Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 95 (2026) 337–344

Transportation Research Procedia

www.elsevier.com/locate/procedia

Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

Modeling train arrival variability: Methodological approaches and data-driven insights for railway systems

Bruna G. Palm[a,*], Mohamed Samy Massoum[a], Henrik Fredriksson[a,b], Mattias Dahl[a], Carolina Bergeling[a]

[a]*Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, 371 79, Karlskrona, Sweden*
[b]*Department of Science and Technology, Communications and Transport Systems, Linköping University, 601 74, Norrköping, Sweden*

Abstract

This study presents a statistical analysis of train delays in the Swedish railway system. The focus of the study is to identify the best-fitting probability distributions for train arrival times across different stations and travel directions. Using the Kolmogorov-Smirnov (K-S) test, we evaluate the goodness of fit for common distributions—gamma, log-normal, and inverse Gaussian—to capture delay patterns at ten stations. Our findings reveal significant variability across stations, with the log-normal distribution providing the best fit for 70% of cases. However, some stations exhibited direction-specific deviations, emphasizing the need for localized analysis. Traditionally, train delays in Sweden have been assumed to be uniformly distributed across the network, an oversimplification frequently used in generating synthetic datasets for AI-based timetable rescheduling systems. This study challenges that assumption, demonstrating that delay distributions vary by station and direction. By incorporating station- and direction-specific modeling, our results contribute to the development of more accurate synthetic datasets. These insights support data-driven approaches to predictive modeling, operational efficiency improvements, and increased reliability in railway networks. Based on the best-fitting distributions identified through statistical testing, we generate synthetic data using maximum likelihood estimates and direct sampling. Our study systematically assesses the distributional characteristics of train arrivals across stations and directions in the southern Swedish railway network, aiming both to understand operational variability and to generate realistic synthetic data for AI-based rescheduling. Building on this analysis, our method produces datasets that preserve the statistical characteristics of real train delays, ensuring they are more suitable for training and evaluating AI-based rescheduling algorithms.

Keywords: Train delays; Statistical modeling; Train arrival; Railway network; Synthetic data

1. Introduction

Efficient transportation is vital for societal development, and the railway offers a sustainable transportation mode for people and goods. An operational challenge in the railway system is maintaining timetable punctuality while meeting

* Corresponding author: Bruna G. Palm
E-mail address: bruna.palm@bth.se

10.1016/j.trpro.2026.02.043

planned demand, which is crucial for making the railway a competitive and attractive transportation mode. However, the railway system is sensitive to disruptions. Minor delays in one part of the network can affect other parts. This may result in severe delays and a lower level of service. The Swedish Transport Administration (Trafikverket) aims for 95% of all scheduled trains to reach their destinations with delays of less than five minutes in the Swedish railway system. However, the new traffic planning system introduced for the 2023 train schedule also contributed to disruptions, as did issues with rolling stock and driver shortages. The overall punctuality of passenger trains in 2023 was just under 88%, while that of freight trains was just over 71% in Sweden (Järnvägsbranschens samverkansforum (JBS), 2023).

When a railway operation is delayed for a train (e.g., increased travel time between stations or extended dwell time at the station), its trip will potentially deviate from what is specified in the timetable. Such delays can be caused by a range of factors, including issues related to train operations, the railway network and infrastructure, or external conditions such as weather and time of day (Berger et al., 2011; Tiong et al., 2023). Efficiently rescheduling the timetable is, therefore, a crucial task for dispatchers to reduce the impact of disturbances and to recover from disruptions. Timetable rescheduling has traditionally been approached through mathematical optimization techniques, which are well-suited for incorporating system constraints such as station capacity and block occupancy. While linear models can effectively capture certain aspects of railway operations—and are sometimes used to approximate non-linear behavior—they are not always the most practical or accurate choice. Many operational characteristics of railway systems are inherently complex and non-linear, making them difficult to model with conventional linear methods. Moreover, optimization techniques, despite their strengths, often rely on rigid predefined rules and can lack transparency, limiting their usability for practitioners. In contrast, data-driven approaches, including models based on artificial intelligence (AI), offer a more flexible and interpretable alternative that can better capture the variability of real-world operations (Huang et al., 2023). Therefore, data-driven methods may be worthwhile to consider. AI-based rescheduling methods utilize historical railway data on arrivals, departures, and schedules to predict potential disruptions and suggest rescheduling strategies. AI-based systems provide greater flexibility and scalability compared to traditional methods, as they are capable of learning, improving, and handling the complexities of the railway system (Tang et al., 2022; Zhang and Zhang, 2023). The effectiveness of AI systems relies on training and learning with realistic scenarios, which emphasizes the need for accurate simulations of railway network conditions. The railway is a complex system, and many scenarios may be needed to understand its variations and different conditions.

Railway operations are inherently stochastic processes (Şahin, 2017). Probability distributions provide valuable insights into train operations compliance with the scheduled timetable. For example, they may reveal that late arrivals and delays are more likely at certain stations. This variability across different stations makes it difficult to identify a single distribution suitable for all scenarios (Yuan, 2006). Studies on probability delay distributions in the Swedish railway network, using data from 2008 and 2009, analyzed passenger and freight train delays separately. The results suggested that for passenger trains, a power law distribution is suitable for modeling delays at the daily level, while an exponential distribution better represents delays at individual stations. In contrast, freight train delays were found to follow an exponential distribution (Bergström and Krüger, 2013; Krüger et al., 2013). These findings indicate that for passenger trains, a large portion of delay time is concentrated in the tail of the distribution, highlighting the significant impact of extreme delays on overall railway network performance (Bergström and Krüger, 2013).

In Sweden, train delays have traditionally been modeled as being uniformly distributed across the network (Palmqvist et al., 2017). This simplification has been widely used when generating synthetic datasets for AI-based timetable rescheduling systems. However, the effectiveness of AI-based scheduling systems depends on their ability to learn from realistic scenarios, underscoring the importance of accurate modeling of railway network conditions. Since delay distributions vary by train type and location, a more refined approach is necessary to ensure realistic simulations (Yuan, 2006). This issue is not unique to Sweden. A punctuality analysis conducted at a station in The Hague, Netherlands, revealed that arrival delays follow an exponential distribution, while excess dwell times follow a normal distribution (Yuan and Hansen, 2002). Similarly, a study of Eindhoven station in the Dutch railway network provided evidence that train delays tend to follow either normal or negative-exponential distributions (Goverde et al., 2001). However, in most cases, train delays exhibit skewed distributions, such as gamma, log-normal, or Weibull distributions (Minbashi et al., 2020; Huang et al., 2019; Jiang and Persson, 2016; Wen et al., 2017). Using maximum likelihood estimation (MLE) to estimate distribution parameters and conducting goodness-of-fit tests reveals that train delays do not follow a single universal distribution. Instead, multiple distributions may provide suitable models depending on the railway context (Yang et al., 2019).

The purpose of this paper is to examine the statistical behavior of train arrivals at stations within a railway network and to utilize this analysis for generating synthetic datasets. The study focuses on a rail line connecting the east and west coasts in southern Sweden. While this particular rail line serves as a case study, the methodology and findings have broader applicability for analyzing train performance and arrival patterns across the railway network. We conduct a statistical analysis of train arrival times and identify the best-fitting probability distributions for different stations. This process involves performing statistical tests, such as the Kolmogorov-Smirnov test, to evaluate the goodness of fit for common distributions, including gamma, log-normal, and inverse Gaussian. By fitting appropriate models to historical data, we aim to capture variations in punctuality across different locations. Building on this analysis, we generate a synthetic dataset to support AI-based timetable rescheduling systems. The data generation method relies on the best-fitting distribution identified from the observed data, after excluding outliers. The parameters of the selected distribution are estimated using the MLE method. Using these fitted parameters, a new dataset is generated through direct sampling. This approach produces a realistic dataset that preserves the statistical characteristics of real train arrivals, making it valuable for AI-driven railway optimization.

2. Statistical analysis

This study uses train arrival records from ten stations along a bidirectional Swedish railway line. The data is provided and maintained by the Swedish Transport Administration (Trafikverket). Each station was analyzed in both travel directions: from Karlskrona to Malmö and from Malmö to Karlskrona. The analyzed period spans from early February to April 2024. As a preliminary step, the robustness of the analysis was ensured by preprocessing the dataset. This includes the removal of duplicate entries and the detection and exclusion of outliers using the interquartile range (IQR) method. Following this, histograms of train arrival times were constructed for each station to visualize the empirical distribution. These histograms represent the minute within the hour when each train arrived at the respective station.

Four candidate probability distributions (gamma, log-normal, Rayleigh, and inverse Gaussian) were evaluated to better understand the underlying data behavior. Each distribution was fitted to the preprocessed data using the MLE method. The goodness of fit for each distribution was assessed using the Kolmogorov–Smirnov (K-S) test. Since outliers had been removed, the resulting data typically lacked heavy tails, making the K-S test particularly appropriate. The K-S test measures the maximum discrepancy between the empirical and theoretical cumulative distribution functions, emphasizing central tendencies, which are especially relevant for modeling train arrival patterns. Model selection was based on the distribution yielding the lowest K-S test statistic, indicating the closest alignment with the observed data. While the histogram plots focus on visualizing the empirical distributions, the specific K-S test statistics for each distribution will be presented in a separate summary table for clarity. Scheduled arrival times were added as vertical reference lines in the histograms, enabling visual comparison with observed arrivals to highlight delays and variability.

After identifying the best-fitting distribution, we used its estimated parameters to generate synthetic data matching the original sample size. While the method allows generating any number of values, we chose this size to preserve the structure and behavior of the observed data. To assess how well the generated data preserved the characteristics of the original data, both observed and synthetic datasets were visually compared using histograms, and their statistical properties were summarized in terms of mean, standard deviation, skewness, and kurtosis. The close agreement between the statistical moments of real and generated data demonstrates the effectiveness of the chosen method in replicating key distributional features. A deliberately misspecified distribution—the log-normal—was also used to generate synthetic data to evaluate the effect of model choice on simulation quality. Although it is commonly employed in the literature for modeling timetable data (Yuan et al., 2010; Jin et al., 2009; Kim, 2016), the log-normal distribution was not selected as the best fit by the K-S test for the tested dataset. By comparing simulations based on the best-fitting and incorrect models, we were able to assess the influence of distributional assumptions on the representativeness and accuracy of generated data, as well as identify potential biases introduced by model misspecification.

3. Numerical results

This section presents the numerical results of this study. First, the conclusions from the statistical analysis are discussed, followed by the introduction of the synthetic data based on these findings. We study the railway corridor from Malmö to Karlskrona. It consists of two main railway lines: the Southern Main Line (Södra stambanan) between Malmö and Hässleholm, and the Blekinge Coastal Line (Blekingekustbanan) from Kristianstad to Karlskrona. Along this route, the key stations are Malmö, Lund, Hässleholm, Kristianstad, Sölvesborg, Mörrum, Karlshamn, Bräkne-Hoby, Ronneby, and Karlskrona. Along this line, we study the ten stations in both travel directions. The segment between Karlshamn and Sölvesborg, located along the Blekinge Coastal Line, forms a critical part of the east-west rail connection across southern Sweden, supporting both regional mobility and national transport resilience. For visual illustration of the considered railway corridor, Figure 1 shows the locations of the stations included in the analysis.

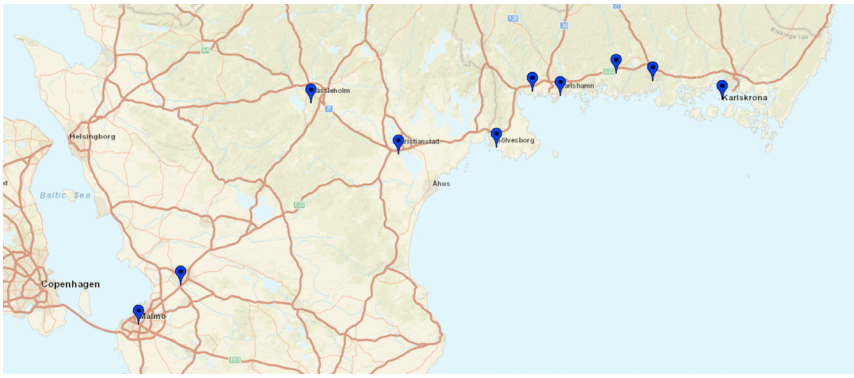


Fig. 1. Geographical locations of the analyzed stations.

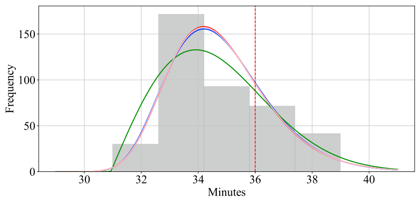
3.1. Arrival time distribution analysis

The results presented in this section highlight variations in train arrival patterns and lay the groundwork for a more detailed analysis of station-specific and direction-specific behaviors. Fig. 2 presents four histograms selected as representative samples of the results obtained in this study. These histograms were included to visually illustrate the differences in behavior identified among the stations, reinforcing the observed statistical variations in travel patterns. They exemplify the diversity of delay distributions across different locations, helping to motivate the necessity of individualized analysis for each station. For brevity, the remaining histograms are not presented; however, the selected four are representative of the broader trends observed across all analyzed stations. In the histograms, the gamma-, log-normal-, Rayleigh-, and inverse Gaussian-based models are presented by the curves in blue, red, green, and pink, respectively. In cases where the train arrived close to the 59th minute, the data were transformed into the range of 60 to 120 for better visualization of the histogram.

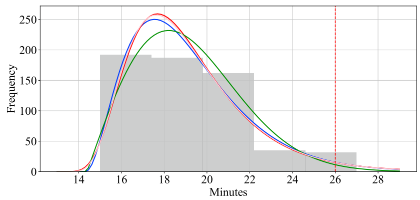
The K-S test statistics further support these findings, as shown in Table 1. The best-fitting model for each station was selected based on the lowest test statistic, indicated in bold. The log-normal distribution, which is widely used in the literature to describe train arrivals and departures, emerged as the best fit for Karlshamn and Sölvesborg (Malmö to Karlskrona direction). However, the gamma distribution yielded the lowest test statistic for Kristianstad, suggesting a better fit in this case. Furthermore, for Sölvesborg (Karlskrona to Malmö direction), the inverse Gaussian model provided the best fit, reinforcing the need for individualized modeling at each station and direction.

Across all 20 stations analyzed (ten stations in two different travel directions), the log-normal distribution emerged as a viable candidate in 14 cases. This result highlights its general suitability in capturing the variability of travel times or delays in this dataset, aligning with its widespread use in the literature for modeling train arrivals. However, despite its frequent applicability, the log-normal distribution is not universally optimal, as evidenced by stations where gamma or inverse Gaussian distributions provided a better fit. Among the ten analyzed stations, only Ronneby and Karlshamn exhibit similar statistical behaviors in both directions, suggesting the presence of underlying factors that warrant deeper exploration. Since these stations are geographically close to each other, this similarity was expected compared to more distant stations. This proximity may influence train arrival patterns, hinting at the need for further investigation into local environmental or operational conditions that could contribute to deviations from the timetable.

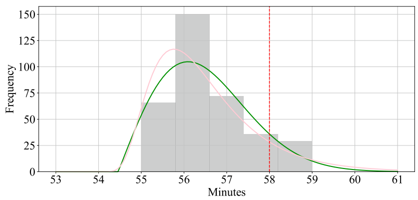
The results further demonstrate that train arrival behavior varies significantly between different travel directions. For instance, from Sölvesborg to Malmö, the log-normal distribution does not fit the data well, whereas in the opposite direction (Malmö to Sölvesborg), it proves to be a strong candidate for modeling arrival times. Additionally,



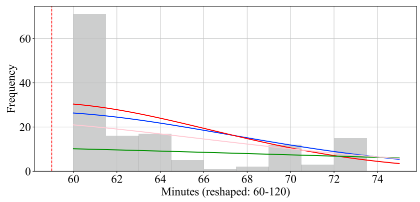
(a) Karlshamn – From Karlskrona to Malmö



(b) Kristianstad – From Karlskrona to Malmö



(c) Sölvesborg – From Karlskrona to Malmö



(d) Sölvesborg – From Malmö to Karlskrona

Fig. 2. Histogram, fitted distributions, and schedule time for three stations. The gamma-, log-normal-, Rayleigh-, and inverse Gaussian-based models are presented by the curves in blue, red, green, and pink, respectively. The scheduled arrival time is represented by the vertical red dashed line.

		Distribution			
Station	Travel direction	gamma	log-normal	Rayleigh	inverse Gaussian
Karlshamn	From Karlskrona to Malmö	0.1350	**0.1349**	0.1827	0.1354
Kristianstad	From Karlskrona to Malmö	**0.0877**	0.0916	0.1200	0.0887
Sölvesborg	From Karlskrona to Malmö	—	—	0.2541	**0.2278**
Sölvesborg	From Malmö to Karlskrona	0.3575	**0.3442**	0.5339	0.4353

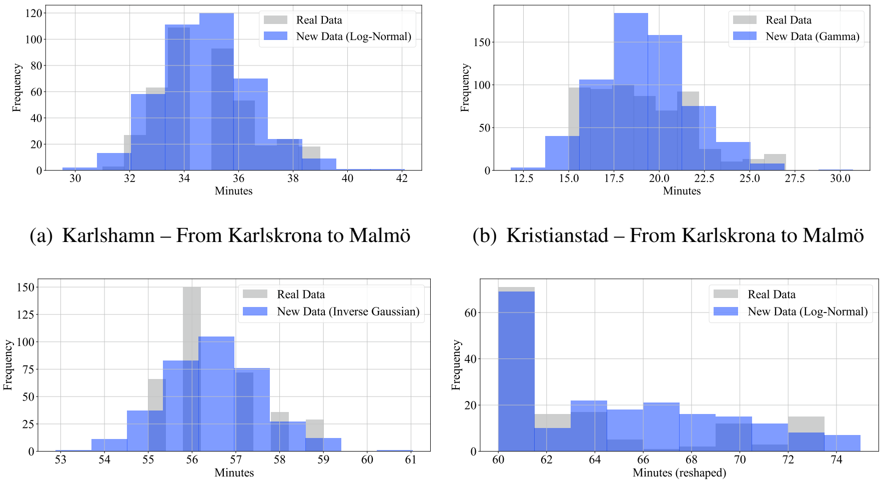
Table 1. Kolmogorov-Smirnov (K-S) test statistics for fitting different probability distributions to train arrival data at various stations. Bold values indicate the lowest test statistic for each station, representing the best-fitting distribution.
Note: — indicates that the hypothesis of model fit was rejected based on the K-S test.

the variability in delays heading toward Sölvesborg is considerably greater than in the opposite direction, further emphasizing the necessity of direction-specific modeling. Another notable case is Kristianstad, which exhibits distinct delay patterns due to its role as a connection hub where trains attach or detach. This operational characteristic significantly impacts stop-time behavior, which, in turn, affects delay patterns. The observed discrepancies underscore the importance of station-specific modeling to better understand and mitigate these anomalies, ensuring a more accurate representation of railway network dynamics.

3.2. Synthetic data

With the pre-analysis conducted, it is possible to identify distinct statistical behaviors both between stations and within stations across different travel directions. A deeper understanding of the real data characteristics allows for incorporating this information into algorithms for generating synthetic data. This ensures that the generated data more accurately reflects the real-world situation at each studied station rather than presenting a generalized, aggregated behavior. Figure 3 presents four histograms comparing the real and generated data. In these plots, the gray bars represent the real data, while the blue bars correspond to the newly generated data. The synthetic data were generated by following the statistical properties of the real observations, ensuring that the ranges of values remained similar. This allows for a realistic comparison between observed and simulated data distributions.

The histograms illustrate how the generated data effectively captured the observed data properties. In Karlshamn (Figure 3(a)), the generated data follows a log-normal distribution, which was identified as the best fit based on the Kolmogorov–Smirnov test. The distribution aligns well with the real data, capturing the overall shape and variability of



(a) Karlshamn – From Karlskrona to Malmö (b) Kristianstad – From Karlskrona to Malmö

(c) Sölvesborg – From Karlskrona to Malmö (d) Sölvesborg – From Malmö to Karlskrona

Fig. 3. Real and simulated data from direct sampling method.

arrival times. In Kristianstad (Figure 3(b)), where a gamma distribution was found to be the best fit, the generated data follows a similar structure, preserving the observed delay patterns. For Sölvesborg (Malmö to Karlskrona direction) (Figure 3(c)), the log-normal model was used, reflecting its suitability for this specific case. In contrast, for Sölvesborg (Karlskrona to Malmö direction) (Figure 3(d)), the inverse Gaussian distribution was found to be the best fit. The generated data successfully replicates the characteristics of the real arrival times, further demonstrating the importance of selecting appropriate distributions for synthetic data generation.

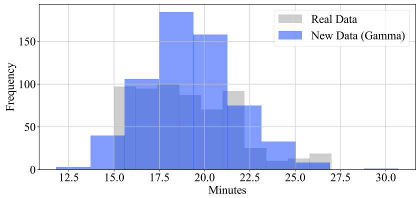
In addition to the visual inspection, a quantitative comparison of the main statistical moments—mean, standard deviation, skewness, and kurtosis—is provided in Table 2. Across all tested cases, the synthetic datasets demonstrate a high level of agreement with most of the observed data, capturing not only the central tendency and dispersion but also the asymmetry and tail behavior of the distributions. This consistency reinforces the conclusion that a statistically grounded approach to distribution selection results in synthetic data that faithfully reflects the underlying patterns of train arrival times. One of the most challenging cases was Sölvesborg, in the direction from Malmö to Karlskrona. This station is known for being a point where train connection and disconnection operations occur, which naturally brings more uncertainty and variability to arrival times. This is reflected in the real data, which shows a high standard deviation (25.66) and a low kurtosis value (–1.58), meaning that the distribution is very spread out and flatter than a normal distribution. Because of that, this scenario is one of the hardest to model realistically. Even so, the synthetic data generated using the best-fitting distribution was able to reproduce the overall behavior of the real data. The model successfully captured the negative skewness and negative kurtosis, which are key characteristics of the original distribution. Although the mean in the generated data was slightly higher than in the real data (56.34 vs. 37.82), the general shape and variability of the distribution were maintained. This shows that using a station-specific modeling approach allows us to produce realistic simulations, even in more complex operational contexts like Sölvesborg. By applying station- and direction-specific distributions, the synthetic dataset effectively mimics real-world variations, making it a valuable tool for AI-based timetable rescheduling and scenario simulations.

To further evaluate the impact of model selection in synthetic data generation, we conducted an experiment using an incorrect distribution. Specifically, we applied the log-normal distribution to cases where it was not the best fit according to the K-S test. This allowed us to assess how the choice of an unsuitable distribution affects the statistical properties of the generated data and its alignment with real observations. Figure 4 illustrates this effect for Kristianstad, where one histogram presents the synthetic data generated using the best-fitting distribution (gamma distribution), while the other shows the synthetic data generated using the incorrect log-normal distribution. In both histograms, the gray bars represent the real data, while the blue bars correspond to the generated data. On the one hand, when the correct distribution (gamma) was used (Figure 4(a)), the generated data effectively replicated the real distribution's shape, peak,

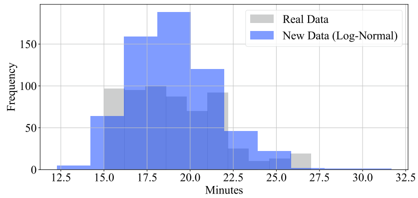
Table 2. Comparison of statistical moments between real and generated data

Station	Travel direction	Mean	Std. Dev.	Skewness	Kurtosis
Karlshamn	From Karlskrona to Malmö	34.77 / 34.79	1.76 / 1.67	0.60 / 0.35	0.02 / 0.69
Kristianstad	From Karlskrona to Malmö	19.10 / 19.19	2.65 / 2.49	0.83 / 0.36	0.35 / 0.49
Sölvesborg	From Karlskrona to Malmö	56.47 / 56.47	1.15 / 1.09	0.72 / 0.23	−0.22 / 0.65
Sölvesborg	From Malmö to Karlskrona	37.82 / 56.34	25.66 / 19.82	−0.58 / −0.68	−1.58 / −0.78

Note: Each pair of values refers to (Real / Generated).



(a) Direct sampling – Gamma distribution



(b) Direct sampling – Log-Normal distribution

Fig. 4. Real and simulated data based on the best distribution model and the Log-Normal for Kristianstad – From Karlskrona to Malmö

and spread, capturing the overall delay characteristics at Kristianstad. The kurtosis values for this case—0.35 for the real data and 0.49 for the generated data—confirm that the shape of the synthetic distribution remained close to the observed one, including the tail behavior and the peakedness of the data. On the other hand, when the log-normal distribution was incorrectly applied (Figure 4(b)), clear distortions appeared in the synthetic data. The tail behavior is misrepresented, with a much higher kurtosis of 1.02, compared to the original 0.35. This overestimates the concentration of data near the center and exaggerates the tail, creating a mismatch with the true pattern of arrival times. Such deviation leads to biased results and may negatively affect the performance of models that rely on synthetic data. This process was repeated for all other stations where the log-normal distribution was not statistically significant, ensuring a broader validation of these findings. However, for brevity, only Kristianstad's case is presented here.

These results emphasize the importance of proper distribution selection in data generation. While the log-normal distribution is widely used in the literature, it is not universally applicable, and assuming an incorrect model can introduce bias, misrepresentation of delays, and unrealistic scenario modeling. Selecting the statistically best-fitting distribution ensures that synthetic data retains the empirical characteristics of real observations, leading to more accurate simulations for applications such as AI-based railway rescheduling and predictive modeling.

4. Conclusions

This study investigated the statistical behavior of train arrivals across different stations and travel directions in southern Sweden. The results confirm that train delays follow diverse statistical patterns depending on location, train type, and direction, making it difficult to apply a single universal distribution for modeling all cases. The K-S test was used to identify the best-fitting probability distributions, revealing that while the log-normal distribution was a viable candidate in 14 out of 20 cases, alternative distributions such as gamma and inverse Gaussian provided better fits in certain stations. These findings highlight the necessity of station- and direction-specific modeling to accurately capture railway dynamics. Building on these insights, we developed a synthetic dataset based on the best-fitting distributions. The generated data effectively preserved the characteristics of real train arrivals, ensuring that the range and variability of arrival times were well-represented. This approach provides realistic scenarios for AI-based timetable rescheduling, supporting better decision-making in railway operations. While the analysis is based on a specific regional line, the methodology applies to other railway corridors with similar data availability and operational contexts. This study underscores the importance of data-driven approaches in railway research, particularly in enhancing AI-based scheduling systems. By using rigorous statistical testing and distribution selection, synthetic datasets can more accurately reflect real-world railway dynamics. Future research will explore the application of these findings in real-time timetable rescheduling and AI-driven predictive modeling, further improving railway efficiency and punctuality.

This work has limitations that must be acknowledged. It focused primarily on statistical patterns in train arrival times, without including detailed technical or infrastructural features of each station—such as signaling systems, control systems, station layouts, or local operational constraints. These elements can play an important role in shaping delay patterns. While leaving them out makes the model simpler, it may reduce how well the results apply to other train networks where such features have a strong influence. Our method captures the variability in train arrivals using probability distributions, but it does not directly consider the physical or operational causes behind these patterns. As a result, caution is needed when applying these findings to networks with very different infrastructure or operations. Future work should consider integrating infrastructure-specific parameters and metadata about station operations to better link statistical observations to operational causes.

References

Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M., 2011. Stochastic delay prediction in large train networks, in: 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, pp. 100–111.

Bergström, A., Krüger, N., 2013. Modeling passenger train delay distributions: evidence and implications. Technical Report. Centre for Transport Studies Stockholm, Swedish National Road & Transport.

Goverde, R.M., Hansen, I., Hooghiemstra, G., Lopuhaa, H., 2001. Delay distributions in railway stations. Proceedings of the 9th WCTR .

Huang, P., Li, Z., Zhu, Y., Wen, C., Corman, F., 2023. Train traffic control in merging stations: A data-driven approach. Transportation Research Part C: Emerging Technologies 152, 104155.

Huang, P., Wen, C., Li, J., Peng, Q., Li, Z., Fu, Z., 2019. Statistical analysis of train delay and delay propagation patterns in a high-speed railway system, in: 2019 5th International Conference on Transportation Information and Safety (ICTIS), IEEE. pp. 664–669.

Jiang, S., Persson, C., 2016. Malfunction in railway system and its effect on arrival delay, in: Current Trends in Reliability, Availability, Maintainability and Safety: An Industry Perspective, Springer. pp. 243–252.

Jin, X., Zhang, Y., Wang, F., Li, L., Yao, D., Su, Y., Wei, Z., 2009. Departure headways at signalized intersections: A log-normal distribution model approach. Transportation Research Part C: Emerging Technologies 17, 318–327.

Järnvägsbranschens samverkansforum (JBS), 2023. TTT – tillsammans för tåg i tid: En redovisning av 2023 års arbete.

Kim, M.S., 2016. Analysis of short-term forecasting for flight arrival time. Journal of Air Transport Management 52, 35–41.

Krüger, N.A., Vierth, I., Fakhraei Roudsari, F., 2013. Spatial, temporal and size distribution of freight train delays: evidence from Sweden. Technical Report. The Swedish National Road and Transport Research Institute.

Minbashi, N., Bohlin, M., Kordnejad, B., 2020. A departure delay estimation model for freight trains. The 8th Transport Research Arena 2020:, Helsinki, Finland .

Palmqvist, C.W., Olsson, N., Hiselius, L., 2017. An empirical study of timetable strategies and their effects on punctuality, in: 7th International Conference on Railway Operations Modelling and Analysis, pp. 1–21.

Şahin, İ., 2017. Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. Journal of Rail Transport Planning & Management 7, 101–113.

Tang, R., De Donato, L., Besinović, N., Flammini, F., Goverde, R.M., Lin, Z., Liu, R., Tang, T., Vittorini, V., Wang, Z., 2022. A literature review of artificial intelligence applications in railway systems. Transportation Research Part C: Emerging Technologies 140, 103679.

Tiong, K.Y., Ma, Z., Palmqvist, C.W., 2023. Analyzing factors contributing to real-time train arrival delays using seemingly unrelated regression models. Transportation Research Part A: Policy and Practice 174, 103751.

Wen, C., Li, Z., Lessan, J., Fu, L., Huang, P., Jiang, C., 2017. Statistical investigation on train primary delay based on real records: evidence from wuhan–guangzhou hsr. International Journal of Rail Transportation 5, 170–189.

Yang, Y., Huang, P., Peng, Q., Li, J., Wen, C., 2019. Statistical delay distribution analysis on high-speed railway trains. Journal of Modern Transportation 27, 188–197.

Yuan, J., 2006. Stochastic modelling of train delays and delay propagation in stations. Phd thesis. Delft University of Technology.

Yuan, J., Goverde, R., Hansen, I., 2010. Evaluating stochastic train process time distribution models on the basis of empirical detection data. WIT Transactions on State of the Art in Science and Engineering 40, 95–104.

Yuan, J., Hansen, I.A., 2002. Punctuality of train traffic in dutch railway stations. Traffic And Transportation Studies (2002) , 522–529.

Zhang, J., Zhang, J., 2023. Artificial intelligence applied on traffic planning and management for rail transport: A review and perspective. Discrete Dynamics in Nature and Society 2023, 1832501.