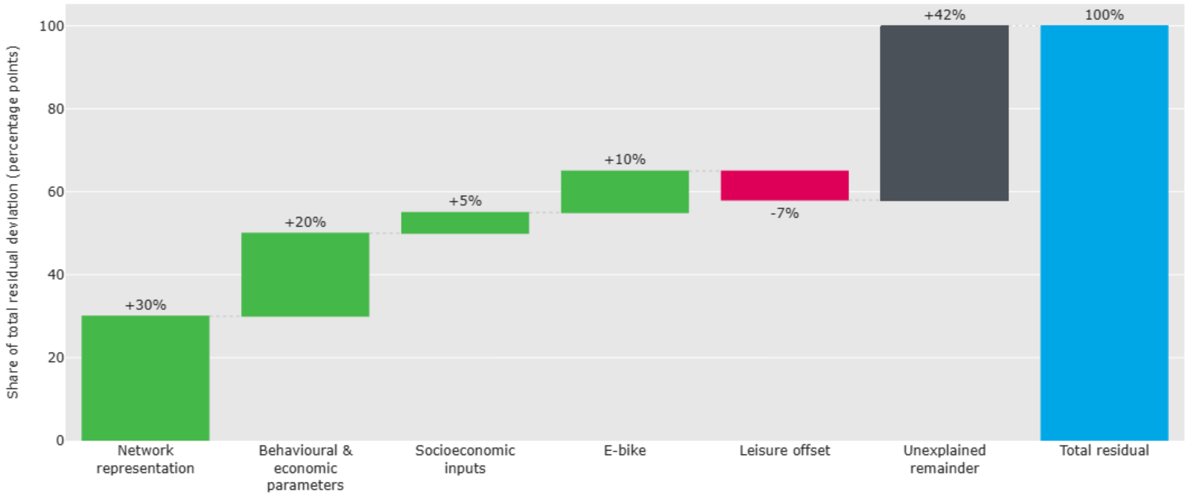


Figure 1: Waterfall graph showing attribution to the total residual forecast deviation.

Beyond the findings of the case study, the broader implication is that forecasts should treated as indicative rather than as fixed targets as forecasts are inherently uncertain and prone to bias. Their primary value is to indicate direction and scale and to support comparisons between alternatives, rather than providing guaranteed outcomes. Forecasts should therefore be communicated in outcome ranges linked to specific scenarios and should reflect heterogeneous demand. On that basis, this study proposes four concrete strategies:

- **Systematic ex-post validation:** Establish standardised continuous ex-post validation schemes using using automated data sources, such as smart card data, to create feedback loops between forecasting practice and observed outcomes, going beyond traditional one-time evaluations.
- **Adaptive parameter calibration:** Implement adaptive parameter management that regularly updates behavioural and economic parameters using real-time data sources, addressing the systematic overestimations that caused by outdated assumptions in the model.
- **Scenario-based uncertainty management**: Adopt scenario-based forecasting methodologies that systematically explore uncertainty across major assumptions, providing probability-based outcome ranges rather than single-point estimates that are more vulnerable to external disruptions like COVID-19.
- **Integrated network planning**: Implement network planning that evaluates how new infrastructure could reshape existing service networks and assesses financial sustainability of service assumptions.

Implementing these changes requires a shift in institutional culture, moving from “predict-and-forget” to “predict-and-learn”. Ex-post evaluations should be mandated and published as part of the funding conditions, turning forecasting into a continuous cycle of planning, evaluation and improvement, helping to mitigate the effects of optimism bias. Data management is fundamental in this process. Ideally, a governmental-led database would provide open, aggregated indicators to support routine validation. For each major project, the project owner, relevant operators and regional authorities should form a small, temporary working group to agree on service scenarios and confirm matching budgets before forecasts are finalised. Finally, roles must be clearly defined, such as who initiates evaluations, who updates parameters and who coordinates across the different agencies. This ensures that continuous learning becomes and integral part of forecasting practice, rather than being viewed as an occasional activity.

This study underscores the value of smart card data as a tool for ex-post validation. The continuous, system-wide nature of this dataset enables a granular understanding of temporal and spatial travel patterns. However, several limitations persist: incomplete coverage of user segments due to alternative payment methods, lack of direct insight into trip purpose or traveller demographics, and the challenge of inferring demand indicators using a secondary model to evaluate the accuracy of the original forecast outputs.

This research demonstrates that systematic ex-post evaluation using smart card data can significantly enhance the accountability and accuracy of transport demand forecasting by revealing systematic biases that would otherwise remain hidden in planning practice. The Hoekse Lijn case study illustrates how traditional forecasting approaches, while achieving some strategic objectives such as improved leisure accessibility, can substantially overestimate commuter demand due to outdated model parameters, unrealistic service assumptions, and inability to anticipate external disruptions such as e-bike adoption and pandemic-induced behavioural changes. The four-step methodological framework developed in this study provides transport planners with a replicable approach for continuous model validation and improvement, while the proposed strategies offer concrete pathways toward more robust forecasting methodologies. As transport systems face increasing complexity from technological innovation, changing mobility preferences, and unexpected societal disruptions, the integration of automated data sources with systematic evaluation frameworks becomes essential for evidence-based infrastructure investment decisions. This study contributes to a more mature and accountable transport planning discipline, where forecasting accuracy is continuously monitored, systematic biases are transparently identified, and lessons learned are systematically incorporated into future practice, ultimately supporting more informed and sustainable transport infrastructure development.

Contents

List of Figures

List of Tables

1 Introduction

Transport demand forecasts play an important role in infrastructure planning, investment decisions, and public transport operations (de Vries & Willigers, 2011). Multimodal transport models are widely used to estimate ridership for new public transport connections. However, despite their importance, transport demand forecasts are often subject to significant deviations, with actual ridership figures eventually differing from projections (Flyvbjerg et al., 2005). Furthermore, ex-post evaluations of transport demand forecasts remain uncommon, making it difficult to systematically assess their accuracy (Tempert et al., 2010).

The need for reliable transport forecasts had become increasingly important as cities continue to expand and prioritise sustainable mobility solutions. Overestimated demand can result in inefficient allocation of resources and financial losses, while underestimation can lead to insufficient capacity and suboptimal infrastructure planning. As a result, improving the accuracy of forecasting models is an important issue in transport planning (Flyvbjerg et al., 2005; Tempert et al., 2010).

1.1. Research gaps

Despite the extensive use of transport demand forecasts in infrastructure planning, research has shown that these forecasts frequently suffer from systematic inaccuracies, particularly in rail projects. Large-scale studies such as those by Flyvbjerg et al. (2005) indicate that rail is demand is often overestimated, with an average overestimation of 106%, leading to substantial financial and operational consequences. Similarly, research on Dutch projects like RandstadRail and the Noord/Zuidlijn has found that forecast models often rely on unrealistic socio-economic assumptions, contributing to deviations between projected and actual ridership (Bojada, 2014; Brands et al., 2020). These inaccuracies persist despite advancements in forecasting methodologies, highlighting a need for more systematic validation.

A major limitation in forecasting research is the scarcity of systematic ex-post evaluations - assessments that compare forecasted ridership with actual usage after project implementation. While some retrospective studies exist, they rely heavily on manual passenger counts or survey data, which are often expensive, time-consuming, and limited in scope (Brands et al., 2020; Hussain et al., 2021).

One of the most promising but underutilised resources for ex-post validation is smart card data, which provides a continuous, large-scale record of real-world passenger movements (Dixit et al., 2024; Pelletier et al., 2011; Van Oort et al., 2015a). While smart card data has already been widely used for OD estimation, route choice analysis, and operational planning, its application in long-term transport forecasting validation remains largely unexplored (Dixit et al., 2024; Tempert et al., 2010). Additionally, data accessibility issues, privacy concerns, and incomplete check-in/check-out data pose methodological challenges that limit its full potential (Bakker, 2019; Vignetti et al., 2020).

Another significant limitation in forecasting accuracy stems from the difficulty of incorporating external influences into demand models. Traditional models assume stable travel behaviour based on socio-economic projections, yet real-world ridership is shaped by policy interventions, urban planning decisions, and technological advancements. Research has shown that policies such as parking restrictions, cycling infrastructure improvements, and congestion pricing can significantly impact public transport demand, yet these effects are often not systematically accounted for in forecasting models (Kolkowski & de Boer, 2023).

The COVID-19 pandemic further exposed the rigidity of conventional models, as long-term shifts in commuting patterns, remote work adoption, and modal preferences led to significant forecasting errors (Gkiotsalitis & Cats, 2021). While some studies have explored post-pandemic mobility trends, there is still limited research on how such disruptive events should be integrated into long-term forecasting methodologies.

While forecasting inaccuracies have been documented in research, there is a lack of systematic, data-driven ex-post evaluations that quantify and explain deviations between projected and actual ridership using empirical smart card data.

1.2. Research objectives

To address the research gap as outlined in the previous section, this study aims to assess the accuracy of past transport forecasts and identify factors contributing to deviations between projected and actual ridership. Therefore, the primary objectives of this research are:

1. Evaluate the effectiveness of smart card data for ex-post validation.
2. Assess the forecast accuracy of existing the transport demand forecasts.
3. Investigate the causes of forecasting discrepancies.
4. Propose improvements for future multimodal forecasting methodologies.

To address the research objectives that are outlined above, this study uses the conversion of the Hoekse Lijn as relevant case study to explore this gap in research, providing an opportunity to analyse how past projections compare with observed passenger behaviour and to assess how external factors may have influenced demand. Using *OV-chipkaart* smart card data, this research aims to bridge the gap by introducing a systematic approach to validating forecast accuracy, distinguishing between model error and external influences, and proposing improvements for future multimodal forecasting methodologies.

To support the decision to convert the Hoekse Lijn from a railway connection to part of the Rotterdam metro network, transport demand forecasts were conducted in 2006, 2011, and 2015 (Goudappel Coffeng, 2011, 2015). Now that the metro has been operational for several years, it presents an opportunity to assess how well these projections align with actual ridership levels. Studies on comparable projects, such as RandstadRail, have revealed significant discrepancies between forecasted and observed demand, raising the question whether similar deviations exist for the Hoekse Lijn. Indentifying these discrepancies and understanding their causes can provide valuable insights for improving forecasting models and supporting better-informed transport planning decisions.

1.3. Research questions

To reach the objectives of this research to systematically analyse the deviations in transport demand models and propose improvements to forecasting methods, using the Hoekse Lijn as a case study, the following research question will be answered:

How can multimodal transport models be improved to enhance the accuracy of future public transport demand forecasts?

To answer the main research question, the following sub-questions are composed:

1. How can *OV-chipkaart* smart card data be used to evaluate the transport demand for the Hoekse Lijn, and what are its limitations?
2. How does the actual transport demand on the Hoekse Lijn compare to the forecasted values from 2011 and 2015, and what systematic deviations can be identified?
3. To what extent do discrepancies between forecasted and observed demand stem from biases or structural limitations within the transport models?
4. What external factors, such as socioeconomic developments, policy changes, and the COVID-19 pandemic, contributed to deviations between forecasted and actual transport demand?

1.4. Research approach

This study will use a variety of methods to evaluate the accuracy of transport demand forecasts for the Hoekse Lijn and determine the causes of any discrepancies. The methodology will integrate a literature review, data-driven analysis and expert consultation in order to evaluate discrepancies in the forecasts.

First, a review of forecasting accuracy studies and best practices in smart card data analysis will establish the theoretical context. Exploratory data analysis (EDA) will then be conducted to assess data quality and identify patterns or inconsistencies. Time-series visualisation and error metrics will then be used to compare forecasted and observed ridership, highlighting any systematic biases.

Next, observed demand will be systematically compared to original forecasts using visual and statistical techniques, with cluster analysis being applied to assess accuracy across different station typologies. Any deviations will be explained through an analysis of the assumptions underlying the model, distinguishing between model limitations and external factors. This will be supported by a literature review and expert consultation.

This research starts with a literature review on forecast accuracy evaluation, smart card applications and external factors influencing transport demand in chapter 2. Chapter 3 describes the methodology that is used to evaluate the forecast accuracy. Next, chapter 4 provides the empirical context of the Hoekse Lijn and analyses observed ridership patterns, while chapter 5 presents the comparative analysis of forecasted and observed demand and explores the causes of discrepancies. Chapter 6 concludes with a synthesis of findings, discussion and recommendations for improving transport forecasting practices.

2
Literature Review

This literature review examines the current state of knowledge regarding the accuracy of transport demand forecasting, focusing particularly on the use of smart card data for systematic forecast evaluation and on external factors that influence forecasting performance. It addresses three research areas that are essential for improving transport forecasting practices: (1) Existing approaches to evaluating forecast accuracy in transport models, including methodological frameworks and systematic biases identified in previous studies; (2) the emerging role of smart card data as a tool for ex-post evaluation. This section highlights the advantages of smart card data over traditional data collection methods, as well as the methodological challenges it presents; and (3) the influence of external factors on public transport demand, with particular attention to the COVID-19 pandemic and the emerging adoption of electronic bicycles.

These three topics are included as they directly align with the research's aim and research questions: (1) defines how forecasts accuracy will be judged and where systematic bias typically arises; (2) assesses the suitability of smart card data as the basis for ex-post evaluation and outlines the necessary processing steps; and (3) identifies the external factors that could explain some of the difference between forecasts and actual outcomes, enabling the analysis to distinguish between errors in the model and external effects.

For the literature research, Google Scholar, ScienceDirect, and the TU Delft Repository were used to identify studies on transport demand forecasting accuracy, ex-post evaluations, smart card data applications, and external effects. The following keywords were used in various combinations: *smart card/OV-chipkaart data, ex-post evaluation, transport/traffic forecast accuracy/reliability, RandstadRail, OD matrix, effects COVID-19.*

2.1. Evaluation of forecast accuracy of transport models

Despite their widespread use, transport demand forecasts are rarely subject to ex post evaluation, making their actual accuracy often uncertain (Tempert et al., 2010). However, large-scale studies, such as those by Flyvbjerg et al. (2005), have shown systematic overestimations, especially in rail transport forecasts. Unlike weather forecasts, which can be quickly tested against real-world data, transport forecasts often project demand 15 years or more in the future, making direct validation a long-term challenge. This limitation means that the applicability and reliability of transport models are often assumed rather than subjected to strict verification.

Statistical validation methods are used to quantify forecast accuracy. Common metrics include Mean Absolute Percentage Error (MAPE) and Percent Root Mean Squared Error (PRMSE) to summarise deviations between predicted and observed values. These range from simple ratio-based calculations, such as computing the actual observed value as a proportion of the forecasted value and subtracting one, to more complex statistical measures, each providing different insights into forecast deviations between predicted and observed values (Macfarlane, 2024).

One of the most consistent findings in the evaluation of transport forecasts is the tendency to overestimate demand, particularly for new infrastructure projects. Tempert et al. (2010) analysed forecasts made in the 1990s and found that, on average, transport models overestimated traffic intensities, with discrepancies being much larger for inner-city roads than for main roads. The main cause of these inaccuracies was found to be overly ambitious socioeconomic development assumptions, such as overestimated new housing projects and employment growth.

Research by Flyvbjerg et al. (2005) provides a broader statistical analysis of the accuracy of transport demand forecasts. Using a dataset of 210 major transport projects across 14 countries, the study found:

- Rail forecasts are particularly prone to overestimation: in 90% of rail projects passenger demand was overestimated, with an average overestimation of 106%.
- Road forecasts show a more balanced distribution of discrepancies, with similar rates of over- and underestimation. However, for 50% of all road projects, the difference between forecasted and actual traffic was more than ±20%.
- The accuracy of forecasts has not improved over the last 30 years, despite claims that forecasting techniques have advanced.

These findings suggest that forecasting inaccuracies are not just the result of technical limitations, but may also be the consequence of political and strategic biases. Many projects face political pressure to justify investments, leading to systematically inflated demand estimates. This effect is especially apparent in rail projects, where strong political or ideological motivations, such as encouraging a modal shift from cars to public transport, can lead to deliberate overestimation of demand (Flyvbjerg et al., 2005).

A KiM (*Kennisinstituut voor Mobiliteitsbeleid*) study strengthens this argument by analysing four major Dutch urban transport projects, including RandstadRail and the Noord/Zuidlijn. The results show that forecasting discrepancies did not consistently lead to over- or underestimations, but rather varied depending on changes in project scope, economic conditions and model assumptions (Kolkowski & de Boer, 2023). Interestingly, the study also highlights that while spatial development assumptions were often too optimistic, other factors such as parking restrictions, improved cycling infrastructure and policies to discourage car use played a larger than expected role in shaping actual ridership. This suggests that transport models may not fully capture behavioural responses to urban policy changes, leading to deviations between forecasts and reality.

In response to these systematic issues, several academics and practitioners advocate the use of bandwidths and scenario-based forecasting rather than single-point estimates. Both Tempert et al. (2010) and Bojada and Clerx (2014) emphasise the importance of providing realistic ranges of high and low demand to account for uncertainty in future developments. Furthermore, risk management approaches, as suggested by Bojada and Clerx (2014), could improve the robustness of transport models by explicitly incorporating uncertainties related to spatial and economic development.

2.2. Smart card data as a tool for ex-post evaluation

Ex-post evaluations of transport forecasts are instrumental in assessing the accuracy of forecasting models and improving their reliability for future transport planning. In recent years, smart card data, such as the *OV-chipkaart* system in the Netherlands, has emerged as a valuable resource for conducting these evaluations, providing insights into travel behaviour that traditional survey-based methods cannot match. Unlike manual passenger counts or surveys, smart card data enables large-scale validation by providing real-world passenger movements and multimodal travel patterns (Dixit et al., 2024; Pelletier et al., 2011; Van Oort et al., 2015b). However, despite these advantages, the use of smart card data in forecasting validation faces challenges related to data accessibility, trip interference, and behavioural assumptions (Bakker, 2019; Vignetti et al., 2020).

One of the uses of smart card data is origin-destination (OD) matrix estimation. Hussain et al. (2021) discusses recent advances and future challenges in OD matrix estimation, emphasising the importance of data cleansing and transfer detection in ensuring accurate trip reconstruction. Data cleansing is critical to correct for discrepancies caused by equipment failure and human error, which can affect up to 2% of transactions (Translink, 2016). Besides filtering out missing or faulty records, setting upper and lower limits for travel time and distance helps identify discrepancies caused by factors such as unrealistic trip durations, fare evasion, or incorrect tap-ins (Hussain et al., 2021). The paper goes on to describe the necessity of transfer detection in public transport systems where passengers tap in for each boarding, as it helps distinguish between single trips and multi-trip journeys. Effective transfer detection algorithms distinguish between true intermodal transfers and activities that may indicate trip-end locations.

Expanding on this, Fu and Gu (2018) illustrate how OD matrix estimation can also be used to assess travel time changes following the introduction of a new metro line. Their study of the Nanjing metro system shows how smart card data can be used to quantify the redistribution of passenger flows across multiple stations and whether the new line has successfully reduced congestion on previously congested corridors.

Other ex-post studies have demonstrated the value of smart card data in validating route choice models and travel behaviour assumptions, particularly in assessing the impacts of service frequency changes and network modifications through empirical scenario testing. For example, Van Oort et al. (2015a) applied smart card data in The Hague to refine demand modelling and evaluate service adjustments, demonstrating how data-driven simulations can support decision-making in public transport planning. Similarly, Dixit et al. (2024) evaluated the predictive performance of a multimodal route choice model in Amsterdam before and after the introduction of the *Noord/Zuidlijn* metro line. Their findings illustrate the challenges of applying existing models to new conditions, showing that while smart card data can validate general demand trends, deviations at the individual route level remain difficult to predict. This suggests that ex-post validation should not only focus on aggregate demand levels but also assess how travellers adapt their routes and transfer behaviour in response to new infrastructure, as existing models may not fully account for these behavioural adjustments. By incorporating these individual-level adaptations, model accuracy and applicability to future scenarios can be improved.

Building on these insights, Brands et al. (2020) contribute to the discussion by demonstrating how smart card data can be used for ex-post evaluations of public transport network changes. Their study on the introduction of the Noord/Zuidlijn in Amsterdam highlights how automated data sources enable the quantification of ridership changes, modal shifts, and variations in travel times across the network. By comparing smart card transactions before and after the metro line's introduction, they reveal a 4% increase in network-wide ridership, largely driven by shifts from tram and bus to metro. Moreover, their findings underscore that besides a reduction in travel times and improved reliability for some passengers, others were confronted with increased tranfers times or longer journeys due to network restructuring. This is in line with Dixit et al. (2024), who found that while smart card data effectively captures aggregate demand trends, individual route choices remain difficult to predict due to behavioural adaptations. The evaluation of the *Hoekse Lijn* can follow a similar approach, using smart card data to assess whether expected demand shifts occur, how travel behaviour adapts, and how network-wide effects like induced demand or transfer penalties play out. Additionally, assessing changes in travel time reliability, as seen with the *Noord/Zuidlijn*, is key to understanding the broader service quality impacts.

Fu and Gu (2018) further highlight the importance of travel time reliability as a metric in evaluating new metro lines. Their study shows that, besides ridership fluctuations, the introduction of new infrastructure can change the consistency of travel times across the network. Using smart card data, they quantified how the opening of Line 4 of the Nanjing Metro reduced travel times for certain OD pairs, while increasing variability for others. This could particularly be attributed to shifts in transfer locations and station congestion. Such findings underscore the importance of assessing not only travel time variability but also broader service quality factors that influence passenger behaviour.

In addition to travel time reliability, smart card data enables the ex-post analysis of service quality impacts on the number of passengers. Van Oort et al. (2015b) examines how comfort and capacity constraints affect travel behaviour, highlighting that traditional transport models often overlook comfort effects, despite their influence on passenger choices. Their analysis shows that frequency increases not only reduce waiting times, but also improve comfort, leading to 20–30% higher predicted passenger growth than models without comfort effects predict. This suggests that fluctuations in travel time consistency, as observed by Fu and Gu (2018), may be linked to broader service quality improvements that shape passenger preferences and overall network demand. These findings emphasise the value of smart card data in capturing both behavioural responses to service changes and deviations from forecast demand estimates.

Furthermore, smart card data can also be used in assessment of the impact of policies. Wang et al. (2018) uses smart card data to analyse the effects of transit fare changes in the Beijing Metro, showing how demand elasticities vary across different passenger groups and times of day. Comparing the results to an ex-ante evaluation shows the significant exaggeration in passengers' responses to fare increases coming from a stated preference survey. Simirlarly, Bakker (2019) emphasises the importance of smart card data for justifying investment decisions and ex-post evaluations. By providing insights into actual travel behaviour, smart card data improves demand forecasting models and ensures better project selection. It also enables a network-wide evaluation of new transit projects, distinguishing between genuine ridership growth and demand shifts, which supports long-term policy assessments.

Finally, ex-post evaluations using smart card data can be integrated with cost-benefit analysis (CBA) to assess long-term project impacts. Vignetti et al. (2020) proposes a methodology combining retrospective CBA with qualitative stakeholder analysis to provide more comprehensive evaluation of transport investments. While traditional CBAs often rely on projected benefits, smart card data allows for real-world validation of passenger growth and accessibility improvements. This approach aligns with a broader goal of transport forecasting: not just to predict demand but to ensure that investments achieve their intended societal benefits.

2.3. Smart card data compared to traditional methods

Smart card data offers several advantages over traditional data collection methods such as travel surveys, manual counts, and GPS tracking, though it also has some limitations. Its broad coverage and scale are key strengths: each fare transaction generates a data point, resulting in large sample sizes across the entire system on a daily basis. This passive data collection occurs continuously and automatically, in contrast to costly surveys that are conducted infrequently (sometimes only once every few years) with limited sample sizes (Lee et al., 2014). Traditionally, data was limited to one-day manual counts or small-scale surveys, but now system-wide, daily OD flows and trends can be continuously monitored. Transport for London highlights that fare data, combined with inferred missing pieces, provides a comprehensive view of trips across the network, enabling effective planning and impact analysis of service changes, insights that were previously only accessible through expensive manual surveys (TfL, 2015).

However, traditional methods offer details that smart cards may lack. Travel surveys collect demographics (age, income, etc.) and trip purpose information, which smart card data do not contain inherently (Bagchi & White, 2005). Also, surveys cover all travel modes (car, bike, on foot), providing a complete picture of the entire journey, while smart card data only reflects the public transport leg of a journey. Manual counts and automatic passenger counters on vehicles can sometimes record the total amount of passengers, including those who might be missed by the smart card system, such as passengers who try to evade paying or those who still use paper tickets (Hussain et al., 2021). GPS-based tracking of volunteer participants (via smartphone apps or dedicated devices) can provide very high-resolution trajectories and mode detection for all trips, but these tend to involve much smaller sample sizes and participant burden (Marra et al., 2022).

2.4. Limitations and considerations of smart card data

As partially already shown in table 2.1 comparing smart card data to traditional methods, while smart card data is extremely useful, it presents some major limitations and biases that must be considered. Data privacy and accessibility issues are the primary concerns (Bakker, 2019). Because smart card data can potentially trace and individual's movements, data is often anonymised or aggregated before analysis. Privacy principles (e.g., data minimisation and use limitations) often restrict using personally identifiable trip sequences (Fan & Chen, 2018). Even anonymised data can raise concerns when combined with other datasets, as noted by Dempsey (2015), 'The greater the data base, and the more extensively it is correlated with other data bases, the less privacy the individual enjoys.'

Another limitation, that also followed from the comparison with traditional methods, is the absence of trip purpose in smart card data. Knowing *why* people travel is important for policy decisions (e.g., commuting trips might respond differently to fare changes or service cuts than leisure trips). Without purpose, forecasts can only be checked on volumes, not on whether the mix of trip types match the expectations. Researchers are experimenting with machine learning methods to infer purposes, such

Table 2.1: Advantages and disadvantages of smart card data compared to traditional methods

Advantages	Disadvantages
Large scale and frequency - continuous data from virtually all users, enabling granular temporal analysis (hourly, daily, seasonally)	**Limited trip context** - no direct information on trip purpose or socio-demographics.
Spatial detail on transit network - precise data on stations or routes used, facilitating accurate OD matrices and transfer analysis that surveys often approximate.	**Partial journey visibility** - only captures public transport legs of the journey; first/last mile and non-PT trips are unobserved.
Reduced respondent bias - automatically collected data avoids issues of survey under-reporting	**Potential data gaps** - missing records due to technical failures, missed check-ins/check-outs, or fare-evading passengers.
Cost efficiency - once the system is in place, data is gathered as a by-product of operations, whereas surveys and counts require dedicated effort and funding	**User coverage bias** - certain user groups might be under-represented, e.g., tourists using single-use tickets, or very occasional riders.

as (Liu et al., 2018) who describe a method of using a naïve Bayes probabilistic model, however, this still seems to be an emerging field of research.

Representation biases in smart card data are also an important aspect to consider, meaning that data may not perfectly represent all user segments or travel patterns. For instance, cash-paying passengers, paper ticket users, tourists, and very occasional riders may be under-represented or absent in the data. Mahajan et al. (2022) emphasise that smart card data ”might not be fully representative of public transport behaviour, since some users do not own or regularly use a smart card”. This can be minimised, as many cities have done, by making smart cards nearly mandatory, but tourists and occasional users might still use single-fare tickets. Even among smart card users there can be biases, for example, some people may carry multiple cards or share the same card, which can confuse individual travel pattern analysis. For example, a single person might own separate cards, one for private trips and one for business trips. This illustrates the possible inability of smart cards to cleanly map unique persons. As a result, any analysis at the individual level (e.g. inferring personal travel habits) can be distorted.

Because of these biases, analysts often strengthen smart card data with rider surveys or manual counts (Mahajan et al., 2022). Surveys can capture those who pay by cash or rarely use public transport, providing the extra context that automated fare data lacks. In The Netherlands, for example, the importance of maintaining ticket alternatives alongside the *OV-chipkaart* has been highlighted to ensure inclusivity (Durand & Zijlstra, 2020). So, while smart card data provide a lot of information in a quick and cheap way, the limited picture of the passenger population must be acknowledged and the over representation of frequent users smart card against ’invisible’ riders like tourists, occasional travellers, or those sharing cards must be accounted for.

There is also the fact that smart card data only reflects actual transit usage; it does not capture latent demand or people who did not travel because of inconvenience. If the forecasts considered some car drivers would switch to using public transport, but they did not, the smart card data would not give any insights about those missing passengers, as it only shows who did use public transport. The Australian Transport Assessment guidelines point out that passive data sources like smart card counts show ”the actual behaviour that is occurring” but do not give any information about those unable or unwilling to travel. If a bus is overcrowded and some riders are left behind, or if poor service quality leads to someone to take a private mode of transport instead, the smart card data only record the passengers who did board. To understand latent demand, additional data sources are required, such as travel surveys, interviews, or stated preference studies that reach non-users (Australian Transport Assessment and Planning, n.d.).

2.5. Data processing and preparation steps

Before it can used for analysis or model validation, smart card data needs to be processed. The first step is data cleaning and validation to ensure the accuracy of the data. The smart card data can contain errors due to software bugs, hardware malfunctions, or user mistakes (e.g., missing check-ins/check-outs, duplicate entries, or implausible timestamps). Different studies describe the causes of erroneous records and developed filters to remove or correct flawed records (Fan & Chen, 2018).

Robinson et al. (2014) proposes a framework that outlines basic checks on raw data, identification of bad check-in / check-out data, aggregation of rides into complete trips, and flagging of any remaining faulty data. In practice, this might involve steps like: removing obviously invalid entries (e.g., negative travel times, or duplicate check-ins), merging consecutive ride segments into a single journey, and matching transaction timestamps with schedule data to identify inconsistencies.

Another important processing step is transfer detection. A single passenger journey may involve multiple check-ins and check-outs (for each vehicle or mode change), which need to be recognised as a single journey. Typically, rules are applied such as: if the same card has another check-in within e.g. 30 minutes of checking out, consider it a transfer rather than a new independent journey. These rules can be refined with knowledge of fare policies (e.g., free transfer windows) and network topology (to ensure that the second check-in is on a line reasonably connected to the first). The output of this stage is a set of complete trips (origin and destination, with any intermediate transfers merged) for each traveller (Hussain et al., 2021).

In addition, White et al. (2010) describes the importance of aggregating the data spatially and temporally to produce indicators that are needed for analysis. Depending on the application, trips might be aggregated to the zone level or station level to compare with modal zones, and to time periods (peak hours, daily totals, etc.). Common spatial-temporal aggregation products include: Origin-Destination matrices, route load profiles, and hourly ridership profiles.

Several studies also demonstrate the use of interquartile range (IQR) filtering to remove outliers in smart card data. For example, Tian et al. (n.d.) applied the standard 1.5×IQR rule to travel times in the Singapore MRT dataset, filtering out extreme values likely caused by missed tap-outs or system errors. This statistical filter removed approximately 5.3% of records, significantly improving data quality for their metro crowding prediction model.

A related approach can be found in Wood (2015), which also applies IQR-based filtering to smart card–derived journey times in the Hong Kong MTR. While Wood does not follow the conventional 1.5×IQR rule, the study removes outliers by defining trip durations as outlier when they fall significantly outside the interquartile range, within each OD pair and time interval. This approach is designed to reflect realistic passenger variability while excluding distortions that could bias reliability estimates. The cleaned data are then used to compute the Individual Reliability Buffer Time (IRBT), a metric for passenger-experienced travel time reliability.

Dixit et al. (2019) describes a more tailored and advanced cleaning methodology, focusing specifically on metro journeys in multimodal networks. Theirs procedure addresses abnormal travel times caused by odd passenger behaviour (e.g., taking the wrong train, extended platform waiting). They propose a two-step method to remove extreme records while preserving genuine large disturbances. For each OD pair, records are first flagged as outliers based on a threshold deviation value from the median, whereafter, for each flagged record, overlapping journeys are checked to ensure that naturally long journeys are not mistakenly removed.

2.6. External factors influencing public transport demand

Beyond model limitations, transport demand is significantly influenced by external factors that are often not caputred in traditional forecasting mdoels. These factors can create substantial deviations beween forecasted and observed ridership, particularly during periods of sudden societal change (ITF, 2021).

2.6.1. Impact of COVID-19 on transport forecasting

The COVID-19 pandemic disrupted transport forecasting as traditional models based on stable demand patterns struggled to account for the sudden and unprecedented changes in mobility behaviour. Demand for public transport fell sharply as a result of lockdowns, social distancing measures and the widespread shift to remote working. These disruptions have created new challenges for forecasting accuracy and ex-post evaluation of transport models (Gkiotsalitis & Cats, 2021).

The primary effect of the pandemic was a significant drop in public transport ridership, with 80-90% declines in major cities during the lockdown periods (Marra et al., 2022). While demand has partially recovered, long-term behavioural changes are still evident, particularly in the form of reduced commuting and a shift towards private and active modes of transport. Data from the Netherlands show that public transport ridership has not fully returned to pre-pandemic levels, partially due to the continued popularity of remote and hybrid working (KiM, 2024).

Another challenge introduced by COVID-19 is the non-linear recovery of transport demand. Unlike previous disruptions, where travel demand returned to predictable patterns, the pandemic has created ongoing uncertainty in transport behaviour. Gkiotsalitis and Cats (2021) identify three major forecasting difficulties in the post-pandemic context:

- Shifts in demand fluctuations, where traditional peak hours have been weakened due to flexible working arrangements.
- Changes in mode preferences, with increased reliance on cycling, walking and the use of private vehicles.
- A reduced willingness to use public transport, due to continuing concerns about hygiene and crowding.

The *Kennisinstituut voor Mobiliteitsbeleid* incorporates structural behavioural changes caused by the COVID-19 pandemic into its transport forecasts using calculated adjustment ranges based on a model that projects from a 2018 base year. These effects are integrated by adjusting the alternative-specific constants within the model, which are used to account for shifts in traveller preferences across different transportation modes.

The KiM Mobility Report 2023 highlights the structural nature of post-pandemic changes in travel behaviour. According to this report, nearly 20% of former public transport users expect to use it less frequently than before the pandemic, while only 6–9% expect to use it more often. These shifts, driven by increased homeworking, digital conferencing, home schooling, and modal shifts to alternative modes of transport, have led to an estimated 7–15% decline in bus, tram, and metro demand, with a baseline scenario suggesting a permanent 12% reduction in distance travelled (KiM, 2023).

Additionaly, the model incorporates the *constant travel time budget*, which suggests that while working from home reduces commuting trips, individuals may reallocate the saved time towards other travel purposes, such as leisure activities or errands. This compensation effect could result in an overall stable or even increased amount of travel times amongst all modalities, despite reductions in work-related travel. Faber et al. (2023) support this approach by highlighting how post-pandemic structural changes have led to significant decreases in commuting trips, particularly affecting public transport use. Their research estimates a decrease in distances travelled by train (-3% to -9%) and by bus, tram, and metro (-1% to -5%), with smaller effects on car travel (-1% to -5%) and potential increases for walking and cycling due to compensatory leisure trips.

2.6.2. E-bike adoption and modal substitution

The widespread adoption of electric bicycles is another significant external factor influencing demand for public transport. In the Netherlands, 20% of the population owned an e-bike by 2021, with 3% purchasing one annually (Huang et al., 2024). This mode of transport directly competes with public transport for medium-distance trips, as Huang et al. (2024) states that e-bikes are most competitive for trips in the 5 to 15 kilometre range, with average e-bike journeys being 5.9 kilometres as compared to 3.6 kilometres for conventional bicycles.

The influence on public transport ridership has been researched by Sun et al. (2019) who found that it leads to a clear decline in public transport use across most trip distances and purposes in the Netherlands. Following the adoption of an e-bike, the proportion of trips made by public transport compared to other modes decreased from 10.8% to 2.0% for journeys of 5–10 km, and from 12.5% to 8.0% for journeys of 10–15 km. Similarly, modal split data by trip purpose shows a substantial decline in public transport use for commuting (from 14.7% to 11.4%), shopping (from 16.2% to 5.3%), and leisure activities (from 19% to 15.0%) following the adoption of e-bikes. These results highlight the growing tendency for e-bikes to substitute public transport, particularly for medium-distance trips and everyday activities.

2.7. Conclusion

This literature review confirms that smart card data has become a valuable tool for ex-post evaluation, providing detailed insights into actual passenger flows. The potential of smart card data to support the systematic validation of transport demand forecasts is particularly relevant given the current lack of such evaluations. However, the reviewed studies also highlight important methodological challenges, including privacy constraints, incomplete transaction records and a lack of information on trip purpose or traveller demographics.

These challenges directly informed the data processing strategy adopted in this thesis. Several safeguards were implemented in response, including outlier detection and correction factors for alternative payment methods, as well as validation using supplementary model outputs. The literature also provided a theoretical basis for identifying and interpreting external explanatory factors, such as the increase in e-bike usage and the ongoing effects of the pandemic, which are considered alongside model-based explanations when analysing forecast deviations.

The methodological choices in the next chapter build on these insights, ensuring a rigorous use of smart card data and a contextual understanding of potential demand shifts.

3 Methodology

The objective of this thesis is to present a replicable, modular framework for diagnosing discrepancies in multimodal transport demand forecasts, identifying their causes, and proposing improvements for future modelling practices for public transport projects. This methodology chapter outlines a case-agnostic framework for conducting ex-post evaluations of multimodal transport demand forecasts using smart card data. The aim of this approach is that it can be applied to any public transport project with comparable data conditions. This chapter presents this methodology that consist of four main stages: (1) demand reconstruction and exploration, (2) forecast comparison, (3) diagnostic analysis of discrepancies and (4) model improvement. This approach is applied to the Hoekse Lijn case.

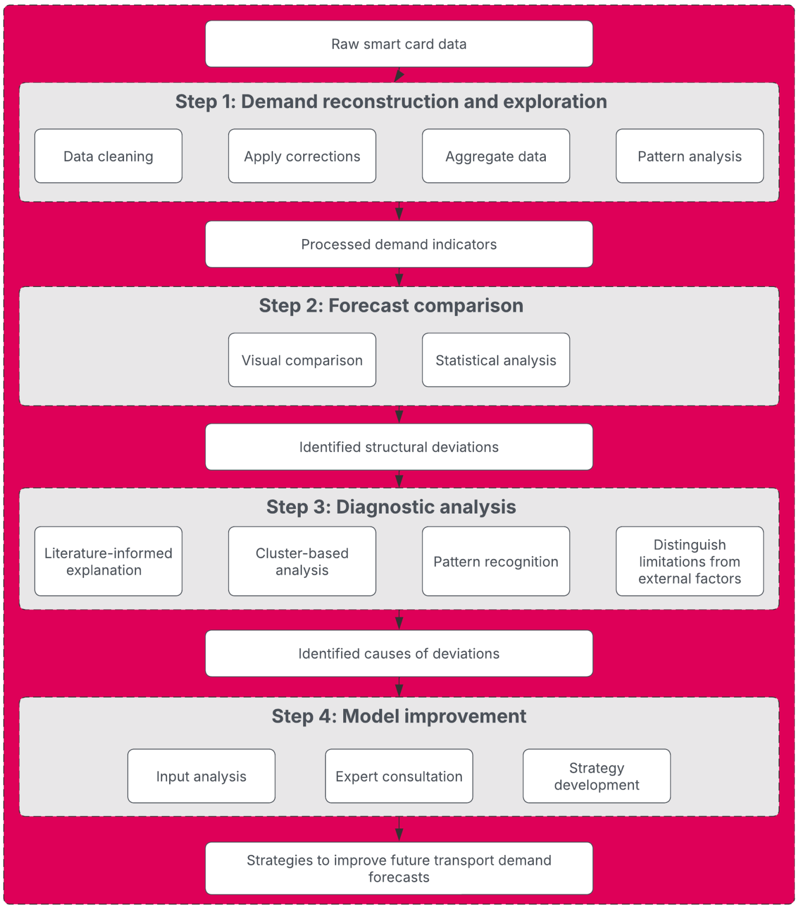


Figure 3.1: Four-step methodological framework for evaluating and improving transport demand forecasts

3.1. Data sources

Due to the high spatiotemporal granularity, scale and objectivity in capturing passenger behaviour, smart card data was used as the empirical backbone for this evaluation. This data typically consist of (anonymised) card IDs, timestamps, boarding and alighting locations, and transport modes.

However, smart card data lack key attributes that are commonly obtained through traditional data collection methods such as household travel surveys. Specifically, they provide no direct information on traveller demographics (e.g., age, income) or trip purposes (e.g., commuting, leisure). Additionally, smart card systems only capture the public transport segment of multimodal journeys, neglecting walking, cycling, or private car trips. Gaps might also exist in the data for passengers that are missed by the system, such as passengers with missed check-ins or check-outs (e.g., due to user error or fare evaders) and passengers who still use paper tickets.

To account for these shortcomings, the analysis used additional data sources including monthly correction factors for the use of mobile/bank card payments and socioeconomic data.

3.2. Evaluation approach

A structured, four-step evaluation approach is applied to assess the accuracy of transport demand forecasts and identify the causes of any discrepancies. Firstly, observed demand is reconstructed and analysed using smart card data to establish a reliable empirical basis. Secondly, the observed demand is compared with the forecasted figures using visual and statistical techniques across various spatial and temporal aggregations. Thirdly, discrepancies are analysed to distinguish between model limitations and external influences, providing insight into how forecasting practices can be improved.

3.2.1. Step 1: Demand reconstruction and exploration

Data processing steps

The first step involved importing and cleaning the data. Trips with invalid attributes should be removed, including those with missing stop names or check-in/out times, negative travel times, and records where check-in and check-out stops are identical.

Although the detailed data cleaning steps outlined below were not applied to the data used in this research, as it was already aggregated upon delivery, they are included here to serve as a general guideline for comparable studies using disaggregated smart card records. This methodological overview allows the study to be replicated in future and provides a basis for interpreting the reliability and limitations of the observed demand patterns.

To clean the data of these anomalies, a two-step outlier detection method based on Dixit et al. (2019) can be applied. This OD-specific approach is designed to preserve genuine service disruptions while filtering out implausible journeys.

A more common method used to clean smart card data of anomalies is the interquartile range method (IQR). Both the two-stage method described above and the IQR method were applied on a small dataset of smart card data with metro trips in Rotterdam on one day in February 2019. Both methods produced similar descriptive statistics after the filters, however the IQR approach flagged and removed a significantly larger portion of the dataset (3.49%) compared to only 0.26% for the two-stage method. Given the small relative difference in the resulting distributions and the risk that IQR filtering method might remove valid but uncommon trips, the more conservative two-stage method is preferred.

The two-stage method is applied like this:

- **Step 1: Duration based flagging** – A journey is flagged as a potential outlier if its travel time exceeds a context-specific threshold. For example, Dixit et al. (2019) apply a threshold of twice the OD-specific median journey plus an additional 15 minutes. This step captures trips that are clearly atypical compared to others on the same OD route.
- **Step 2: Overlap verification** – Each flagged trip is then assessed for the presence of overlapping trips made by other passengers on the same OD pair. If another passenger made the same trip within the time window of the long trip and their check-out time was more than 10 minutes before the check-out of the flagged trip, the flagged trip is confirmed to be an outlier and removed. The

10-minute threshold is based on the typical headway of metro services and should be adjusted accordingly.

Since not all public transport trips are captured in smart card data, increase factors were be applied to correct for underreported trips, such as those made with paper tickets, alternative payment methods, or unrecorded due to fare evasion. These factors can be based on administrative statistics, survey data, or external counts.

Following cleaning and correction, the smart card data should be aggregated into forms that are suitable for comparison with the forecast outputs. The specific form of aggregation depends on the structure of the forecasts that are being evaluated. Where possible, aggregation should match the spatial and temporal resolution of the original forecasts in order to be able to make a meaningful comparison. As mentioned earlier, the data for this research was already aggregated, but if this is not the case the following aggregation should be considered:

- If forecasts are provided at the station level (e.g., average daily boardings), observed data are aggregated by stop and day type.
- If forecasts include OD matrices, smart cards trips are aggregated into origin-destination pairs using inferred transfers.
- If segment-level or vehicle load profiles are forecasted, smart card data must be aggregated to estimate onboard occupancy across links and time periods.
- If forecasts are expressed in passenger-kilometres, OD flows are multiplied by segment distances to compute comparable totals.

Demand pattern analysis

In this phase, descriptive statistics, distribution plots and trend analyses were used to gain initial insights, detect anomalies and understand temporal usage patterns. The aim was to identify station typologies and assess the completeness and usability of the dataset. This analysis established the current demand levels by:

- temporal usage pattern identification (hourly, daily, seasonal);
- station hierarchy and typology development based on usage characteristics.
- assessment of data quality and completeness.
- special event analysis (particularly weather-dependent recreational travel).

3.2.2. Step 2: Forecast comparison

Once observed demand has been established, it was compared with forecasted figures through a combination of visual and statistical techniques, structured around the indicators that are used in the forecasts. Depending on the available aggregation level available in the forecasts, this comparison was done on station, segment, or OD-pair level, and across different time horizons. Time-series visualisations were used to evaluate daily and seasonal trends, and structural deviations.

To quantify deviations, standard error metrics were applied:

- Mean Absolute Error (MAE) for absolute discrepancy:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

  where y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the total number of observations.

- Mean Absolute Percentage Error (MAPE) for relative accuracy:

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

  This metric expresses the error as a percentage of the observed value.

- Percentage Error to assess direction and scale of deviations for individual observations:

$$\text{Percentage Error}_i = \frac{y_i - \hat{y}_i}{y_i} \times 100\%$$

Positive values indicate overestimation, negative values indicate underestimation.

The selection of which metric to use depends on its intended use: MAE is useful for operational implications, as it expresses the average forecast error in actual units, making it intuitive for resource planning. On the other hand, MAPE provides a scale-independent measure, useful for comparing relative errors across stations or segments with varying demand levels, but may become unreliable when actual values are near zero (Kim & Kim, 2016).

In the Hoekse Lijn analysis, Weighted Mean Absolute Percentage Error (WMAPE) addresses the issue where low-volume stations could disproportionately influence overall accuracy metrics. By weighting errors according to actual ridership volumes, WMAPE provides a more meaningful assessment of forecast performance for the cluster-based evaluation.

3.2.3. Step 3: Diagnostic analysis of discrepancies

The third step involves identifying and explaining the causes of observed discrepancies between forecasted and actual transport demand. These deviations may result from structural limitations in the forecasting model or from external, unforeseen influences. This phase integrates a combination of exploratory techniques and literature review to find plausible causes of the identified deviations.

Literature-informed attribution

Literature was used to contextualise deviations by comparing them with known systematic forecasting issues in comparable projections. For example, studies have shown that demand models often overestimate commuting flows due to idealised assumptions on socio-economic development, while underestimating leisure-related or induced demand effects (Brands et al., 2020; Flyvbjerg et al., 2005). Additionally, the literature was used to construct informed methods for quantifying external influences and handling of data limitations. For example, it grounded the correction factor that was used to compensate for structural effects of the pandemic and helped by contextualising modal shifts due to rising e-bike demand in the Netherlands.

Cluster analysis

Clustering techniques are applied to group stations based on three specific behavioural characteristics derived from smart card data to identify systematic patterns in forecast accuracy across different station typologies. The clustering variables are defined as follows:

Peak-hour usage ratio (P_i): The proportion of daily boardings at station i occurring during peak periods (07:00-09:00 and 16:00-18:00), calculated as $P_i = \frac{B_{peak,i}}{B_{total,i}}$, where $B_{peak,i}$ represents boardings during peak hours and $B_{total,i}$ represents total daily boardings.

Seasonality indicator (S_i): The coefficient of variation of monthly average weekday boardings for station i in 2024, calculated as $S_i = \frac{\sigma_{month,i}}{\mu_{month,i}}$, where $\sigma_{month,i}$ is the standard deviation and $\mu_{month,i}$ is the mean of monthly boardings.

Daily variability indicator (D_i): The coefficient of variation of hourly boardings at station i on an average weekday in February 2025, calculated as $D_i = \frac{\sigma_{hour,i}}{\mu_{hour,i}}$, where $\sigma_{hour,i}$ is the standard deviation and $\mu_{hour,i}$ is the mean of hourly boardings.

These three variables create a feature vector $\mathbf{x}_i = [P_i, S_i, D_i]$ for each station i, enabling the identification of stations with similar temporal usage patterns, seasonal variations, and daily demand distributions.

For this research, k-means clustering is used to partition the stations into k clusters by minimizing the within-cluster sum of squares: $\min \sum_{j=1}^{k} \sum_{\mathbf{x}_i \in C_j} |\mathbf{x}_i - \boldsymbol{\mu}_j|^2$, where C_j represents cluster j and $\boldsymbol{\mu}_j$ is the centroid of cluster j. This unsupervised algorithm is particularly effective for identifying repeated spatial and temporal patterns in smart card data and distinguishing between frequent and infrequent usage profiles, as demonstrated in several public transport studies using smart card data(Agard et al., 2006; Kieu et al., 2013). The algorithm iteratively assigns each station to the cluster with the nearest centroid and updates centroids until convergence, enabling systematic grouping of stations with similar behavioural characteristics for targeted forecast accuracy analysis.

3.2.4. Step 4: Model improvement

The final step of the approach focusses on translating the findings from the earlier steps into concrete suggestions for improving future forecasting practices. This includes the integration of sensitivity analysis and expert consultation.

Model input analysis

A systematic evaluation of the original model inputs was conducted to assess their influence on demand outcomes and identify potential sources of forecast deviations. This analysis was informed by the findings from the earlier methodology steps, where systematic over- and underestimation patterns were identified.

In this case, it turned out not to be feasible to re-run the original forecasting model with new data. However, for models that can be accessed and updated, re-running them with corrected inputs can provide insights into possible systematic biases or misestimations. Alternatively, as model access is limited, a qualitative assessment of original assumptions was conducted. The sensitivity analysis focused on three main areas:

- Transport network configuration: Evaluating whether the modelled transport network, including service frequencies, network topology, and transfer connections, accurately reflected the system post-implementation.
- Behavioural parameters and policy settings: Parameters that influence travel choice behaviour within the model - such as car ownership, monetary cost coefficients, value of time, and perceived penalties such as transfer resistance and parking costs – are evaluated based on publicly available data.
- Socioeconomic forecasts: Projected socioeconomic data within the model, such as number of inhabitants, employment figures, number of students, and retail area, that are used to calculate trip production/attraction between different zones, are reviewed against current data from public sources.

The goal of this input evaluation method was to understand how outdated or inaccurate assumptions may have affected forecast outcomes and identify which parameters have disproportionate effects on the accuracy of projections. Using this systematic assessment of model inputs against empirical evidence, areas where improved parameter estimation or calibration could help to enhance future forecasting reliability were highlighted.

Expert Consultation

To complement the technical assessment from the model input analysis, the research findings were discussed with two experts in the field of transport modelling and operations. These consultations were designed as semi-structured interviews but evolved into open conversations, allowing for in-depth exploration of the identified deviations and their practical implications.

The goal of these expert consultations was primarily to validate the findings from the model input analysis. The experts were able to contextualise the observed deviations within the broader experience of the industry, highlighting how factors such as service changes, evolving travel behaviour and external shocks like the COVID-19 pandemic have influenced forecast accuracy. Additionally, their perspectives helped identify further considerations, such as the impact of emerging mobility trends and the importance of robust scenario planning, enriching the overall interpretation and recommendations of the study.

The following experts were interviewed. The transcripts of the interviews can be found in appendix B.

Expert 1: Jeroen Henstra (RET)
Role: Operations and Planning, Rotterdam public transport operator

Expertise: Operational experience with the Hoekse Lijn conversion and practical knowledge of ridership patterns
Focus: Operational validation of findings and network-specific insights

Expert 2: Dr. ir. A.J. (Adam) Pel (TU Delft)
Role: Academic researcher in transport modelling
Expertise: Theoretical foundations of transport demand modelling and forecasting accuracy
Focus: Methodological validation and broader implications for transport forecasting practice

3.3. Application to the Hoekse Lijn case study

This methodological framework is applied to the Hoekse Lijn case study to evaluate the accuracy of transport demand forecasts made in 2015 for the train-to-metro conversion project. Using OV-chipkaart smart card data from 2020-2025, the four-step evaluation approach systematically compares observed demand patterns with the original RVMK3.1 model projections, accounting for external disruptions including COVID-19 impacts and evolving travel behaviour. The following chapters present the empirical application of this methodology: Chapter 4 establishes the case study context and analyses observed ridership patterns, while Chapter 5 presents the comparative analysis of forecasted versus actual demand and explores the systematic causes of identified discrepancies.

4
Empirical context: understanding demand on the Hoekse Lijn

This chapter establishes the empirical basis for assessing the accuracy of transport demand forecasts for the Hoekse Lijn, which serves as a case study to find an answer on how transport forecast models can be improved. First, it outlines the strategic objectives that motivated the line's conversion from a regional railway to a metro service. Then, it provides an overview of the corridor and its role within the Rotterdam public transport network. Using smart card data, the chapter analyses observed ridership patterns in spatial and temporal detail. This offers initial insights into how the line is used in practice, and indicates to what extent it aligns with policy goals. These insights will form the basis for the comparative analysis in the next chapter.

4.1. Strategic objectives of the Hoekse Lijn Conversion

The decision to convert the Hoekse Lijn from a regional rail line to a metro line was based on multiple strategic objectives, including improving regional accessibility, increasing ridership, and optimising the integration of the line within the Rotterdam metropolitan area (Gemeente Rotterdam, 2018). While some of the set objectives are broad in scope, several of them were translated into explicit quantitive expectations that will be evaluated in this case study.

One of the main objectives of the conversion was to increase public transport ridership by providing a more frequent and reliable service that would seamlessly integrate into the RET network. Demand forecasts estimated a 52% increase in average boardings on the Schiedam Nieuwland - Hoek van Holland Strand section and a 28% increase in total passenger kilometres compared to a scenario where the existing rail service is continued. These figures, derived from the last forecast using the RVMK3.1 model (Goudappel Coffeng, 2015), will serve as the most important indicators for evaluating the success of the project and its forecasts.

A second objective was to optimise the use of the line throughout the day and both directions, addressing the strong peak/off-peak imbalance and directional asymmetry that characterised the former rail service (Gemeente Rotterdam, 2015). Before the conversion, passenger flows where concentrated in the morning towards Rotterdam and in the evening in the opposite direction, while the rest of the day the line saw only limited usage. By converting the Hoeke Lijn into a metro service with higher frequencies and direct integration into the RET network, it was expected that the line would become more attractive for a wider range of journeys. This broader use was intended to make better use of the corridor's capacity and improve its financial performance. While broader spatial development goals were also pursued by the addition of Maassluis Steendijkpolder, the primary focus of the project was enhancing existing travel flows, improving the quality and safety of station facilities, and increasing accessibility to make the service more appealing throughout the day.

Another objective was to improve access to Hoek van Holland Strand by extending the metro line to a new station location directly at the beach. In the old situation, the terminus was located 1.2 kilometres inland, meaning passengers had to walk the last part to reach the beach. The extension aimed to reduce the total travel time from the centre of Rotterdam by 26% and increase the attractiveness of the beach as a recreational destination (Gemeente Rotterdam, 2015). According to the 2015 projections, this improvement of accessibility was expected to lead to a 10-20% increase in beach visitors arriving by public transport (Goudappel Coffeng, 2015). In addition, the extension was intended to contribute to the long-term development of Hoek van Holland as a four-season destination, supporting tourism, local economic activity, and housing development (Gemeente Rotterdam, 2015).

This case study assesses these objectives though an ex-post analysis of actual ridership figures using smart card data. The evaluation focuses on three indicators that were also used in the 2015 forecasts (Goudappel Coffeng, 2015):

- Occupancy at arrival: Assessing forecasted versus actual load factors at stations to determine if passenger distribution was accurately predicted.
- Boarding passengers: Identifying deviations in expected versus actual passenger entries per station and analysing demand shifts over time.
- Total passenger kilometres: Comparing projected and observed passenger travel distances to evaluate network-wide demand accuracy.

By comparing these indicators to the original forecasts, this research aims to get a better understanding of the degree to which the policy objectives were achieved and how accurately they were projected by the forecasting models.

4.2. Forecasting history

Before the conversion of the Hoekse Lijn to a metro service, multiple forecast were made to support planning and investment decisions. The intial forecasts were done in 2006 and later updated in 2011 and 2015 (Goudappel Coffeng, 2011, 2015). The 2015 update formed to basis for the eventual planning of the project and is evaluated in this study.

The 2015 projection used the RVMK 3.1 model, which simulated multimodal demand across the Rotterdam region. This model estimated transport demand based on expected socio-economic developments, including population growth, employment levels, and educational enrolment across production and attraction zones, combined with assumptions about the availability and attractiveness of various transport modes.

Several alternative scenarios were evaluated with the model and formed the basis of the 2015 forecasts:

- **Base year (2010):** This scenario depicts the Hoekse Lijn as an NS train service in 2010, with the existing spatial and transport infrastructure that was in place at that time. It served as a calibration year to aligh the model with observed traffic counts and measured flows.
- **Reference scenario (2025):** This scenario projects ridership if nothing is changed on the Hoekse Lijn, assuming that it continues as a regional train service without any operational changes. However, it does take into account broader regional developments and infrastructural changes.
- **Conversion scenario (2025):** In this alternative, the train service is replaced by a metro service that connects Hoek van Holland Strand and Nesseland (Line B) with higher frequencies and full integration into the RET network. THe existing Line A is extended eastward during peak hours. A new station, Maassluis Steendijkpolder, is also added.
- **Conversion + extension scenario (2025):** This final alternative includes all elements of the previous metro conversion scenario, plus an additional westward extension of the line with the relocation of Hoek van Holland Strand next to the beach.

The model's input assumptions included:

- **Service levels:** Metro line B would run three times per hour, and during peak hours, an extended metro line A would serve Vlaardingen West six times per hour.
- **Modal split:** The model predicted a shift from car and bicycle to metro, particularly in Maassluis and Vlaardingen, with a 12.3% increase in public transport usage for inter-municipal trips.
- **Seasonal demand:** Recreational travel to Hoek van Holland Strand was not included in the model due to its irregular and weather-dependent nature. The alternative method used to include this demand is explained in section 4.4.4.

The model is described in more detail in section 5.3, where the exact model input and assumptions are compared and evaluated.

4.3. Case study background: the Hoekse Lijn corridor

4.3.1. Rotterdam and its public transport network

Rotterdam and the surrounding areas are served by a public transport network of metro, tram and bus lines operated by the RET. This network covers Rotterdam and neighbouring municipalities (e.g., Capelle aan den IJsel, Schiedam, Vlaardingen etc.). In total, RET operates 5 metro lines, 9 tram lines, and around 58 bus lines. The metro system alone has 71 stations across the lines A to E.

RET operates under concessions that are granted by the Rotterdam-The Hague metropolitan transport authority (*Metropoolregio Rotterdam Den Haag - MRDH*). Notably, there are two transit concessions: "Rail Rotterdam" which encompasses the metro and tram services, and "Bus Rotterdam" which encompasses the bus services, both currently held by RET. The RET is owned by the municipality of Rotterdam and the MRDH.

Figure 4.1: Overview of the Rotterdam metro network

The RET follows the national *OV-chipkaart* smart card system, which uses a stored-value smart card for all transit. Since 2023, passenger can also use their bank cards to check-in and out of public transport. The fare is composed of a base boarding tariff plus a distance-based charge (€1.12 + €0.171/km). If passengers transfer to another tram, bus or metro operated by RET within 35 minutes after checking out, no new base fee is charged. RET also offers day passes and 2-hour passes. These can be loaded on an OV-chipkaart or bought in the RET app.

The layout of the Rotterdam metro network is characterised by a radial structure, with all lines passing through the city centre in east-west or north-south direction (specifically station Beurs), branching out to urban areas surrounding the city (see figure 4.1). Line E (RandstadRail) connects Rotterdam with the city of The Hague. The network is integrated with tram, bus and national rail services, serving as the primary rapid transit backbone of the metropolitan region.

4.3.2. Description of the Hoekse Lijn corridor

The Hoekse Lijn is a former heavy railway line that has been transformed into a light rail/metro extension of the Rotterdam metro system which opened service in 2019. The 24-kilometre line runs from Schiedam to the North Sea coast at Hoek van Holland, and was originally part of the Dutch national railway network, being served by trains from the Dutch National Railways (*Nederlandse Spoorwegen - NS*). It is now fully integrated as the western branch of metro line B, which runs from Nesselande in the east to Hoek van Holland Strand in the west, following a one-kilometre extension completed in 2023. As part of the conversion, a new station, Maassluis Steendijkpolder, was added to the line.

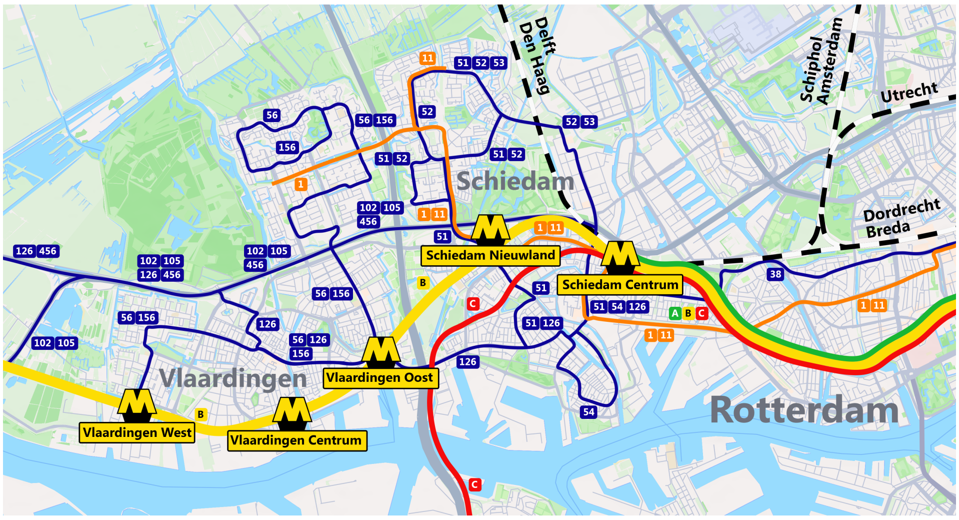


Figure 4.2: Overview of the Hoekse Lijn: eastern section (Schiedam to Vlaardingen)

The ten stations that are part of the Hoekse Lijn are listed from east to west in table 4.1. These stations serve a mix of urban centres, suburban neighbourhoods, as well as recreational areas. The land use around each station can be described as follows:

- **Schiedam Centrum** - Located just outside of Rotterdam, this is an important interchange between metro, tram, bus and national rail. The area features a mix of high-density housing, shopping streets, museums, and restaurants in the historic centre. On the north side of the station are educational institutions, offices, and industrial zones, along with various cultural and entertainment venues.
- **Schiedam Nieuwland** - Situated in a residential area with bus and tram connections (about 200 metres away). The station also serves the Franciscus Vlietland Hospital and the newly developed Park Harga neighbourhood (550 homes). The surrounding area includes several schools and sports facilities.
- **Vlaardingen Oost** - Serves the eastern part of Vlaardingen. This area is mainly residential, with industry and businesses located near the port area on the southern side of the station.
- **Vlaardingen Centrum** - Located near the historic town centre of Vlaardingen, with shops and restaurants. The surrounding area has a mix of housing and commercial properties, with some industry to the south.
- **Vlaardingen West** - Serves the high-density Westwijk neighbourhoud of Vlaardingen, as well as the industry and business parks located to the south of the station.

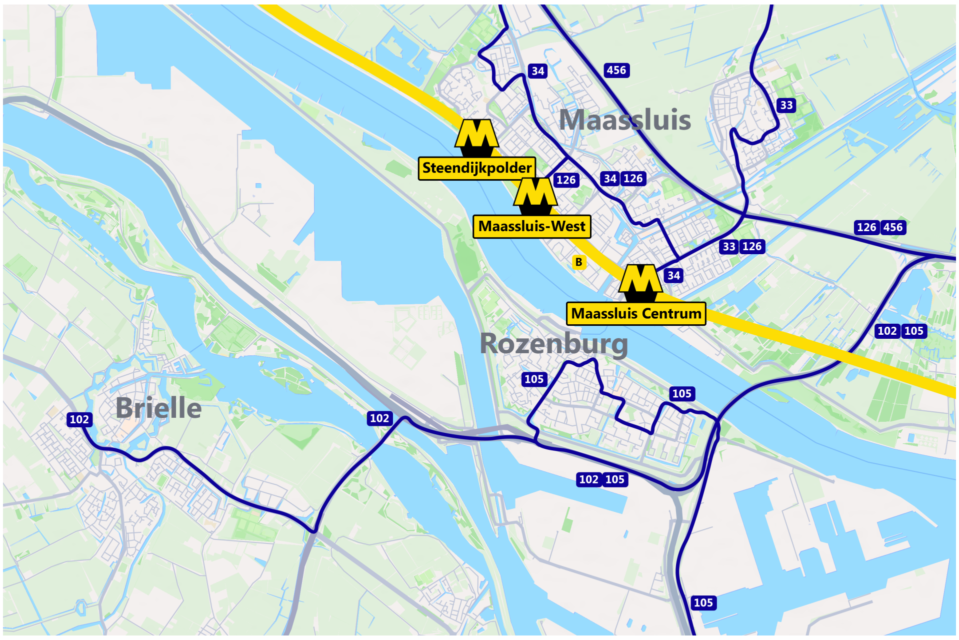


Figure 4.3: Overview of the Hoekse Lijn: central section (Maassluis area)

- **Maassluis Centrum** - Serves the historic centre of Maassluis, as well as a lot of residential areas including the newly developed ’De Kade’ neighbourhood on the shore of the Maas (840 homes), and some businesses to the west of the station.
- **Maassluis West** - Surrounded mainly by residential areas, including the newly developed ‘Het Balkon van Maassluis’ (1,000 homes). The station also provides access to the large ’Koningshoek’ shopping mall and Maassluis’s municipal office.
- **Maassluis Steendijkpolder** - Newly built station that serves the neighbourhood with the same name.
- **Hoek van Holland Haven** - Serves the Stena Line ferry to England, as well as the town centre of Hoek van Holland with shops, restaurants and residential streets.
- **Hoek van Holland Strand** - Relocated station situated directly at the beach. It mainly serves visitors to the beach, dunes, and surrounding hospitality venues. Due to the protected nature of the dunes, there is little residential development in this area.

Figure 4.4: Overview of the Hoekse Lijn: western section (Hoek van Holland)

Table 4.1: Stations on the Hoekse Lijn and their surroundings

Station	Main Land Use	Transport Connections
Schiedam Centrum	Residential, commercial, institutional	Metro lines A, B, C (shared tracks); NS trains to Rotterdam Central, Amsterdam, The Hague, Dordrecht; RET tram 1, 11; RET bus 38, 51, 52, 53, 54, 126; EBS bus 102, 105, 456, 826
Schiedam Nieuwland	Residential, institutional	RET tram 1, 11; RET bus 51
Vlaardingen Oost	Residential, industrial	RET bus 56, 126, 156; EBS bus 826
Vlaardingen Centrum	Residential, commercial, industrial	No additional connections
Vlaardingen West	Residential, industrial, institutional	RET bus 56, 156
Maassluis Centrum	Residential, commercial	EBS bus 33, 34, 133
Maassluis West	Residential, commercial, institutional	RET bus 126
Steendijkpolder	Residential	No additional connections
Hoek van Holland Haven	Residential, commercial	EBS bus 31; Stena Line ferry to Harwich (UK)
Hoek van Holland Strand	Recreational	No additional connections

4.4. Exploratory data analysis: observed ridership trends and data considerations

The following sections present temporal and spatial patterns observed in the boarding data, which form the empirical foundation for the comparison with forecasted demand in chapter 5.

4.4.1. Data sources and resolution

This section describes the available datasets used to analyse observed demand on the Hoekse Lijn. While the original intention was to work with disaggregated, passenger-level smart card data, the dataset ultimately provided by RET was more aggregated than anticipated. As a result, additional modelling steps were required to infer certain demand indicators. The forecast data used for comparison — including assumptions, modelling approach, and scenario definitions — has already been discussed in Section 4.1.1.

Smart card data (OV-chipkaart)

The dataset provided by RET consists of aggregated *OV-chipkaart* smart card data for all station on the Hoekse Lijn (being Hoek van Holland Strand to Schiedam Centrum). Specifically, for the Hoekse Lijn the following data was made available:

- **Daily boardings per station:** number of boardings per station for each day from 01-01-2020 until 31-03-2025.
- **Averaged boardings per day type per month:** precomputed average boardings for weekdays, Saturdays, and Sundays for each month from 01-2020 until 03-2025.
- **Hourly boardings for:**
  - Weekdays, Saturdays, and Sundays for all stations on the Hoekse Lijn, averaged over the period from 01-02-2025 until 21-02-2025 (prior to the Dutch spring holiday).
  - A full-day in- and outflow profile per station for Wednesday, June 26, 2024, representative of a warm and sunny working day outside of the summer holiday.

- **One-directional eastbound OD-matrix:** for weekdays in February 2025 passengers departing from the Hoekse Lijn and aggregated at the station-to-station level. While only departure flows are recorded, it can be considered symmetrical on daily basis.

Table 4.2: Monthly increase factors for *OVpay* (Jan 2023 – Mar 2025)

Month	Factor	Month	Factor
January 2023	1.000	January 2024	1.110
February 2023	1.000	February 2024	1.117
March 2023	1.000	March 2024	1.125
April 2023	1.008	April 2024	1.148
May 2023	1.016	May 2024	1.155
June 2023	1.035	June 2024	1.164
July 2023	1.067	July 2024	1.192
August 2023	1.080	August 2024	1.201
September 2023	1.077	September 2024	1.167
October 2023	1.086	October 2024	1.166
November 2023	1.091	November 2024	1.176
December 2023	1.109	December 2024	1.194
		January 2025	1.188
		February 2025	1.199
		March 2025	1.195

Since the dataset only includes *OV-chipkaart* data, increase factors were provided for each month in which *OVpay* (a payment method using bank cards or mobile phones) was active. The hourly boardings and one-directional OD-matrix have already been adjusted using these correction factors. No increase factor was applied for passengers travelling with day tickets or fare evaders.

Table 4.3: Comparison between boardings for 26-06-2024 from the daily boardings and hourly boardings dataset

Stop	Boardings from data aggregated per day, including OVpay	Total daily boardings from the hourly data for 26-06-2024	Difference
Schiedam Nieuwland	2593	2836	9%
Vlaardingen Oost	2664	2907	9%
Vlaardingen Centrum	2209	2450	11%
Vlaardingen West	2123	2300	8%
Maassluis Centrum	1703	1842	8%
Maassluis West	1422	1601	13%
Steendijkpolder	1368	1513	11%
Hoek van Holland Haven	1198	1406	17%
Hoek van Holland Strand	8213	9728	18%

It is important to note that these correction factors are monthly averages. An analysis of the hourly data (where the increase factor was already applied) for 26 June 2024 revealed that the total number of boardings per station was between 9% and 18% higher compared to the daily boarding per station data multiplied by the increase factor for June 2024 (see table 4.3). RET confirmed that it is expected that *OVpay* usage is significantly higher on days that there are much more incidental travellers who tend to use OVpay instead of personal smart cards, typically beachgoers, tourists and other occasional users. As 26 June 2024 was a warm, sunny weekday outside school holidays, it likely attracted a disproportionate number of such users, particularly to Hoek van Holland Strand. This explains the difference between the two sets of data.

RET had indicated that no additional data cleaning was performed before aggregating the smart card data into the available formats. As a results, certain transaction anomalies remain present in the dataset and should be taken into account when interpreting the results:

- Incomplete transactions (for example journeys without a check-out) are included in the one-directional OD-matrix. These appear with destination station code [0]. This group is relatively small as most trips take place between stations with gates.
- There are also journeys with the same origin and destination in the OD-matrix. These may reflect activities by RET service or security staff, people dropping off or picking up someone from the station, or possible data errors. This category is somewhat larger thane expected (according to RET), but still represents a minor portion of the total dataset.
- Barcode tickets (such as group or event tickets) are not included in the data but RET estimates that these count for less than 1% of the total usage.

Although the exclusion of barcode tickets introduces a small undercount, it is assumed that the overestimations caused the other categories may partially cancel this underestimation out, at least to remain within the margin of error.

OV-Lite dataset

Due to the fact that the RET dataset does not contain any information about disembarking passengers or on-board occupancy, it is not possible to directly compare the full set of demand indicators, such as occupancy upon arrival or passenger kilometres, or to disinguish between travel directions on the Hoekse Lijn. To overcome this limitation, a dataset extracted from the OV-Lite model was used.

OV-Lite is a model that uses elasticity based demand estimation to compute origin-destination matrices specifically for public transport, and it is used to compute the effects of changes to the public transport network or services. The base OD-matrix is constructed using *OV-chipkaart* data and assigned to the network using a Zenith-algorithm (Goudappel, personal communication, April 2025). The OV-Lite output that is used for this study contains the following data for all line B metro services that use the Hoekse Lijn for the average weekday in 2023:

- The occupancy upon arrival for all stations on the Hoekse Lijn.
- The total number of boarding and alighting passengers per station.
- The distances between stations.

The following steps were followed to infer the other two demand indicators using the OV-Lite dataset:

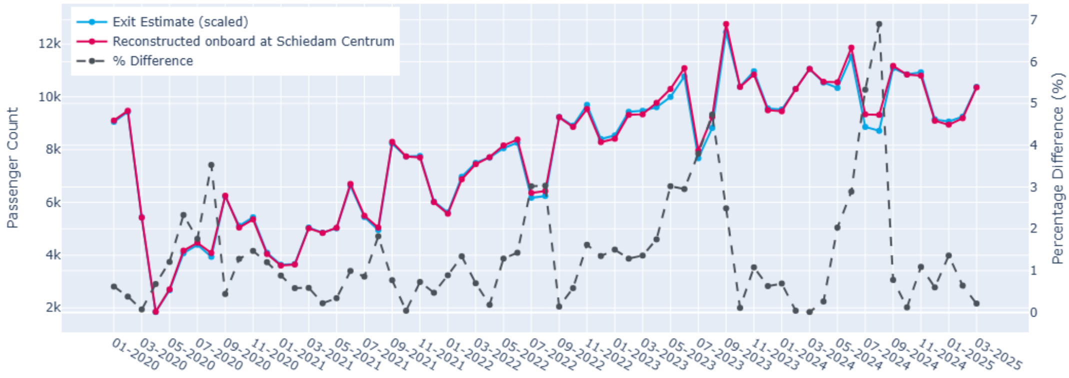
1. The proportion of boardings per direction per station was derived from OV-Lite data. These proportions were subsequently applied tot the total number of boardings from the RET data to estimate direction-specific boardings.
2. For each direction, the ratio of alightings to boardings per station, as observed in the OV-Lite data, was used to estimate alightings from the directional boardings.
3. The number of passengers entering or exiting the Hoekse Lijn network at Schiedam Nieuwland and Schiedam Centrum was obtained from OV-Lite. These values were then scaled per month based on the ratio between RET boardings and OV-Lite boardings over the corridor between Schiedam Nieuwland and Hoek van Holland Strand.

$$E_{month}^{\text{RET}} = E^{\text{OV-Lite}} \times \frac{\sum Boardings_{month}^{\text{RET}}}{\sum Boardings^{\text{OV-Lite}}}$$

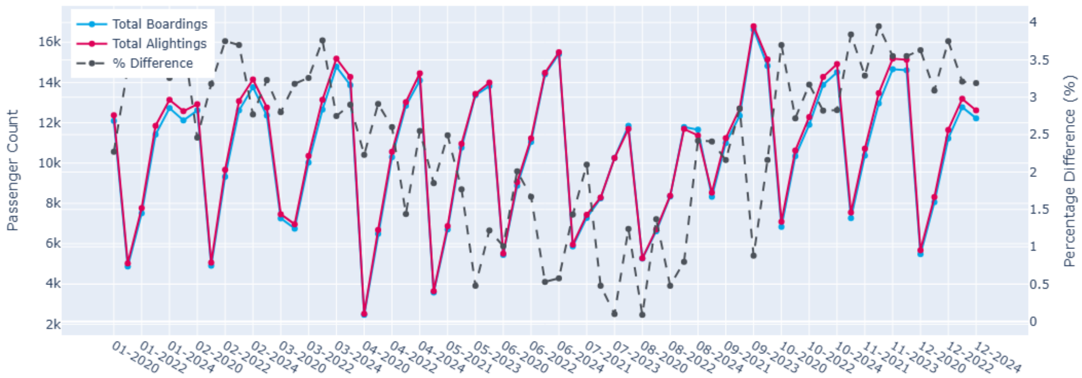
4. Using the scaled entry point as an intitial condition, onboard occupancy at arrival was calculated recursively. At the terminus (either Hoek van Holland Strand or Hoek van Holland Haven depending on the month), remaining passengers were assumed to all alight.
5. Passenger kilometres per segment were computed by multiplying occupancy by segment length and then summed.

To verify the internal consistency and reliability of the approach, two validation checks were performed:

- The scaled exit estimate was compared to the cumulative onboard occupancy at the terminus that was constructed recursively. Absolute and percentage differences were computed per month.
- The total number of boardings and alightings per month were compared.



(a) Comparison of scaled exit estimate and reconstructed onboard occupancy at the terminus (Schiedam Centrum)



(b) Comparison of total boardings and alightings per month

Figure 4.5: Internal consistency validation of the directional split and alighting estimation method

Figure 4.5 visualises the validation checks that were applied to the reconstructed demand indicators.

(a) shows the scaled exit estimate against the recursively reconstructed onboard occupancy when arriving at Schiedam Centrum (end of the Hoekse Lijn system). The two time series closely align, with percentage differences generally remaining below 2%. Notably, higher differences between reconstructed occupancy and scaled estimates are observed during the summer months. This suggests that the model-based scaling approach may slightly underestimate higher seasonal demands.

(b) displays the total number of boardings and alightings per month. The curves show strong similarities, with percentage differences varying between 1.5% and 3.5%. This confirms that the alighting estimation method does not introduce any significant bias.

To support these observations, statistical measures were also used. The MAPE for both (a) and (b) were below 2% (1.35% and 1.96% respectively), indicating that the average deviation between the two methods represents only a small fraction of total monthly demand. The Pearson correlation coefficients of 0.998 (a) and 0.999 (b) confirms a strong linear relationships, meaning that the fluctuations in both time-series follow the same pattern. However, both comparisons resulted in statistically significant differences in the paired t-test ($p<0.05$) indicating that the mean difference is not zero, so there is a slight systematic over- or underestimation.

4.4.2. Demand indicators directly derived from the smart card data

The first step consists of assessing the readily available station-level boardings over time in order to reveal usage patterns, seasonal fluctuations, and possible anomalies. Hourly boarding distributions help to be able to group station into categories based on their peak patterns.

To establish an initial picture of boarding patterns across all stations on the Hoekse Lijn, figure 4.6 displays the average weekday boardings per station in March and September 2024, compared to the annual average. March and September were chosen as both are relatively stable months in terms of travel behaviour as they are not affected by extreme weather or major holidays, and they represent different parts of the year with different weather patterns, while also avoiding seasonal extremes. Furthermore, the American Public Transport Association (APTA) describes that ridership often increases around September and October each year being the end of summer and the resumption of regular work and school routines (APTA, 2024).

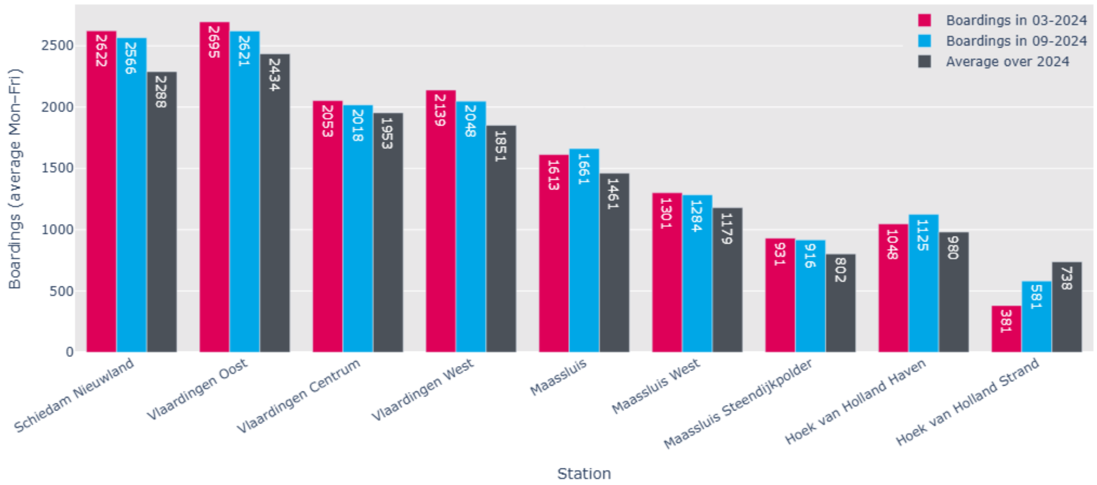


Figure 4.6: Number of boardings per station for March 2024 and September 2024 compared to the 2024 yearly average

The graph in figure 4.6 shows a consistent station hierarchy with Vlaardingen Oost en Schiedam Nieuwland being the most used stations for all three time periods, with a declining trend in boardings the further the line goes west. Maassluis and Hoek van Holland Haven are the only two stations with higher boardings in September than in March. A possible explanation for this might be touristic demand that is still relatively high post-summer. On the other hand, Maassluis Steendijkpolder and Maassluis West show relatively stable demand between March and September, reflecting their commuter-oriented function. Furthermore, most stations show higher boardings in both March and September compared to the annual average, except for Hoek van Holland Strand. This stations shows significantly lower boardings in September, and especially March, confirming the seasonal, leisure-oriented usage with peaks during summer.

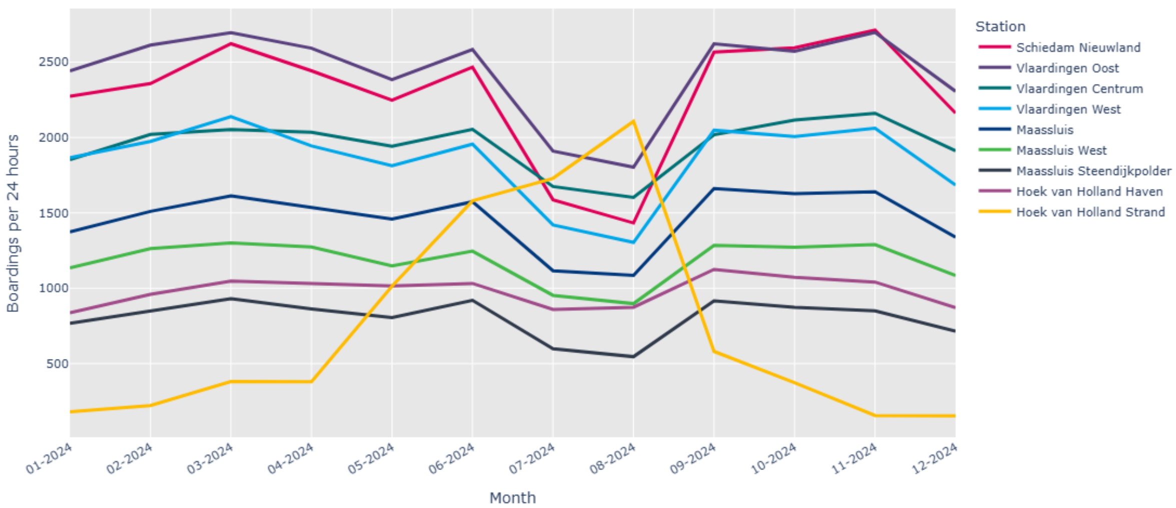


Figure 4.7: Monthly average weekday boardings per station in 2024, showing seasonal variation across the Hoekse Lijn

Figure 4.7 further illustrates the yearly trends per station. Most station follow the same yearly trends, with a clearly visible dip in the summer and Christmas holiday periods, highlighting the predominant commuter-oriented nature of most stations. The passenger dips are especially apparent for Schiedam Nieuwland, possibly explained by proximity of the station to several educational institutions, which have pre-determined, fixed holidays, not only during summer and Christmas, but also in February, multiple during spring, and one week during the end of October. Figure 4.7 also clearly indicates the opposite trend of Hoek van Holland Strand, with peak usage occurring during the summer months, and only a fraction of that the rest of the year. Table 4.4 supports these conclusions by quantifying the monthly variability per stations, showing relatively low variations for commuter-oriented stations and a remarkebly high variation for Hoek van Holland Strand. Interestingly, according to this table, Vlaardingen Centrum seems to exhibit the most stable boarding trend throughout the year, indicating that it is not solely commuter oriented. Notably, these monthly travel patterns confirm that overall patterns are consistent across the line, but station specific variants emphasise the need to further explore more disaggregated travel patterns to identify the different functions the stations fulfil.

Table 4.4: Monthly variation in boardings per station in 2024

Station	Mean	Standard Deviation	Coefficient of Variation (%)
Schiedam Nieuwland	2289	400	17.47
Vlaardingen Oost	2435	296	12.16
Vlaardingen Centrum	1953	170	8.70
Vlaardingen West	1851	259	13.99
Maassluis	1461	197	13.48
Maassluis West	1179	138	11.70
Maassluis Steendijkpolder	803	125	15.57
Hoek van Holland Haven	981	97	9.89
Hoek van Holland Strand	738	695	94.17

Looking at the hourly patterns for weekdays, three types of stations can be identified: commuter oriented stations with sharp morning peaks; stations with balanced and spread demand; and stations with a clear afternoon peak. Vlaardingen Oost, Maasluis Centrum, Maassluis West, Steendijkpolder and Hoek van Holland Haven display a clear peak between 07:00 and 09:00, consistent with commuters heading towards work or school. Besides the morning peak, Vlaardingen Centrum also shows a slight increase in boardings during the evening rush hour. This double-peaked profile is even more clearly visible at Maassluis West, indicating the presence of places people work at. Schiedam Nieuwland and Vlaardingen West exhibit an atypical, more spread out, double-peaked profile with the second peak being earlier in the afternoon between 14:00 and 16:00. This is possibly explained by the fact that there are several educational facilities around these stations, which tend to finish their activities earlier and at different times during the day. Finally, the solely leisure oriented function of Hoek van Holland Strand is again confirmed with boardings peaking in the afternoon, an no distinct peaks during commuter travel times.

Figure 4.9 presents the development of total daily boardings from January 2020 to March 2025, alongside a yearly average. The sharp decline in March 2020 aligns with the COVID-19 outbreak and the accompanying lockdowns and other measures, severely affecting public transport ridership. Recovery only starts to set in at the end of 2021, and by March of 2022 all measures for public transport are lifted. This resulted in substantial growth of passenger numbers during 2022, albeit not reaching the levels of pre-COVID times. March 2023 marks another milestone that affected the total number of boardings on the Hoekse Lijn, as the station of Hoek van Holland Strand, located right at the beach, was opened. This is immediately reflected in passenger peaks during the summer of 2023 (outside of the summer holiday), and another substantial increase of the yearly average over 2023.

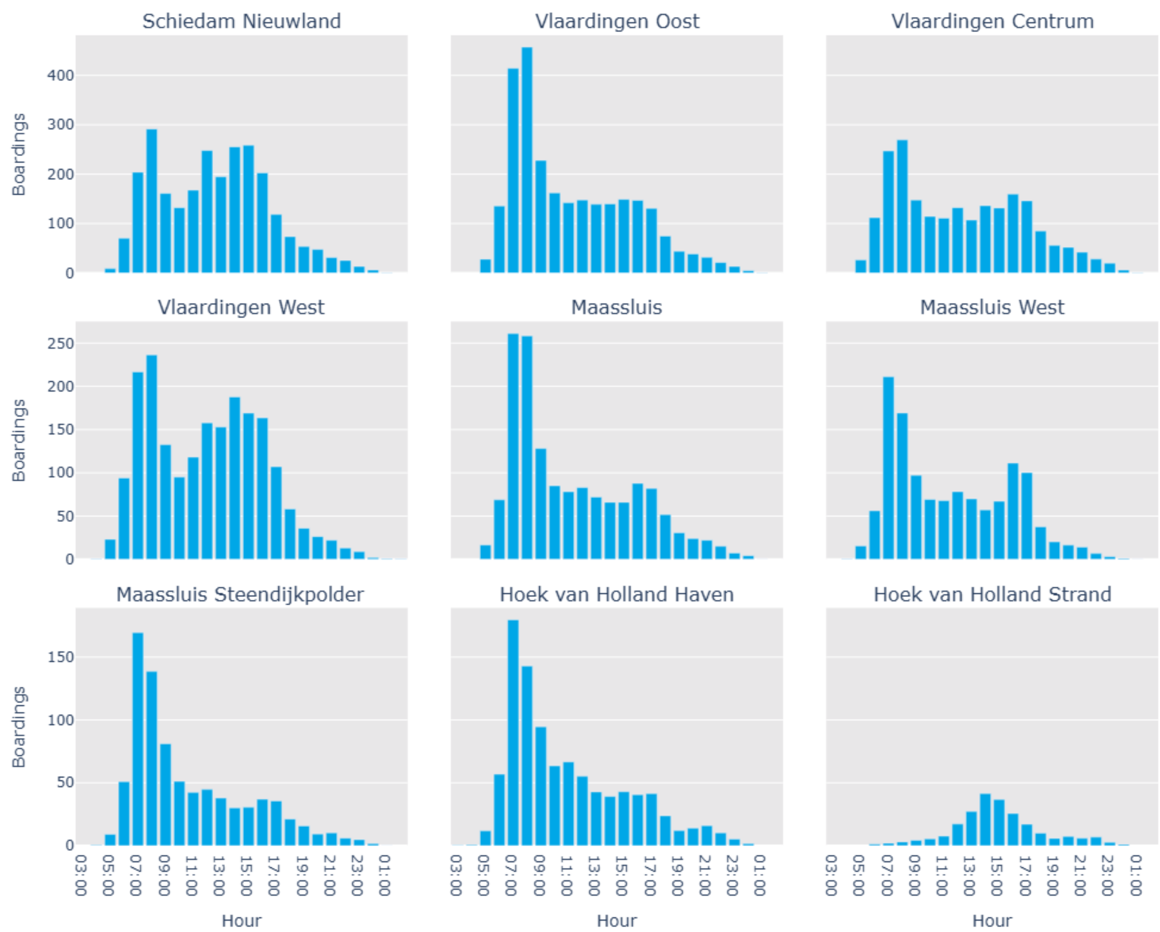


Figure 4.8: Average boardings per hour per station

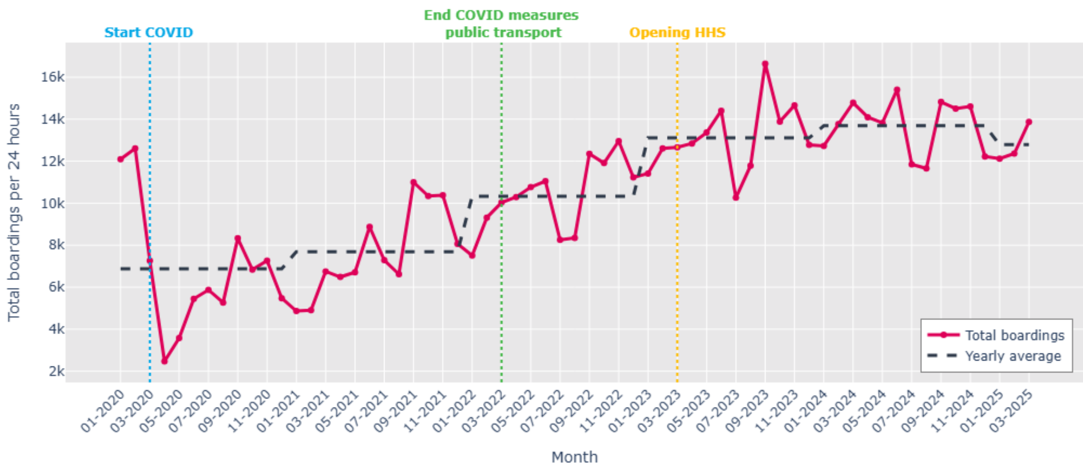


Figure 4.9: Number of boardings for all stations from 01-2020 to 03-2025, for the average weekday

4.4.3. Inferred indicators

To make some sort of comparison with the other two indicators that were used in the projections, occupancy at arrival and total passenger kilometres, these were inferred using a combination of the RET aggregated data and the data extracted from the OV-Lite model. This process is described in section 4.4.1.

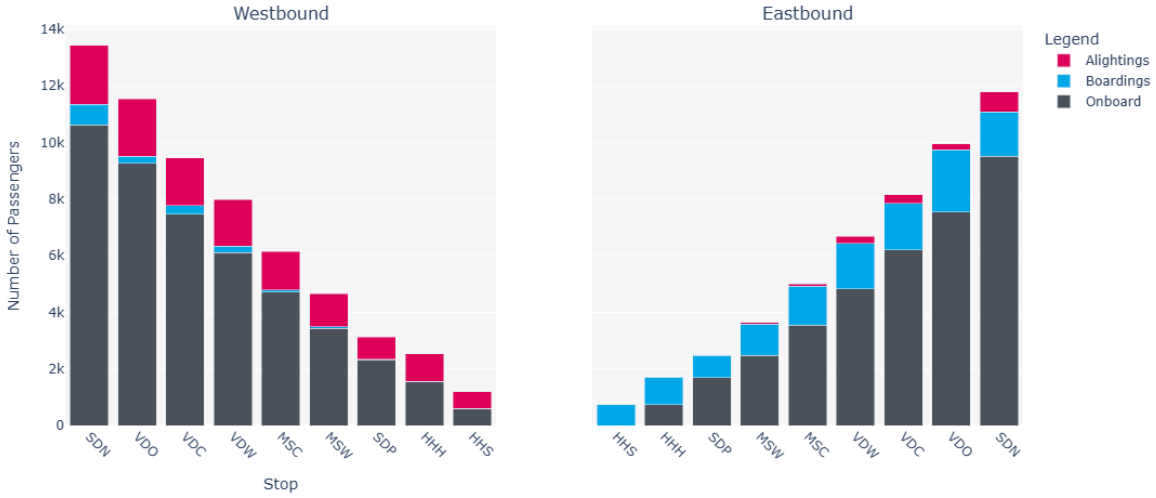


Figure 4.10: Occupancy upon arrival at each station in both directions for the average weekday in 2024, as well as a visualisation of the passengers boarding and alighting at each station

Figure 4.10 presents the average weekday boarding and alighting patterns along the Hoekse Lijn in 2024. The occupancy in westbound direction is slightly higher than the occupancy in eastbound direction. Occupancy is generally highest in the eastern part of the line and decreases towards the western terminus at Hoek van Holland Strand. This spatial gradient reflects a dominant commuting pattern: in the morning, passengers board in the west to travel to eastward towards work in Schiedam, Rotterdam, or further into the Randstad. In the evening, this trend reverses, with higher westbound alightings as passengers return home.

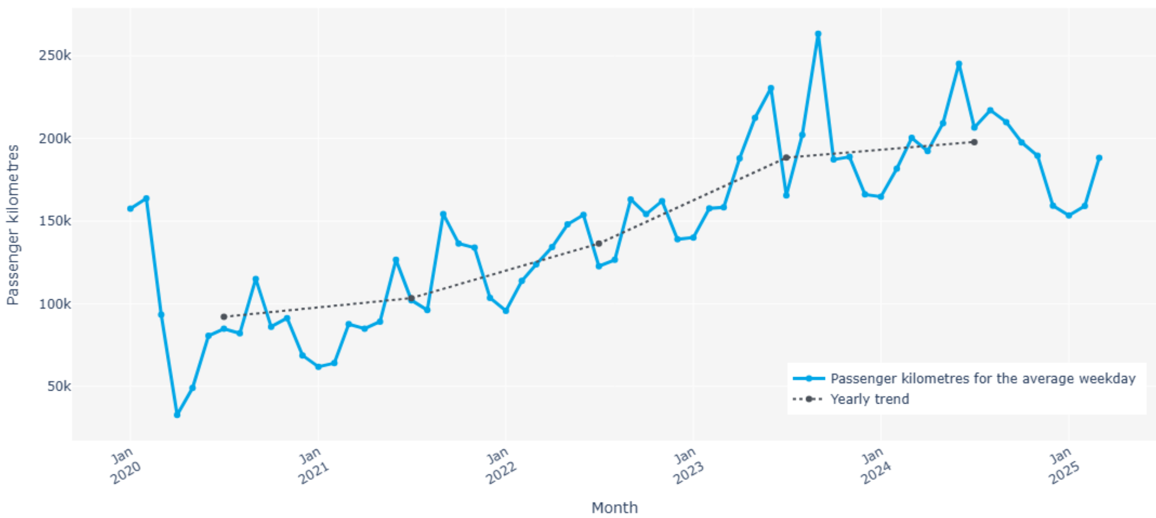


Figure 4.11: Passenger kilometres for the average weekday for every month between 01-2020 to 03-2025

Figure 4.11 shows the development of passenger kilometres per average weekday for each month since January 2020 on the Hoekse Lijn. A sharp decline as a result of the COVID-19 pandemic is observable at the beginning of 2020. The passenger numbers gradually recover in the years that follow and this incline more or less stabilises in 2024, with only a slight increase compared to 2023 but passenger kilometres exceeding the February 2020 levels. The passenger kilometres peak during summer months that are not in the summer holiday, indicating that besides normal commuter traffic there is a high recreational use of Hoek van Holland Strand that is furthest west on the line.

4.4.4. Beach ridership

One of the objectives of converting the Hoekse Lijn into a metro service was to improve accessibility to Hoek van Holland Strand and thereby increase the amount of people who choose to travel to the beach using public transport (Gemeente Rotterdam, 2015). Due to its highly weather-dependent and irregular demand, beach-bound travel was excluded from the multimodal transport model used in the 2015 forecasts. Instead, a separate estimation methodology was used to account for beach visitors in the overall demand forecast (Goudappel Coffeng, 2015).

To evaluate the accuracy of these assumptions, observed beach-related ridership was analysed using smart card data of the year 2024. Based on daily maximum temperatures obtained from the KNMI (Royal Netherlands Meteorological Institute), beach days were identified and classified into three categories (similar to the forecast methodology):

- Top beach days with maximum temperatures higher than 30°C.
- Busy beach days with temperatures between 25°C and 30°C.
- Moderate beach days with temperatures between 20°C and 25°C.

Table 4.5 shows the average boardings at the stations by beach day category, compared to non-beach weekdays during the same year. There are clear differences in average boardings across the four day types at Hoek van Holland Strand. On the busiest beach days, over 10,000 boardings are recorded, 27 times as much as on regular non-beach weekdays. There are also substantial increases on busy weekdays with over 3,000 boardings, while moderate beach days show a more modest increase in the number of boardings.

Table 4.5: Average boardings by beach day category (weekdays only, 2024)

Station	Top beach day	Busy beach day	Moderate beach day	Non-beach day
Schiedam Nieuwland	1,693	1,980	1,800	2,417
Vlaardingen Oost	1,690	2,170	2,093	2,528
Vlaardingen Centrum	1,585	1,825	1,765	2,003
Vlaardingen West	1,476	1,651	1,555	1,929
Maassluis Centrum	1,228	1,334	1,271	1,510
Maassluis West	999	1,089	1,039	1,215
Steendijkpolder	609	801	689	826
Hoek van Holland Haven	1,301	1,021	982	943
Hoek van Holland Strand	10,068	3,667	1,157	364

This extreme variation confirms the recreational function of Hoek van Holland Strand and demonstrates the weather-dependent nature of beach travel demand. The monthly average increase factors for OVpay usage are applied to these boardings, though as explained in section 4.4.1, OVpay usage on high recreational demand days can result in underestimation of up to 18%.

4.4.5. Other relevant data

The one-directional OD-matrix for weekdays in February 2025, provided by RET, was used to get an overview of the most used destination station for passengers boarding on the Hoekse Lijn. The data shows that there is a substantial share of passengers that use the Hoekse Lijn to reach one of the stations in or near the city centre of Rotterdam. Some of the most popular destinations are:

- Schiedam Centrum - The nearest major interchange with possibilities to transfer to other metro lines, various tram and bus services, and national rail services.
- Beurs - Major interchange that allows for transfer to other north-south and east-west metro lines without the need to check-out and back in, but also centrally located near Rotterdam's main shopping and commercial district.
- Dijkzicht, Coolhaven, Marconiplein, Eendrachtsplein and Blaak - All located within or on the edge of the city centre. These stations provide access to important employment centres, educational facilities, cultural institutions, and retail areas, making them important for both work and leisure trips.

Notably, three of the Hoekse Lijn stations are among the ten most used destinations highlighting that intraline use of the Hoekse Lijn should not be disregarded. For example, for Schiedam Nieuwland, Vlaardingen West, Vlaardingen Centrum, and Vlaardingen Oost take the third, fifth and sixth place in most popular destinations for people departing from the station. These patterns suggest that the lines also serves important local travel needs, such as connections for school, healthcare, or shorter commutes.

4.5. Reflections on observed demand and data usability

The goal of this chapter was to provide the empirical basis for evaluating forecast accuracy, which is presented in the next chapter. Three key elements were combined: (1) the strategic objectives that motivated the conversion of the Hoekse Lijn into a metro line; (2) the spatial and functional characteristics of the line; and (3) the observed ridership patterns derived from the smart card data.

The ridership data confirms several spatial patterns that align with the functional roles of the station as described earlier. Schiedam Nieuwland and Vlaardingen Oost are important stations on the line, as both serve as origin station for commuters and as interchanges to other parts of the RET network, confirmed by having the highest number of boardings. While both stations show a clear morning peak, there is also a notable amount of boardings during the afternoon, which can partially be attributed to the nearby educational facilities with a more varied schedule. Although less pronounces, Vlaardingen Centrum and Maassluis West also show some afternoon usage. At the remaining commuter-oriented stations, demand is clearly concentrated in the morning peak hours, suggesting a dominant commuter function with limited off-peak usage.

These patterns offer mixed evidence for the objective of improving utilisation of the line throughout the day and in both directions. While some stations do appear to support more varied travel purposes, other remain highly peak-oriented. Whether this represents a meaningful improvement on the situation before conversion, and whether this aligns with the modelled expectations, will be explored in the next chapter through direct comparison with the 2015 demand forecasts.

One of the project's most significant impacts in relation to its strategic objecitves is observed at Hoek van Holland Strand, where beach-related travel causes exterme seasonal fluctuations. On peak beach days, boardings can exceed 10,000, and the daily profile shows a clear pattern of afternoon arrivals and early evening departures, confirming the station's strong recreational function. While this suggests that extending the line directly to the beach has made the service more appealing for leisure trips, as intended, the next chapter will provide a full evaluation of whether this aligns with the projected increase in beach ridership.

OV-chipkaart data proved to be useful in evaluating the development of passenger demand on the Hoekse Lijn. Despite being aggregated, it enables the reconstruction of the boardings per station and hourly flow patterns. With support from external model assumptions, it is also possible to infer occupancy upon arrival and passenger kilometres, providing a strong empirical basis for assessing the accuracy of the demand forecasts made in 2015.

However, the data also has important limitations. Check-out data is missing, so origin-destination flows must be estimated using directional ratios from OV-Lite. The dataset also excludes barcode tickets and fare evasions. While monthly correction factors for *OVpay* were applied, these may underestimate usage on days with many incidental travellers, such as beach days. Furhermore, the dataset lack any socio-demographic information or insights into user motivations, which limits the ability to explain behaviour beyond the observed patterns.

Nevertheless, despite these limitations, the smart card data offers a sufficient basis for evaluating realised demand on the Hoekse Lijn, enabling the reconstruction of key demand indicators that align with those used in the 2015 forecasts. Building upon this, the next chapter assesses the extent to which observed ridership matches forecasted expectations.

5 Comparison of forecasted and observed demand

5.1. Comparison of demand indicators

5.1.1. Boardings per station

Figure 5.1 illustrates the development of the average number of weekday boardings from 2020 to 2024, alongside the original 2015 forecast and two projections for 2025, to enable a comparison between the forecasts and the recent data. The first projection assumes yearly growth in passenger numbers of 3% from 2024 onwards, based on RET transport plans and transport committee meeting documents from the Rotterdam council. The second projection corrects for the structural decline in public transport demand following the pandemic. This correction is applied directly to the original 2025 forecast from the 2015 study, reflecting how projected demand would have evolved if the pandemic's structural impacts had been considered in the initial forecast.

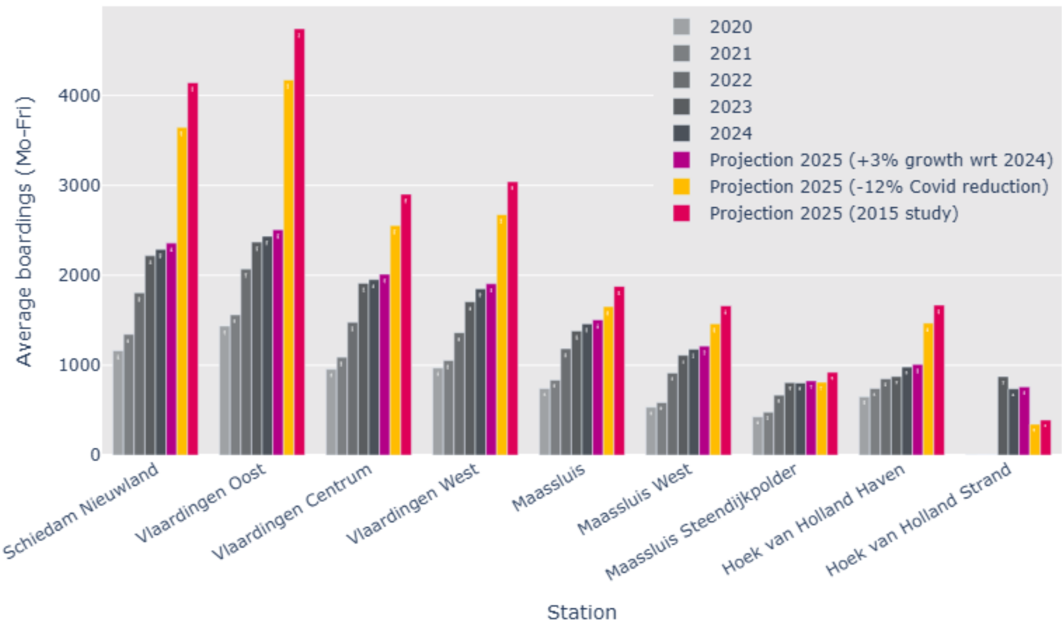


Figure 5.1: Observed boardings per station (2020-2024) and projected demand for 2025, including COVID-19 correction (−12%) applied to the original 2015 forecast and baseline growth projection (+3%) based on 2024 data

To account for structural change in public transport usage following the COVID-19 pandemic, a correction factor of -12% was applied to the original forecasts for 2025. This factor was derived from the *KiM* Mobility Report for 2023, which provides projections based on empirical evidence of continued declines in public transport demand compared to pre-pandemic levels (KiM, 2023). Specifically, KiM estimates that the total structural effects of the COVID-19 pandemic result in a decrease 6% to 16% of distance travelled in bus, tram, and metro, depending on the scenario. For the baseline scenario, this factor is -12%.

The visual comparison reveals a structural overestimation of boardings in the 2015 forecast for most stations on the Hoekse Lijn. While the observed data shows a steady recovery from the COVID-19 pandemic and further growth between 2023 and 2024, it remains significantly lower than the projected values. The deviations are particularly high for the stations between Schiedam Nieuwland and Vlaardingen West. In contrast, the newly built Maassluis Steendijkpolder station performs better than expected when the COVID correction is taken into account, suggesting that the forecast may have underestimated demand at this location. Hoek van Holland Strand even exceeds the projections, highlighting the challenge of accurately modelling leisure travel, which is more sensitive to external factors such as weather and changing recreational patterns

Table 5.1 presents both the absolute and percentage errors between forecasted and actual boardings for 2025, shown separately for eastbound and westbound directions. The projections have been adjusted for the structural COVID-19 impact by applying the 12% reduction, that was described earlier, to the original 2015 forecasts.

Table 5.1: Forecast error per station in 2025 (COVID-corrected projection). Positive percentage error indicates more actual travellers than forecasted; negative error indicates overestimation.

Stop	2025 Observed		2025 Projection		Absolute Error		Percentage Error	
	East	West	East	West	East	West	East	West
Schiedam Nieuwland	1605	753	2382	1265	-777	-512	-33%	-40%
Vlaardingen Oost	2254	254	3142	1032	-888	-778	-28%	-75%
Vlaardingen Centrum	1688	323	2096	459	-408	-136	-19%	-30%
Vlaardingen West	1658	248	1798	878	-140	-630	-8%	-72%
Maassluis	1416	89	1447	205	-31	-116	-2%	-57%
Maassluis West	1158	57	1246	215	-88	-158	-7%	-73%
Maassluis Steendijkpolder	806	21	610	202	196	-181	+32%	-90%
Hoek van Holland Haven	999	10	1437	31	-438	-21	-30%	-68%
Hoek van Holland Strand	760	0	344	0	416	0	+121%	–

Both the absolute and percentage errors are presented as they provide different insights into the forecast accuracy. The percentage error provides a proportional understanding of the deviation relative to the forecasted value, which is useful for comparing stations with different demand levels. However, it can be misleading at stations with low projected boardings, where small absolute deviations result large percentage values. That is why the absolute error is important for the understanding of the practical impact of the forecast inaccuracies on operations.

The results show that the overestimation for the westbound trains is significantly higher than for the eastbound trains, with deviations exceeding 100% on all stations apart from Vlaardingen Centrum. However, due to boardings being higher in eastbound direction caused by a commuter morning peak towards Rotterdam and the eastbound deviations being more manageable, the total percentage error does not exceed 100% at any of the station. The smallest deviations can be observed at the newly constructed Maassluis Steendijkpolder station, possibly explained by no accessible prior knowledge of passenger number to compare with or an underestimation of people choosing for metro when it became better accessible. The underestimation for Hoek van Holland Strand can be explained by the more complicated modelling of leisure passenger flows as these are highly weather dependent.

On the other hand, the largest overestimations were observed at stations such as Vlaardingen Oost, Vlaardingen West, Maassluis West, and Steendijkpolder, particularly for westbound trains. These substantial deviations suggest that fundamental assumptions in the original model may not have realised as anticipated.

Analysing peak-hour travel provides further insight into the discrepancies between the predicted and actual demand on the Hoekse Lijn. The visualisation of projected and observed boardings per station for peak periods in figure 5.2 reveals that the overestimations identified in overall daily boardings are largely concentrated during the peak periods. In particular, the morning peak shows significant dis-

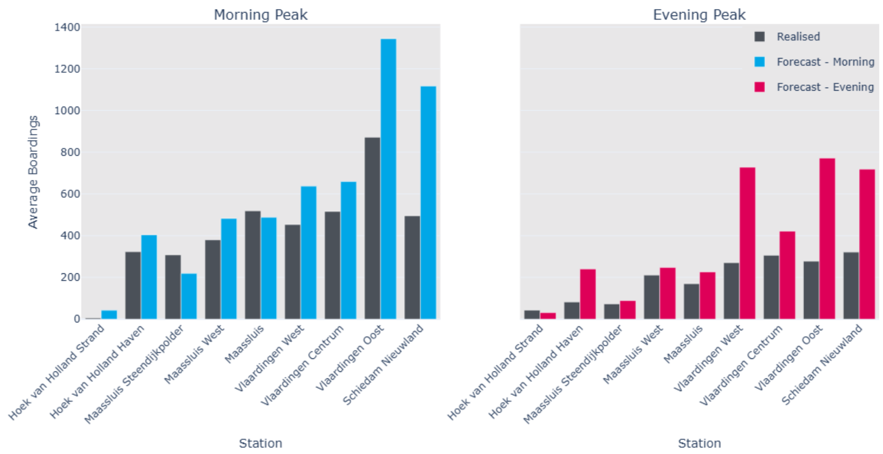


Figure 5.2: Forecasted and observed boardings during morning and evening peak periods (Feb 2025)

crepancies at stations such as Schiedam Nieuwland, Vlaardingen Oost and Vlaardingen West, where forecasted peak demand is much higher than actual boardings. This is consistent with the findings of overall boardings, in which these stations displayed the largest absolute and percentage errors in the westbound direction, underscoring the overestimations of commuter flows in the original model.

Furthermore, the proportion of daily boardings during peak periods, as displayed in figure 5.3, provide additional insights. While the model substantially overestimated afternoon peak travel demand for almost all stations, the situation is more mixed for the morning peak. Stations such as Hoek van Holland Haven, Maassluis Steendijkpolder, Maassluis and Vlaardingen Oost had a higher share of daily boardings than projected during the morning peak. In contrast, Schiedam Nieuwland shows a clear underperformance during the morning peak relative to the forecasts, consistent with its lower-than-expected overall daily boardings. Similarly, Hoek van Holland Strand underperformed during the morning peak despite performing better than expected on a daily basis. This highlights a strong leisure-driven demand outside of traditional commuting hours, which the model did not correctly capture.

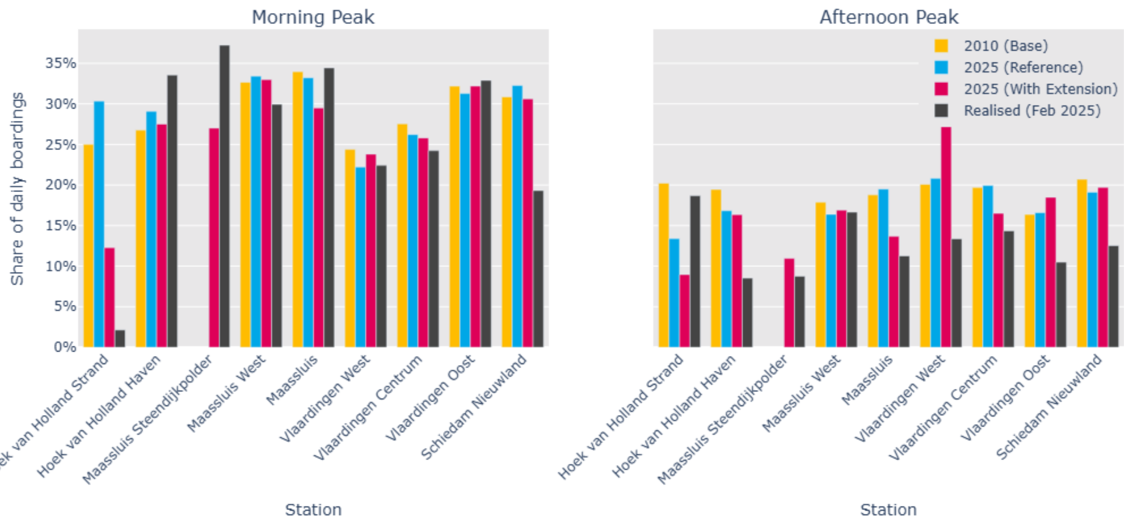


Figure 5.3: Share of daily boardings during morning and afternoon peak periods by station, compared across forecast and observed scenarios.

Interestingly, Maassluis West presents a different pattern: despite the general trend of overestimating the afternoon peak, this station did not show a significant deviation in its afternoon peak share. This suggests that the model accurately captured the commuting patterns of people working near the station, with afternoon demand aligning more closely with the forecast.

5.1.2. Occupancy upon arrival

Besides using boardings per station as an indication of demand mismatches, it is also important to assess how well the model predicted load distribution along the Hoekse Lijn by looking at the occupancy upon arrival at each station. Deviations between forecasted and actual occupancy between stations may suggest incorrect assumptions about trip length, distribution of alightings, or route choice patterns. Figure 5.4 presents the absolute error in occupancy upon arrival per station in both travel directions for 2025, compared to the forecast. Here, negative values indicate fewer passengers on board than predicted, while positive values indicate more passengers than forecasted.

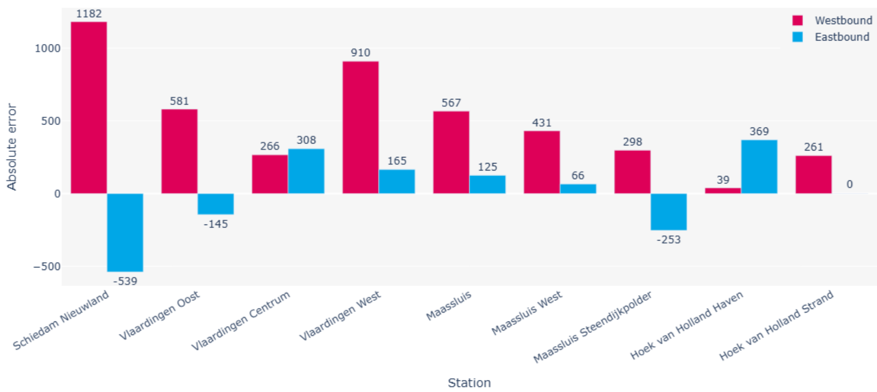


Figure 5.4: Absolute error between forecasted and observed occupancy upon arrival per station (2025)

In contrast to the boardings per station, deviations in occupancy mainly underestimate passenger on-board, with most stations showing relatively higher passengers on board, especially in westbound trains. The most significant underestimations occur at Hoek van Holland Haven and Hoek van Holland Strand with the number of passenger onboard upon arrival being more than 40% higher than expected, which is consistent with earlier findings that ridership to and from the beach was underpredicted.

On the other hand, Maassluis Steendijkpolder stands out with an overestimation in eastbound occupancy, explained by the relatively high eastbound overestimation of boarders at Hoek van Holland Haven (table 5.1). While the number of passengers boarding at Hoek van Holland Strand was significantly higher than forecasted, did this not fully compensate for the overestimation this. Furthermore, in eastbound direction occupancy upon arrival at stations up to Vlaardingen Oost remained higher than forecast, likely due to underestimated boarding at Maassluis Steendijkpolder, and possibly due to more passengers staying on board for longer than the model assumed. However, the substantial overestimation of boardings at Vlaardingen Centrum and Vlaardingen Oost led to the actual occupancy falling below the forecasted levels at these stations, as the earlier compensating effects were insufficient to offset the overestimations of local boardings.

Overall, the analysis of the occupancy figueres reveals a notable trend: although boardings at most stations were lower than predicted, occupancy upon arrival was often higher than forecasted. This suggest that more passengers than expected travelled longer distances along the line, while the model overestimated the number of shorter trips. The higher-than-expected recreational demand for Hoek van Holland Strand was an important factor in this, as these longer leisure trips originating outside of the Hoekse Lijn helped to maintain higher train occupancy levels along the route, compensating the reduced commuter ridership.

5.1.3. Passenger kilometres

The number of passenger kilometres for the average weekday for 2024 was computed to be 197,809 kilometres. Applying the 3% growth to scale this to 2025 resulted in 203,743 kilometres per average weekday. This is 18% higher then projected in 2015. This discrepancy is consistent with earlier findings from the occupancy analysis, which pointed to higher-than-anticipated long-distance travel, particularly toward Hoek van Holland Strand. These trips, typically spanning the entire corridor, contribute disproportionately to total passenger kilometres.

5.1.4. Beach ridership analysis

The analysis of beach ridership provides crucial insights into the accuracy of recreational demand forecasting and helps explain the 121% overperformance of Hoek van Holland Strand compared to original projections. This section examines the broader patterns of beach travel across the corridor and compares observed demand with the original forecast assumptions.

In the original forecasts, beach-bound travel was added as an additional demand component outside the multimodal model. The method estimated the average number of beach-related public transport trips on a typical warm weekday. This estimate was based on historical ridership patterns for beach visitis and adjusted for expected travel time improvements due to the metro extension. Specifically, it was assumed that 850 public transport trips (in both directions) would be made to and from Hoek van Holland Strand on an average beach day. Of these, 100 trips per day were attributed to the accessibility improvements resulting from the metro extension. These values were integrated into the overall forecast outcomes by distributing the estimated beach ridership accross three station: Maassluis Centrum, Vlaardingen Centrum, and Schiedam Centrum. This was done in proportion to their original share of boardings. This way the effect of beach travel and the metro extension is implemented in the reported model results, even though the demand itself was not simulated within the model itself (Goudappel Coffeng, 2015).

Notably, the data on one of the busier beach days in this dataset (26 June 2024) revealed an underestimation of 18% in boardings at Hoek van Holland Strand when the aggregated daily boarding figures were corrected using the June OVpay factor, as opposed to the actual hourly boarding total. This suggests that the values observed in table 4.5 may still underestimate the true demand on the busiest beach days, particularly among incidental travellers, who are more likely to use OVpay than regular commuters.

Other stations show more varied effects. While Hoek van Holland Haven also experiences higher ridership on beach days, most of the more commuter-oriented stations record lower boardings on the top beach days. This is not necessarily due to passengers changing their mode of transport or destination, but is more likely to reflect a decline in commuter demand due to the summer holiday period, as it is important to note that the busiest beach day in the analysis, and the only top beach day (12 August 2024), falls during the Dutch summer holidays. Therefore, the ridership figures at intermediate stations should be interpreted cautiously.

To better understand how beach travel interacts with holidays and typical travel patterns, table 5.2 categorises average daily boardings into four scenarios: whether it is a beach day, and whether the days falls inside or outside a holiday period. This disaggregation allows for a more nuanced interpretation of the beach day effects. Again, the clearest effect is visible at Hoek van Holland Strand, where boardings rise from 360 on non-beach days outside the holidays to 1,716 on comparable beach days, and even further to 2,285 on beach days during the summer holidays.

Table 5.2: Average boardings by holiday and beach day condition (weekdays only, 2024)

Station	Avg 2024	No holiday no beach	No holiday beach	Holiday no beach	Holiday beach
Schiedam Nieuwland	2,281	2,512	2,346	1,352	1,288
Vlaardingen Oost	2,429	2,603	2,515	1,669	1,655
Vlaardingen Centrum	1,950	2,045	2,003	1,509	1,523
Vlaardingen West	184	1,997	1,918	1,183	1,212
Maassluis Centrum	1,846	1,555	1,571	984	980
Maassluis West	1,176	1,252	1,245	866	833
Steendijkpolder	800	855	934	477	475
Hoek van Holland Haven	979	1,009	1,087	806	845
Hoek van Holland Strand	739	360	1,716	996	2,285

At all other stations, the patterns are more varied and are likely to be driven by fluctuations in commuter demand. All ridership figures decrease during the holidays, regardless of beach day status, indicating that the effects of changes in work and school routines have dominate over incidental beach traffic effects. Closer to the coast, Steendijkpolder and Hoek van Holland Haven show natable increases in boardings on beach days outside the holiday periods. This may suggest that some beach visitors access the metro within the corridor itself. However, it cannot be firmly concluded whether these increases reflect recreational travel or unrelated fluctuations in demand.

Furthermore, when comparing ridership patterns along the line, no significant increase in boardings is evident at the other stations, either between beach and non-beach days outside the holiday periods or during the holiday periods itself. This suggests that most beachgoers probably enter the metro network upstream, outside of the Hoekse Lijn corridor. While the current dataset does not include upstream stations or transfer flows, the absence of a measurable uplift at stations such as Schiedam Nieuwland, Vlaardingen Centrum or Maassluis Centrum supports the interpretation that the Hoekse Lijn is primarily used by beach visitors as the final leg of their journey rather than as their point of entry.

Figure 5.5 helps in getting a better understanding of the daily passenger dynamics of beach-related travel, showing the hourly number of people boarding and alighting at Hoek van Holland Strand on 26 june 2024 - a sunny weekday with a maximum temperature of 25.6°C and outside the school holidays. This day was classified as a busy beach day and serves as a representative example of incidental recreational demand.

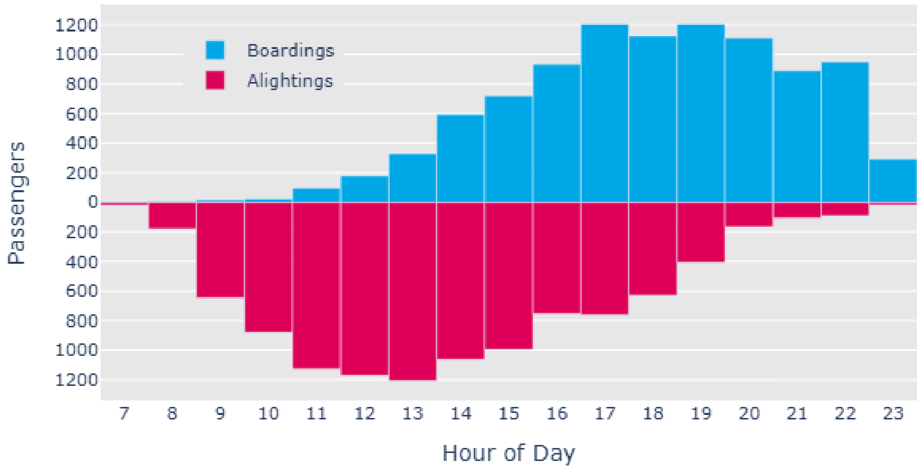


Figure 5.5: Number of boardings and alightings for Hoek van Holland Strand on 26 June 2024

The figure clearly shows the temporal seperation between arrivals and departures, as alightings dominate the morning and early afternoon, peaking around 13:00. Boardings increase steadulty from the afternoon and peak between 17:00 and 20:00, reflecting the end-of-day return flow of passengers. The volume of both alightings and boardings, each exceeding 1,000 passengers per hour during peak periods, illustrates the significant scale of beach-related travel on warm, non-holiday weekdays.

The extreme weather-dependent variability observed at Hoek van Holland Strand, where peak beach days generate 27 times normal weekday boardings, demonstrates the challenges of modeling irregular recreational demand within traditional transport models. The original approach of handling beach travel as a separate component outside the multimodal model proved inadequate for capturing both the magnitude and temporal concentration of actual demand. The findings suggest that recreational demand modeling requires more sophisticated approaches that can account for weather variability, seasonal patterns, and the interaction between recreational and commuter travel patterns.

5.2. Cluster-based analysis of forecast accuracy

This section evaluates the accuracy of the forecasts by clustering the stations and comparing observed and forecast ridership using the weighted mean absolute percentage error (MAPE), to reveal variations in model performance across station types and passenger groups. The objective of clustering is to group stations with similar demand profiles, enabling forecast performance to be assessed by type rather than per station. Using scale-free features such as peak-hour share, seasonality, and intra-day variability means the clusters describe how demand is distributed over time rather than how large it is.

This makes it easier to identify whether errors are systematic, for instance, whether the model performs differently at commuter-dominant, mixed-use, or leisure-oriented stations, and assists in interpreting the MAPE results by cluster. It also provides targeted input for explaining deviations and refining the model assumptions.

5.2.1. Station typologies from cluster analysis

Using k-means clustering, three station types are identified based on the following information from the smart-card data:

- Peak-hour usage ratio - share of daily boardings between 07:00-09:00 and 16:00-18:00.
- Seasonality indicator - coefficient of variation between monthly average weekday boardings for 2024.
- Daily variability indicator - coefficient of variation of hourly boardings on an average weekday in February 2025.

Based on the elbow test and the silhouette score shown in figure 5.6, it was concluded that three clusters were the most appropriate solution. The elbow test shows that the reduction in variance within clusters significantly levels off after three clusters, suggesting that there are fewer benefits to be gained from adding more clusters. While the silhouette score reaches its maximum at two clusters, it remains reasonably high at three, offering a good balance between cluster separation and interpretability.

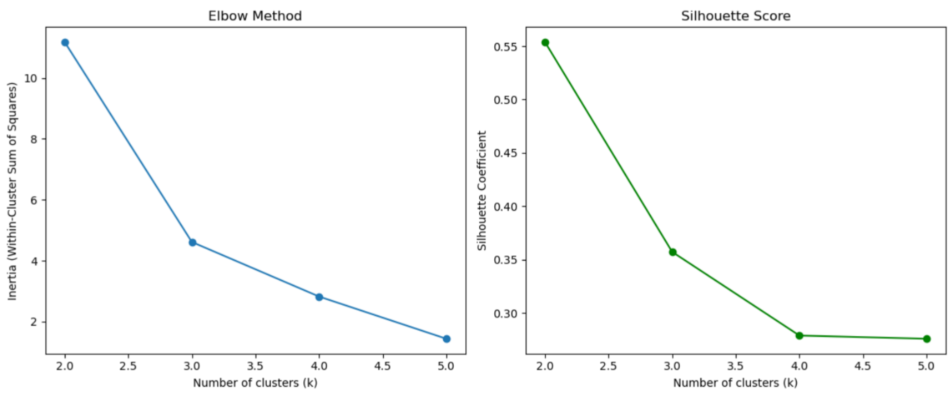


Figure 5.6: Elbow method showing within-cluster variance and silhouette score as a function of the number of clusters

The characteristics of the three clusters that were identified are summarised in table 5.3. Each clusters groups stations that share similar patterns in terms of peak hour usage, seasonality, and daily variability.

Table 5.3: Cluster characteristics of stations on the Hoekse Lijn

Cluster	Stations	Description	Statistics
C1 – Commuter dominant stations	Vlaardingen Centrum, Maassluis Centrum, Maassluis West, Steendijkpolder, Hoek van Holland Haven	Stations with relatively high peak hour usage, moderate seasonality, and high daily variability. Typical commuter stations with concentrated morning peaks.	Peak ratio ∼0.42–0.47; seasonality ∼9–16; daily variability ∼94–112
C2 – Mixed-used urban stations	Schiedam Nieuwland, Vlaardingen Oost, Vlaardingen West	Stations with lower peak hour usage, moderate seasonality, and lower daily variability. Mixed-use stations with more balanced ridership across the day.	Peak ratio ∼0.32–0.43; seasonality ∼12–17; daily variability ∼70–80
C3 – Leisure-oriented and seasonal stations	Hoek van Holland Strand	Station with a distinct seasonal pattern, very low peak ratio, and high seasonality and daily variability. Strongly associated with leisure trips.	Peak ratio ∼0.21; seasonality ∼94; daily variability ∼103

5.2.2. Forecast accuracy by station typology

Table 5.4 displays the weighted mean absolute percentage error (MAPE) of the ridership forecasts for each station cluster. The results highlight clear differences in forecast accuracy across station typologies. *Commuter-dominant stations* (C1) achieved the lowest overall weighted MAPE (27.0%), indicating that the forecasting model performed relatively well in capturing the stable and predictable demand patterns typical of these stations. In contrast, *mixed-use urban stations* (C2) exhibit a significantly higher total MAPE of 55.0%. This suggests that the model struggles to account for the more complex and diverse trip purposes, and temporal variations that characterise the urban environment surrounding these stations. The *leisure-oriented and seasonal stations* (C3), only including Hoek van Holland Strand, also demonstrate poor forecast accuracy, with a weighted MAPE of 54.7%. This emphasises the difficulty of predicting irregular, weather-dependent and highly seasonal travel demand.

Table 5.4: Forecast accuracy by cluster: Weighted Mean Absolute Percentage Error (MAPE)

Cluster	**Stations**	**MAPE Total**	**MAPE Eastbound**	**MAPE Westbound**
C1 – Commuter dominant stations	Vlaardingen Centrum, Maassluis Centrum, Maassluis West, Steendijkpolder, Hoek van Holland Haven	27.0%	19.1%	122.4%
C2 – Mixed-used urban stations	Schiedam Nieuwland, Vlaardingen Oost, Vlaardingen West	55.0%	32.7%	152.9%
C3 – Leisure-oriented and seasonal stations	Hoek van Holland Strand	54.7%	54.7%	–

The directional MAPE values emphasise the overestimation of demand for westbound commuting during the evening peak hours, with westbound errors being particularly at the commuter-dominant and mixed-use urban stations. Another factor that could explain this discrepancy is assumption of a greater number of shorter westbound trips between station on the line than actually occurred. In reality, a larger proportion of travellers travelled from station outside this line, or took alternative transportation.

5.3. Explaining forecast deviations through model input analysis

In order to understand the causes behind the discrepancies between forecasted and observed demand on the Hoekse Lijn, this section evaluates whether the deviations can be explained by inaccuracies in the original model imputs. A comparative analysis was conducted across three dimensions: transport network configuration, behavioural parameters, and socioeconomic forecasts. This was done using by analysing the inputs into the RVMK3.1 model in OmniTRANS.

5.3.1. Model description

The RVMK3.1 model follows a four-step transport modelling framework implented in the OmniTRANS software platform, using 2010 as its base year for calibration, which is desribed in the model documentation (Goudappel Coffeng, 2013). The model begins with trip generation, computing the total number of trips produced and attracted in each zone based on socioeconomic data such as number of inhabitants, employment and land use. These trip generation rates are calibrated to known travel survey data. Next, RVMK3.1 distributes the generated trips between origin□destination (O□D) zone pairs using a gravity□type distribution model. Specifically, the model uses a simultaneous gravity model (SGM): the likelihood of trips between any two zones decreases with the generalised travel costs between them. Crucially, the RVMK3.1 model treats trip distribution and mode choice as a single simultaneous step rather than sequential processes, meaning destination and mode choices are solved simultaneously for car, bicycle, and public transport options.

This simultaneous, multimodal, gravity-based approach sets RVMK apart from traditional, sequential, four-step models, as it incorporates the accessibility of each destination via each mode directly into the trip distribution calculation. Mode choice is determined by logit-type allocation rules that assign trip proportions based on the relative generalised costs of each origin-destination pair. Finally, the model allocates trips to network routes using standard static assignment for car traffic and the Zenith algorithm for public transport. Rather than single shortest paths, Zenith enumerates multiple viable routes between each origin-destination (OD) pair, assigning passengers across alternatives based on their relative attractiveness, vehicle capacity, and transfer possibilities up to specified cost thresholds.

5.3.2. Transport network assumptions

The inclusion of outdated assumptions concerning the configuration connecting public transport services to the Hoekse Lijn could be contributing to the deviations in the forecasts. There are some major changes to the transport network that was included in the 2025 model scenario as compared to the current situation, possibly leading to an overestimation of station accessibility and therefore the number of boardings.

An important difference in the representation of the Hoekse Lijn network concerns the inclusion of Metro Line A. According to the forecast report, Line A was implemented as a peak-hour service between Schiedam Centrum and Binnenhof, operating with a frequency of six trains per hour, and an off-peak frequency of 4 trains per hour (Goudappel Coffeng, 2015). In the RVMK3.1 model a five-train-per-hour service is included for the entire day to represent this service.

In reality, as mentioned before, this service has been terminated since October 2020 due to personnel shortages. The inclusion of this service overstated the frequency and perceived reliability of the Hoekse Lijn in the forecasts. Because waiting time in the model is inversely related to frequency, this led to an underestimation of perceived travel cost for trips by metro. The model applies a wait time route factor of 1.25 to reflect passenger sensitivity to frequency, meaning that the attractiveness of frequent services is disproportionally higher in the generalised cost calculation in the model. As a result, the forecast assigned too many trips to the Hoekse Lijn both during peak and off-peak hours. This effect is also visible in figure 5.1, which shows particularly large deviations between the observed and forecasted number of boardings at the first four stations (Schiedam Nieuwland, Vlaardingen Oost, Vlaardingen West and Vlaardingen Centrum) which would have been served by Metro Line A throughout the day according to the model.

Another significant difference lies in the modelling of the bus network. A detailed comparison between the model's input and the services that are present in 2025 (see table 5.5) reveals several major differences.

Table 5.5: Differences in bus service assumptions per station (model vs. observed)

Station	Lines in model	Actual lines	Differences
Hoek van Holland Strand	-	-	No bus connection
Hoek van Holland Haven	31 & 35	31	Lines 35 and 31 combined, removal of loop through Hoek van Holland
Maassluis West	137, 30, 126	Only 126 (half frequency)	137 and 30 removed; 126 reduced in frequency and coverage
Steendijkpolder	137, 126	Only 126 (shortened + halved)	137 discontinued before opening; 126 truncated
Maassluis Centrum	–	33, 34 (low frequency)	Feeder services not included in model
Vlaardingen Centrum	57	None	Line 57 discontinued in 2019; replaced by on-demand line 557
Vlaardingen West	56	56, 156	Line 156 not in model; route/frequency changes
Vlaardingen Oost	56, 126	56, 126	Line 126 half frequency during peak hours
Schiedam Nieuwland	1, 11, 51, 57	1, 11, 51 (half frequency)	Line 57 removed; frequency of line 51 halved

While the model assumed a well-connected and frequent bus network, in the current situation the bus network is more stripped down. According to RET, these changes were primarily motivated by cost-efficiency considerations and the aim of better aligning services with observed demand as it was expected that the metro would absorb much of the demand that was previously carried by parallel bus services (RET, 2019).

These changes directly imapact the ridership forecasts, as the RVMK3.1 model uses the Zenith multi-routing method to allocate travellers across available routes based on total "impendance" (generalised cost), which includes in-vehicle time, waiting time (derived from frequency), access time and transfer time penalties (Goudappel Coffeng, 2018). A feeder bus link with short acces times and a high frequency, as assumed at some stations in the model, results in a low impendance, making the metro route choice more appealing and therefore attracting more demand. However, in case of the absence of the feeder service or a reduced frequency of the service, the actual impendance is higher and fewer travellers use the metro.

Conversely, the removal of parallel bus services at some stations may have resulted in additional passengers using the metro, since in the model these trips were divided between bus and metro. This could explain the smaller forecast deviations at Steendijkpolder, where local buses were removed in reality, but were included in the model predictions.

Overall these changes show that the accuracy of the forecasts is sensitive to assumptions about the metro service frequency and the outdated bus network, which could explain part of the overestimations in the model predictions for ridership on the Hoekse Lijn. In particular, the assignment of demand in the model was heavily influenced by optimistic assumptions about feeder accessibility and service frequency. Where these assumptions were not realised in practice, the perceived travel cost was underestimated, resulting in inflated demand forecasts, particularly at stations that relied on these feeder services or were expected to be served by higher metro frequencies.

These network representation errors directly explain the concentrated overestimations at the four easternmost stations on the Hoekse Lijn. The inclusion of Metro Line A service in the model, with six trains per hour during peak periods and five trains off-peak, made these stations appear significantly more accessible than they actually became after the service termination in 2020. Since the model applies a wait time route factor of 1.25, the projected high frequencies were disproportionally attractive in the generalised cost calculation, resulting in inflated demand forecasts for Schiedam Nieuwland, Vlaardingen Oost, Vlaardingen Centrum, and Vlaardingen West.

The degraded bus network further reduced actual accessibility compared to the RVMK3.1 model assumptions. Feeder services with short access times and high frequencies, as assumed at some stations in the model, resulted in low impedance calculations that made the metro route choice appear more appealing than it was in reality. Where these feeder services were discontinued, rerouted, or reduced in frequency, the actual impedance was higher and fewer travellers used the metro than forecast.

5.3.3. Behavioural parameters and policy settings

The accuracy of transport demand forecasts is dependent on the behavioural parameters that that determine how travellers make modal and route choices in the model. Analysis of the RVMK3.1 model inputs reveals several parameters that may have contributed to the deviations in forecasted and realised ridership.

Value of time

The model uses different cost coefficients for various trip purposes. This is apparent in the skim-building process, where generalised costs for car and bicycle travel are calculated using the same structure with $g\text{Cost} = (\text{distance} \times \text{distance\_cost\_factor}) + (\text{time} \times \text{time\_cost\_factor})$ for each motif (commuting, business, shopping, education, and other). The distance cost factors for cycling are substantially lower though, with €0.025/km representing mainly wear and tear.

Public transport calculations are considerably more complex, including multiple cost components: $g\text{Cost}_{\text{PT}} = (\text{in-vehicle time} \times \text{time cost factor}) + (\text{wait time} \times \text{wait cost factor}) + (\text{access/egress time} \times \text{time cost factor}) + (\text{transfer time} \times \text{transfer penalty}) + (\text{fare} \times \text{fare cost factor})$. Each component is again also differentiated by the same trip purposes as for car and bicycle trips. This structure enables the model to distinguish, for example, between a short metro journey with a long waiting time at the platform and a longer journey on a high-frequency tram service with no waiting time, as passengers generally perceive waiting time as more burdensome than in-vehicle time, even when the total journey duration is the same.

While the framework used in the model in conceptually robust, the values of time (VoT) embedded in RVMK3.1 are based on CPB (*Centraal Planbureau*, Dutch Bureau for Economic Policy Analysis) 2006 estimates projected to 2010, meaning they no longer represent current user preferences (Goudappel Coffeng, 2013). Recent national research commissioned by the Netherlands Institute for Transport Policy Analysis (KiM) shows significantly higher and more differentiated valuations, even after these values were converted to 2022 price levels using the same method as the KiM study: an inflation correction (factor 1.3258) combined with a real income correction (factor 0.9771), yielding a combined adjustment factor of 1.2953 (Kouwenhoven et al., 2023). Table 5.6 presents the original RVMK3.1 values, the adjusted RVMK3.1 values, and the KiM values from 2022.

Table 5.6: Comparison of value of time (VoT) by mode and purpose: RVMK3.1 (original and adjusted to 2022) vs. KiM (€/hour, 2022)

Mode	Commute				Business				Other			
	RVMK	Adj	KiM	Diff (%)	RVMK	Adj	KiM	Diff (%)	RVMK	Adj	KiM	Diff (%)
Car	€9.09	€11.77	€10.78	-8	€31.47	€40.76	€21.20	-48	€6.28	€8.13	€9.60	18
Train	€9.13	€11.83	€12.05	2	€19.31	€25.01	€17.96	-28	€6.46	€8.37	€8.64	3
BTM	€9.13	€11.83	€7.62	-36	€19.31	€25.01	€14.39	-42	€6.46	€8.37	€6.66	-20
Cycling	€9.09	€11.77	€10.17	-14	€31.47	€40.76	€11.20	-73	€6.28	€8.13	€10.43	28

The differences in assumptions of value of time, as summarised in Table 5.6, are likely to have had an impact on how generalised costs are perceived in the model and on the projected modal split, as these differences affect relative attractiveness between alternatives. When value of time parameters change proportionally across alternatives, the relative attractiveness between modes remains the same, not impacting the modal split. However, when VoT ratios between modes shift, as is the case in the comparison between RVMK3.1 values and the 2022 benchmarks, the balance between transport alternatives is fundamentally altered, directly influencing travellers' choice behaviour. This can be attributed to the logit-based modal choice formulations used in the model, in which the probability of selecting a particular mode depends on the difference in relative utility between alternatives rather than on absolute utility (Train, 2009). Thereby, three systematic effects were identified.

First, for commuting trips, the model applied similar VoT values for car, train, and BTM, which reduced the perceived differences in travel experience across these modes. The latest empirical values from KiM research show that, for commuters, time savings in train travel are valued the most, followed by car, and time savings in bus, tram, and metro are valued the least. According to the study, this can be attributed partly to changes in the mix of travellers and trip types, such as a higher share of higher-educated commuters, evolving comfort expectations, and self-selection effects of travellers. Specifically, higher VoT commuters may have shifted to faster modes such as e-bikes, leaving a lower VoT group within public transport, which reduces the average VoT for bus, tram and metro in reality. (Kouwenhoven et al., 2023). This implies that the model overvalued BTM time, making it appear more sensitive to travel time changes, compared to other modes, than it is in reality.

Second, for business travel, the model applied substantially higher VoTs across all modes than the recent KiM benchmarks specified. In practice, business travellers appear to be less time-sensitive than the model assumes. This difference is largely explained by methodological updates: the 2022 study moved from productivity-based valuations to willingness-to-pay approaches and also reflects COVID-19 effects such as more remote work and fewer in-person meetings, both of which reduced average business VoTs (Kouwenhoven et al., 2023). This means that actual demand may be higher and more evenly distributed across modes, including slower but cheaper alternatives such as BTM or cycling, as the model overestimated time sensitivity for this group.

Third, for discretionary and educational trips, the model applied a higher VoT for BTM than the empirical KiM benchmark, while simultaneously underestimating the value of time savings for car, train, and especially cycling. This means that the model overstated the disutility of BTM time, making BTM appear less attractive than it actually is in reality for longer trips. However, for short, local trips, travellers' preference for flexible, direct options, particularly due to the time savings by e-bikes, lead to greater use of these modes in reality than the model predicts.

Distance cost factors

Besides the value of time, the RVMK3.1 model also uses distance cost factors as part of the general cost calculation. These cost factors have also become increasingly different with real-world transport economics since 2010. These outdated assumptions could also have distorted the relative attractiveness of the different modes of transport included in the model.

The distance cost factors consist of fuel prices, parking costs, and public transport ticket prices (Goudappel Coffeng, 2013). These factors are indexed for the different future scenarios that are included in the model (2015, 2020 and 2030). To account for trends in fuel efficiency of cars, the model applied a separate efficiency index to the fuel price indices. For this analysis, the model values were compared with realised consumer price indices from CBS (Dutch Central Statistical Office), which track actual price developments for fuel and public transport (CBS, 2025a). A corresponding efficiency index was created for the CBS price indices to allow for consistent comparison. This index was based on estimates of the average fuel consumption of the car fleet and the emergence of electric vehicles.

Figure 5.7 shows the indices used in the model, as well as the realised indices for train and BTM tariffs relative to fuel costs (2010 = 100). Although the model predicted that public transport would become more expensive than fuel, the actual price developments show that public transport became even more expensive: train fares increased by around 25% more than fuel costs, and BTM fares by around 46% more, between 2010 and 2024.

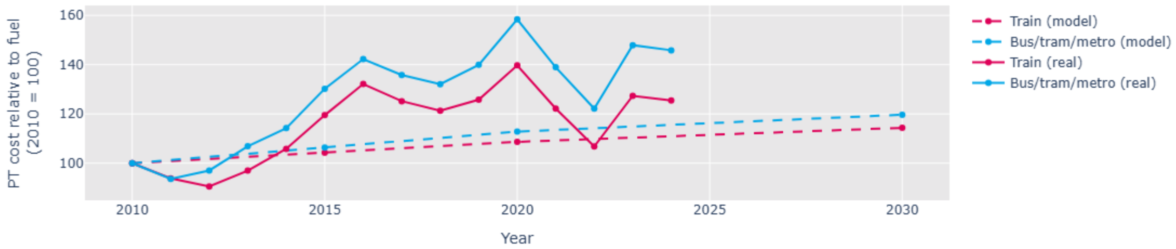


Figure 5.7: Public transport cost relative to fuel cost (2010 = 100). Observed data from CBS consumer price index (CBS, 2025a)

Unlike ticket prices and fuel prices, the model assumes no change in parking charges over time. In practice, however, parking rates have risen considerably, with the highest rate increasing from €3.33 per hour in 2015 to €6.18 in 2025, an increase of almost 100% (Gemeente Rotterdam, 2025). Additionally, the paid parking zones have expanded more than anticipated in the model, particularly in areas in Rotterdam South. This evolution has made car use to these areas significantly less attractive and has increased the relative competitiveness of public transport. Around the Hoekse Lijn, paid parking now applies in the city centres of Vlaardingen and Schiedam, as well as in almost all parts of Rotterdam.

These discrepancies between the model's assumptions and the actual developments mainly affected the relative attractiveness of public transport compared to private transport modes. The model underestimated the increase in public transport fares compared to fuel costs. Consequently, the forecasts overstated the appeal of public transport and understated the competitiveness of car use. For shorter journeys, this made cycling also a more appealing alternative than predicted by the model. Although higher parking costs dampened the shift towards car use to some extent, overall, the model overstated the attractiveness of public transport.

Synthesis

These outdated behavioural and economic parameters systematically overestimated public transport attractiveness relative to emerging alternatives, contributing to the overestimations ranging from 28% to 75% across most stations. The frozen 2010-level value of time parameters, which applied similar time sensitivities across car, train, and BTM for commuters, overstated BTM's sensitivity to travel time changes compared to recent empirical evidence showing lower valuations for bus, tram, and metro travel.

The distance cost factor discrepancies further intensified these effects. Public transport fares increased by 25-46% more than fuel costs between 2010 and 2024, making the metro less competitive than projected in the model. While higher parking costs partially counteracted the shift away from public transport, the combined effect of outdated parameters overstated the appeal of metro services and understated the competitiveness of alternatives, particularly e-bikes for medium-distance trips.

5.3.4. Socioeconomic forecasts

An important element of the RVMK3.1 model is the use of socioeconomic data to calculate the number of departing and arriving trips for each model zone (trip generation). Therefore, socioeconomic forecasts are essential for the model to be able to estimate future transport demand. Discrepancies between these forecasts and actual developments in population, housing or employment can therefore lead to significant differences between predicted and observed travel patterns.

Population and housing growth

The socioeconomic forecasts in the model are made at the level of the former Stadsregio Rotterdam, which comprises 15 municipalities. Between 2010 and 2020, the population of this region increased by 7.26%, which is significantly higher than the 3.82% growth assumed by the (Goudappel Coffeng, 2013). A closer look at the municipalities served by the Hoekse Lijn reveals a similar pattern. Maassluis, Vlaardingen and Schiedam each saw an increase in population of around 4–5%, while Rotterdam's population grew by 10% (an increase of almost 60,000 inhabitants). This contrasts starkly with the model's socioeconomic forecasts, which predicted a 6% increase for Rotterdam and a population decline of around 3–4% for Maassluis, Vlaardingen and Schiedam.

This stronger-than-expected population growth is reflected in the increase in the housing stock. Maassluis, Vlaardingen and Schiedam added 274, 3,083 and 2,084 homes respectively between 2010 and 2020 (CBS, 2025b), whereas the model assumed that the housing stock in these municipalities would remain stable during this period. Several spatial developments have contributed to this growth, notably projects such as Park Haga, Parkweg Noord, and Park Vijfsluizen in Schiedam and Vlaardingen. Furthermore, ongoing development in Maassluis, including Wilgenrijk and De Kade, will add more than 2,000 additional homes in the coming years (Gemeente Maassluis, 2025; Limmen, 2017).

Employment trends

The model assumed an increase in employment in Vlaardingen, Schiedam and Maassluis between 2010 and 2020. However, according to CBS data, there was a net loss of around 4,000 jobs accross there three municipalities, compared to the anticipated increase of 5,000 jobs (CBS, 2025b). Conversely, employment growth in Rotterdam was substantially higher than expected with 22,200 jobs instead of 14,227. This discrepancy suggests a stronger concentration of employment in Rotterdam, likely resulting in increased commuting to the city centre and fewer eastbound trips in the during the afternoon peak on the Hoekse Lijn.

At Schiedam Nieuwland, the model correctly identified significant employment at the Franciscus Vlietland Hospital, however, it appears that the model assumed standard 09:00-17:00 working hours for these jobs, neglecting the fact that the hospital operates day and night. This means that the actual travel demand is likely to be more evenly distributed throughout the day and night, with less concentration during peak periods than the model estimated.

Synthesis

The employment distribution errors particularly explain the systematic overestimation of westbound demand, with deviations ranging from 57% to 90% at most stations. While the model predicted an increase of 5,000 jobs in Schiedam, Vlaardingen, and Maassluis, actual employment in these areas declined by 4,000. Conversely, Rotterdam gained 22,200 jobs versus the projected 14,227. This employment concentration in Rotterdam rather than dispersal along the corridor meant that the expected reverse commuting flows during the afternoon peak did not materialise.

The population and housing growth errors, while initially appearing to support higher ridership, were offset by fundamental changes in travel behaviour. Despite 4-5% population growth in corridor municipalities (versus projected 3-4% decline) and 10% growth in Rotterdam (versus projected 6%), the expected ridership increases did not materialise. This indicates that the emergence of the e-bike as a viable alternative played a decisive role in travel behaviour shifts. By 2021, 20% of the Dutch population owned an e-bike, competing directly with metro services for the 5-15 kilometre trips typically served by the Hoekse Lijn. Research shows e-bikes replaced around 12% of public transport trips, helping explain why population growth did not translate to proportional ridership increases.

These socioeconomic forecasting errors interacted with the network and behavioural parameter errors identified in the previous subsections. The inclusion of the terminated Metro Line A service made westbound travel appear disproportionally attractive for the anticipated (but unrealised) reverse commuter flows, while outdated value of time parameters overestimated business and commuting travellers' sensitivity to travel time savings, further inflating the forecasted afternoon return flows.

5.4. Concluding synthesis

After applying the -12% structural correction for the impact of the COVID-19 pandemic, the remaining deviations between forecasts and actual passenger figures can be attributed to the factors set out earlier in this chapter. The figures below express shares of the total residual deviation and are intended as educated estimates to support the waterfall chart in figure 5.8.

The first component that was identified are differences in network assumptions. The 2015 scenario assumed an additional service by Metro Line A of 6 trains per hour during peak hours and 4 trains per hour off-peak (modelled as 5 trains per hour during the entire day). In reality this service was not operating in 2025 and feeder bus links were also more limited than anticipated due to reduced frequencies and line discontinuations. Furthermore, the 2015 forecasts were based on a fully integrated service by 2025, however the conversion to a metro line was delayed by two years with the beach extension only operational in 2023, giving less time for demand to build up. At the four eastern stations, which account for roughly 70% of the forecasted boardings (the sum of the east and west projections from table 5.1), the assumed service improvement implies an effective frequency change of approximately 70% across the day. Using a conservative service elasticity of 0.5 (Paulley et al., 2006), this results in effects at station level of a 35% decrease in boardings. Translating this with the 70% share in boardings suggests a corridor impact of 25%. Adding to that the effects of the alteration of the surrounding bus network, plausibly brings the effects of the deviation in the transport network to around 30%.

Second, behavioural and economic parameter alignment was also identified as an important factor contributing to the deviations between forecasted and realised passenger numbers. The key parameters (value of time and distance cost factors) were calibrated on patterns and preferences from 2010 that were assumed to remain stable, while costs for public transport rose faster that costs for the use of private cars and commuting during peak hours weakened. In quantitive terms, the distance-cost aspect can be approximated by applying standard short-run public transport elasticities to the generalised cost (approximately -0.4, (Paulley et al., 2006)) and the observed relative change in prices. With public transport becoming approximately 25% to 46% more expensive then cars between 2010 and 2024 (see figure 5.7, this implies a demand effect of around 10% to 18%. Regarding the value of time, the public transport against car VoT ratio for commuting decreases from about 1.0 in RVMK3.1 to about 0.71 in the KiM 2022 benchmarks (table 5.6; (Kouwenhoven et al., 2023)). If is assumed that time is roughly half of generalised costs on an average trip, this 29% decrease in the weight of public transport time savings equates to 15% less effective time weight. Taking the same price elasticity, this implies an extra 6% demand effect, bringing the overall attribution of the misalignment of behavioural and economic parameters to approximately 20%.

Furthermore, the model over-allocated job growth to Schiedam, Vlaardingen en Maassluis (+5,000 forecasted versus -4,000 observed) and under-allocated to Rotterdam (+22,200 observed versus +14,227 forecasted), which weakened reverse-commuting flows and reduced westbound boardings. Using published destination-employment elasticities for commuting of about 1.3 (Transport Scotland, 2015), and applying them directionally to the observed reallocation of jobs (i.e., with a larger eastbound commute share than westbound) yields an estimated corridor-level effect of 1% in absolute demand. Relative to the total residual deviation, this equates to 5% attributable to socioeconomic inputs.

Additionally, e-bikes turned oud to compete directly with public transport for trips of 5 to 15 kilometres, which are common along the corridor, and where not represented in the original model. Using this estimate, the implied replacement is around 2% of total PT demand (roughly a quarter of adults owning an e-bike by 2024 (Huang et al., 2024)), with about a quarter to a third of e-bike trips drawn from public transport and a significant proportion of corridor trips in the 5–15 km range (Sun et al., 2019)). Relative to the total residual deviation, this corresponds to around 10% that can be attributed to unmodelled e-bike competition. In this percentage, an overlap factor of 0.8 is taken into account, as some e-bike shift is already explained by the other factors.

On the other hand, Hoek van Holland Strand significantly outperformed the forecast on an average working day, with 760 eastbound boardings observed versus a forecast of 344 (+416, i.e. +121%). Set against the corridor's residual underperformance of 5,302 boardings, this surplus reduces the deficit by 416/5,302, which is approximately 7.9%. As a cross-check, Strand accounts for around 5.4% of corridor boardings (approximately 738 out of 13,690 on a monthly basis). Applying the 121% over-performance to this proportion gives 5.4% × 121% ≈ 6.5%. Taking this into account, a conservative average attribution of 7% (range 6–8%) is used.

Taken together, the attributions suggest that approximately 60% of the residual deviation can be explained by differences in network representation (30%), misalignment of behavioural and economic parameters (20%), socioeconomic inputs (5%) and unmodelled competition from e-bikes (10%). This is partially counterbalanced by stronger leisure demand at Hoek van Holland Strand (-7%). The remaining 40% represents as a combination of interaction effects and the inherent methodological limitations of smart card data analysis, rather than a single missing cause.

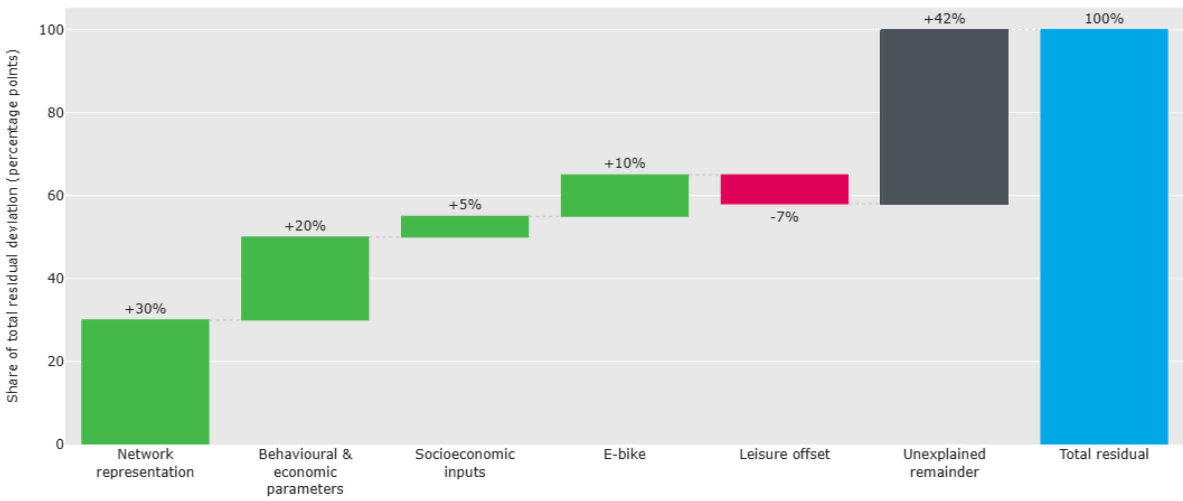


Figure 5.8: Waterfall graph showing attribution to the total residual forecast deviation.

Interaction effects occur when multiple factors reinforce each other (e.g., reduced service frequencies combined with increased generalised costs and e-bike adoption), resulting in a combined impact that is greater than the sum of the individual effects, which cannot be cleanly split (Koppelman & Bhat, 2006). Ramp-up effects also play a role: ridership usually increases gradually in the first years after opening (and after the beach extension in 2023), as passengers familiarise themselves with the service, feeder links become established, and timetables stabilise (Shinn & Voulgaris, 2019). Therefore, early counts tend to understate the eventual steady-state level. Further explanations could be seasonality and weather not being captured by working-day averages, temporary timetable changes and disruptions, and endogeneity, whereby service levels respond to observed demand, resulting in reductions that then lead to further declines in demand (Rahman et al., 2019).

These do not only apply to the Hoekse Lijn. The next chapter will continue to translate these findings into broader recommendations with the ultimate goal to reduce the unexplained gap in future projects by minimising untested interactions and improving the evidence base used to calibrate and evaluate models.

6
Implications and recommendations

The Hoekse Lijn case study revealed four systematic deviations that demonstrate structural limitations in current modelling approaches rather that isolated forecasting errors. These deviations were spatial asymmetry in forecast errors, consistent direcitonal bias overestimating westbound travel, systematic trip purpose misalignment that overestimated commuter demand while underestimating leisure demand, and fundamental underestimation of trip length distributions. These patterns highlight the broader challenges that are faced in transport planning influenced by technological advancement such as e-bike adoption, change working patterns including remote or hybrid working, and external shocks such as the COVID-19 pandemic.

In this chapter, the case-specific insights from the Hoekse Lijn evaluation are translated into actionable strategies for planners and policymakers to improve the reliability and robustness of future transport demand forecasting practices.

6.1. Implications for practice

The case study provides a broader insight into the challenges of transport demand forecasting and highlights several implications for forecasting practice. The first important lesson is that the outputs of forecasts are inherently uncertain and biased as even well-established models can deviate significantly form real-world outcomes, as demonstrated by the the case study's findings of systematic overestimation of commuter demand and underestimation of leisure travel. These discrepancies are not just random fluctuations, but reveal systematic biases, often leading towards optimistic future ridership figures. This reinforces the established observation that inaccurate forecasts can have serious consequences for infrastructure investment and policy decisions (Flyvbjerg et al., 2005). In other words, forecasts should be used to indicate direction and scale, and support comparisons between alternatives and they should not be interpreted as guaranteed outcomes or precise targets.

Another implication for those working with transport models is the need to critically examine the scope of the model and its assumptions. As traditional transport models tend to focus on regular commuting patterns, irregular travel demand may be poorly represented (Fulman et al., 2023). The findings from the Hoekse Lijn case study illustrate this issue as recreational trips were affected by factors such as the weather in ways that the model failed to predicts, resulting in peaks in demand that were far beyond what was projected. Broadly speaking, the treatment of occasional leisure travel as an additional static component outside of the core model is inadequate, as prior work shows that models tend to prioritise commuting and underrepresent leisure or induced demand (Flyvbjerg et al., 2005). This pattern was also evident in the case study, highlighting a broader limitation in current forecasting practice: the assumption of stable, homogeneous travel behaviour while in reality different types of trips behave fundamentally differently. Commuter travel follows predictable patterns while leisure trips are highly dependent on weather, events or seasonal factors. When models treat everything the same using average conditions without considering these extremes, they will systematically fail to predict certain types of demand. Effective forecasting needs to be flexible enough to handle both the steady flow of daily commuters as well as the unpredictable surges of recreational travellers.

The case study also shows how external factors can completely disrupt carefully planned forecasts. Demand forecasts typically assume that historical patterns and current data can accurately predict future usage. However, sudden societal changes and unexpected events often happen outside what models can capture. The COVID-19 pandemic perfectly illustrates this as it caused massive shifts in how people travel that no traditional model could have predicted. Beyond pandemics, policy changes, economic fluctuations and climate events can also create significant discrepancies between forecasts and actual outcomes. This means that there is a need for forecasting approaches that are more resilient and flexible. Rather than treating forecasts as fixed predictions, transport planners and policymakers should incorporate scenario-based analysis and sensitivity testing to consider different possible future scenarios, as earlier studies already suggested (Bojada, 2014; Tempert et al., 2010).

Finally, the Hoekse Lijn case study has demonstrated the value of systematic evaluation and data-driven feedback for enhancing transport forecast accuracy. Smart card data enabled detailed comparisons to be made between predicted and observed ridership patterns that revealed systematic biases that would otherwise remain undetected. The absence of standardised evaluation frameworks is a critical gap that allows forecasting errors to persist and prevents institutional learning, as otherwise model assumptions could be refines continuously as planners learn which predictions were accurate and which were not. Regular evaluations should become standard practice in order to refine model assumptions based on empirical outcomes. Integrating automated data sources with standardised evaluation frameworks is necessary in order to develop more robust forecasting methodologies that reflect the complexity of modern multimodal systems and support better informed decision-making.

6.2. Actionable recommendations

The detailed analysis of forecast deviations and model input shows that improving the accuracy of multimodal transport forecasting requires changes across multiple interconnected dimensions. Rather than addressing individual model components in isolation, the evidence from the Hoekse Lijn case study shows that forecast inaccuracies stem from multiple, reinforcing sources that must be addressed systematically. This section presents strategies for forecast improvement that build directly on the findings presented in the case study.

Systematic ex-post validation

While a growing number of studies have begun to explore the ex-post evaluation of transport forecasts, these analyses remain relatively scarce and are often limited in scope (Brands et al., 2020; Hussain et al., 2021). The evaluation of ridership for the Hoekse Lijn showed that eight out of ten stations underperformed the forecast with significant differences and clear directional biases. These findings highlight the importance of routine ex-post evaluations in identifying structural issues in forecasting that would otherwise remain hidden.

International experience demonstrates the value of systematic evaluation frameworks. The Post Opening Project Evaluation (POPE) in the United Kingdom and Norway's post-opening evaluations provide include systematic transport demand validation as a core component of their transport project evaluation schemes (Jong et al., 2019). However, there is still significant variation in the quality, consistency and scope of these systems within individual countries. This lack of standardisation limits the comparability of findings, reducing the potential for shared learning and continuous improvement in forecasting practices (Nicolaisen & Driscoll, 2016).

Strategic approach:

- **What:** Implement standardised, continuous ex-post evaluation as a routine practice in transport planning, rather than relying on one-time evaluations.
- **How:** Use automated and continuing data sources (e.g. smart card data, automated passenger counts) to continuously compare forecasted demand with actual ridership. This creates a standardised feedback loop in which discrepancies are regularly identified and used to improve forecast models, with the goal to go beyond traditional one-time evaluations.
- **Who:** Transport authorities should implement and oversee these evaluation programs. They can collaborate with transit operators (for data provision) and research institutions to analyse the results. Additionally, the government must play a strong role to ensure data availability by establishing a mandatory database that openly accessible to those who need it. This would force transit operators to share ridership and operational data, which they might otherwise withhold due to concerns about competition.

Adaptive parameter calibration

The evaluation of the model's input parameters revealed that outdated behavioural assumptions had led to a systematic overestimation of demand forecasts for the Hoekse Lijn. More recent research shows that travellers now prioritise time savings differently across modes and for different trip purposes than was assumed in the original model. Additionally, the distance cost factors did not reflect the shift in relative travel costs, particularly the disproportionate increase in public transport fares compared to car travel. This further distorted the projections for mode choice.

Strategic approach:

- **What:** Develop adaptive parameter management systems that regularly update behavioural and economic parameters using current data and evolving travel patterns, ensuring that the model's assumptions remain aligned with reality.
- **How:** Set up a schedule for periodically recalibrating important parameters, such as value of time, fare elasticity, mode-specific constants and transfer penalties. This should be done using up-to-date data sources, such as recent smart card travel data, travel survey results and macro-economic indicators (e.g., fuel prices, transit fares and inflation rates). By continuously integrating real-world data into the model, planners can adjust the parameters to reflect changing traveller preferences and economic conditions.
- **Who:** The organisations that are responsible for transport models (e.g., transport authorities, transport operators, consultancy firms) should be leading the adaptive calibration. Model developers and analysts must collaborate closely with providers of data, such as transport agencies and operators, to gather up-to-date data.

Scenario based uncertainty management

The permanent reduction in public transport demand caused by the pandemic is the type of structural change that traditional forecasting methods, which rely on fixed assumptions, cannot anticipate. Other external factors, including the adoption of e-bikes and policy changes, also fundamentally altered the competitiveness of different modes of transport after the original forecasts were done.

Decision Making under Deep Uncertainty (DMDU) approaches have emerged as a response to such challenges. Rather than relying on a few sensitivity tests of individual parameters, DMDU frameworks assess policies across a wide range of plausible future scenarios by systematically varying key input parameters within defined ranges. This allows planners to identify robust strategies that can withstand different assumptions about future conditions (Engholm & Kristoffersson, 2024).

As also noted in the literature review, several academics and practitioners advocate the use of bandwidths and scenario-based forecasting rather than single-point estimates, with emphasis on providing realistic ranges of high and low demand to account for uncertainty in future developments (Bojada, 2014; Tempert et al., 2010).

Strategic approach:

- **What:** Adopt scenario-based forecasting methodologies that systematically explore future conditions and provide probability-based outcome ranges for demand indicators.
- **How:** Integrate scenario analysis and DMDU techniques into forecasting processes. In practice, this involves defining a range of potential future conditions, such as scenarios with higher or lower economic growth, different rates of remote working adoption, technological changes like increased e-bike usage and policy shifts in pricing or regulation. For each scenario, the model outcomes should be communicated as bandwidths or confidence intervals for key indicators (such as ridership, revenue or modal share) rather than as a single number.
- **Who:** Transport planning agencies and policymakers should require scenario-based uncertainty analysis as part of model forecasting (and project evaluation). This process needs to be carried out by analysts and modelling teams, who may require support from experts in scenario planning or academic researchers familiar with DMDU methodologies. In the end, organisations that use forecasts, such as governments and transport authorities, need to accept uncertainty. Rather than expecting one exact prediction, they should get used to working with forecast ranges and base their decisions on different possible outcomes.

Integrated network planning

The exploration of transport network assumptions in the model identified substantial forecast deviations at eastern Hoekse Lijn stations resulting from unrealistic service assumptions, including terminated Metro Line A service and scaled-back feeder bus networks. These network representation errors demonstrated how forecasting models failed to anticipate the systemic effects of new infrastructure on existing transport networks and the vulnerability of service assumptions to policy and budgetary changes.

Strategic approach:

- **What:** Adopt integrated network planning whenever new transport infrastructure is introduced. This means planning the new line or service alongside the existing network rather than treating it as an isolated addition, with the aim to ensure that the broader network's effects and operational realities are considered. This includes anticipating different service scenarios, for instance, planning not just for a best-case high-frequency service, but also for more constrained scenarios where budget or policy might limit the service level.
- **How:** Systematically evaluate and adjust the surrounding network as part of the network planning process, incorporating scenario planning into the forecasts. In practice, this involves coordinating service plans. For example, when a new metro or rail line is introduced, determine how it will affect other routes (e.g. metro lines and bus lines) and plan for those changes in advance. Furthermore, the financial sustainability of assumed service levels under varying policy scenarios should also be tested. Ultimately, iterate between infrastructure design, service scheduling and budget constraints to ensure that the modelled accessibility (frequency, coverage and connectivity) matches what can actually be delivered.
- **Who:** Planning authorities and transport agencies should coordinate operators, municipalities, and consultants, setting and enforcing integrated, cross-modal planning requirements. Consultants and modellers embed network-wide effects and scenario-based service assumptions in forecasts, working with operators and authorities to test long-term affordability so proposed frequencies and coverage remain sustainable under varying budget scenarios.

6.3. Implementation considerations

Implementing the four strategies does not only require technical adjustments, but also institutional and governance changes. Firstly, these practices must be made routine rather than optional through clear mandates. Transport authorities should incorporate ex-post validation into funding agreements and project lifecycles, similar to the UK's Post Opening Project Evaluation (POPE), which requires systematic reviews one and five years after opening (Jong et al., 2019). This would ensure that standardised forecast evaluations become a condition of investment, with transparent reporting of results to project funders and policymakers, transforming forecasting practices from a one-off deliverable into the beginning of an institutional learning cycle.

Secondly, adaptive parameter calibration and integrated network planning are highly dependent on data sharing and collaboration. Ideally, a government-led data bank should provide open aggregate figures and controlled access to anonymised smart-card data to support this. Transport operators, (regional) authorities and planning agencies should agree on clear data-sharing rules, including common formats, timeliness and privacy guidelines, as well as who is responsible for managing the data. For each major project, a small, temporary working group should be set up existing of parties like the project owner, relevant transport operators and regional authorities. This group should agree on service scenarios to be modelled, as well as confirming the matching budget and policy commitments, before the forecast finalised. This keeps the process manageable and helps prevent unrealistic service assumptions from being included in the assessment.

Finally, there must be a shift in organisational culture, moving from a ”predict-and-forget” approach to a ”predict-and-learn” one. Leadership should encourage open discussion of forecast errors to improve their credibility, rather than assigning blame. Clear lines of communication and defined roles are essential, for example, designating who triggers ex-post evaluations, who oversees parameter updates, and who ensures coordination across stakeholders. Without these foundations, even well-designed strategies will fail to be embedded in everyday practice.

7 Conclusion and discussion

7.1. Conclusion

The aim of this research was to improve multimodal transport demand forecasts through an ex-post analysis of the Hoekse Lijn train to metro conversion project, using *OV-chipkaart* smart card data as the primary empirical basis. By systematically comparing ridership forecasts from 2015 with observed passenger flows from 2020 to 2025, this study sought to identify systematic biases in transport models and distinguish between model limitations and external influences, with the ultimate aim to propose improvement for future forecasting practices.

In this conclusion, the research questions that structured this study will be addressed. Each sub-question will be answered in turn, followed by an integrated synthesis that answers the main research question.

How can OV-chipkaart smart card data be used to evaluate the transport demand for the Hoekse Lijn, and what are its limitations?

The primary advantage of smart card data over traditional data collection methods is the enabling of large-scale, continuous data collection (Lee et al., 2014). Unlike manual passenger counts or survey data, which only provide snapshots of demand, the provided *OV-chipkaart* data for this research offered daily boarding figures from 2020 to 2025, revealing temporal patterns that would be far more strenuous to detect using conventional approaches, such as peak versus off-peak usage, seasonal variations and irregular patterns such as beach travel. When the aggregated data lacked certain indicators, directional ratios from the OV-Lite model enabled reconstruction of complete trip patterns, load profiles, and passenger kilometres for direct comparison with original forecasts.

In addition, standardising and making evaluations comparable builds accountability and helps counteract overly optimistic forecasts. As demonstrated in (Flyvbjerg et al., 2005), systematic overestimation frequently results from optimism bias or strategic behaviour. Publishing comparable ex-post results for each project and indicator increases the cost of unrealistic assumptions from the beginning and encourages the provision of more balanced inputs and explicit uncertainty ranges. In settings without such a system in place (including the Netherlands), the national government should make a standardised evaluation framework a condition of funding to ensure that lessons are learned and to actively counter optimism bias.

However, significant limitations emerged that must be considered in future applications. Missed check-outs, barcode tickets, and OVpay adoption introduced systematic undercounting, with monthly correction factors underestimating ridership by up to 18% on leisure-heavy days. Additionally, the absence of behavioural attributes and potential representation biases limit the explanatory power of smart card data. Without trip-purpose or socio-demographic information, travel motivations and user profiles must be inferred indirectly, restricting insight into why forecast discrepancies occur (Liu et al., 2018). Furthermore, smart card data often under-represents certain user groups, such as tourists and occasional riders, who use different payment methods, such as paper tickets or bank cards (Mahajan et al., 2022).

In summary, when properly processed, smart card data provides a robust, low cost foundation of ex-post demand validation. The Hoekse Lijn case demonstrates how the systematic processing of temporal data can deliver valuable insights into demand patterns that would otherwise be difficult to obtain.

However, incomplete transactions, payment heterogeneity and the absence of behavioural attributes must be considered when interpreting the results. Therefore, to maximise the analytical value, smart card data should be combined with other sources of information to provide context on trip purpose and traveller motivations, providing a more complete and reliable evaluation of transport demand.

How does the actual transport demand on the Hoekse Lijn compare to the forecasted values from 2011 and 2015, and what systematic deviations can be identified?

The Hoekse Lijn conversion was designed to achieve several strategic objectives that formed the basis for the transport demand projections. The 2015 forecasts estimated a 52% increase in average boardings on the section of the line between Schiedam Nieuwland and Hoek van Holland Strand, and a 28% increase in total passenger kilometres compared to the scenario where the existing rail service was continued. The project also aimed to optimise the utilisation of the line throughout the day and in both directions, addressing the strong imbalance between peak and off-peak travel that characterised the former rail service. Additionally, the extension to Hoek van Holland Strand was expected to reduce the total travel time from the centre of Rotterdam by 26%, thereby increasing beach visitors arriving by public transport with 10-20% (Gemeente Rotterdam, 2015; Goudappel Coffeng, 2015). These strategic objectives provided clear performance indicators against which the accuracy of the forecasts could be evaluated, particularly regarding ridership growth, improvements in temporal distribution, and enhancement of recreational travel.

The comparison revealed substantial systematic deviations from the 2015 forecasts. Eight of ten stations experienced underperformance ranging from -9% to -40% for total weekday boardings relative to COVID-corrected projections, with the largest errors concentrated at the four easternmost stations. Westbound boardings fell more than 50% below forecasted levels at most stations, while exceptions included Maassluis Steendijkpolder (+2%) and Hoek van Holland Strand (+121%). These percentage deviations should be considered alongside absolute flows. Low forecast base figures, especially in westbound direction, can result in significant percentage errors from modest absolute differences. For operational purposes, the absolute gaps are the most important factor.

Despite lower station-level boardings, total passenger kilometres exceeded forecasts by 18% due to longer trip lengths and unexpected recreational travel patterns. Beach-related travel generated up to 27 times normal boardings on peak days, demonstrating the project's success in enhancing leisure accessibility.

These findings highlight the mixed performance of the Hoekse Lijn against its strategic objectives. While the target of 52% ridership increase was not achieved at most stations, the project far exceeded its goals for beach accessibility and total passenger kilometres as a result of strong long-distance, recreational travel. Furthermore, the analysis revealed four systematic deviations in the forecasts:

- **Spatial asymmetry:** Easternmost stations showed the largest absolute errors in number of boardings.
- **Directional bias:** Westbound boardings were consistently overestimated across nearly all stations.
- **Trip purpose misalignment:** Commuter demand being overestimated and leisure travel being underestimated.
- **Trip length distribution change:** Average trip length were fundamentally underestimated, with a higher proportion of longer-distance trips and integration into the bigger Rotterdam metro network attracting travellers from outside the corridor.

So while the project achieved mixed succes against its strategic objectives, the forecast deviations demonstrated the systematic limitations in capturing the full complexity of demand patterns on the Hoekse Lijn, particularly in the balance between regular commuter flows and irregular recreational travel.

To what extent do discrepancies between forecasted and observed demand stem from biases or structural limitations within the transport models?

Using a systematic analysis of the model inputs, it was revealed that the majority of the forecast deviations is due to structural model errors, particularly unrealistic service supply and outdated behavioural and economic parameters, rather than random noise. Four categories of structural limitations that distorted the forecast outcomes for the Hoekse Lijn in particular have been identified.

The first limitation was inaccurate service assumptions: the model included high service frequencies for Metro Line A that were never realised, resulting in inflated demand forecasts for stations in the east of the city. The second limitation was that the behavioural and cost parameters were outdated, such as the value of time and fare sensitivity. These no longer aligned with current user preferences or the shift in relative transport costs. Thirdly, errors in the network representation arose from the model's assumption of an extensive feeder bus network, despite this having already been scaled back by 2019, which reduced actual accessibility. Finally, socioeconomic inaccuracies affected spatial and directional demand estimates: the model overestimated job growth in Schiedam, Vlaardingen and Maassluis while underestimating employment concentration in Rotterdam.

These structural limitations in the model input show that the forecast discrepancies did not arise from random variation or minor calibration errors, but from a fundamental misalignment between the model's assumptions and the actual conditions. Together, these factors help explain why deviations were not isolated or incidental, but mostly consistent across stations, directions and demand indicators.

What external factors, such as socioeconomic developments, policy changes, and the COVID-19 pandemic, contributed to deviations between forecasted and actual transport demand?

While structural model limitations explain a large part of the forecast deviations on the Hoekse Lijn, several external factors that emerged after 2015 have also contributed significantly to the discrepancy between the predicted and observed demand. These factors have altered travel behaviour and the competitiveness of public transport in unforeseen ways.

Key external influences included the lasting behavioural shifts caused by the pandemic, such as increased homeworking and reduced commuting, as well as a broader move towards private transport. The adoption of e-bikes further disrupted traditional travel choices by competing directly with metro services for medium-distance trips, which are typical of the Hoekse Lijn corridor. Meanwhile, public transport fares increased considerably more than fuel prices, reducing competitiveness, even as Rotterdam introduced stricter parking policies. Spatial developments also diverged from the model's assumptions: population and employment growth was concentrated in Rotterdam rather than in the surrounding municipalities, which weakened the expected reverse commuting flows. On the other hand, the 2023 extension of the line to Hoek van Holland Strand generated beach-related demand that far exceeded expectations, with peak boardings reaching 27 times the weekday average.

These external factors did not operate in isolation, but rather reinforced each other to create a cumulative effect that amplified their impact on demand for public transport. The shift to remote working during the pandemic significantly reduced commuter travel. At the same time, the pandemic changed people's mode preferences, switching from public transport to private modes of transportation (Gkiotsalitis & Cats, 2021). This decline in ridership happened at the same time as a rapid rise in e-bike usage, as these offered a convenient and affordable alternative at a time when public transport fares were rising. Although parking rates in central Rotterdam also rose during this period, potentially making public transport more attractive, the combined effect of higher ticket prices and the growing appeal of e-bikes meant that public transport usage did not increase substantially.

These exogenous shocks altered mode preferences and travel volumes in ways that the original 2015 forecasts could not have anticipated. The success of the beach extension partially compensated for losses caused by the pandemic, but it also created new seasonal volatility that challenged traditional demand modelling approaches. Together, these external factors demonstrate how vulnerable long-term transport forecasts are to unforeseen technological, policy and societal disruptions that can reshape travel behaviour within a short timeframe.

How can multimodal transport models be improved to enhance the accuracy of future public transport demand forecasts?

To improve multimodal transport forecasting, four strategies were proposed to address the fundamental limitations identified in the Hoekse Lijn case study.

- **Systematic ex-post validation:** Establish standardised continuous ex-post validation schemes using using automated data sources, such as smart card data, to create feedback loops between forecasting practice and observed outcomes, going beyond traditional one-time evaluations.

- **Adaptive parameter calibration:** Implement adaptive parameter management that regularly updates behavioural and economic parameters using real-time data sources, addressing the systematic overestimations that caused by outdated assumptions in the model.
- **Scenario-based uncertainty management**: Adopt scenario-based forecasting methodologies that systematically explore uncertainty across major assumptions, providing probability-based outcome ranges rather than single-point estimates that are more vulnerable to external disruptions like COVID-19.
- **Integrated network planning**: Implement network planning that evaluates how new infrastructure could reshape existing service networks and assesses financial sustainability of service assumptions.

The Hoekse Lijn case study shows that improving forecast accuracy means addressing multiple interconnected dimensions at the same time, rather than looking at individual model components separately. The systematic nature of the identified deviations highlights significant opportunities to enhance current forecasting practices and better capture the complexity of modern multimodal transport systems.

However, implementing these strategies requires fundamental institutional and governance changes, rather than just technical adjustments. Most critically, these must be a shift in organisational culture, moving from a ”predict-and-forget” approach to a ”predict-and-learn” methodology. Ex-post evaluations should be mandated and published as part of funding conditions to turn forecasting in a continuous cycle of planning, evaluation and improvement, which also helps to mitigate the effects of optimism bias.

Data management infrastructure is essential for this transformation. Ideally, a government-led database would provide open, aggregated indicators and controlled access to anonymised smart card data to support routine validation. Transport operators, regional authorities, and planning agencies must agree on clear data-sharing protocols, including common formats, timeliness requirements, and privacy guidelines. For each major project, temporary working groups comprising project owners, relevant transport operators and regional authorities should agree on service scenarios and confirm matching budgets before forecasts are finalised.

Leadership must encourage transparent discussion of forecast errors to improve institutional credibility, rather than assigning blame. Clear roles must be defined regarding who initiates evaluations, who updates parameters and who coordinates across agencies. Without these institutional foundations, even well-designed technical strategies will fail to become embedded in everyday practice.

Transport planners can develop more robust forecasting frameworks that account for uncertainty, external disruptions and evolving travel behaviour by implementing these four strategies in combination with the necessary institutional changes. Ultimately, this supports more informed infrastructure investment decisions, more effective transport planning and a more mature, accountable transport planning discipline, where forecasting accuracy is continuously monitored, systematic biases are transparently identified and lessons learned are incorporated into future practice.

7.2. Discussion

Although transport demand forecasting is a critical component of evidence-based infrastructure planning, systematic post-implementation evaluation is surprisingly scarce in the academic literature (Flyvbjerg et al., 2005; Tempert et al., 2010). This discussion examines the empirical findings from the Hoekse Lijn case study within the broader theoretical framework of ex-post transport evaluation.

7.2.1. Limitations

Although this research was supported by an extensive dataset, several data-related constraints limited the depth of the analysis and the conclusion that could be drawn. The primary limitation stemmed from the smart card dataset provided by RET being aggregated, only containing boarding information at daily and hourly levels per station rather than the disaggregated passenger-level check-in and check-out transactions that were originally anticipated. This aggregation required adaptions to the methodology that was used, such as the development of a proportional scaling technique using the OV-Lite model outputs to reconstruct directional flows, occupancy profiles and passenger kilometre estimates.

The absence of the alighting data posed a challenge as it prevents the direct measurement of origin-destination flows, trip length distributions, and network-level movement patterns. While the proportional reconstruction method proved statistically robust, it did introduce model-dependent assumptions into the empirical validation process. This circular dependency, whereby model outputs are used to validate the model itself, poses a fundamental challenge to the credibility of transport forecasting evaluations.

Limitations concerning data coverage further constrained the robustness of the findings. The exclusion of barcode tickets and fare evasion resulted in a systematic undercount, with unknown temporal and spatial variation. Most notably, the underestimation of 18% on leisure-heavy days demonstrated how heterogeneity in payment methods can distort demand validation, particularly for travel patterns that are irregular and deviate from traditional commuter flows. This finding suggest that conventional smart card datasets may systematically under-represent precisely those user segments whose behavioural responses most challenge traditional forecasting assumptions, namely tourists and occasional travellers (Fulman et al., 2023).

The complete absence of socio-demographic attributes and trip purpose information related to the trips made by smart card users in the dataset is another significant analytical limitation. This caused this research to only validate behaviour without giving precise explanations on why it occurred. This prevented a more in-depth analysis of how different passenger groups responded to infrastructure changes, economic conditions and other external disruptions.

7.2.2. Contributions to transport research

This research contributes to the growing academic literature on the structural impacts of the COVID-19 pandemic on transport system, building directly on the foundational review of pandemic adaptiations in public transport planning by Gkiotsalitis and Cats (2021). Expanding on general ridership declines that were researched by (KiM, 2023), the analysis of Hoekse Lijn demand reveals how the effects of the pandemic unfold differently across different station types, trip directions and temporal patterns within a single corridor. The application of -12% correction factor for structural effects of the pandemic, derived from the KiM study, enabled systematic separation of pandemic-induced behavioural changes from endogenous model errors

These findings demonstrate that the impacts of the pandemic goes beyond a simple reduction in demand and also involve fundamental changes in travel distribution patterns. The continued recreational demand for Hoek van Holland Strand, despite a decline in commuter demand, illustrated how leisure travel has shown greater resilience than work-related travel patterns. This finding challenges conventional assumptions about trip purpose hierarchies in crisis conditions that (Gkiotsalitis & Cats, 2021) noted required empirical validation. Furthermore, the shift towards private and active modes during the pandemic, coupled with parallel disruptions such as the increase in transit fares, has amplified forecasting deviations. This underscores the necessity of models to account for the cumulative impact of these disruptions rather than treating them in isolation.

The extreme temporal variability that was observed at Hoek van Holland Strand, where beach ridership on peak days was 27 times the amount of average weekday levels, contributes to the documentation of weather-dependent recreational demand dynamics. This builds upon the research on weather effects on Dutch transport demand that was conducted by Sabir (2010). While this research already demonstrated that recreational trips are more sensitive to weather changes than commuting and business trips, the Hoekse Lijn analysis provides further evidence of the scale of these fluctuations and their implications for the entire corridor. Most significantly, the research suggests that current transport forecasting methodologies may inadequately capture recreational demand patterns. Even though the 2015 forecasts treated beach travel separately and estimated it using historical patterns rather than integrated behavioural modelling, the still underestimated the magnitude and temporal concentration of weather-dependent recreational demand.

This study also adds to the limited methodological literature on systematic forecast validation by demonstration how automated data sources can support comprehensive ex-post analysis, building directly on the works of Brands et al. (2020) and Dixit et al. (2024) in smart card-based ex-post evaluation. While

Brands et al. (2020) demonstrated the value of smart card data for evaluating the Noord/Zuidlijn introduction and Dixit et al. (2024) validated multimodal route choice models using automated data sources, the Hoekse Lijn study extends their methodological frameworks by developing a cluster-based accuracy assessment approach that provides a replicable method for identifying systematic biases across different station typologies.

7.2.3. Recommendations for future research

Based on the findings and limitations of this research, several directions for future research emerged that would advance both the theoretical understanding and practical application of transport forecast validations.

First, the extreme weather-dependent variability observed at Hoek van Holland Strand, where peak beach days generated 27 times normal weekday boardings, reveals a critical gap in current forecasting methodologies. Future research should focus on developing advanced modelling approaches for irregular recreational demand, taking into account weather variability, seasonal patterns and the temporal concentration of leisure travel. This should include the development of probabilistic forecasting models that incorporate meteorological forecasts and climate projections, as well as an understanding of how recreational demand interacts with regular commuting patterns throughout transport networks. Treating beach travel as a static add-on outside multimodal models has proved to be an inadequate way of capturing the magnitude and operational implications of weather-sensitive demand.

Furthermore, future research should move beyond applying a flat correction factor to account for the long-term impacts of societal disruptions like COVID-19, as the findings in this study reveal that such events have far more complex and differentiated effects on travel behaviour. Instead, research should focus on understanding how pandemic-induced changes, such as remote and hybrid work, evolving preferences for private and active transport modes, and heightened sensitivity to crowding, work differently across trip purposes, locations and user groups. This calls for the development of methodologies that can distinguish between temporary and permanent behavioural shifts, as well as the creation of dynamic forecasting frameworks capable of anticipating and adapting to a range of hypothetical shock scenarios.

The primary limitation encountered in this research was the reliance on aggregated smart card data. This meant that reconstruction techniques had to be developed using external model outputs, creating a circular dependency whereby the model outputs were used to validate the model itself. Future research should not only focus on obtaining disaggregate passenger-level data, but should additionally try to find techniques, such as machine learning approaches, that can automatically infer trip purposes from individual travel patterns, possibly strengthening it with supplementary data, such as travel survey data, mobile phone location data and land-use information. This would allow to capture the complete multimodal journey patterns significantly more accurately.

References

Agard, B., Morency, C., & Trépanier, M. (2006). MINING PUBLIC TRANSPORT USER BEHAVIOUR FROM SMART CARD DATA. *IFAC Proceedings Volumes*, *39*(3), 399–404. https://doi.org/10.3182/20060517-3-FR-2903.00211

APTA. (2024). *APTA Public Transportation Ridership Update* (tech. rep.). American Public Transportation Association (APTA). https://www.apta.com/research-technical-resources/research-reports/transit-savings-report/.

Australian Transport Assessment and Planning. (n.d.). Australian Transport Assessment and Planning Guidelines | 4. Data collection. https://www.atap.gov.au/tools-techniques/travel-demand-modelling/4-data

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474. https://doi.org/10.1016/J.TRANPOL.2005.06.008

Bakker, P. (2019, September). *Toegang tot ov-data* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid (KiM).

Bojada, M. (2014). *The contribution of Risk Management to cost, time and travel volume performances of RandstadRail* (tech. rep.).

Bojada, M., & Clerx, W. (2014, August). *Het succes van RandstadRail in de regio Rotterdam: Zijn de vervoersprognoses uitgekomen?* (Tech. rep.).

Brands, T., Dixit, M., & Van Oort, N. (2020). Impact of a New Metro Line in Amsterdam on Ridership, Travel Times, Reliability and Societal Costs and Benefits. *Issue*, *20*(4), 335–353. https://doi.org/10.18757/ejtir.2020.20.4.4084

CBS. (2025a, June). StatLine - Consumentenprijzen; prijsindex. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83131NED/table?ts=1751280068830

CBS. (2025b, June). StatLine - Regionale kerncijfers Nederland. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072ned/table?ts=1751896403914

de Vries, B., & Willigers, J. (2011). On the State-of-the-Art Demand Forecasting Model Developed by Netherlands Railways (NS).

Dempsey, P. S. (2015). Privacy Issues with the Use of Smart Cards. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2692652

Dixit, M., Brands, T., van Oort, N., Cats, O., & Hoogendoorn, S. (2019). Passenger Travel Time Reliability for Multimodal Public Transport Journeys. *Transportation Research Record*, *2673*(2), 149–160. https://doi.org/10.1177/0361198118825459/ASSET/IMAGES/LARGE/10.1177{_}0361198118825459-FIG7.JPEG

Dixit, M., Cats, O., van Oort, N., Brands, T., & Hoogendoorn, S. (2024). Validation of a multi-modal transit route choice model using smartcard data. *Transportation*, *51*(5), 1809–1829. https://doi.org/10.1007/s11116-023-10387-z

Durand, A., & Zijlstra, T. (2020). *The impact of digitalisation on the access to transport services: a literature review* (tech. rep.). KiM | Netherlands Institute for Transport Policy Analysis.

Engholm, A., & Kristoffersson, I. (2024). *MUST Managing Deep Uncertainty in Planning for Sustainable Transport: Project report: phase 1* (tech. rep.).

Faber, R., Hamersma, M., De Haas, M., Krabbenborg, L., & 't Hoen, A. (2023). Estimating post-pandemic effects of working from home and teleconferencing on travel behaviour. *European Journal of Transport and Infrastructure Research*, *23*(1). https://doi.org/10.18757/ejtir.2023.23.1.6733

Fan, W., & Chen, Z. (2018, August). *Estimation of Origin-Destination Matrix and Identification of User Activities Using Public Transport Smart Card Data* (tech. rep.). Center for Advanced Multimodal Mobility Solutions and Education. https://orcid.org/0000-0001-9815-710X

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2005). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association*, *71*(2), 131–146. https://doi.org/10.1080/01944360508976688

Fu, X., & Gu, Y. (2018). Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data. *Journal of Advanced Transportation*, *2018*. https://doi.org/10.1155/2018/9247102

Fulman, N., Marinov, M., & Benenson, I. (2023). Exploring Non-Routine Trips Through Smartcard Transaction Analysis. *AGILE: GIScience Series*, *4*, 1–9. https://doi.org/10.5194/AGILE-GISS-4-25-2023

Gemeente Maassluis. (2025, April). Wilgenrijk. https://www.maassluis.nl/wilgenrijk

Gemeente Rotterdam. (2015, May). *Milieueffectrapport Hoekse Lijn Hoofdrapport* (tech. rep.). https://www.commissiemer.nl/projectdocumenten/00004584.pdf#page=11.13

Gemeente Rotterdam. (2018, February). *Raadsonderzoek Hoekse Lijn - Eindrapport* (tech. rep.). Gemeente Rotterdam. Rotterdam.

Gemeente Rotterdam. (2025). Betaald parkeren | rotterdam.nl. https://www.rotterdam.nl/betaald-parkeren

Gkiotsalitis, K., & Cats, O. (2021). Public transport planning adaption under the COVID-19 pandemic crisis: literature review of research needs and directions. *Transport Reviews*, *41*(3), 374–392. https://doi.org/10.1080/01441647.2020.1857886

Goudappel Coffeng. (2011, July). *Vervoerwaardestudie Hoekse Lijn* (tech. rep.). www.goudappel.nl

Goudappel Coffeng. (2013, December). *Verkeersmodel RVMK 3 -Actualisering RVMK naar basisjaar 2010* (tech. rep.). www.goudappel.nl

Goudappel Coffeng. (2015, April). *Verkeerseffecten en vervoerwaardestudie* (tech. rep.). Goudappel. www.goudappel.nl

Goudappel Coffeng. (2018, October). *Verkeersmodel MRDH 2.0* (tech. rep.).

Huang, B., Wüst, H., & de Haas, M. (2024). Assessing the E-bike trends and impact on sustainable mobility: A national-level study in the Netherlands. *Journal of Cycling and Micromobility Research*, *2*, 100027. https://doi.org/10.1016/j.jcmr.2024.100027

Hussain, E., Bhaskar, A., & Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, *125*. https://doi.org/10.1016/j.trc.2021.103044

ITF. (2021). *Travel Transitions: How Transport Planners and Policy Makers Can Respond to Shifting Mobility Trends* (tech. rep.). ITF Research Reports, OECD Publishing. Paris. www.itf-oecd.org

Jong, G. D., Vignetti, S., & Pancotti, C. (2019). Ex-post evaluation of major infrastructure projects. *Transportation Research Procedia*, *42*, 75–84. https://doi.org/10.1016/J.TRPRO.2019.12.008

Kieu, L. M., Bhaskar, A., & Chung, E. (2013). Mining temporal and spatial travel regularity for transit planning. *Australasian Transport Research Forum 2013 Proceedings*. http://www.atrf.info/papers/2013/index.aspx

KiM. (2023). *Mobiliteitsbeeld 2023* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid (KiM). Den Haag.

KiM. (2024). *Kerncijfers Mobiliteit 2024* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid (KiM). Den Haag.

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kolkowski, L., & de Boer, L. (2023, March). *Boven verwachting? Prognoses en realisatie van reizigersaantallen bij grote stedelijke ov-projecten* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid (KiM).

Koppelman, F., & Bhat, C. (2006, June). *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models* (tech. rep.). U.S. Department of Transportation, Federal Transit Administration. https://www.caee.utexas.edu/prof/bhat/courses/lm_draft_060131final-060630.pdf

Kouwenhoven, M., Muller, J., Thoen, S., Willigers, J., & de Jong, G. (2023, November). *Values of Time, Reliability and Comfort in the Netherlands 2022 New values for passenger travel and freight transport* (tech. rep.). Significance.

Lee, R. J., Sener, I. N., & Mullins, J. A. (2014). Emerging Data Collection Techniques for Travel Demand Modeling: A Literature Review. http://www.ntis.gov

Limmen, R. (2017, October). OC Park Harga Midden bouwt 380 woningen in Schiedam. https://propertynl.com/Nieuws/OC-Park-Harga-Midden-bouwt-380-woningen-in-Schiedam/24286e5e-5fc7-4ad5-a9b2-9a273a6b5fe7

Liu, Z., Li, Q. Q., Zhuang, Y., Xiong, J., & Li, S. (2018). Inferring travel purposes for transit smart card data using. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10699 LNCS*, 11–18. https://doi.org/10.1007/978-3-319-73830-7{_}2/FIGURES/5

Macfarlane, G. (2024, October). Empirical Accuracy of Travel Forecasts. https://tfresource.org/topics/Empirical_Accuracy_of_Travel_Forecasts.html

Mahajan, V. ; Kuehnel, N. ; Intzevidou, A. ; Cantelmo, G. ; Moeckel, R. ; & Antoniou, C. (2022). General rights Data to the people a review of public and proprietary data for transport models. https://doi.org/10.1080/01441647.2021.1977414

Marra, A. D., Sun, L., & Corman, F. (2022). The impact of COVID-19 pandemic on public transport usage and route choice: Evidences from a long-term tracking study in urban area. *Transport Policy*, *116*, 258–268. https://doi.org/10.1016/J.TRANPOL.2021.12.009

Nicolaisen, M. S., & Driscoll, P. A. (2016). An International Review of Ex-Post Project Evaluation Schemes in the Transport Sector. *Journal of Environmental Assessment Policy and Management*, *18*(1). https://doi.org/10.1142/S1464333216500083 ; PAGE : STRING : ARTICLE / CHAPTER

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., & White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, *13*(4), 295–306. https://doi.org/10.1016/j.tranpol.2005.12.004

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Rahman, M., Yasmin, S., & Eluru, N. (2019). Controlling for endogeneity between bus headway and bus ridership: A case study of the Orlando region. *Transport Policy*, *81*, 208–219. https://doi.org/10.1016/J.TRANPOL.2019.07.004

RET. (2019, October). Wijziging buslijnen rond Hoekse Lijn per maandag 4 november - RET. https://corporate.ret.nl/nieuws/wijziging-buslijnen-rond-hoekse-lijn-per-4-november

Robinson, S., Narayanan, B., Toh, N., & Pereira, F. (2014). Methods for pre-processing smartcard data to improve data quality. *Transportation Research Part C: Emerging Technologies*, *49*, 43–58. https://doi.org/10.1016/J.TRC.2014.10.006

Sabir, M. (2010, March). *Impact of weather on daily travel demand* (tech. rep.). VU University, Department of Spatial Economics. Amsterdam. https://edepot.wur.nl/306452

Shinn, J. E., & Voulgaris, C. T. (2019). Ridership Ramp-Up? Initial Ridership Variation on New Rail Transit Projects. *Transportation Research Record*, *2673*(10), 82–91. https://doi.org/10.1177/0361198119844462

Sun, Q., Feng, T., Kemperman, A., & Spahn, A. (2019). Modal shift implications of e-bike use in the Netherlands: moving towards sustainability? https://doi.org/10.1016/j.trd.2019.102202

Tempert, B., van Meulen, A., Bezembinder, E., & de Graaf, S. (2010). Komen verkeersprognoses uit? *Verkeerskunde*. https://www.verkeerskunde.nl/artikel/komen-verkeersprognoses-uit

TfL. (2015, October). *Safety, Accessibility and Sustainability Panel Item: Leveraging Our Data for Maximum Impact* (tech. rep.). Transport for London (TfL).

Tian, X., Zhang, C., & Zheng, B. (n.d.). Crowding Prediction of In-Situ Metro Passengers Using Smart Card Data.

Train, K. E. (2009, January). *Discrete choice methods with simulation, second edition* (Vol. 9780521766555). Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Translink. (2016). *Translink Tracker October-December 2016 Q2* (tech. rep.). South East Queensland (SEQ), Australia.

Transport Scotland. (2015). *Larkhall-Milngavie Railway Project Evaluation Study: Final Report* (tech. rep.). Transport Scotland. https://www.transport.gov.scot/media/32433/j372238.pdf

Van Oort, N., Brands, T., & De Romph, E. (2015a). Short-Term Prediction of Ridership on Public Transport with Smart Card Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2535*(1), 105–111. https://doi.org/10.3141/2535-12

Van Oort, N., Ebben, M., & Kant, P. (2015b, August). *Innovatieve toepassingen van OV chipkaartdata: buiten de lijntjes van data-gedreven OV onderzoek* (tech. rep.). CVS Conference.

Vignetti, S., Giffoni, F., Pancotti, C., & Pagliara, F. (2020). Analytical framework for ex-post evaluation of transport projects: Lessons learnt on retrospective CBA. *Papers in Regional Science*, *99*(3), 683–703. https://doi.org/10.1111/pirs.12501

Wang, Z. j., Chen, F., Wang, B., & Huang, J. l. (2018). Passengers' response to transit fare change: an ex post appraisal using smart card data. *Transportation*, *45*(5), 1559–1578. https://doi.org/10.1007/s11116-017-9775-1

White, P., Bagchi, M., & Bataille, H. (2010). The Role of Smart Card Data in Public Transport. *Proceedings of the 12th World Conference on Transport Research*, 1–16.

Wood, D. A. (2015). A Framework for Measuring Passenger-Experienced Transit Reliability Using Automated Data.

A

Appendix A: Overview of beach days

Table A.1: Top beach days in 2024 (≥ 30°C)

Date	Max. temperature (°C)	Day type
12-08-2024	33.1	Weekday
01-09-2024	30.5	Weekend

Table A.2: Busy beach days in 2024 (25–30°C)

Date	Max. temperature (°C)	Day type
12-05-2024	26.6	Weekend
14-05-2024	27.0	Weekday
26-06-2024	25.6	Weekday
27-06-2024	26.6	Weekday
09-07-2024	25.9	Weekday
15-07-2024	25.1	Weekday
19-07-2024	29.7	Weekday
20-07-2024	29.9	Weekend
29-07-2024	27.6	Weekday
30-07-2024	25.7	Weekday
31-07-2024	25.1	Weekday
06-08-2024	27.0	Weekday
11-08-2024	26.4	Weekend
13-08-2024	25.5	Weekday
24-08-2024	25.0	Weekend
27-08-2024	25.3	Weekday
28-08-2024	29.9	Weekday
02-09-2024	27.5	Weekday
05-09-2024	29.2	Weekday
06-09-2024	25.3	Weekday
07-09-2024	26.7	Weekend
21-09-2024	25.1	Weekend

Table A.3: Moderate beach days in 2024 (20–25°C)

Date	Max. temperature (°C)	Day type
06-04-2024	23.0	Weekend
13-04-2024	20.7	Weekend
30-04-2024	21.3	Weekday
02-05-2024	22.3	Weekday
06-05-2024	20.2	Weekday
11-05-2024	22.7	Weekend
13-05-2024	21.4	Weekday
21-05-2024	21.8	Weekday
26-05-2024	21.0	Weekend
04-06-2024	22.1	Weekday
23-06-2024	21.7	Weekend
24-06-2024	20.8	Weekday
25-06-2024	24.9	Weekday
29-06-2024	21.5	Weekend
30-06-2024	20.0	Weekend
08-07-2024	20.5	Weekday
10-07-2024	21.2	Weekday
11-07-2024	20.0	Weekday
14-07-2024	20.5	Weekend
17-07-2024	20.4	Weekday
18-07-2024	24.5	Weekday
21-07-2024	21.5	Weekend
22-07-2024	21.9	Weekday
23-07-2024	20.8	Weekday
24-07-2024	20.0	Weekday
25-07-2024	22.3	Weekday
26-07-2024	21.3	Weekday
27-07-2024	22.2	Weekend
28-07-2024	21.5	Weekend
01-08-2024	22.9	Weekday
02-08-2024	24.8	Weekday
03-08-2024	23.7	Weekend
04-08-2024	21.2	Weekend
05-08-2024	23.4	Weekday
07-08-2024	23.2	Weekday
08-08-2024	24.5	Weekday
09-08-2024	23.9	Weekday
10-08-2024	24.0	Weekend
14-08-2024	24.1	Weekday
15-08-2024	24.2	Weekday
16-08-2024	22.9	Weekday
17-08-2024	24.4	Weekend
18-08-2024	22.1	Weekend
19-08-2024	24.3	Weekday
20-08-2024	22.6	Weekday
21-08-2024	20.0	Weekday
22-08-2024	22.5	Weekday
23-08-2024	22.7	Weekday
25-08-2024	20.4	Weekend
26-08-2024	21.6	Weekday
29-08-2024	22.3	Weekday
30-08-2024	22.5	Weekday
31-08-2024	23.8	Weekend
03-09-2024	22.1	Weekday
04-09-2024	22.0	Weekday
08-09-2024	22.2	Weekend
09-09-2024	20.9	Weekday
16-09-2024	20.0	Weekday
17-09-2024	21.3	Weekday
18-09-2024	23.4	Weekday
19-09-2024	23.8	Weekday
20-09-2024	23.7	Weekday
22-09-2024	22.9	Weekend
23-09-2024	21.4	Weekday
16-10-2024	21.1	Weekday
26-10-2024	20.8	Weekend

B

Appendix B: Transcripts of expert consultation

Interview 1: Jeroen Henstra

Date: June 18, 2025
Position: Operations and Planning, Rotterdam Public Transport (RET)
Format: Semi-structured interview (conversational format)

Q1: What is your general impression of the deviations between forecasts and realized demand (boardings, occupancy, passenger kilometres)? What patterns do you notice?

Jeroen acknowledges that occupancy and boardings per station in practice can deviate from the model, something that is also known from previous studies. Notably, Hoek van Holland Strand performs better than predicted on average (higher occupancy), especially due to peak days in summer. At the same time, the total number of boardings at many stations is lower than in the forecast, with the exception of specific cases. The longer average travel distance appears to be partly explained by beach travelers who travel the entire line.

Short trips (local traffic) have declined more strongly post-COVID than longer trips, but this plays less of a role on the Hoekse Lijn than in urban areas.

Q2: How do you assess the role of model inputs and assumptions (e.g., Metro A daily service, bus network changes, behavioral parameters) in these deviations?

- The excessive modeling of line A service during off-peak hours (5 instead of planned peak frequency) has certainly influenced results, potentially explaining overly high forecasts during off-peak hours and at western stations.
- The changes in bus supply (such as discontinued parallel lines) led to different passenger flows than in the model, which included these lines.
- Overestimation of the evening peak direction toward Rotterdam likely results from overly optimistic assumptions about commuting behavior from the west, whereas it is a more locally oriented area in practice.

Q3: Which external influences (network changes, COVID, changing travel behavior) do you see as the most important explanation for the found deviations?

- COVID-19 is recognized as the dominant influence, structurally changing travel behavior with fewer peak travelers and stronger recovery of recreational transport.
- The rise of the e-bike contributes to the decline in local public transport use.
- Operational changes (e.g., discontinuation of peak reinforcement on line A) and seasonal influences (e.g., beach days) also impacted occupancy and passenger kilometres.

Q4: What lessons do you see for future modeling of such lines or comparable cases?

- Use more robust input and scenarios for service frequencies and network configuration, especially when plans are still flexible prior to commissioning.

- Improve modeling of post-COVID behavior, including altered peak structures and increased recreational mobility.
- Explicitly include parallel network components (e.g., bus lines) that influence demand distribution.
- Pay more attention to local factors (e.g., educational and hospital locations) in predicting daily patterns and peak ratios.
- Monitor and recalibrate forecast models based on new empirical data as trends stabilize post-COVID.

Interview 2: Adam Pel (TU Delft)

Date: June 19, 2025
Position: Academic Researcher in Transport Modelling, TU Delft
Format: Semi-structured interview (conversational format)

Q1: How do you generally view the reliability of public transport passenger forecasts / multimodal transport models?

For new or strongly modified lines (like the Hoekse Lijn after COVID), it is relevant to examine whether demand and supply equilibria have been established. Feedback loops (e.g., habituation, chain behavior) affect the pace at which this happens.

Q2: Which factors, in your experience, usually cause predicted and actual passenger numbers to differ?

In addition to network changes and policy assumptions, the rise of private alternatives (e.g., e-bikes and micromobility) plays a major role—particularly for shorter trips.

Q3: To what extent do new transport models account for changes in passenger behavior in recent years (post-COVID)?

It is not so much the value of time that has changed, but rather trip generation: simply fewer work-related trips are made than before COVID. A relevant method is to determine what share of original demand has structurally disappeared in the past six years.

Q4: Are trends such as working from home, the rise of new modalities (e-bike), or changing preferences for public transport vs. car taken into account?

These trends—especially substitution by e-bike for short distances—are still only limitedly incorporated in models, despite their significant impact on metro networks with many suburban stations.

Q5: To what extent can the political/policy context at the time of the forecast influence the forecast itself?

With prestige projects, forecasts are sometimes presented optimistically. Objectivity is preserved through transparency of assumptions and use of sensitivity analysis.

Q6: It is sometimes suggested that for prestigious public transport projects, forecasts are sometimes presented more optimistically to get them approved. Do you recognize this mechanism?

See above—transparency and sensitivity analysis are key safeguards.

Q7: Where do you see the main opportunities to improve transport forecasts with current knowledge and data?

- Use of dynamic, data-driven models that incorporate real-time data such as smart card transactions.
- Improved modeling of latent demand.
- Accounting for modal shifts toward emerging mobility forms.

C
Appendix C: AI acknowledgement

Artificial intelligence tools have been used to provide supplementary support for this thesis. The utilisation of ChatGPT was intended to enhance creativity, clarify concepts and improve the quality of written language. Furthermore, AI was used for the assistance in programming and and creating visualisations using Python. The analytical reasoning, methodological decisions and conclusions in this thesis are all based on personal knowledge and critical judgment.