Denne artikel er udgivet i det elektroniske tidsskrift
Artikler fra Trafikdage på Aalborg Universitet
(Proceedings from the Annual Transport Conference
at Aalborg University)
ISSN 1603-9696
https://journals.aau.dk/index.php/td

Measuring the service reliability and the passenger experience in frequence-based transport systems

Fabrizio Cerreto, fac@m.dk
The Copenhagen Metro and Greater Copenhagen Light Rail

Abstract

This article explores the application and limitations of operational Key Performance Indicators (KPI) in frequency-based metro systems, specifically focusing on the Copenhagen Metro and Greater Copenhagen Light Rail. Recognizing that currently used KPIs, which primarily measure the frequency of station departures, fails to capture crucial aspects such as travel time reliability and system responsiveness during disruptions, the study proposes a reevaluation of the assessment focus. It delves into alternative measures that consider both the reliability of departures and the quality of the passenger experience. The research methodology evaluates metro performance through a variety of new KPIs, emphasizing time metrics like deviation and stability, thereby providing a more nuanced understanding of service reliability. This preliminary exploration seeks not only to enhance operational assessments but also to aid strategic planning and the execution of improvements within high-frequency public transportation systems. The article underscores the need for more adaptive, comprehensive KPIs that reflect the true operational realities of metro systems, suggesting that such advancements could significantly improve passenger satisfaction and system efficiency.

Introduction

In the dynamic landscape of urban transportation, metro systems distinguish themselves by offering frequent and reliable service without adhering to traditional timetabled schedules. This operational model presents unique challenges, especially in how performance is measured. Typically, metro networks like Copenhagen's Metro utilize a frequency-based approach, necessitating the use of distinct key performance indicators (KPIs). A primary KPI in use is Service Availability (SA), which emphasizes the frequency of station departures rather than fixed schedules.

Service Availability primarily gauges the frequency of actual departures at stations, which is especially relevant where there is no public timetable for comparison. However, this KPI does not account for linked travel times between stations. As a result, SA might not fully capture delays that do not disrupt the departure frequency, potentially presenting an incomplete picture of system efficiency. Furthermore, the SA metric, due to its complex, non-linear calculation method and its notable stability even in times of service disruptions, poses inherent challenges.

Furthermore, metro systems are inherently closed systems, which creates a significant operational challenge not typically encountered in road-based public transportation systems like buses. In metro systems, if one segment experiences a disruption, it can halt the entire line, unlike buses where detours or overtaking another stalled vehicle might mitigate delays. This aspect of metro operations emphasizes the need for highly responsive and robust KPIs that can quickly reflect disruptions in service and help in swift recovery and management strategies.

Given these considerations, this paper introduces a preliminary exploration into alternative KPIs that could potentially offer a more comprehensive analysis of metro system performance. It acknowledges the pre-existing limitations associated with the SA metric but does not delve into these intricacies further within this scope. Instead, this study looks to identify other indicators used globally, evaluating their suitability and adaptability to the specific context of the Copenhagen Metro.

This exploration is crucial for future strategic developments, as accurately assessing and improving metro service quality relies heavily on the robustness of the KPIs used. Improvements in travel times, for instance, should reflect in performance assessments, which SA currently fails to capture. Moreover, during unusual operating conditions—such as night services or major disruptions—the inability of SA to differentiate between planned and actual service levels becomes a significant oversight.

In navigating through these challenges, this research will set a foundation for more in-depth future studies that might directly compare these new findings against the existing SA metrics. The ultimate goal is to enhance not only the operational assessments but also the strategic planning and execution of improvements within high-frequency public transportation systems.

The need for a clear, more encompassing KPI that can effectively mirror the operational realities of a metro system—both in real settings and future scenario simulations—is critical. This paper aims to ignite a broader discussion on how best to encompass and quantify the complexities of metro operations through more tailored and comprehensive performance indicators.

Literature Review

Measuring and improving the reliability and comprehensive performance of metro systems, which typically operate on a frequency-based model without fixed timetables, are central to ensuring efficient urban transit. Service Availability (SA) is commonly used to assess these systems, primarily measuring the frequency of train departures rather than their punctuality or consistency (Ovenden, 2013). However, while SA provides insight into the high-frequency nature of metro operations, it overlooks critical factors such as the reliability of journey times between stations, leading to a partial view of operational efficiency (Schwandl, 2018).

Further analyses highlight the shortcomings of SA in accommodating off-peak fluctuations or major disruptions, which can lead to discrepancies between reported performance and the actual service quality experienced by passengers (Jerome et al., 2016). Reynolds (2019) points out the challenges in setting realistic performance targets with SA, noting its complex, non-linear calculation and minimal variability during disruptions, thus questioning its robustness as a performance measure.

In response, there is a growing trend towards integrating data-driven approaches for more adaptable and responsive KPIs. Park and Kim (2021) support the inclusion of real-time data processing to more accurately reflect operational conditions, particularly in dense urban environments. Furthermore, Harrison and Beck (2020) promote the adoption of composite metrics that encompass both service frequency and aspects like average delays per trip and passenger satisfaction, aiming for a holistic view of system performance.

The critical role of reliability in influencing passenger satisfaction and system performance is underscored by various studies. Wilson et al. (2018) emphasizes the need for precise measurement techniques to assess metro system reliability, particularly for those without fixed timetables. Cats et al. (2019) discuss how reliability affects passenger perceptions and choices, advocating for metrics that capture these effects, while Yerra et al. (2020) advocate for a comprehensive framework that encapsulates the multifaceted nature of reliability in dynamic transit systems.

Methodologies suggested by Zhao et al. (2021) utilize advanced data analytics and modeling to enhance the predictiveness and efficiency of evaluating service reliability. Such approaches underscore the need to develop adaptable and comprehensive KPIs that truly reflect the operational realities of metro systems, aiding in strategic decision-making and optimizing service delivery.

In the industry, standardized practices for measuring quality of service in metro systems focus significantly on headway management and the punctuality of service. Notable transit authorities like the Transport for London (2023), San Francisco Municipal Transportation Agency (2023), and the Metropolitan Transportation Authority (2023) publish regular performance dashboards and service reliability reports. These documents often detail metrics such as on-time performance, headway consistency, and service delays, providing stakeholders with a transparent view of operational effectiveness.
The Transit Capacity and Quality of Service Manual (2003) provides a foundational approach used across the industry, emphasizing the importance of reliable headway as a critical factor that affects passenger wait times and satisfaction. Metro Transit (2022) and other bodies like the Massachusetts Bay Transportation Authority (2023) routinely assess their performance against these standardized metrics, aiming to identify areas for improvement and manage service capacity strategically.
These practices highlight a common theme across metro systems worldwide: the balance between maintaining frequent, reliable service and adapting to real-time operational challenges. The industry's focus on such metrics aligns with academic findings, validating the need for resilient and adaptive management strategies in high-frequency public transportation systems. Such alignment between theory and practice is essential for developing future innovations in metro system management.

Methodology

This research adopts a comprehensive methodology to evaluate metro system performance while addressing the complexities of networks with branching lines and station-dependent headways. Because the frequency requirement varies between the common trunk and the branches, traditional approaches to performance measurement need critical adaptations.

Key Approach Elements

Frequency-Based Measures at Individual Stations

- **Fundamental Unit**: We continue to focus on frequency-based measures that are relevant at the individual station level, ensuring the analysis remains granular and actionable.
- **Transport Supply – Quantity and Quality**:
  - Departures and Trips: Both the number of departures at individual stations and complete trips from end-to-end are counted.
  - Quality Measures: This encompasses more than just the frequency of trips but also the reliability and consistency with which services are delivered.
- **Time Metrics – Deviation and Stability**:
  - Running Time Analysis: We not only measure the deviation of the running time from the planned schedules but also analyze the stability of this running time, utilizing metrics such as standard deviation, coefficient of variation, and the inter-quantile range. The reliability of realized running times is especially relevant for habitual passenger, who do not necessarily plan their transport according to public schedules or trip planners, but often define their route according to their previous experience and acquired knowledge of the system.
- **Variability Considerations**: Most of the indicators are defined in this paper as deviations from planned performance, with zero indicating performance as planned. Negative deviations denote lower absolute values while positive deviations indicate higher absolute value. performances that exceed planned objectives. Notably, higher and lower absolute values do not inherently correspond to better or worse outcomes. For instance, a higher supply level is considered better, whereas higher headway and running times indicate poorer performance.

Challenges

Branched Network Specifics: The main challenge arises from the varying frequency requirements between different sections of the network. This necessitates tailored indicators that can account for these variations without losing the context of overall system performance.

Data Structure and Analysis: Initial data exploration is unstructured to identify any inherent patterns or anomalies within the operational data. Following this, a structured approach provides a temporal comparison of KPIs.

Data Comparison and Analysis

Correlation Matrix: To synthesize the relationships between different KPIs, a correlation matrix is employed. This helps in understanding how different measures relate to one another and their collective impact on operational performance.
Ranking Operational Days: By ranking the operational days according to various KPIs, the research aims to pinpoint which types of operational disturbances are most effectively captured by each indicator. This ranking also aids in identifying outlier days where performance deviates significantly from the norm.
Through this methodology, the research aims to offer a nuanced understanding of metro system performance, particularly in systems where operational complexity is heightened by network design. The comprehensive analysis of both quantity and quality of transport supply, coupled with a deep dive into time-related performance metrics, seeks to provide a robust framework for assessing and improving metro service delivery.

Case Presentation: Copenhagen Metro Performance Analysis

This analysis focuses on Copenhagen's Metro Lines M1 and M2, which converge between Vanløse and Christianshavn across nine stations, then diverge towards Copenhagen Airport (M2, 7 stations) and Vestamager (M1, 6 stations). The Y-shaped network provides a unique framework for performance evaluation.
Data was collected from January 1 to June 30, 2024, on weekdays, excluding holidays and exceptional operation days like the week prior to Easter and January 17 due to data issues. This refined the dataset to regular operational days only.
The analysis targets weekday afternoon peak hours from 14:00 to 18:00. During this period, the operational headway is 101 seconds in the central section and 202 seconds on the branches, compared to the publicly advertised 2 and 4 minutes respectively. This scheduling allows for a detailed examination of metrics such as departure frequency, headway stability, and running time deviations across 491,686 stops and 29,981 complete trips.
The system is under review for increasing the number of trains and decreasing headway to enhance passenger capacity management. This prospective adjustment emphasizes the need for this detailed performance baseline to inform future operational upgrades.
Included network schematics (Figure 1 and Table 1) highlight operational sections and logistical intricacies, aiding understanding of the system's dynamics. This concise study assesses high-frequency urban rail performance during peak hours, paving the way for operational improvements amid capacity expansion.

Figure 1 - Metro network schematics

Table 1 - Station names and abbreviations of the Copenhagen Metro network

Name	Code	Metro Lines
Vestamager	VEA	M1
Ørestad	ORE	M1
Bella Center	BC	M1
Sundby	KHS	M1
DR Byen	UNI	M1
Islands Brygge	ISB	M1
Christianshavn	KHC	M1/M2
Kongens Nytorv	KGN	M1/M2/M3/M4
Nørreport	KN	M1/M2
Forum	FOR	M1/M2
Frederiksberg	FB	M1/M2/M3
Fasanvej	SOT	M1/M2
Lindevang	LIT	M1/M2
Flintholm	FL	M1/M2
Vanløse	VAN	M1/M2
Københavns Lufthavn	CPH	M2
Kastrup	KSA	M2
Femøren	FEO	M2
Amager Strand	AMS	M2
Øresund	OSV	M2
Lergravsparken	LGP	M2
Amagerbro	AMB	M2

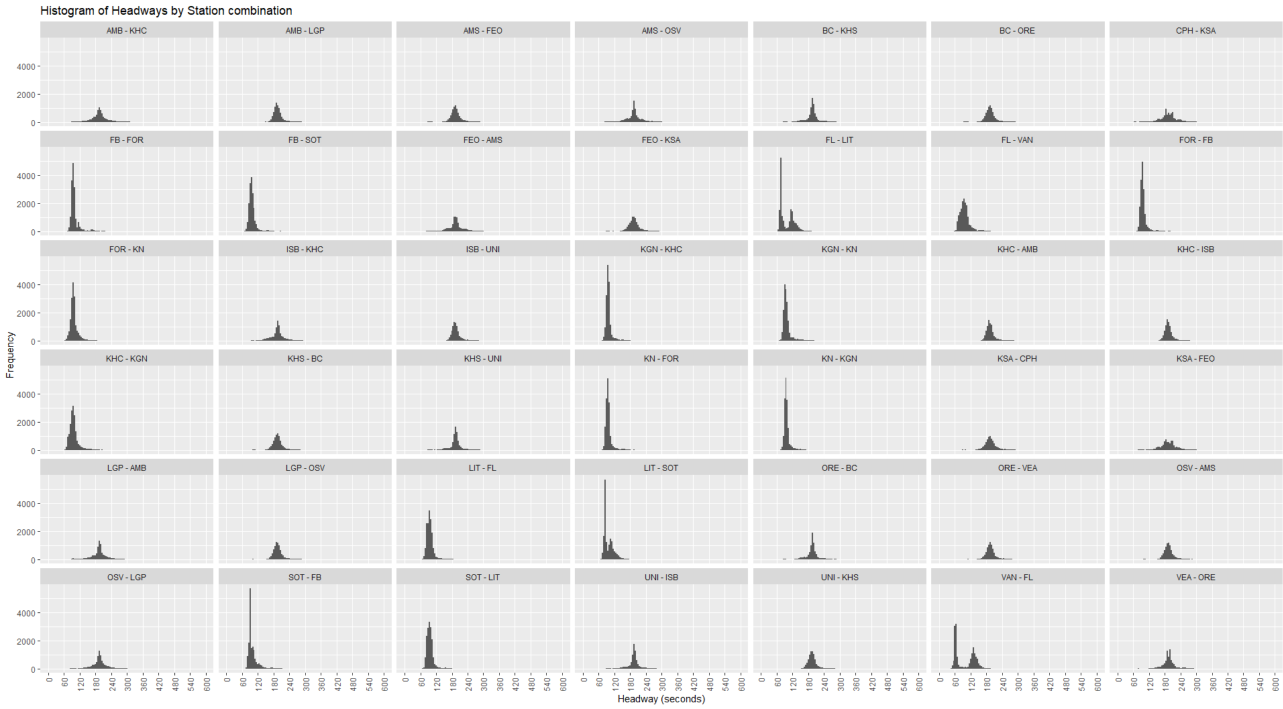

Figure 2 - Realized headways in the Copenhagen Metro in the first half of 2024. Each box represents one station in one direction

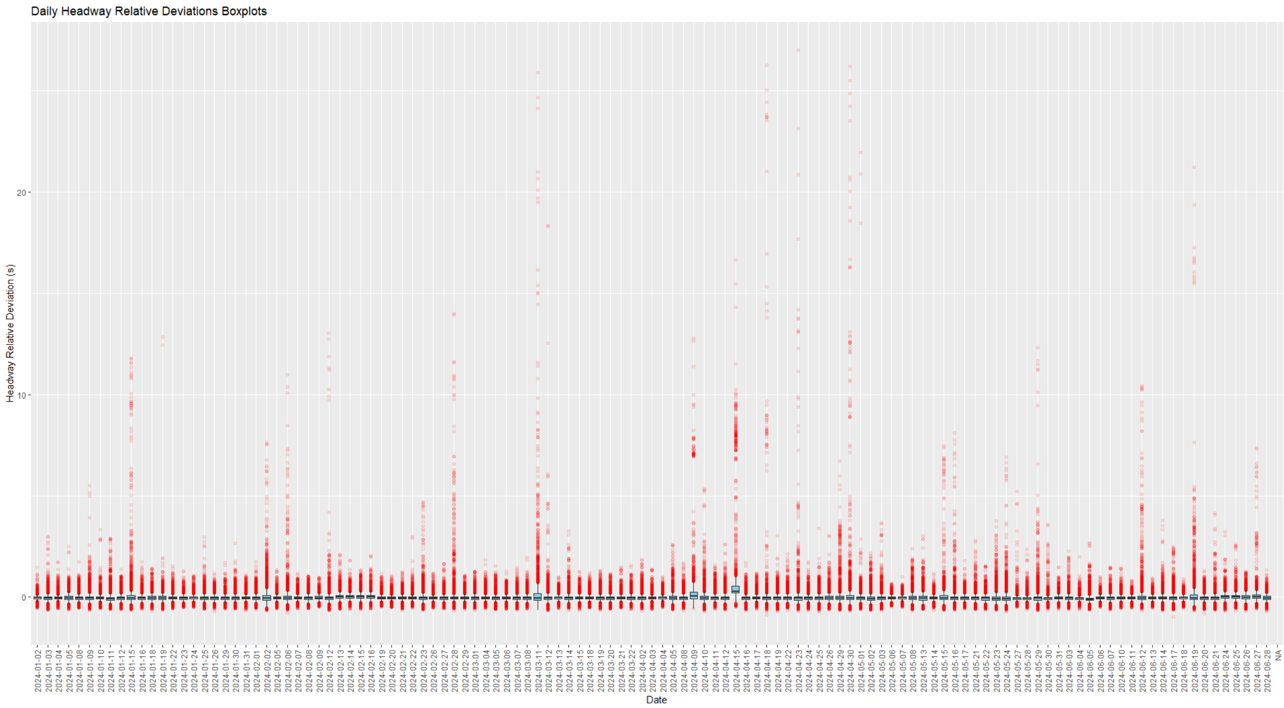

Figure 3 – Boxplot of realized headways on the Copenhagen Metro network in first half of 2024

Results

Headways

The analysis of headways across the network, as shown in Figure 2, indicates that the headway distribution generally follows a standard unimodal pattern, centering around 100 seconds in the central section and 200 seconds on the branches. However, an exception is observed from Vanløse to Fasanvej, where the headway distribution appears bimodal, with peaks around 60 seconds and slightly over 120 seconds. This pattern diminishes towards Frederiksberg, with the peaks converging around 100 seconds. The bimodal distribution is likely due to operational bottlenecks at Vanløse, where train inversion at the end-station creates minimum headway dependencies and results in longer wait times for departure in the opposite direction. This specific phenomenon is not the focus of the current study but warrants detailed analysis in future research.

Figure 3 presents a daily boxplot of realized headway relative deviations, highlighting specific days with significantly higher average headways and variances, such as 2024-04-15, 2024-04-09, 2024-03-11, and 2024-01-15. This visualization aids in identifying days with particularly poor performance, allowing for simultaneous analysis of both the central section and the branches.

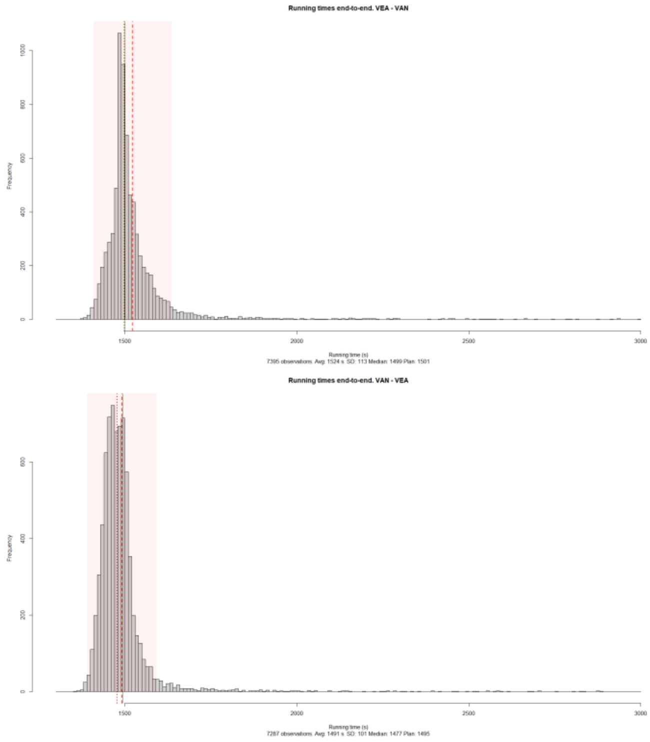

Figure 4 - Realized running times on the different lines, divided by direction. Green line: scheduled running times. Red dashed line: average realized running time. Red box: +- sd realized running time. Dark red dotted line: media realized running time.

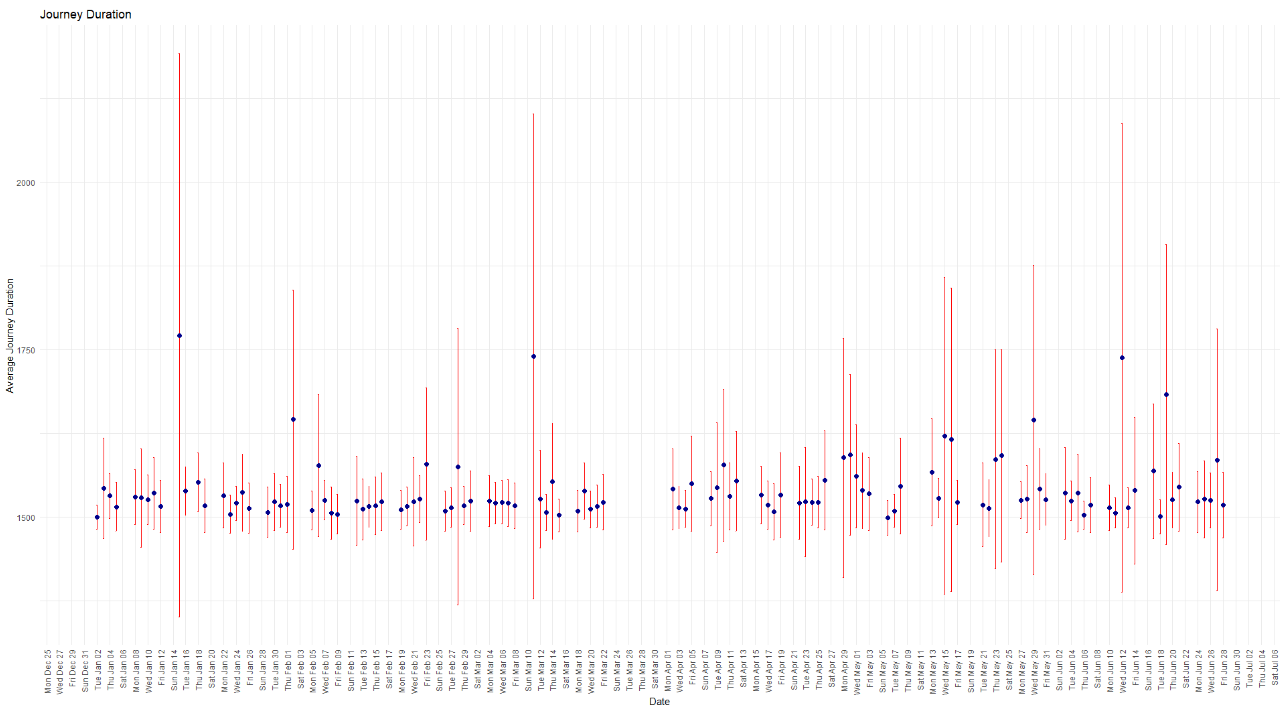

Figure 5 -Running time for complete trips. Blue dot: average. Red error bars: standard deviation.

Running times

The distribution of realized running times (Figure 4) reveals challenges in maintaining scheduled running times, particularly approaching Vanløse on both lines. The median realized running times are close to or exceed the scheduled times, indicating congestion issues. Vanløse station acts as a bottleneck, primarily due to conflicting train movements, causing delays from Frederiksberg to Vanløse.
Figure 5, which details the daily distribution of running time relative deviations, identifies the worst performing days such as 2024-1-15, 2024-3-11, 2024-6-12, and 2024-6-19, with high average deviations and variances.

KPI Examination and Relevance

Static KPIs: Many planned figures such as Headway Plan, Planned Departures, and Trains in operation (planned) were excluded due to their invariant nature over the analyzed period, offering limited dynamic operational insights.
Departure Production Deviation and **Complete End-to-End Journeys Deviation**: These KPIs were preferred over simple counts due to their ability to juxtapose actual operations against planned figures, revealing the operational efficiency and adaptability during normal and disrupted conditions.
Trains in Operation: Preserved for its simple yet essential representation of actual service capacity, a crucial metric for real-time operational assessment.
Punctuality Metrics (Punctual and Unpunctual): Directly reflect service reliability and are crucial for gauging passenger satisfaction, justifying their retention in the performance analysis framework.
SD_delay (Standard Deviation of Delays): Maintained due to its relevance in assessing the variability and reliability of service, especially in contrast to other redundant delay-related metrics.
CD Run_Time (Coefficient of Deviation of Running Time): Excluded due to redundancy with other time deviation measures but important to note that it holds a weakly negative correlation with the number of departures indicating closer attention required during disruptions.
Mean HW Relative Deviation, **SD HW Relative Deviation**, and **Range P5-P95 HW Relative Deviation**: These KPIs related to headway deviations were included to provide a detailed analysis of the consistency and reliability of service intervals, essential for effective traffic management and operational planning.
Figure 6 serves as a comprehensive correlation matrix and includes scatterplots depicting the interaction between various daily calculated KPIs during the analysis period. This visualization allows for the identification of both expected and anomalous correlations within the day-to-day operations of the Copenhagen Metro.

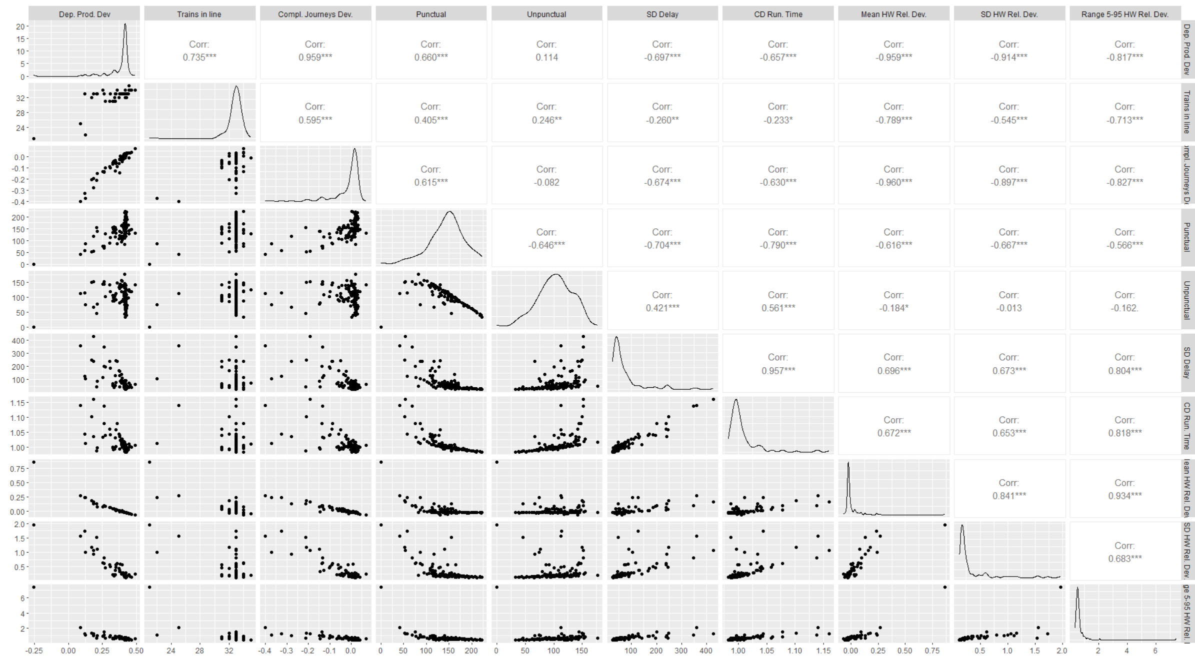

Figure 6 - Correlation matrix of selected KPIs measured on the Copenhagen Metro in the first half of 2024

Departure Production vs. Complete Trips

The correlation matrix shows that departure production does not match complete trips on a 1:1 basis. During significant disruptions, the Metro system adapts by initiating shuttle services around the affected areas, resulting in departures that do not contribute to complete end-to-end journeys. This adaptive strategy ensures continued partial service rather than a complete halt, thus maintaining a higher level of service reliability under disrupted conditions.

Punctual Services and Production of Departures

There is a weak positive correlation between the production of departures and punctual services. When disruptions occur, the control strategy often involves redistributing or increasing departures to prevent bunching and maintain a semblance of schedule adherence. This highlights an operational focus on managing punctuality, even during periods of increased challenge.

Correlation between Punctual and Unpunctual Metrics

Punctual and unpunctual trips are only weakly negatively correlated. This pattern is primarily due to the operational policy during major disruptions, where the standard service operation (punctual or minor delays) shifts to shuttle mode or trip cancellation, leaving a significant number of trips unoperated. This weak correlation suggests that the distinctions between punctual and unpunctual services can blur under stress, which complicates straightforward performance assessment.

Headways vs. Departures

A strong negative correlation exists between the number of departures and headways. More frequent departures result in shorter headways, demonstrating a fundamental operational principle aimed at boosting service capacity and reducing passenger wait times.

Standard Deviation of Headways

This indicator also shows a strong negative correlation with the number of departures and completed journeys. This relationship indicates that achieving operational homogeneity—less variability in headway—is key to accommodating more trips, thereby enhancing service regularity and reliability.

Impact of Disruptions on Service Metrics

The correlation between SD_delay and CD_Run_Time and the number of departures underlines how disruptions impact service performance. Both show a weak negative correlation with departures, suggesting that disruptions not only reduce the frequency of service but also increase the variability of running times and delays.

Date ranking

Bad days

Table 2 shows the 5 worst days according to different KPIs.

Days Consistently Performing Poorly:

- April 15, 2024 (15-04-2024)
  - **Incidents**: The day was plagued with extensive technical issues on the M1/M2 lines during morning, midday, and afternoon. Operational shutdowns occurred between Vanløse and Frederiksberg at different times, and there were limited trains in service.
  - **Affected KPIs**: Practically all KPIs were negatively impacted, indicating a catastrophic effect on the metro operation—departure production, train operation, journey completion, punctuality, and all related deviations.
- March 11, 2024 (11-03-2024)
  - **Incidents**: A cascade of incidents including a faulty door at Christianshavn stopping all trains, followed by repeated periods of delays and service interruptions impacting the whole network.

    - **Affected KPIs**: Featuring as the second or third worst day for multiple KPIs, this day highlights pervasive delays and substantial disruptions affecting nearly every aspect of metro operation.
- April 30, 2024 (30-04-2024)
    - **Incidents**: Technical problems with a train at Forum in the afternoon caused a complete halt of operations, further extending to no train service between Vanløse and Nørreport.
    - **Affected KPIs**: This day ranks among the worst in terms of departure production deviation, complete journeys deviation, and several punctuality-related KPIs.
- April 9, 2024 (09-04-2024)
    - **Incidents**: There was no operation between Frederiksberg and Vanløse due to a train with technical issues at Lindevang, causing significant delays and forced train changes for many passengers.
    - **Affected KPIs**: Ranks similarly poor to April 30, again highlighting significant issues particularly with operations and scheduling disruptions impacting train departure and headway consistency.
- June 19, 2024 (19-06-2024)
    - **Incidents**: Technical problems in the afternoon led to no train operation between Vanløse and Nørreport and caused delays across the M1/M2 lines.
    - **Affected KPIs**: This day consistently ranks as the fifth worst, slightly better than others highlighted but still significantly impactful on service reliability and headway management.

The days mentioned exhibit a range of operational issues from technical failures to disruptions due to emergency shutdowns. April 15, 2024, stands out as the most impacted day, severely affecting nearly all KPIs, suggesting a major systemic failure or a series of compounded issues.

Table 2 - Worst performing days according to different KPIs in the Copenhagen Metro in first half of 2024. Each day is assigned a color for better readability.

Rank	Departure Production Deviation	Trains in operation	Complete end to end Journeys Deviation	Punctual	Unpunctual	SD delay	Weighted CoeffDev Run Time	Mean HW Rel Deviation	SD HW Rel Deviation	Range P5 P95 HW Rel Deviation
1	15-04-2024	15-04-2024	15-04-2024	15-04-2024	18-01-2024	15-04-2024	15-04-2024	15-04-2024	15-04-2024	15-04-2024
2	11-03-2024	09-04-2024	11-03-2024	11-03-2024	13-05-2024	15-01-2024	15-01-2024	11-03-2024	30-04-2024	11-03-2024
3	30-04-2024	11-03-2024	09-04-2024	19-06-2024	26-04-2024	11-03-2024	11-03-2024	09-04-2024	11-03-2024	02-02-2024
4	09-04-2024	13-02-2024	30-04-2024	15-01-2024	15-01-2024	12-06-2024	12-06-2024	30-04-2024	18-04-2024	19-06-2024
5	19-06-2024	14-02-2024	28-02-2024	30-04-2024	12-06-2024	19-06-2024	19-06-2024	19-06-2024	19-06-2024	30-04-2024

Good days

The data outlined in Table 4 highlights variability in what constitutes a 'good day' across different KPIs for the Copenhagen Metro, showcasing that positive performance can vary significantly depending on the metric in question. This variability suggests that while some days may excel in one area (e.g., punctuality), they might not necessarily perform as well in others (e.g., headway reliability).

Table 3 - Best performing days according to different KPIs in the Copenhagen Metro in first half of 2024. Each day is assigned a color for better readability.

Rank	Departure Production Deviation	Trains in operation	Complete end to end Journeys Deviation	Punctual	Unpunctual	SD delay	Weighted CoeffDev Run Time	Mean HW Rel Deviation	SD HW Rel Deviation	Range P5 P95 HW Rel Deviation
1	05-06-2024	30-05-2024	05-06-2024	02-01-2024	15-04-2024	02-01-2024	06-05-2024	05-06-2024	07-05-2024	04-06-2024
2	04-06-2024	02-01-2024	04-06-2024	06-05-2024	02-01-2024	06-06-2024	02-01-2024	04-06-2024	11-06-2024	02-01-2024
3	28-05-2024	06-02-2024	28-05-2024	06-06-2024	06-06-2024	11-06-2024	18-06-2024	28-05-2024	06-05-2024	07-05-2024
4	30-05-2024	21-05-2024	26-02-2024	15-03-2024	06-05-2024	24-01-2024	15-03-2024	27-05-2024	04-06-2024	27-05-2024
5	27-05-2024	27-05-2024	22-05-2024	11-06-2024	15-03-2024	07-05-2024	06-06-2024	30-05-2024	18-06-2024	06-05-2024

Consistently Good Days:

- January 2, 2024 (02-01-2024)
    - **Best KPIs**: Trains in operation, Punctual, SD delay, Trains in operation, Range P5-P95 HW Rel Deviation
    - **Summary**: This day consistently ranks as one of the best across multiple KPIs. It falls shortly after New Year's Day, which could mean that while operational capacity is still high, ridership might be lower, reducing congestion and leading to smoother operations.
- June 5, 2024 (05-06-2024)
    - **Best KPIs**: Departure Production Deviation, Complete end-to-end Journeys Deviation, Mean HW Rel Deviation
    - **Summary**: This day stands out in terms of maintaining scheduled departures, journey completions, and headway management. It could indicate less external disruption (e.g., technical issues or external events).
- May 30, 2024 (30-05-2024)
    - **Best KPIs**: Trains in operation, Complete end-to-end Journeys Deviation, Mean HW Rel Deviation
    - **Summary**: High performance in both operational readiness and journey completions suggests excellent control over both train availability and service consistency.

Specific Dates for Specific KPIs: Some days excel uniquely for certain operations. For instance, February 6, 2024, is noted for optimal train operation, while unique KPIs like Range P5-P95 HW Rel Deviation favor dates such as February 1, 2024, and June 4, 2024.

Unlike the consistency seen on the worst days, the best days demonstrate that large-scale disruptions tend to impact all operational indicators, whereas positive performance can be more isolated depending on the specific efficiencies or external factors affecting that day.

Conclusions

The analysis of Copenhagen Metro's operational performance underscores a nuanced challenge: while "bad" operational days are readily identified due to their widespread impact on multiple KPIs, pinpointing "good" days is more complex because their performance may not uniformly exceed across all metrics (Reynolds, 2019). This finding resonates with the insights provided by Jerome et al. (2016), who emphasize the limitations inherent in traditional KPIs which may not capture the subtleties of daily operational excellence. It suggests the necessity for enhanced analytical tools that can depict a broader spectrum of operational quality, capturing both the highs and lows with equal clarity.

Additionally, the focus on punctual trips as a metric closely aligns with passenger satisfaction, mirroring findings by Wilson et al. (2018) that reliability and timeliness form the cornerstone of user experience in metro systems. This perspective is crucial as it reflects a passenger-centric approach to performance assessment, which is vital for maintaining and increasing rider confidence and patronage. However, relying primarily on visual tools for performance analysis, while effective for identifying glaring issues, falls short in quantifying the subtler aspects of daily operations. As Schwandl (2018) points out, visual representations often fail to highlight operational successes which are less conspicuously distinct but equally important for overall service quality.

Further expanding on this, merely analyzing the frequency and punctuality of trips proves insufficient for a holistic operational analysis. Including data on trip duration is imperative as highlighted by Cats et al. (2019), who show that trip reliability impacts passenger perceptions and behavioral intentions. This aligns with research by Park and Kim (2021), suggesting that integrating real-time data improves the accuracy and comprehensiveness of performance metrics. Delving deeper into trip times offers a richer, more complete view of service quality, reflecting not just the frequency of trains but also the actual time passengers spend during their travel, thus capturing a critical element of the transit experience which, according to Zhao et al. (2021), greatly influences satisfaction levels.

In synthesizing these observations, it becomes clear that for a high-frequency metro systems, evolving beyond traditional metrics to more dynamic and inclusive KPIs is essential. This resonates with the industry trends mentioned in the Transit Capacity and Quality of Service Manual (2003) which prioritize balancing punctuality with reliable headway management, underscoring a need to adopt multifaceted and nuanced approaches that address all aspects of the passenger experience. These adjustments can provide more accurate reflections of system performance and drive strategic improvements that align with both operational goals and passenger needs, ensuring the system remains responsive to the demands of urban transit environments.

Further Research

Further comprehensive research that integrates Service Assurance metrics could deepen the understanding of operational effectiveness. Exploring the relationship between these operational KPIs and passenger satisfaction would pinpoint which improvements would markedly enhance the commuting experience. Advancing our analytical tools could overcome the existing limitations of visual tools, enabling a more detailed and quantitative evaluation of operational performance across a continuum and refining the distinction between 'good' and 'bad' operational days.

Longitudinal studies would unveil patterns and causes behind daily operational variability, improving forecasting and preparation for potential disruptions. Implementing real-time passenger feedback mechanisms, especially concerning trip times and responsiveness to service disruptions, could further refine operational strategies by linking direct user feedback to service quality assessment.

It is also essential to investigate the KPIs more thoroughly in terms of their sensitivity to various disturbances. Understanding which KPIs are most affected by specific types of disruptions can help in crafting more resilient operational strategies and in developing targeted interventions to mitigate these effects.

Reference List

1. Cats, O., Wilson, J., & Yerra, S. (2019). *The impact of reliability on passenger perceptions and travel behavior*. Journal of Public Transportation.
2. Fujiyama, T. (2017). *Comparative analysis of KPIs across various international metro systems*. Transportation Research Part A.
3. Harrison, A., & Beck, P. (2020). *Adapting composite metrics for metro system performance assessment*. Urban Transport Planning Journal.
4. Ingvardson, J.B., Jensen, J.K., & Nielsen, O.A. (2017). *Analysing improvements to on-street public transport systems: a mesoscopic model approach*. Public Transport, 9(3), 385-409. https://doi.org/10.1007/s12469-016-0151-x.
5. Jerome, C., Reynolds, L., & Schwandl, R. (2016). *The limitations of current KPIs in metro systems*. International Journal of Urban Planning.
6. Liao, C. (2017). *Metro transit service reliability measures assessment*. Retrieved from the University of Minnesota Digital Conservancy: https://hdl.handle.net/11299/226966.
7. Massachusetts Bay Transportation Authority. (2023). *Service reliability report*. Retrieved from https://www.mbta.com/.
8. Metropolitan Transportation Authority. (2023). *System performance report*. Retrieved from https://new.mta.info/.
9. Metro Transit. (2022). *Annual performance report*. Retrieved from https://www.metrotransit.org/.
10. Ovenden, N. (2013). *Service availability as a performance indicator in metro systems*. Public Transport International.
11. Park, Y., & Kim, S. (2021). *Data-driven approaches to metro system KPIs*. Journal of Transportation Technologies.
12. Reynolds, L. (2019). *Challenges of setting performance targets using service availability*. Transportation Research Record.
13. San Francisco Municipal Transportation Agency. (2023). *Service quality and reliability report*. Retrieved from https://www.sfmta.com/.
14. Schwandl, R. (2018). *Critical evaluation of metro system efficiency*. International Journal of Metro Planning and Design.

15. Transit Capacity and Quality of Service Manual. (2003). Washington, D.C.: Transportation Research Board, National Research Council.
16. Transport for London. (2023). *Performance dashboard: quarterly report*. Retrieved from https://tfl.gov.uk/.
17. Wilson, J., Cats, O., & Yerra, S. (2018). *The role of reliability in passenger satisfaction and system performance*. Transportation Science.
18. Yerra, S., Wilson, J., & Zhao, X. (2020). *Factors influencing metro service reliability*. Urban Transit Journal.
19. Zhao, X., Park, Y., & Kim, S. (2021). *Advanced methodologies for evaluating service reliability in metro systems*. Journal of Advanced Transportation.