Data-driven Train Delay Prediction

Tiong, Kah Yong

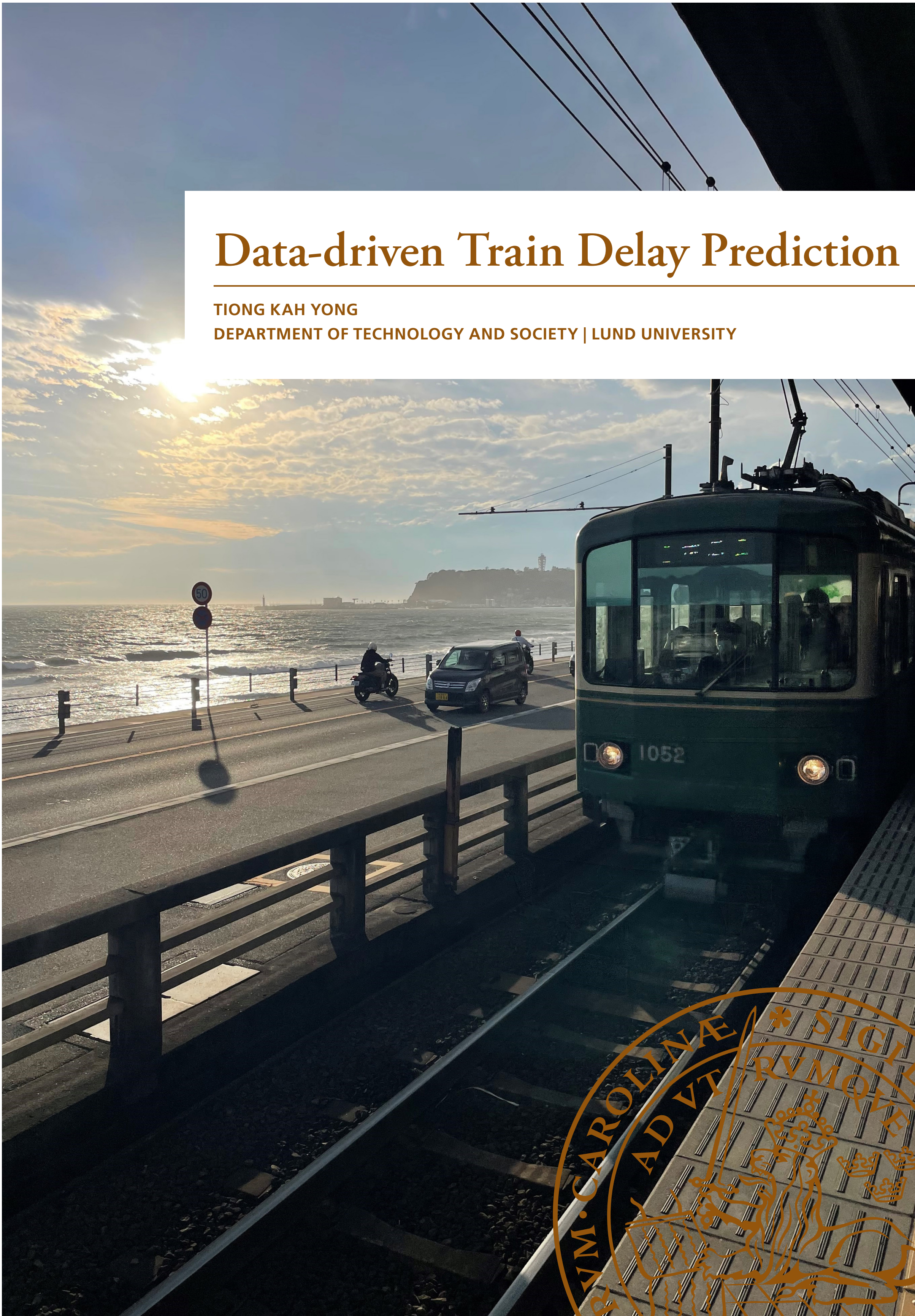
2024

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Data-driven Train Delay Prediction

TIONG KAH YONG
DEPARTMENT OF TECHNOLOGY AND SOCIETY | LUND UNIVERSITY



In Sweden, the sensitivity of the interconnected railway system is heightened by the increased capacity utilisation and inherent heterogeneity in train traffic, which can lead to delays that easily propagate throughout the railway network. Technological advancements enable the adoption of data-driven approaches such as predictive models to tackle train delay issues by improving train traffic management and passenger planning. The thesis examines the existing train delay prediction models and introduces innovative approaches to enhance the model's performance. The ultimate goal is to increase the understanding of data-driven train delay prediction models, thereby accelerating the adoption of data-driven methods in the railway research community.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Engineering
Department of Technology and Society

Bulletin – Lund University, Faculty of Engineering,
Department of Technology and Society, 333
ISBN 978-91-8039-970-8
ISSN 1653-1930

9 789180 399708

Data-driven Train Delay Prediction

Data-driven Train Delay Prediction

Tiong Kah Yong

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering, Lund University, Sweden. To be defended at the Faculty of Engineering, John Ericssons väg 1, Auditorium V: A, 8 May 2024 at 14:00.

Faculty opponent
Professor Ronghui Liu, University of Leeds

Organization: Lund University

Document name: Doctoral dissertation

Date of issue: 8 May 2024

Author(s): Tiong Kah Yong

Sponsoring organization: The Swedish Transport Administration

Title and subtitle: Data-driven Train Delay Prediction

Abstract:

In Sweden, extensive use and the inherent heterogeneity of train traffic have significantly increased the sensitivity of the train system so that the delay of one train can easily propagate to others. Repeated experiences with train delays can lead passengers to perceive that train transportation is unreliable. Despite being considered a green mode of transport, this environmental benefit truly comes into effect when passenger volumes are sufficiently high. Therefore, minimising train delays becomes crucial to promote a modal shift from private vehicles to railways.

The advent of advanced technologies facilitating the collection and storage of extensive train operation data has paved the way for addressing train delay issues from a data-driven perspective, thus leading to a predominant focus on train delay prediction research. To develop theoretical and practical knowledge for the continuous advancement of decision-support tools, this thesis aims to explore and understand data-driven train delay prediction. The thesis is grounded in the findings of six papers. Paper 1 systematically reviews existing literature on data-driven approaches for predicting train delays, captures commonly adopted technical solutions and identifies weaknesses in current models. It suggests promising directions for future research in this area while highlighting under-researched prediction issues. To ascertain useful input variables, Papers 2 and 3 employ statistical regression to quantify the relationship between various explanatory variables and train delays. Papers 4 and 5 address the development of robust data-driven train delay prediction models and introduce dynamic multi-output models capable of continuously predicting train arrival delays for multiple downstream stations at arbitrary prediction times. To enhance performance, the studies further introduce error adjustment strategies that continuously correct predictions based on observed train traffic information. To ensure real-world effectiveness, Paper 6 seeks to construct an evaluation framework for a thorough assessment of train delay prediction models.

The main contribution of the thesis is twofold. Firstly, it sheds light on the current practices in data-driven train delay prediction studies, synthesising progress in various aspects of model development and highlighting the limitations of existing modelling techniques. Secondly, the thesis introduces innovative approaches to enhance model performance. For example, it identifies limitations in current evaluation processes and introduces an evaluation framework to address these gaps. Recognizing the limitations of the current focus on one-step-ahead prediction for practical applications, the thesis introduces a dynamic multi-output modelling framework that generates predictions for all downstream stations at arbitrary times. Overall, the thesis helps to bring greater transparency to this growing field of research, with the ultimate goal of accelerating the adoption of data-driven approaches in the railway research community.

Key words: Predictive models, Machine learning, Data-driven, Railways, Trains, Delays

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title: 1653-1930
Bulletin – Lund University, Faculty of Engineering, Department of Technology and Society, 333

ISBN: 978-91-8039-970-8 (print)
978-91-8039-971-5 (electronic)

Number of pages: 112

Signature

Date 2023-03-21

Data-driven Train Delay Prediction

Tiong Kah Yong

LUND UNIVERSITY

Cover photo by Michelle Ochsner
Back cover photo by Manying Zhao

Lund University
Faculty of Engineering
Department of Technology and Society

ISBN 978-91-8039-970-8 (print)
978-91-8039-971-5 (electronic)
ISSN 1653-1930

Printed in Sweden by Media-Tryck, Lund University, Lund 2024

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

Printed matter 3041 0903

MADE IN SWEDEN

To my family

Table of Contents

List of tables 12
List of figures 12
List of abbreviations 13
Abstract 15
Popular science summary 16
Acknowledgements 18

1 Introduction 20

1.1 Operational challenges arising from train delays 21
1.2 Passenger dissatisfaction caused by train delays 21
1.3 Economic ramifications of train delays 22
1.4 Train delay-induced environmental impacts 23
1.5 Toward an intelligent transport system 23
1.6 Thesis outline 24

2 Background 26

2.1 Railway timetable 26
2.1.1 Scheduled running time 26
2.1.2 Scheduled dwell time 27
2.1.3 Scheduled headway 28
2.1.4 Capacity utilisation 29
2.2 Train delays 29
2.2.1 Cause of primary delays 30
2.2.2 Cause of secondary delays 31
2.3 Train delay prediction models 31
2.3.1 Data-driven train delay prediction approaches 32
2.4 Use cases of train delay prediction models 33
2.4.1 Decision-support tool for investment planning 34
2.4.2 Decision-support tool for timetable planning 34
2.4.3 Decision-support tool for real-time train management 35
2.4.4 Reliable passenger information system 36

2.5 Research gaps 37
2.5.1 Insufficient model understanding 37
2.5.2 Innovation for practical applications 38

3 Aim 40

3.1 RQ1 What factors need to be taken into account when building a train delay prediction model? 40

3.2 RQ2 How are selected input variables improving the performance of the train delay prediction model? 41

3.3 RQ3 What approaches can enhance the train delay prediction model? 41

3.4 RQ4 How can train delay prediction models be evaluated? 41

4 Research design 42

4.1 List of included papers 42
4.1.1 Author's contribution to the included papers 43
4.1.2 Related publications not included in the thesis 44

4.2 Relationship between included papers 45

4.3 Relationship between the research questions and included papers 47

5 Methods 49

5.1 Content Analysis 50

5.2 Regression Analysis 51
5.2.1 Logistic regression 52
5.2.2 Seemingly unrelated regression 53

5.3 Predictive Analysis 54
5.3.1 Light gradient boosting machine 54
5.3.2 Gradient boosting regression 55
5.3.3 Random forest regression 55
5.3.4 Linear regression 55

5.4 Model evaluation 56
5.4.1 Area Under the Curve of the Receiver Operating Characteristic 56
5.4.2 Coefficient of determination 56
5.4.3 Root-mean-square error and mean absolute error 57

6 Data preparation and case study 58

6.1 Datasets 58
6.1.1 Reviewed literature 59
6.1.2 Train Operation Data 59
6.1.3 Weather data 60
6.1.4 Trackwork plan 60

6.2 Data preprocessing 61
6.2.1 Data cleaning 62
6.2.2 Feature Engineering..62
6.2.3 Feature selection 64
6.3 Study areas and scope of data..64
6.3.1 The global perspective in Papers 1 and 6 64
6.3.2 Swedish cases in Papers 2–5..65

7 Summary of Papers 68
7.1 Paper 1: A Review of Data-driven Approaches to Predict Train Delays 68
7.2 Paper 2: The Effects of Train Passes on Dwell Time Delays in Sweden 69
7.3 Paper 3: Analysing Factors Contributing to Real-time Train Arrival Delays using Seemingly Unrelated Regression Models.......................70
7.4 Paper 4: Real-time Train Arrival Time Prediction at Multiple Stations and Arbitrary Times 71
7.5 Paper 5: Real-time Train Arrival Time Prediction along the Swedish Southern Mainline..72
7.6 Paper 6: AP-GRIP Evaluation Framework for Data-driven Train Delay Prediction Models: A Systematic Literature Review................73

8 Answering the research questions..74
8.1 Factors to consider when building a train delay prediction model.......74
8.1.1 Scope determination 75
8.1.2 Model Input 76
8.1.3 Data quality..76
8.1.4 Methodologies 76
8.1.5 Model outputs 77
8.2 Input variables improving the train delay prediction model................77
8.2.1 Operational variables..78
8.2.2 Weather variables 78
8.2.3 Calendar variables 79
8.2.4 Maintenance variables 79
8.3 Approaches to enhance the train delay prediction model.....................80
8.3.1 Location-conditioned concept 80
8.3.2 Multi-output framework 81
8.3.3 Error adjustment strategies 82

8.4 Evaluating train delay prediction models ... 84
8.4.1 Evaluation aspect: Accuracy ... 84
8.4.2 Evaluation aspect: Precision ... 85
8.4.3 Evaluation aspect: Generalisability ... 86
8.4.4 Evaluation aspect: Interpretability ... 86
8.4.5 Evaluation aspect: Robustness ... 87
8.4.6 Evaluation aspect: Practicability ... 88
8.4.7 Evaluation dimensions ... 88

9 Synthesis ... 90

9.1 Scope ... 91
9.2 Input ... 91
9.3 Methodologies ... 92
9.4 Output ... 93
9.5 Evaluation ... 93

10 Conclusion ... 95

10.1 Contribution ... 96
10.1.1 Enhancing the long-term train delay prediction model ... 96
10.1.2 Enhancing the short-term train delay prediction model ... 97
10.1.3 Uncovering the model development process ... 97
10.1.4 Improving the model selection process ... 98
10.2 Recommendations to put into practice ... 98
10.2.1 Timetable planning ... 98
10.2.2 Train management systems ... 99
10.2.3 Passenger information system ... 99
10.3 Limitations ... 100
10.4 Future research ... 101

References ... 103

List of tables

Table 1: Connection between the research questions and the included papers 47
Table 2: Overview of the different method used 50
Table 3: Overview of data used 58
Table 4: Case study for the Paper 2–5 66

List of figures

Figure 1: Connection between the six papers 45
Figure 2: The data preprocessing process.. 61
Figure 3: Capacity use in Sweden 2020 (Trafikverket, 2023)............................... 65
Figure 4: Components in existing train delay prediction models 75
Figure 5: The factors examined in the thesis for
their impact on train arrival delays 78
Figure 6: Enhancements for existing train delay prediction models 80
Figure 7: Conditional multi-outputs framework... 82
Figure 8: (a) One-step before prediction error correction;
(b) Upstream prediction error correction 83
Figure 9: Measures for different evaluation aspects and dimensions................... 84
Figure 10: Train delay prediction framework. Source:
Adapted from Tiong et al. (2023) 90

List of abbreviations

AIC	Akaike information criterion
AP-GRIP	Accuracy, Precision-Generalisability, Robustness, Interpretability, and Practicality
AUC-ROC	Area Under the Curve of the Receiver Operating Characteristic
BIC	Schwarz's Bayesian Information Criterion
BO	Bayesian optimisation
CNNs	Convolution Neural Networks
CO2	Carbon Dioxide
CST	Stockholm Central Station
CV	Coefficient of Variation
DGM	Deep Generative Models
FCNN	Fully Connected Neural Networks
GAN	Generative Adversarial Networks
GBR	Gradient Boosting Regression
GHG	Greenhouse Gases
GMRAE	Geometric Mean Relative Absolute Error
HSR	High-Speed Train
ITS	Intelligent Transport Systems
KPH	Copenhagen Central Station
LIME	Local Interpretable Model-agnostic Explanation
LightGBM	Light Gradient Boosting Machine
LORE	Local Rule-based Explanations
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAPLE	Model Agnostic Supervised Local Explanations
OHE	One Hot Encoding
OLS	Ordinary Least Squares Regression

R^2	Coefficient of Determination
RFR	Random Forest Regression
RMdSE	Root Median Squared Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
RQ	Research Question
SDC	Sundsvall Central Station
SHAP	Shapley Additive Explanations
SMHI	Swedish Meteorological and Hydrological Institute
SUR	Seemingly Unrelated Regression
SVR	Support Vector Regression
SML	Southern Main Line
Trafikverket	Swedish Transport Administration
VAE	Variational Autoencoders
WMAE	Weighted Mean Absolute Error
XGBoost	eXtreme Gradient Boosting

Abstract

In Sweden, extensive use and the inherent heterogeneity of train traffic have significantly increased the sensitivity of the train system so that the delay of one train can easily propagate to others. Repeated experiences with train delays can lead passengers to perceive that railway transportation is unreliable. Despite being considered a green mode of transport, this environmental benefit truly comes into effect when passenger volumes are sufficiently high. Therefore, minimising train delays becomes crucial to promote a modal shift from private vehicles to railways.

The advent of advanced technologies facilitating the collection and storage of extensive train operation data has paved the way for addressing train delay issues from a data-driven perspective, thus leading to a predominant focus on train delay prediction research. To develop theoretical and practical knowledge for the continuous advancement of decision support tools, this thesis aims to explore and understand data-driven train delay prediction. The thesis is grounded in the findings of six papers. Paper 1 systematically reviews existing literature on data-driven approaches for predicting train delays, captures commonly adopted technical solutions, and identifies weaknesses in current models. It suggests promising directions for future research in this area while highlighting under-researched prediction issues. To ascertain useful input variables, Papers 2 and 3 employ statistical regression to quantify the relationship between various explanatory variables and train delays. Papers 4 and 5 address the development of robust data-driven train delay prediction models, introducing dynamic multi-output models capable of continuously predicting train arrival delays for multiple downstream stations at arbitrary prediction times. To enhance performance, the studies further introduce error adjustment strategies that continuously correct predictions based on observed train traffic information. To ensure real-world effectiveness, Paper 6 seeks to construct an evaluation framework for a thorough assessment of train delay prediction models.

The main contribution of the thesis is twofold. Firstly, it sheds light on the current practices in data-driven train delay prediction studies, synthesising progress in various aspects of model development and highlighting the limitations of existing modelling techniques. Secondly, the thesis introduces innovative approaches to enhance model performance. For example, it identifies limitations in current evaluation processes and introduces an evaluation framework to address these gaps. Recognizing the limitations of the current focus on one-step-ahead prediction for practical application, the thesis introduces a dynamic multi-output modelling framework that generates predictions for all downstream stations at arbitrary times. Overall, the thesis helps to bring greater transparency to this growing field of research, with the ultimate goal of accelerating the adoption of data-driven approaches in the railway research community.

Popular science summary

The demand for train services has increased steadily in Sweden over the years. Increased capacity utilisation combined with heterogeneous train traffic makes the railway network in Sweden susceptible to delays. The widespread nature of train delays creates substantial adverse impacts on train operations, passengers, the environment and the economy. For instance, train delays increase the risk of a train exceeding its scheduled track occupation time, which hinders other trains or even causes route conflicts. Repeated experiences with delays can lead passengers to perceive that train transportation is unreliable, thus generating negative word-of-mouth publicity and inevitably deteriorating the image of railway transport. These phenomena demotivate the modal shift from private vehicles to railways, which results in lost revenue from train operations while causing the transport sector to continue contributing a larger share of total $CO2$ emissions, the major source of greenhouse gases. This, consequently, leads to climate issues.

One solution to address train delays is to expand railway capacity, but constructing new railways is capital-intensive and problematic in terms of time and environmental constraints. An alternative solution is to provide decision-support tools such as train delay prediction models for passengers and train operators. For train dispatchers, predictive models provide essential information on expected train trajectories and conflicts, thus enabling necessary adjustments to be made for resolving conflicts between train paths and enhancing the operational efficiency and capacity of the railway transportation system. For passengers, predicted train delays make it easier to identify valid alternatives when planning the best connections for their trips. In current practice, train dispatchers play an important role in real-time train management by solving any potential train conflicts and recovering disruptive events based on their experience. Similarly, the current passenger information system in Sweden relies on manual forecasts generated based on the scheduled timetable and updated by dedicated staff. These practices are heavily experience-oriented and have the limitation of assuming the expected arrival delays are equal to the current upstream delays, neglecting the fact that some trains recover from delays by running in the maximum performance regime while others become more delayed due to possible time loss in route conflicts. The growing availability of data makes it possible to address train delay problems from a data-driven perspective. However, there remains a lack of clarity regarding the specific requirements for developing such models. Therefore, the goal of the thesis is to improve clarity in the development process of data-driven train delay prediction models.

To identify useful input variables for the train delay prediction models, the thesis examines the heterogeneous impact of various factors derived from multiple data sources on train delays. My research shows that the variables derived from the train operation data (e.g. dwell time, running time, and delays from previous trains and upstream stations) have the most significant impacts on train arrival delays at downstream stations. We also found that it is important to consider the temporal and spatial perspectives of explanatory variables given that the variables from the nearest stations in real-time have a greater impact on subsequent train delays. More specifically, the consecutive upstream station delay of the same train has a greater impact on the current station train delay than delays from further upstream stations because the consecutive upstream stations are more reflective of current traffic conditions since the impacts of further upstream delays are absorbed by the consecutive upstream station delay.

Although many existing studies focus on building one-step-ahead prediction models, the train delay prediction problem is more complex than it may at first seem. For instance, passengers have varied concerns about stations and thus have varying needs for train arrival time information based on their current locations. To build a prediction model with strong predictive capabilities that is useful in practical applications, we propose a dynamic multi-output machine learning model that can generate predictions for multiple downstream stations at arbitrary times. We further introduce an innovative approach—adaptive error adjustment strategies—in which both real-time and historical observed information is leveraged to enhance the model's performance. Our findings show that this approach enables the model to continuously correct its prediction output, thus improving prediction accuracy.

To date, it has been shown that existing studies heavily emphasise prediction accuracy when evaluating the train delay prediction models, which has led to scepticism as to their practical effectiveness. At the same time, we need to know when, where and why certain models excel or perform poorly under certain circumstances. We simply need more knowledge regarding their predictive capabilities in diverse circumstances and the inadequacy of their prediction patterns. Thus, we proposed a model performance evaluation framework that assesses a model based on six key aspects (accuracy, precision, generalisation, interpretability, robustness and practicality) and multiple dimensions, including overall, spatial, temporal and train-specific dimensions. The comprehensive assessment bridges the gap between the theoretical and real-world implementation of train delay prediction models.

Acknowledgements

Finally, I have now reached the moment of composing my PhD dissertation. Undertaking a PhD has proven to be one of the best decisions I've ever made. Throughout this unique journey, I've gained invaluable experience and found immense joy in the process. Therefore, I would like to take this opportunity to thank the people who have been important to me during my PhD journey.

First of all, I would like to thank Carl-William Palmqvist, my main supervisor, for always being so supportive, approachable, compassionate and understanding during my supervision. My sincere thanks also go to my co-supervisors, Lena Hiselius and Nils Olof Emanuel Olsson, for their continuous support and encouragement. I feel so thankful for having the opportunity to work with Zhenliang Ma. I am very grateful for having had the chance to learn from you and for the knowledge and experience gained through our work together.

I also give special thanks to the community of researchers and PhD students at the Division of Transport and Roads and K2 for filling my working days with laughter, encouragement and unwavering support. I thank my friends and colleagues at Lund Railway Group. Special thanks go out as well to Ruben Kuipers for his enlightening intellectual challenges, including the concept that downloading emails and having a fully charged laptop can increase a laptop's weight. I appreciate our discussions, especially those on using "inside voices". To Michelle Ochsner, the department's top spy, thank you for keeping me updated. Despite our contrasting personalities, I'm pleasantly surprised at how well we get along. I appreciate Daria Ivina's optimism and her ability to find beauty in everything, as well as the positive energy she brings to the office. Grace Mukunzi's calm wisdom in balancing home and work life is admirable, and I miss our jogging sessions, even though we never managed to use the same route twice. Lastly, a big thank you to Frida Carlvik for her unwavering support in the railway group, and I wish her all the best in her PhD studies. Thanks to the visiting PhD student, Joe Wright, for all the discussions we have had. I extend my sincerest gratitude to the PhD students in the Railway group at KTH: Emil Jansson, Elin Hellblom, Ingrid Johansson and Niloofar Minbashi. The group meetings and manuscript review sessions with all of you have been enjoyable and enriching.

I extend my sincerest thanks to Francesco Corman for inviting me to visit his lab at ETH Zürich. I had a great time with his students and colleagues. Special appreciation goes out to REMESH, Andreas Norrman and Ren Jun for making my research visits to China and Thailand possible. I will never forget these enjoyable journeys. Thank you to Jirapan Liangrokapart and Waressara Weerawat for hosting me at Mahidol University and organising so many great seminars and events during my research secondments. Many thanks to my colleagues and the students at

Mahidol University for the wonderful moments shared in Thailand. Thank you to Ouyang Wu for hosting me at the Wuhan University of Technology, and I extend my appreciation to the students for the memorable experiences in Wuhan. Special thanks to Ping Huang, Zhongcan Li and Jie Luo at Southwest Jiaotong University for the warm welcome and engaging discussions on train delay prediction studies. I hope we can author some papers together in the future. Thanks to Pengling Wang and your research group at Tongji University for sharing insights into the Chinese railway system and the wonderful time we had in Shanghai. I am particularly thankful to Zhongcan Li for the insightful tour of Tsinghua University. Thank you to Shukai Li for the warm welcome at Beijing Jiaotong University and to your research group for providing me with happy memories of Beijing.

During my time in Sweden, I had the pleasure of getting in touch with saints in Copenhagen and Sweden. I enjoyed all our gatherings and love feasts. A special thanks to Sister Gloria for always remembering me and being ready to lend a helping hand since my arrival in Lund. I feel relieved to have had you by me throughout my PhD experience. I am grateful to Brother Wei and Sister Xiaokang for graciously sharing the word of the Lord, and I particularly enjoyed their delicious dumplings and bao. I also extend my thanks to my Malaysian friends in Sweden for the delightful celebrations and delicious Malaysian food.

I'd also like to express my deepest thanks to my family. I´m pleased that I always know that everything at home is in good care, and I am always free to continue my journey. Thanks to my parents for everything you do and have done for me; there are no words for that! Thanks to Catherine and Aik Xin and Louis and Bethel for providing wise encouragement, peace and joy throughout these years. Kah Swee and Kah Ling, thanks for challenging and supporting me whenever it was needed. It is so impressive that both of you always find the right words! Thanks to Alen for always sharing his interesting thoughts.

1 Introduction

The demand for railway services has experienced a steady increase over the years, with many railways operating at a high degree of capacity utilization. In Sweden, the total number of train trips grew by 15.7% between 2017 and 2019, rising from 229 million to 265 million (Trafikanalys, 2018a, 2020). From 2000 to 2019, the traffic supply, measured by passenger kilometres, surged by 77%, going from 8 243 to 14 617 million passenger kilometres (Trafikanalys, 2020). Following two years of pandemic-related restrictions, there was a strong recovery in train travel, although a full rebound has not yet been achieved. In 2022, there were 244 million railway trips, marking a 35% increase compared to 2021, but this represents an 8% decrease compared to 2019. The total passenger kilometres for railway transport reached approximately 12.8 billion, reflecting a 60% increase from 2021 and a 12% decrease from 2019 (Trafikanalys, 2023a). Besides high-capacity utilisation, the railway system in Sweden faces challenges due to considerable heterogeneity in terms of train types, such as high-speed trains, commuter trains and freight trains, which all have distinct stopping patterns, acceleration capabilities and maximum speeds. The combination of high-capacity utilisation and heterogeneous traffic makes the railway network in Sweden susceptible to delays and disturbances.

Despite the steady growth in Swedish railway traffic and the relatively less steep growth in railway construction, the Swedish Transport Administration (Trafikverket) has the long-term goal of making train services more reliable. This commitment became evident when Trafikverket set a goal for achieving a punctuality level of 95%, where punctuality is defined as a maximum delay of five minutes at the final station (JBS, 2021). From 2013 to 2017, train punctuality hovered around 90%, declining slightly to 88% in 2018 (Trafikanalys, 2018b). After a year of decreased reliability, there was an improvement in 2019, with just over 91% of scheduled trains arriving at their final stations less than five minutes behind schedule. This positive trend continued in 2020, reaching a punctuality level of 94% (Trafikanalys, 2022b). This increase in punctuality could potentially be due to the reduced railway travel during the COVID-19 pandemic. Following a significant drop in demand, railway transport experienced traffic recovery in 2021, returning to normal volumes of train traffic due to the easing of restrictions. Consequently, train punctuality declined to 90% in 2021, a level comparable to the period from 2013 to 2017, and further decreased to 87.2% in 2022, marking the lowest level in the last 10 years (Trafikanalys, 2023b). This suggests that a lack in the number of train services needed to meet daily travel demand is one of the causes of train delays, and the situation is expected to worsen due to the significant increase in train traffic volumes expected in the future.

1.1 Operational challenges arising from train delays

Train delays, defined as the deviation of actual train events from scheduled train events, are one of the most important performance indicators of railway operations (Goverde, 2005). Due to the involvement of many interacting processes dependent on human behaviour, technical devices and the environment, trains are inevitably subjected to delays due to a variety of causes in practice (Kecman, 2014). Infrastructure occupation refers to the time interval that block sections or interlocked routes are exclusively allocated to a train and therefore blocked for other trains (Goverde & Hansen, 2013). Train delays increase the risk of a train exceeding its track occupation time and hindering other trains. When scheduled dwell or run times are exceeded due to hindrance by other trains, the minimum safety distance and headways between two consecutive trains, especially at critical route nodes, becomes insufficient, thus causing route conflicts (Nie & Hansen, 2005). In busy and heavily utilized networks, when the buffer times incorporated into the timetable are not sufficient to absorb the train delays, a slight deviation from the schedule by a single train can easily propagate the delays to other trains that run over the same infrastructure and affect the rolling stock or crew connections, thereby disturbing the entire train network (Carey & Kwieciński, 1994).

1.2 Passenger dissatisfaction caused by train delays

The punctuality of train services is one of the major determinants of passenger satisfaction (Brons & Rietveld, 2008; van Loon et al., 2011). Despite the fact that railway transport offers an effective solution for commuting and intercity travel, its overall attractiveness is constrained by the perceived consequences of delays and unreliability in comparison to the travel times of other modes, such as private vehicles. The disutility arises from the risk of arriving late, which goes beyond the mere costs of the actual delays since it encompasses anxiety costs (uncertainty in itself), decision costs (the adjustments made in the face of uncertainty) and contingency planning costs (the additional time needed to execute the contingency plan) (Börjesson & Eliasson, 2011). This aligns with the observations of Parbo et al. (2016), who highlight that passengers dislike the stress induced by uncertain onward connections. Besides the average network performance, passengers are also concerned with the variability and uncertainty of the travel times during their journeys (Barron et al., 2013).

The response of passengers towards train delays depends on factors such as user expertise, car availability, perception of service recovery time, opinions on passenger information services, available transport services, time constraints, and the moment and place at which communication about the disruption occurs (Adele et al., 2019). Train delays can inconvenience passengers, especially those with tight connections or time-sensitive commitments. In consequence, unscheduled delays may create customer dissatisfaction, especially if they result in missed connections, late arrivals, missed appointments or other inconveniences, all of which contribute to a decline in the public’s perception of railway transportation. Improving the train service’s reliability and minimising delays can lead directly to increased passenger satisfaction, potentially encouraging a modal shift away from private motorised transport in favour of railway travel and fostering higher levels of ridership retention (Monsuur et al., 2021) given that the quality of public transport also has a direct effect on car ownership (Holmgren, 2020).

1.3 Economic ramifications of train delays

Train delays can have economic implications. The unreliability of train arrival times often prompts passengers and companies to adopt conservative measures, such as departing early or maintaining a safety stock of goods, thus incurring additional costs. The “cost” of train delays therefore signifies the foregone benefit that could have been realised if all trains were punctual. On the other hand, train operators take into account the inherent uncertainties when planning timetables and thus incorporate buffer times to facilitate the recovery of the railway system from delays. However, the increase in buffer time correlates with higher capacity consumption, leading to a reduction in maximum potential revenue (Jovanović et al., 2017).

The costs associated with train delays for train operators can be broadly categorised into crew, locomotives, fuel, railcars and lading. The applicability of these cost categories depends on factors such as how trains are operated and where the delay is experienced (Lovett et al., 2015). For instance, due to restrictions on the working hours of train crews, train delays may require hiring more crew rather than relying on existing crews to work overtime. To maintain the minimum safety headway between trains, substantial delays or operating close to theoretical capacity might force the cancellation of some trains, resulting in lost revenue. Furthermore, train operators are subjected to regulatory requirements regarding punctuality, so service reliability punctuality and persistent delays may lead to non-compliance, thus drawing regulatory attention and risking potential penalties. In Sweden, train operators causing delays exceeding five minutes face fines imposed by Trafikverket (Andersson, 2014). The EU Rail Passengers' Rights Regulation grants passengers the right to compensation for delays unless specified otherwise in the Terms and Conditions of Travel. This dual dynamic of compensating passengers and penalising infrastructure managers for delays translates into augmented costs or revenue losses for train operators.

1.4 Train delay-induced environmental impacts

In Europe, the transport sector contributes significantly to a larger share of total carbon dioxide (CO2) emissions, the major source of greenhouse gases (GHG) and, consequently, climate change. For example, approximately 40% of Sweden's total CO2 emissions in 2017 originated from the transportation sector (Blayac & Stéphan, 2021). In this context, railway transportation emerges as a more environmentally friendly mode that offers a potential solution to alleviate congestion on highways and urban roads. It also promotes sustainable development and reduces pollution levels locally and globally. This is emphasised by the fact that every passenger shifting from road to train can save 0.105 kilogrammes of CO2 per passenger per kilometre travelled (Weber et al., 2022). Moreover, the use of electric trains is found to be 22% less costly to society in terms of climate impact compared to diesel train operations (Givoni et al., 2009). The advantage of electric trains is anticipated to grow as more railway networks are electrified, and the CO2 content of electricity is expected to decrease over the coming decade. Presently, around 75% of the total track length in Sweden is electrified (Trafikanalys, 2022a). Despite being considered a green mode of transport, this environmental benefit only really comes into effect when passenger volumes are sufficiently high (Weber et al., 2022). Since punctuality is ranked first among those aspects of customer satisfaction that potentially affect passengers' modal choices (Guirao et al., 2016), minimising train delays becomes crucial to encourage a modal shift from private vehicles to railways.

1.5 Toward an intelligent transport system

One solution to address train delays due to congested railway networks is to expand railway capacity and provide a dedicated infrastructure for all trains. However, constructing new railways is capital-intensive and problematic in terms of time and environmental constraints (Kecman, 2014). An alternative strategy involves real-time adjustments during operations to minimise train delays and restore feasible railway system operations (i.e. preventing deadlocks in which trains are stuck due to occupation and reservation constraints). In current practice, train dispatchers make decisions based on their experience with train delays and recovery, which is a complex process fraught with uncertainty and random factors (Wang et al., 2021).

Furthermore, they are not supported by decision-support tools since the software needed to regulate railway facilities has not kept pace with the necessary technological developments.

In recent years, with the advancement of communications, sensing technologies, and increased computer storage capacities, a vast amount of train operation data has been collected and stored, thus making it possible to address train delay problems from a data-driven perspective. These phenomena consequently give rise to the emerging use of intelligent transport systems (ITS) in the railway field, where predicting train delays has become a dominant area of railway research. The concept of establishing an efficient ITS environment involves predicting the expected train traffic conditions at a future time given a continuous flow of information about the way train traffic conditions evolve over time (Vlahogianni et al., 2004). This is important since train delay estimation is an important input for solving many problems related to train traffic management. For train dispatchers, predictive models provide essential information on expected train trajectories and conflicts, thus enabling necessary adjustments to be made for resolving conflicts between train paths and enhancing the operational efficiency and capacity of the railway transportation system. For passengers, predicted train delays make it easier to identify valid alternatives when planning the best connections for their trips.

The thesis presented here focuses on predictive models for train delays with the overarching aim of increasing our knowledge in the realm of data-driven train delay prediction. The work delves into various components of the prediction models, including the exploration of input variables as well as the methods and data quality that can impact model performance. By providing a thorough understanding of data-driven train delay prediction models, this thesis helps to identify the potential modelling solutions that contribute to the enhancement of existing models. This endeavour ensures the continuous evolution of decision-support tools for both passengers and train operators, thus improving accuracy and efficiency when predicting and managing train delays.

1.6 Thesis outline

Following *Chapter 1 Introduction*, the rest of the thesis is structured as follows: *Chapter 2 Background* introduces the main concepts of railway operation as well as the definitions and terminology needed for understanding the remainder of the thesis. Moreover, a review of the important contributions made in the scientific literature related to train delays, train delay prediction and the use case of the prediction models is presented. The chapter ends by outlining two identified research gaps. *Chapter 3 Aim* introduces the aim of the thesis, the research questions (RQs) and a detailed description of the research gaps to be filled by each question. *Chapter 4 Research design* begins with a presentation of the six papers and how they are oriented and built on each other. This is followed by an overview of the connections between the research questions and the papers. *Chapter 5 Method* first presents an overview of the main ideas necessary for connecting the methods used in the six papers. Consequently, a detailed description is provided of each method used in the thesis. *Chapter 6 Data preparation and case study* provides an overview of the data handling process in the thesis; this encompasses the list of datasets and data preprocessing processes, and continues with a description of the study areas and scope of data used in the thesis. *Chapter 7 Summary of Papers* provides summaries of the six papers in the thesis. *Chapter 8 Answering the research questions* connects the findings from the papers back to the research questions. *Chapter 9 Synthesis* synthesises the findings from the research questions to answer the overarching aim of the thesis, while *Chapter 10 Conclusion* recapitulates the main findings, contributions and recommendations for practice that can be derived from the thesis. This chapter ends by discussing the limitations of the thesis and providing a clear direction for further research.

2 Background

This chapter presents the frame of reference used for the thesis. The chapter begins with a review of extant research on the different aspects regarding railway timetables. Research on train delays is also reviewed to understand the different types of train delays and their causes. The chapter provides an overview of the key methods adopted to model train delays, with a specific focus on data-driven methods given the growing availability of data in the railway field. The practical use cases of the train delay prediction model are discussed in terms of model integration at different stages of train operation management for the convenience of passengers. The research gaps are then summarized from the discussed literature.

2.1 Railway timetable

Railway traffic typically operates according to a timetable outlining process times, including the running and dwell times for each train as well as the headway that needs to be respected. To maximise the probability of adhering to the process time, a buffer time is incorporated to ensure that the timetable remains robust even when encountering variations in internal and environmental conditions. The term “buffer time” specifically refers to time intervals between two events, such as between the arrival of transferring passengers at the connecting train and the departure of this train, or between the release of a crossing by one train and the blocking of the crossing by the next train. In other words, the scheduled process time in the timetable comprises the sum of the process time (running time, dwell time and headway) and a buffer time (running time margin, dwell time buffer and headway margin) that must be maintained to ensure a realisable process time. It is worth noting that the calculation of process times is either based on calculations in normal conditions using mean values of parameters (e.g. running times, dwell time, headway) or by simple experience-based norms depending on local characteristics (Goverde, 2007).

2.1.1 Scheduled running time

Running time refers to the duration of time between the train's departure and its complete halt at the arrival station (Kecman, 2014). In addition to the diversity of traffic and variations in train priority, the physical attributes of trains, including factors such as length, tonnage, and power along with infrastructure-related elements like track grade, track curvature and siding lengths, play a crucial role in influencing the overall runtime of trains, thereby contributing to increased variability and uncertainty in their runtime (Dingler et al., 2009). To increase the reliability of the railway system by reducing the propagation of delays resulting from the interdependencies between trains, Vromans et al. (2006) suggested mitigating these interdependencies by reducing the differences in running time per track section, thus creating more homogeneous timetables. This is crucial because train delays can occur when fast trains are caught behind slower ones, especially in train systems with services travelling at different speeds. Recognising that running slower trains can impact the performance of faster trains, Huisman and Boucherie (2001) proposed reducing the service for interregional and intercity passengers to ensure better service for regional passengers.

Scheduled running times contain a certain amount of running time margin, as outlined by Goverde (2007), for three main reasons. Firstly, the margin is essential to accommodate slower train speeds in adverse conditions that were not accounted for during the running time calculations. Additionally, the running time margin serves as recovery time to reduce departure delays, with drivers operating at maximal speed to reach the next station. Furthermore, the extra running time can be utilised for energy-efficient trains running through a coasting regime, thereby enabling on-time arrivals with minimal energy consumption and contributing to cost-effective operations. Compared to a uniform distribution of running time supplements, Scheepmaker et al. (2020) concluded that the optimal distribution of running time supplements led to higher energy savings; for example, the shorter the distance between two stops, the larger the relative number of running time supplements must be. Parbo et al. (2016) suggested placing smaller-than-average runtime supplements at the earliest and last part of the line since delays occurring at the beginning are often relatively small whereas delays at the last stations do not affect as many stations. However, considering that Trafikverket only measures punctuality at the final station, this might lead to a larger runtime margin in the last part of the trains' journeys to prevent them from being late at the final station, thus neglecting how late they arrived at intermediate stations.

2.1.2 Scheduled dwell time

Dwell time is defined as the difference between the train departure time and train arrival time (Li et al., 2014). Train dwell time is one of the most unpredictable components of railway operations due to the varying volumes of alighting and boarding passengers (Li et al., 2016; Olsson & Haugland, 2004). Certain events during the exchange of passengers can influence dwell times in various ways; for example, late-arriving passengers can extend dwell times by door holding, railway staff might offer extensive passenger service by holding the train for passengers to board, and clustered boarding can extend the time required to complete the alighting and boarding process (Kuipers et al., 2021). However, Harris et al. (2013) emphasised that passengers are not the sole cause of delays, pointing out that other factors such as inappropriate rolling stock, defective rolling stock (particularly regarding doors), slow dispatch, and crew changes can extend dwell time and lead to the propagation of delays.

Scheduled dwell time may include a dwell time buffer to partially or completely absorb arrival delays and/or accommodate seasonal and daily variations in boarding and alighting times, thereby minimising the propagation of train delays. It is important to note that an insufficient dwell time buffer might be a source of delay, especially during peak hours with heavy traffic, while an excessive dwell time buffer results in longer travel times and high station capacity utilization. Peterson (2012) found that the dwell times along the Swedish Southern Mainline were often underestimated. Andersson (2014) recommended allocating time supplements shortly after stations where the most frequent delays occurred, thus increasing the train's chances of recovering from delays. Pedersen et al. (2018) found that long dwell times coincide with times and areas of high-capacity utilisation and thus propose the use of longer trains instead of more trains during rush hour.

2.1.3 Scheduled headway

Headway refers to the time interval between two consecutive trains. Headway is a significant factor contributing to increased delay sensitivity in railway networks since exceeding headway time can lead to congestion interference on subsequent trains. Headway time is crucial not only on mutual routes where the leading train is slower than the following train but also on conflicting routes such as crossing routes at junctions and opposing routes on single-track corridors (Goverde, 2005). Hofman et al. (2006) found that strategies resulting in a substantial increase in headways led to the largest improvement in punctuality. In Sweden, minimum headways vary from two to seven minutes depending on the location, but they are most commonly between three and five minutes (Trafikverket, 2017).

Headway margin is incorporated into scheduled headway times between train pairs to ensure that small delays do not immediately result in secondary delays. The shorter the scheduled headway between trains, the greater the expected knock-on delays and hence, the longer the expected trip times of the following trains (Yuan & Hansen, 2008). Another important purpose of the margin is to provide the train dispatchers with extra time and flexibility to reschedule trains during a disrupted situation. Goverde and Hansen (2013) stated that as long as a train stays within its allocated envelope, maintaining its scheduled process time and having sufficient buffer time over minimum headway times between train paths, it will not hinder other trains, thus avoiding secondary delays due to variations in process times.

2.1.4 Capacity utilisation

Capacity can be defined as the maximum number of trains that can operate on a given railway infrastructure during a specific time interval given the operational conditions (Peterson, 2012). According to Bešinovic (2017), capacity utilization in railways depends on the infrastructure (railway layout, track speed limits, signalling system), rolling stock (braking and acceleration capacity, maximum speed, train composition), and traffic-related factors (train type, use of tracks, mix of services). Despite buffer times being added to the process times to ensure some degree of robustness in the timetables and the punctuality of the train operations, large buffer times may result in longer travel times for passengers, higher operating costs and less efficient infrastructure capacity utilization. Therefore, railway infrastructure managers, timetable designers and train dispatchers need to collaborate to maintain the expected operating costs and revenues and ensure an efficient capacity utilization, all while considering that trade-offs exist regarding the customers' desired level of service.

Homogenous traffic refers to traffic composed of similar types of trains running at the same speed, stopping at the same stations, and maintaining equal headways, such as in a metro system (Palmqvist, 2019). From a purely operational perspective, homogeneous traffic allows for more efficient use of capacity than heterogeneous traffic (UIC, 2004). However, the traffic in Sweden is highly heterogeneous; it comprises various train types, including high-speed trains, regional trains, freight trains and long-distance trains. In heterogeneous traffic, trains utilise the infrastructure unevenly over time, with significant differences in average speed. The major drawback of traffic with varying speeds is that the faster trains risk catching up to the slower trains, which may be required to stand aside for unscheduled overtakings. As a result, buffer times for a heterogeneous railway system are longer than those for homogenous traffic. This is because buffer time is needed for slower trains to make extra stops due to overtaking and for faster trains to have extra running time due to speed homogenization (Andersson, 2014). To reduce heterogeneity, Vromans et al. (2006) proposed several options, including slowing down fast long-distance trains, speeding up short-distance trains, incorporating overtaking, letting short-distance trains have shorter lines or equalising the number of stops, although these options may not always be practically relevant.

2.2 Train delays

Train delays occur due to the variability in process times (i.e. running times and dwell times), capacity and synchronisation processes, as well as dependencies on the availability of infrastructure, rolling stock, and crew (Kecman, 2014). These delays can be categorised into primary and secondary delays. A primary delay is an extension of the scheduled process time due to disruptions within the process (Goverde, 2005). Given the nuanced variability in train running times between stations and dwell times at stations, which are influenced by numerous factors within the railway system and external sources, this thesis groups delay sources into internal and external sources in Section 2.2.1. Secondary or knock-on delays refer to increases in process time caused by hindrances from other trains (Corman, 2010). The detailed causes of these delays are outlined in Section 2.2.2.

2.2.1 Cause of primary delays

Internal sources of train delays include sources directly related to train operations, including factors related to train, crew, and infrastructure. Since trains constitute the fundamental component of the railway system, consideration of the technical malfunctions and impacts of various train types is important. Different train types possess distinct specifications, which leads to diverse infrastructure capacity requirements, operational frequencies, and priorities. For example, the dwell time for longer trains is extended due to the spatial distribution of alighting and boarding passengers, requiring train operators to spend extra time ensuring that no passenger boards prior to departure (Li et al., 2016). Ceder and Hassold (2015) pointed out that crew-scheduling issues significantly impact the reliability of the rail network. The minimum crew on-duty requirement must always be satisfied, even if it involves the considerable expense of stopping a train and transporting a replacement crew to complete the trip (Nabian et al., 2019). Infrastructure-related variables include variables related to the structures, buildings, and equipment designed to support railway systems. As infrastructures are shared facilities among trains, delays can arise due to infrastructure constraints related to either the occupation of infrastructure by another train or infrastructure faults that need to be addressed. Veiseth et al. (2007) reported that infrastructure faults contribute to 30% of the delay hours in Norway. Other sources of delay, such as train engine breakdowns and damage to the overhead catenary system, result in significant delays, but these occurrences are infrequent.

Common external factors contributing to train delays include passengers, weather and maintenance- or roadwork-related factors. Harris et al. (2013) emphasised that the time required for passenger boarding and alighting at stations is a critical element of overall train service performance. Train arrival delays vary with seasonal fluctuations in passenger demand (Laifa et al., 2021), different months of the year (Grandhi et al., 2021) and days of the week (Ceder & Hassold, 2015). For example, trains dwell longer during peak hours to accommodate a higher volume of passengers (Li et al., 2016). Weather-related factors include variations in temperature, wind speed, snow depth and rainfall. Numerous studies indicate that extreme weather has a significant negative impact on train punctuality since severe weather conditions disrupt regular train operations (Brazil et al., 2017; Wang & Zhang, 2019; Zakeri & Olsson, 2017). Maintenance- or roadwork-related factors refer to infrastructure unavailability and the restriction of train movement due to the interference caused by maintenance or roadwork activity. Ivina et al. (2021) discovered a correlation between track work activities and increased train delays.

2.2.2 Cause of secondary delays

Secondary delays are caused by train interactions, as outlined by Daamen et al. (2008), who categorised them into two main classes: hindrance at shared track sections and waiting for a scheduled connection at a station. Hindrance at shared track sections occurs when a delayed train impedes the passage or crossing of other trains, which propagates existing delays as secondary delays. Due to the interdependencies in the timetable and the shared use of infrastructure, extended dwell times cause trains to occupy the platform track and station routes of other scheduled trains, resulting in consecutive delays at busy stations. Therefore, factors hindering trains at shared track sections include route conflicts for an arriving or departing train at a station and headway conflicts for trains following at different speeds, especially when no overtaking track is available.

The second category of secondary delays is associated with the synchronisation of trains at transfer stations to ensure connections. As identified by Higgins and Kozan (1998), three types of connections may result in delays due to late connections: 1) one train is waiting at a station for another to arrive for the transfer of passengers, 2) a predetermined departure order where one train must depart before another at a station (e.g. for the coupling of trains or changing train crews), and 3) turnarounds at terminals, indicating the commencement of a new service using the same physical train after its arrival at the destination. Goverde (2005) emphasises the importance of considering the difference between hard and soft connections in the timetable design process in order to incorporate sufficient buffer time into the connection times. A soft connection allows a connecting train to wait for delayed feeder trains, thus securing the connection, but if the delay is excessive, the connection is cancelled, and the train departs as scheduled. On the other hand, hard connections cannot be cancelled and include situations where train pairs need to be (de-)coupled during coach changes at an intermediate or terminal station or when part of the feeder train's crew must transfer to the connecting train.

2.3 Train delay prediction models

Delay prediction is the process of estimating delay probability based on known data at a specific station, typically measured via arrival (departure) delays. The generated prediction serves as a basis for proactive decision-making by both operators and passengers (Mou et al., 2019). Since establishing a reliable delay prediction system is one of the approaches for addressing current delay issues, extensive research has been dedicated to developing efficient train delay prediction models using various methods. Spanninger et al. (2022) classified these train delay prediction models into event-driven and data-driven models based on their inherent modelling paradigm.

Event-driven approaches explicitly capture dependencies among train events (departure, arrival, and pass-through) by modelling railway operation dynamics, including procedures and restrictions. The event-driven train delay prediction process involves a chain of prediction steps for delays at subsequent stations. Event-driven approaches make certain assumptions, such as a train can arrive at the second station only after the train has departed from the first station, and only one train at a time can occupy a track section or stop at a platform due to capacity constraints. Event-driven approaches are primarily based on either a graph model such as timed event graphs (Kecman & Goverde, 2013), Bayesian networks (Corman & Kecman, 2018), Petri nets (Milinković et al., 2013; Zhuang et al., 2016), Markov chains (Schmidt et al., 2019) and max-plus algebra (Goverde, 2007), or an equation system (Medeossi et al., 2011).

In contrast to event-driven approaches, data-driven approaches do not explicitly model train-event dependency structures nor aim to capture traffic flow dynamics explicitly. These approaches directly predict the delay at target stations without intermediate predictions by mapping the input to the output. Without making assumptions, data-driven approaches utilise explanatory variables to quantify their impact on process times when generating delays at subsequent train stations. This thesis focuses explicitly on data-driven approaches due to the rapidly expanding volume of data in the railway industry and the advancement of computing techniques. Data-driven approaches are increasingly adopted for developing train delay prediction models and conducting in-depth analyses due to their capability to handle large datasets and extract valuable insights from ever-growing train operation databases.

2.3.1 Data-driven train delay prediction approaches

Kecman and Goverde (2015) broadly categorise data-driven train delay prediction models into global and local models based on their prediction horizon in space, whereas Milliet de Faverges et al. (2018) categorise them into long-term and short-term delay prediction models based on their prediction horizon in time. Despite being categorised based on different aspects, the global model and long-term train delay prediction model, as well as the local model and short-term train delay prediction model, are equivalent and have the same functionality. Thus, in this chapter, only the terms global and local models are used to avoid confusion.

The global model is commonly utilised to investigate the impacts of various factors on train delays, particularly at the strategic and tactical levels of railway traffic planning. This model utilises aggregated historical train operation data instead of real-time data to predict delays several days or even months in advance, thus providing train operators with sufficient time to develop train management plans. The key advantage of global regression models lies in the generalisability of their results since the findings derived from the global model can be applied to train lines not included in the training datasets (Chiou et al., 2015; Kecman & Goverde, 2015). Statistical regression models are commonly employed to develop global train delay models. Gorman (2009) indicated that primary congestion-related factors such as train meetings, passing and overtaking have the most significant impact on congestion delays. However, the major drawback of the global models is the masking of time and geographical variation in the relationships between variables. This limitation arises because global models overlook the existence of local variations due to spatiotemporal autocorrelation since they utilise aggregated data and treat all observations independently.

In contrast to the global model, local models use real-time data specific to particular train lines to continuously predict and update predictions in response to the evolution of railway traffic. These models are commonly employed for proactive operational railway traffic management and passenger information(Huang, Wen, Fu, Peng, et al., 2020; Lulli et al., 2018; Taleongpong et al., 2020; Wen et al., 2019). Local models are recognised for providing more accurate predictions of train arrival delays. Moreover, local models consider spatiotemporal characteristics by modelling train delay variations along the train line over multiple stations via the continuous input of real-time train operation data. A study by Bao et al. (2021) demonstrated the importance of features used for train delay prediction changing in time and space as the train moves towards its destination. Advanced methods such as conventional machine learning (Li, Wen, et al., 2020; Taleongpong et al., 2020; Wang & Zhang, 2019), neural network models (Oneto et al., 2017, 2018; Wen et al., 2019) and hybrid models (Huang, Wen, Fu, Lessan, et al., 2020; Huang, Wen, Fu, Peng, et al., 2020; Lulli et al., 2018), which map the complex relationship between input and output, are commonly adopted in order to model local train delay prediction with great accuracy.

2.4 Use cases of train delay prediction models

Based on the planning horizon of railway operation, three levels of railway traffic planning—strategic, tactical and operational—are identified by Dennis et al. (2005). At each of these planning levels, an accurate train delay prediction model plays a well-established role. In the following subsections, each level of railway traffic planning is introduced, and the use cases of train delay prediction models at each level are detailed.

2.4.1 Decision-support tool for investment planning

Strategic planning focuses on the strategic design of the scheduled transport network and the long-term capacity management of resources or traffic means to meet future traffic demand, where resources refer to the sufficient infrastructure capacity, rolling stock and train personnel to accommodate the expected traffic flows (Goverde, 2005).

To aid management in selecting the most effective investment plan within the available budget, long-term train delay prediction models play a crucial role in increasing our understanding of the relationship between railway transportation efficiency and infrastructure investments or facility improvements. For example, Marković et al. (2015) employed support vector regression (SVR) models to analyse the effects of different infrastructure projects on delays, thus assisting planners in comparing various investment alternatives. To assist in the development of long-term investment and reinvestment plans in the Swedish railway transportation system, Jiang et al. (2019) proposed a hybrid model comprising random forest regression (RF) and logistic regression to estimate the consequences of planned investment/reinvestment measures on train punctuality. To investigate the impact of infrastructure on train delays, Shi et al. (2021) took into account both infrastructure information, such as station distance and track allocation, and train operation data when predicting train arrival delays using eXtreme Gradient Boosting (XGBoost) and a Bayesian optimisation (BO) algorithm.

2.4.2 Decision-support tool for timetable planning

Tactical planning is concerned with the capacity allocation of resources to transport services for the intermediate planning horizon. Typical tactical planning problems include the allocation of infrastructure time-distance slots, rolling stock, and crews to trains (Goverde, 2005). Creating feasible and realisable timetables is one of the main tasks at the tactical planning level, and it requires that a trade-off be made between margin time and capacity utilization. Increased allocation of margin times is beneficial for delay recovery and allows a delayed train to catch up and return to its planned timetable, but unused margin time may be wasted since the train cannot depart earlier than its scheduled time; this reduces the capacity available for accommodating more train services and is thus not economical (Andersson et al., 2011). On the other hand, reduced margin times decrease scheduled travel times (running, dwell and transfer times), leading to shorter total journey times and seamless connections. The reduced train occupancy rate enables the railway system to accommodate more train services to meet passenger demand, but this results in a sensitive system where even a slight delay can easily propagate throughout the entire train network, thus increasing the likelihood of train conflicts or even deadlocks (Corman, 2010).

Understanding the impact of various factors on train arrival delays is a prerequisite for effective timetable planning. Long-term train delay prediction models play a significant role in revealing the heterogeneous impact of various factors on train events. This allows train operators to adjust their plans accordingly when anticipating disruptions several days or even months in advance, and it gives train operators adequate time to develop train management plans. Palmqvist et al. (2020) utilised linear regression to quantify the impact of weather, timetable, operational and infrastructure variables on passenger train punctuality in Sweden. By quantifying these impacts on process times, long-term train delay prediction models offer valuable insights into the timetable's structure and dependencies, thus providing comprehensive theoretical support for train operators to explain punctuality variation and help with train schedule planning or adjustment. Additionally, these models contribute to optimising timetable design by detecting potential instabilities and identifying where and how margin time should be inserted for enhanced robustness. Yaghini et al. (2012) presented a highly accurate neural network-based passenger train delay prediction model for Iranian railways that facilitated the scheduling of a suitable timetable by the train operator.

2.4.3 Decision-support tool for real-time train management

Operational planning involves performing rescheduling during operations in response to unforeseen events, disruptive incidents, or accidents. Current real-time train management relies heavily on intense manual control, with dispatching tasks being highly experienced-oriented. The train dispatcher plays an important role in solving any potential conflicts and recovering from disruptive events in a short time by making the necessary adjustments to the personnel and rolling stock plan to comply with the actual train operation situation, all while preventing resources such as track, crew and other shared infrastructure from becoming unavailable. According to D'Ariano (2008), local dispatchers assume expected arrival delays are equal to the current upstream delays since they lack information about possible recovery times except from experience. However, this assumption neglects the fact that some trains recover from delays by running in the maximum performance regime and exploiting the running time supplements in the timetable. However, they can also get more delayed due to possible time lost from route conflicts. Furthermore, the intense communication involved in the dispatching process, with individuals distributed across time and space (dispatchers, train drivers, locomotive engineers and workers on the track), increases cognitive workload and time pressure, thus leading dispatchers to focus on implementing practical but temporary solutions.

Despite the fact that existing technology delivers precise real-time information (such as the exact positioning of trains), a short-term train delay prediction model remains vital as an intelligent decision-support system for train dispatchers since it can provide them with information about future train delays based on the current train operation situation. The model enables the train dispatcher to anticipate conflicts accurately, thus giving them enough time to use conflict-resolution methods to prevent the propagation of delays. As a consequence, the dispatcher has an increased possibility of optimally planning and efficiently controlling the traffic. Based on actual data from the Dutch railway Rotterdam Central-to-Dordrecht section, Wen et al. (2019) employed the long short-term memory (LSTM) model to predict train arrival delays, thus offering decision support to dispatchers. To provide theoretical support to dispatchers in developing rescheduling strategies and adjusting the station work plan when primary delays occur, Li, Huang, et al. (2020) proposed a hybrid model consisting of XGBoost to predict the number of affected trains, which is then input into SVR to predict the total time of affected trains.

2.4.4 Reliable passenger information system

In addition to the three levels of train traffic planning, a train delay prediction model can also be used for passenger information systems. However, the quality of predictions directly impacts the passengers' overall experience, so the discrepancies between expected and actual times need to be mitigated. Providing reliable passenger information involves communicating schedule changes, missed connections, or exceptional delays, thus enabling passengers to set realistic expectations for travel times. Timely dissemination of adverse travel conditions allows passengers to make informed decisions, such as whether to continue to wait for the original train, change tickets, obtain a refund or opt for alternative transportation modes. Additionally, accurate predictions help alleviate any passenger stress associated with uncertain train journeys.

Offering passengers accurate information about the additional waiting time required in the event of delays can mitigate passenger dissatisfaction, even if the passenger still has to wait due to delays in train operation. Although passenger information may have little impact on the actual travel time needed, it indirectly influences the perception of service quality by affecting whether a passenger will accept the waiting time. Informed passengers have control over the time they are willing to spend by choosing to wait for the delayed train or selecting another valid alternative. With the help of a reliable passenger information system, passengers are mentally more prepared for the delays encountered and are likely to be more understanding and tolerant of the delays. To provide accurate predictions of train delays in the Deutsche Bahn passenger services network, Nair et al. (2019) proposed a large-scale hybrid model comprising two statistical models and one simulation-based model. To provide passengers with accurate train delay predictions, Nabian et al. (2019) proposed a bi-level random forest approach in which the classification forest at the primary level predicts whether a train delay will increase, decrease or remain unchanged, and the secondary level regression estimates the actual delay (in minutes) given the predicted delay category at the primary level.

2.5 Research gaps

In the background section, we reviewed existing work on predicting train delays and revealed two gaps that this thesis aims to fill. The first regards the insufficient understanding of existing train delay prediction models, and the second relates to the potential for innovation in developing train delay prediction models with practical applications. We will now briefly examine these two research gaps.

2.5.1 Insufficient model understanding

Current practice in train delay prediction research is dominated by the design and development of increasingly sophisticated train delay prediction algorithms, possibly due to the rapid advancement of computer technology and the maturation of data-driven methods. The overall objective of such research is to enhance the accuracy of the predictive model. A rich variety of algorithms have been proposed and creatively applied to solve the train delay prediction problem, with much weight given to conventional machine learning, neural networks and hybrid methods. This trend is inevitable, especially when most classical analytic and statistical approaches have demonstrated themselves to be “weak” or inadequate when modelling complex train traffic conditions and dealing with extensive datasets. Despite these valuable contributions, scholarly comprehension of the development of train delay prediction models remains limited and fragmented since existing work has not been comprehensively reviewed and synthesised from a technical perspective.

Building data-driven train delay prediction models is a complicated process involving various aspects such as problem definition, input/output representations, data preprocessing, modelling techniques, model evaluation (Vlahogianni et al., 2004) and especially the interdisciplinary nature of predictive analytics, which requires expertise from the computer science and railway domains. Furthermore, there are a variety of key technical problems related to each aspect that must be addressed, such as determining how data from multiple sources should be processed before being incorporated into the predictive model. Since effective solutions for different data-driven train delay prediction modelling problems are scattered throughout existing studies, a comprehensive synthesis of proven techniques from existing research not only serves as a valuable reference for railway researchers facing similar modelling challenges, but it also presents up-to-date techniques for the corresponding modelling components. Notably, the lack of insight from existing studies hinders the speed at which data-driven approaches are adopted, thus limiting progress in train operation management practices.

The ability of a prediction algorithm to capture the dynamics of train traffic and provide accurate delay predictions is heavily dependent on the input variables incorporated into the models (Chollet, 2021). Recent developments in data collection and storage technology, coupled with the widespread use of powerful computers, offer researchers unprecedented access to multiple datasets and the ability to compile a comprehensive list of input variables encompassing operations, infrastructure, maintenance, and external variables. However, only limited studies have been found that examine the impacts of these factors on train arrival delays. Despite many existing studies showing the improved performance associated with the use of real-time variables such as train delays at current and preceding stations, the specific contributions of these real-time variables to train delays remain unexplored. Understanding the input variables is a very important issue, especially in data-driven algorithms, because it directly impacts the quality of the predictions generated and helps avoid the loss of valuable information that could enhance model performance.

2.5.2 Innovation for practical applications

Several diverse modelling efforts have been undertaken to tackle the problem of train delay prediction. However, the predominant focus on achieving high accuracy poses a major challenge in designing modelling schemes with practical applications. Most of the existing literature, such as Barbour et al. (2018) and Li, Huang, et al. (2020), concentrates on one-station-ahead prediction, thus restricting train operations to decisions about the current phase. Apart from continually updating new information about the latest train events, the prediction algorithm should be capable of computing predictions for all train events across the prediction horizon in a single execution (Luo et al., 2022). This is essential because a predictive model serves to continuously provide train operators and passengers with information about the expected traffic conditions and offer a complete picture of the train operation conditions at each station along the designated routes. Nevertheless, existing literature often overlooks the practical application of prediction models.

According to Vlahogianni et al. (2004), the concept of utilising the train operation variables as a function of time and space is theoretically valid considering the fact that temporal and spatial information from previous locations capture the true dynamics of train traffic, thus offering valuable information on how train traffic evolves. The incorporation of information from a location near the location of interest has been found to improve prediction accuracy; conversely, information from a location far away may lead to less reliable predictions. For example, Shi et al. (2021) found that train delays at the current station are the most important variables in enhancing the model's performance. This could probably be because the importance of each variable for train delay predictions changes over time and space as the trains move towards the destination (Bao et al., 2021). Despite acknowledging the contribution of time-lagged information and spatial inputs based on topology, efforts to maximise the benefit of this data characteristic for model refinement remain limited.

The deployment of train delay prediction models is hampered by the absence of a comprehensive assessment test that ensures their effectiveness in real-world scenarios. While existing literature emphasises model accuracy, it should not be the sole determinant in selecting the appropriate prediction methods (Vlahogianni et al., 2014). Andersson (2014) highlighted that many models in academic studies are evaluated in controlled academic settings, which raises concerns about their applicability in real-world industrial projects and signifies that the prediction models may not be usable for industrial projects and will remain purely academic contributions. Despite the fact that the railway transportation field still lacks an extensive evaluation procedure, the performance evaluation methods for prediction models have been extensively studied in various fields, particularly in forecasting research (Davydenko & Fildes, 2013). Surprisingly, few studies have analysed how evaluation practices from other fields can enhance current train delay prediction evaluation practices.

3 Aim

The overarching aim of the thesis is to increase understanding of data-driven train delay prediction models. This involves a detailed exploration of the various components within the prediction models (e.g. the input variables, data quality and performance evaluation methods) in order to identify modelling solutions that contribute to the improvement of existing data-driven train delay prediction models. In accordance with the research gaps identified in the preceding section, four specific research questions have been formulated. The following research questions, which are complementary to one another, are taken into account in this thesis to fulfil the aim of the research.

3.1 RQ1 What factors need to be taken into account when building a train delay prediction model?

This research question led us to scrutinise existing data-driven train delay prediction models to increase our understanding of the various aspects that must be considered when employing data-driven approaches for predicting train delays. This is achievable through a comprehensive review of the existing literature. The research question serves the purpose of identifying technical solutions to key challenges in the development of data-driven prediction models. Through the systematic synthesis of existing research, we seek to answer this research question to help uncover unresolved prediction issues and weaknesses in current data-driven models. This, in turn, provides insights into the methodologies employed, thus enabling us to emphasise areas that remain unexplored or comparatively under-researched in the existing literature and, consequently, provide some promising directions for future train delay prediction research.

3.2 RQ2 How are selected input variables improving the performance of the train delay prediction model?

This research question seeks to ascertain useful input variables by quantifying their impact on train delays, especially when we have accessed multiple datasets to generate a comprehensive list of variables from diverse sources. Quantifying this impact reveals the relationship between the input variables and train delays, facilitating the selection of relevant variables for enhanced model performance. Thus, factor analysis is performed to solve the research question, which involves an interpretable and statistically sound model with robust mathematical foundations to gain insights into the underlying relationship between dependent and independent variables.

3.3 RQ3 What approaches can enhance the train delay prediction model?

This is the primary question to answer in this thesis since it leads to developing a robust train delay prediction model by building upon insights gained from the first two research questions. To be precise, the prediction model will incorporate useful input variables identified from RQ2, while findings from RQ1 will serve as the foundation for the model development process. The question extends to offering innovative technical solutions to the methodological and modelling challenges that are encountered in the current development of prediction algorithms. In other words, this research question requires the formulation of a new methodology to improve the performance of the existing model.

3.4 RQ4 How can train delay prediction models be evaluated?

This research question seeks to construct an evaluation framework for a thorough assessment of train delay prediction models in order to ensure their practical effectiveness. A comprehensive review of existing performance metrics and evaluation methods was undertaken to discern their strengths and weaknesses in assessing prediction model performance. This research question is designed to unveil the efficiency of the model, thus serving as a crucial tool to bridge the gap between the development and real-world implementation of train delay prediction models.

4 Research design

To achieve this work's overarching aim, the thesis includes a total of 6 papers, four of which are journal papers (Papers 1–3 and Paper 6) and two of which are conference papers (Papers 4 and 5), as detailed in Section 4.1. Section 4.2 provides detailed descriptions of the connections between the research conducted in this thesis. All six papers contribute to some extent to the exploration of each of the four research questions described in Section 3 An overview of the connections between papers and research questions is shown in Table 1, with detailed descriptions provided in Section 4.3.

4.1 List of included papers

Paper 1

Tiong, K.Y., Ma, Z. & Palmqvist, C.W. (2023). A review of data-driven approaches to predict train delays. *Transportation Research Part C: Emerging Technologies. 148*, Article104027, https://doi.org/10.1016/j.trc.2023.104027

Paper 2

Tiong, K.Y., Palmqvist, C.W., Olsson, N.O.E. (2022). The effects of train passes on dwell time delays in Sweden. *Applied Sciences*, *12*(6), Article 2775. https://doi.org/10.3390/app12062775

Paper 3

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.* https://doi.org/10.1016/j.tra.2023.103751

Paper 4

Tiong, K.Y., Ma, Z. & Palmqvist, C.W. (2022, September 18-October 12). *Real-time train arrival time prediction at multiple stations and arbitrary times*. 25th IEEE International Conference on Intelligent Transportation Systems, Macau, China. https://doi.org/10.1109/ITSC55140.2022.9922299

Paper 5

Tiong, K.Y., Ma, Z. & Palmqvist, C.W. (2022, September 21-23). *Prediction of train arrival times along the Swedish Southern Mainline*. 18th International Conference on Railway Engineering Design & Operation, Comprail2022, Valencia, Spain. https://doi.org/10.2495/CR220121

Paper 6

Tiong, K.Y., Ma, Z. & Palmqvist, C.W. (2024). AP-GRIP evaluation framework for data-driven train delay prediction models: Systematic literature review. *European Transport Research Review*. Submitted.

4.1.1 Author's contribution to the included papers

Paper 1

Kah Yong Tiong: conceptualisation, methodology, data curation, formal analysis, visualization, writing–original draft, writing–review & editing. Zhenliang Ma: conceptualisation, methodology, formal analysis, supervision, writing–review & editing. Carl-William Palmqvist: conceptualisation, methodology, supervision, writing–review & editing.

Paper 2

Kah Yong Tiong: methodology, formal analysis, writing–original draft preparation, writing–review & editing, visualisation. Carl-William Palmqvist: methodology, formal analysis, writing–original draft preparation, writing – review & editing, supervision, funding acquisition. Nils Olof Emanuel Olsson : supervision, writing–review & editing.

Paper 3

Kah Yong Tiong: conceptualisation, methodology, formal analysis, investigation, data curation, writing–original draft, writing–review & editing, visualization. Zhenliang Ma: conceptualization, methodology, validation, writing–original draft, writing–review & editing, supervision. Carl-William Palmqvist: resources, data curation, writing–review & editing, supervision, project administration, funding acquisition.

Paper 4

Kah Yong Tiong: conceptualisation, methodology, formal analysis, investigation, data curation, writing–original draft, writing–review & editing, visualization. Zhenliang Ma: conceptualisation, methodology, validation, writing–original draft, writing–review & editing, supervision. Carl-William Palmqvist: resources, data curation, writing–review & editing, supervision, project administration, funding acquisition.

Paper 5

Kah Yong Tiong: conceptualisation, methodology, formal analysis, investigation, data curation, writing–original draft, writing–review & editing, visualization. Zhenliang Ma: conceptualisation, methodology, validation, writing–original draft, writing–review & editing, supervision. Carl-William Palmqvist: resources, data curation, writing–review & editing, supervision, project administration, funding acquisition.

Paper 6

Kah Yong Tiong: conceptualisation, methodology, formal analysis, writing - original draft, writing–review & editing, and visualization. Zhenliang Ma: conceptualization, methodology, validation, writing–original Draft, writing–review & editing, and supervision. Carl-William Palmqvist: resources, data curation, writing–review & editing, supervision, project administration, and funding acquisition.

4.1.2 Related publications not included in the thesis

Tiong, K.Y., Palmqvist, C.W., Olsson, N.O.E., Winslott Hiselius, L. (2021, November 3-7) *Train passes and dwell time delays.* 9th International Conference on Railway Operations Modelling and Analysis (ICROMA) – RailBeijing, Beijing, China.

Tiong, K.Y., Palmqvist, C.W. (2022, November 14-17). *Quantitative methods for train delay propagation research*. 9th Transport Research Arena TRA, Lisbon, Portugal. https://doi.org/10.1016/j.trpro.2023.11.326

Tiong, K.Y., Ma, Z. & Palmqvist, C.W. (2023, July 17-21). Real-time High-Speed Train Delay Prediction using Seemingly Unrelated Regression Models. World Conference on Transport Research (WCTR), Montreal, Canada.

Tiong, K.Y., Ma, Z., Palmqvist, C.W. (2023, August 4-6) *Evaluation framework for train delays prediction models.* 7th International Conference on Transportation Information and Safety (ICTIS 2023).

4.2 Relationship between included papers

This section provides a concise overview of the six papers by clearly describing their structures, interconnections, and progressive relationships. It delineates how each paper is strategically positioned to lead into the next, thus providing a coherent and cumulative narrative. The visual representation of these connections is depicted in Figure 1.

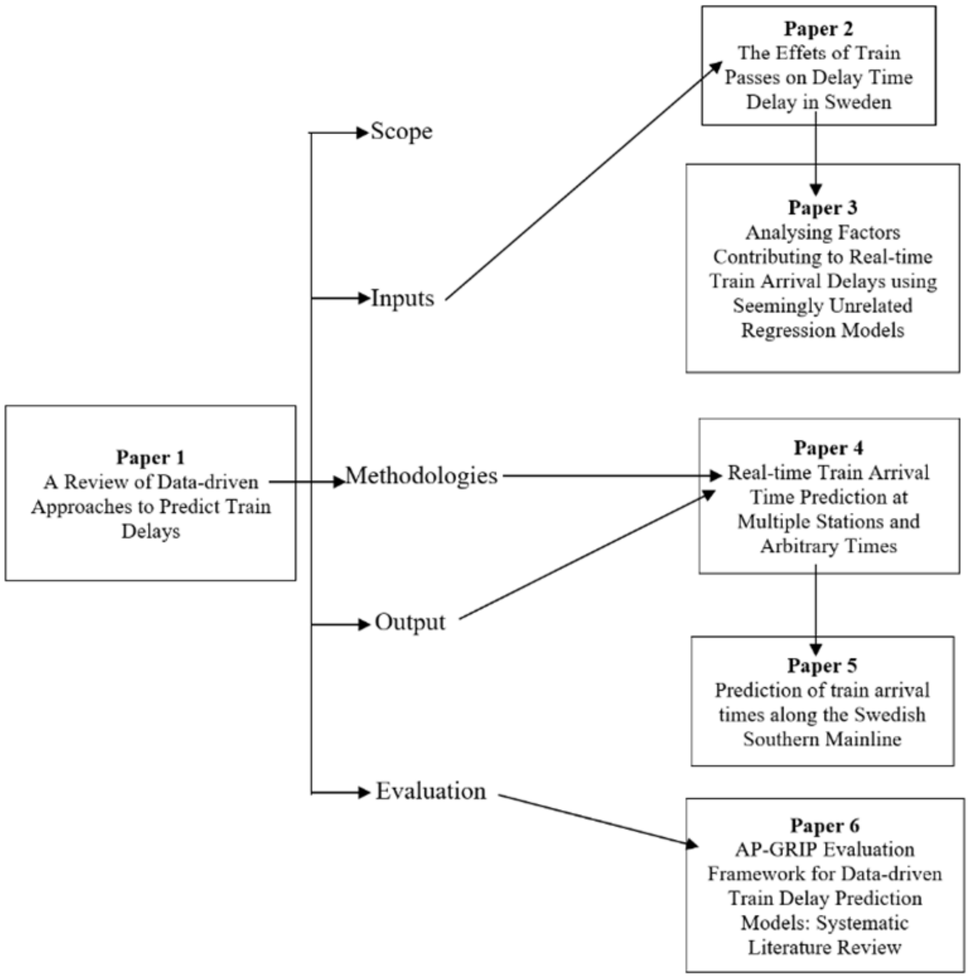


Figure 1: Connection between the six papers

Paper 1 primarily focuses on RQ1, but it also broadly covers all the other research questions in order to lay the groundwork for the subsequent papers in the thesis. Paper 1 delves into data-driven train delay prediction from the model development perspective, including problem definition, input/output representations, modelling techniques and model validation. In contrast, Papers 2 to 6 have a narrower scope, each dealing with specific research questions. Paper 2 and Paper 3 are primarily dedicated to RQ2 and investigate different factors influencing train delays. Paper 1 highlights that train operation data is the primary source of data for the train delay prediction model and underscores the importance of investigating how various factors contribute to train delays from a spatiotemporal perspective. Building on this foundation, Paper 2 extracts relevant factors for analysis directly from the train operation data. Additionally, Paper 2 sheds light on the drawbacks associated with the design of a global model, emphasising the need for a more nuanced and context-specific approach. Building on these insights, Paper 3 strategically places considerable emphasis on local model design by conducting further regression analysis to assess the heterogeneous impacts of spatiotemporal factors on train delays.

Factors identified as significant in Paper 3 are integrated into the prediction models developed in Papers 4 and 5. Paper 4 and Paper 5 specifically address RQ3 by focusing on the development of real-time prediction models with practical applications. Given that Paper 1 identifies the need to generate continuous predictions of train arrival times for multiple downstream stations at arbitrary prediction times, Paper 4 concentrates on constructing a dynamic multi-output train delay prediction model. Paper 5 expands on Paper 4 by incorporating error adjustment strategies into the multi-output real-time prediction framework to enhance the performance of the train delay prediction model. Paper 6 focuses on RQ4 regarding the comprehensive evaluation of the train delay prediction model. Paper 6 responds to the deficiency noted in Paper 1, which emphasised an over-reliance on model accuracy. Paper 6 conducts a thorough exploration of the various evaluation aspects and dimensions of model performance to ensure the effectiveness of the train delay prediction model in practical deployment, drawing insights from the output generated by the real-time prediction model in Paper 5.

4.3 Relationship between the research questions and included papers

Table 1 provides a concise summary of how the papers connect to the research questions. The subsequent subsection delves into a detailed discussion of these connections, clarifying how each paper contributes to addressing the different research questions.

Table 1: Connection between the research questions and the included papers

Research Question	Paper					
RQ1: What factors need to be taken into account when building a train delay prediction model?	1					6
RQ2: How are the input variables improving the performance of the train delay prediction model?	1	2	3			
RQ3: What approaches can enhance the train delay prediction model?	1	2	3	4	5	
RQ4: How can train delay prediction models be evaluated?	1					6

RQ1 What factors need to be taken into account when building a train delay prediction model?

To aid practitioners in developing data-driven prediction models for railway practices, Paper 1 introduces a domain-specific prediction framework for railway operational delays that aligns with the generic data science framework (e.g. CRISP-DM). The paper reviews existing studies from a technical standpoint and disaggregates the train delay prediction process into six components: scope determination, model inputs, data quality, methodologies, model outputs and evaluation techniques. For each component, the important problems and techniques reported are synthesized, and research gaps are discussed. It is important to note that RQ1 serves as the backbone for the thesis, and certain components within the proposed framework are detailed in subsequent questions.

RQ2 How are the input variables improving the performance of the train delay prediction model?

To explore valuable input variables for train delay prediction models, Papers 2 and Paper 3 employ interpretable models such as logistic regression and the seemingly unrelated regression (SUR) model to quantify the impact of various explanatory variables on train delays in the railway system. The previous train delay prediction literature reviewed in Paper 1 helps in the selection of explanatory variables in Paper 2 and Paper 3. Paper 2 investigates the effects of different train passes on dwell time delays, while Paper 3 expands on Paper 2 by examining a comprehensive set of factors influencing train arrival delays, encompassing train operations, network, weather, maintenance and calendar variables. Unlike Paper 2, which only uses historical train operation factors, Paper 3 considers both real-time and historical factors, thus addressing the temporal and spatial perspectives of explanatory variables crucial for capturing traffic dynamics and providing insights into the evolution of future train traffic.

RQ3 What approaches can enhance the train delay prediction model?

To build a prediction model that has strong predictive capabilities and is useful in practical applications, Paper 3 introduces a local model based on the location-conditioned concept to enhance real-time prediction models, especially since Paper 2 found that the global model provides low prediction accuracy. In line with the findings of Paper 1 on the importance of developing dynamic multi-output train delay prediction models at arbitrary prediction times for practical applications, Paper 4 and Paper 5 propose multi-output machine learning models and test various frameworks to determine the most efficient framework for the multi-output prediction task, such as direct multi-output regression and chained multi-output regression in Paper 4 and SUR in Paper 5. Consequently, Paper 5 introduces two error adjustment strategies—one-step before prediction error correction and upstream prediction error correction—to optimise real-time observed information utilisation and enhance the prediction model.

RQ4 How can train delay prediction models be evaluated?

To enhance decision-making processes related to the selection of train delay prediction models, Paper 6 proposes a standardised evaluation framework by incorporating insights from the existing literature. Building on the three main evaluation aspects outlined in Paper 1—prediction accuracy, generalizability and interpretability—Paper 6 introduces three additional evaluation aspects—model precision, robustness and practicality—to advance the effectiveness of train delay prediction models in practical applications. Paper 6 provides an in-depth examination of these six aspects that encompasses current practices, metrics, important considerations and limitations. Furthermore, Paper 6 reveals the importance of evaluating prediction models from multiple dimensions—overall, spatial, temporal and train-specific dimensions—to provide a comprehensive understanding of their performance in various circumstances.

5 Methods

The methodological approaches in Papers 1–6 vary due to differences in the scope and type of each study. Paper 1 comprehensively covers numerous aspects of train model development, while Paper 6 focuses on numerous model evaluation aspects of train delay prediction models. Conversely, Papers 2–5 concentrate on more in-depth analyses of specific aspects of train delay prediction models. These research approaches can be contrasted as extensive versus intensive approaches (Swanborn, 2010). The extensive approaches used in Papers 1 and 6 relied on information from a large number of instances gathered through a comprehensive literature review. In contrast, the intensive approaches used in Papers 2–5 required more in-depth information obtained through case studies in order to emphasise context-specific features.

This thesis utilised both quantitative and qualitative methods. Paper 1 and Paper 6 are literature review studies involving content analysis that combine both qualitative and quantitative methods. The quantitative component entailed reviewing numerous relevant articles to identify common issues in the areas of interest. In contrast, the qualitative component entailed a thorough review of numerous relevant articles to develop an in-depth understanding of a given topic. In Paper 1, the quantitative part offers a general overview of the various approaches used in developing train delay prediction models, while in Paper 6, the quantitative part provides a general overview of the various key components in evaluating the train delay prediction models. The qualitative approaches used in Papers 1 and 6 were valuable for gaining insights into train delay prediction models and understanding the different components constituting the proposed framework.

Papers 2–5 are quantitative studies utilising data-driven approaches, including regression and predictive analyses, to conduct case studies. In Papers 2 and 3, statistical regression methods such as logistic regression and SUR were used to investigate factors influencing train arrival delays. The main distinction between them is the experimental designs used. While Paper 2 used a global model based on aggregated data for train stations in Sweden to ensure the generalisability of the results, Paper 3 employed a local model covering a specific stretch of train line across multiple stations to investigate the spatiotemporal effects of train operation variables on train arrival delays. Papers 4 and 5 emphasise prediction model accuracy for practical applications. To generate real-time predictions in response to real-time railway traffic conditions, the setting with conditional multi-output regression models based on train location was adopted in Papers 4 and 5. Various machine learning algorithms were tested in Paper 4, and error adjustment strategies were introduced in Paper 5 to enhance model performance. An overview of the various methods used in the six papers is shown in Table 2.

Table 2: Overview of the different method used

Paper	Category of Method	Method	**Details of method**
1	Quantitative and qualitative methods	Content Analysis	Conventional and summative content analysis
2	Quantitative method	Regression analysis + Model evaluation	Logistic regression
3	Quantitative method	Regression analysis + Model evaluation	SUR
4	Quantitative method	Predictive Analysis +Model evaluation	LightGBM, GBR, and RFR
5	Quantitative method	Predictive Analysis + Model evaluation	Liner regression, SUR
6	Quantitative and qualitative methods	Content Analysis	Conventional and summative content analysis

SUR = seemingly unrelated regression, LightGBM = Light Gradient Boosting Machine, GBR = gradient boosting regression, and RFR =random forest regression

5.1 Content Analysis

Papers 1 and 6 utilise content analysis to extract key findings from previous literature. Hsieh and Shannon (2005) identified three distinct content analysis approaches: conventional, directed, and summative. The conventional content analysis derives relevant information directly from the text data, the directed approach benefits from existing theory or prior research about a phenomenon to validate or conceptually extend a theoretical framework or theory, and the summative approach involves counting and comparing content followed by interpreting the underlying context. Papers 1 and 6 used both the directed and summative content analysis approaches; conventional content analysis was not relevant since neither paper acquired direct information from study participants through interviews or open-ended questions.

Based on the well-established model development framework from existing literature, the directed content analysis approach enabled Paper 1 to extend the available framework into a domain-specific (train delay prediction) framework by analysing the patterns and approaches commonly adopted in existing literature. By employing the directed approach based on the selected literature, Paper 1 addressed the common issues surrounding data-driven train delay prediction problems and identified promising solutions for each aspect of the framework. The findings from the directed content analysis offer supporting evidence for the proposed novel evaluation framework in Paper 6 by gathering existing evaluation practices for train delay prediction models. Additionally, a comparison of previous literature provides valuable indications of key research gaps since dissimilarities are signs of possible knowledge gaps, i.e. under-researched areas where more in-depth understanding is needed.

The summative content analysis in Papers 1 and 6 identifies and quantifies certain terms in the literature so as to understand the nuances of the concept. In both papers, summative analysis is employed to discover the underlying meanings of the words. For instance, summative analysis explores how the term "generalisability" is perceived by forecasting professionals in order to create a working definition. Summative content analysis also makes it possible to interpret the context associated with the use of the word and provide insights into how a word is normally used. For example, Paper 6 explores "generalisability" in terms of definitions, measuring metrics and important considerations to gain deeper understanding. The quantitative component of summative analysis involves reviewing a large number of relevant articles to identify common issues in the areas of interest. In Paper 1, the quantitative part provides a general overview of the various components of the train delay prediction model and outlines important problems and techniques in the model development framework. Similarly, the quantitative assessment in Paper 6 summarises the key aspects of the evaluation framework and indicates the popular evaluation metrics.

5.2 Regression Analysis

In regression analysis, traditional statistical regressions are commonly used to infer causal relationships between independent and dependent variables. This is because more complex models, although possessing stronger predictive capabilities and being designed to adapt to complicated data relationships, tend to generate difficult-to-understand outcomes. Regression analyses are valuable since they are able to identify statistically significant relationships between dependent and independent variables, assess the strength of these relationships and facilitate prediction (Sarstedt & Mooi, 2014). In the thesis, two types of regression analyses were conducted: logistic regression in Paper 2 and SUR3 in Paper 3. Detailed explanations are provided in subsequent subsections.

Apart from the different models chosen for the regression analyses, Papers 2 and 3 also diverge in terms of study design. Paper 2 adopts the global model design by aggregating data and treating all observations independently, while neglecting the heterogeneity across different routes to ensure the generality of results. Conversely, Paper 3 uses a local model design that captures the heterogeneity of trip-level factors affecting train delays. More specifically, Paper 3 conducts trip-level train delay analysis conditioned on train locations to understand the heterogeneous impact of the same factors in space and time.

5.2.1 Logistic regression

Paper 2 utilised a logistic regression model to investigate the tendency of dwell delays to occur due to train passes. Logistic regression is a statistical modelling technique that estimates the probability of a dichotomous outcome event being associated with a set of explanatory variables. It is the log odds taking the form of Equation 1. It is worth noting that odds represent the ratio of the probability of an event occurring relative to the probability of the event not occurring. The rationale for employing logistic regression lies in its ability to easily compute the odds ratio, a metric for comparing the effects of two different events, using the model coefficients, as demonstrated in Equation 2.

$$\begin{aligned}\log \text{odds} &= \log\left(\frac{\pi_i}{1-\pi_i}\right)\\ &= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\end{aligned} \tag{2}$$

where π_i represents the probability of event *i* occurring

$$\begin{aligned}Odds\ ratio &= \frac{Odds_i}{Odds_j}\\ &= \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}}\end{aligned} \tag{3}$$

where π_i and π_j represent the probabilities of events *i* and *j* occurring, respectively, where $i \neq j$

To assess which dispatching strategy is most effective in preventing delays, Paper 2 quantifies their impacts in terms of the odds ratio of experiencing dwell time delays due to different types of train passes, thus enabling a direct comparison of the efficiency between two different actions. Unlike regression models that predict the amount of delays, logistic regression makes it possible to compare the effectiveness of alternative train passes in reducing the likelihood of delays occurring. It is also important to note that odds and odds ratios, like percentages, do not have units.

5.2.2 Seemingly unrelated regression

To identify factors influencing real-time train delays at the trip level, Paper 3 performs regression analysis using the SUR model. SUR is a system of linear equations tailored to address contemporaneous correlation across the trip-level disturbance terms during model estimation. However, when using traditional ordinary least squares regression (OLS), the general equation for the train delays of a train line with I stations can be expressed in Equation 3 and expanded into Equation 4. It is important to note that the error terms are implicitly assumed to be contemporaneously uncorrelated if the parameters of each equation in Equation 4 are estimated separately via OLS.

$$y_i = X_i\beta_i + \varepsilon_i, i = 1, \cdots, I \tag{3}$$

$$\begin{gathered} y_1 = X_1\beta_1 + \varepsilon_1, i = 1, \cdots, I \\ y_2 = X_2\beta_2 + \varepsilon_2, i = 1, \cdots, I \\ \vdots \\ y_{I+1} = X_I\beta_I + \varepsilon_I, i = 1, \cdots, I \end{gathered} \tag{4}$$

where y_i represents the train arrival delay at station i with $n=1,\cdots,I$ denoting the number of individual observations. X_i denotes the set of explanatory variables, β_i represents the regression coefficient, and ε_i is the error term.

Paper 3 adopts SUR to solve the real-time train arrival delay factor analysis problem while accounting for the limitations of OLS. Despite each equation in the system having its own vectorβ_i parameter, as in Equation 4, SUR assumed that a contemporaneous correlation across the error terms exists instead of disregarding any potential correlation between the equations entirely. In practice, it is impractical to collect all possible feature variables when predicting train arrival delays. The common unexplained model variabilities at the trip level are expected to be consistent for the same train travelling along the same line towards the same destination (e.g. characteristics of rolling stock, tracks and driver styles). Therefore, the error terms in Equation 4 may share the same patterns, indicating the existence of contemporaneous disturbance term correlations across the equations for the factor analysis problem in Paper 3. In other words, although these equations are seemingly unrelated, they actually have unobserved properties in common and should be treated as a system for parameter estimation. Another advantage of utilising SUR is that it allows each regression model to benefit from the information contained in other regression equations by accounting for correlated errors among different equations caused by unobserved train operation characteristics.

5.3 Predictive Analysis

Predictive analytics, a branch of statistics, focuses on extracting insights from data to forecast trends and behavioural patterns, typically with an emphasis on future events (Ongsulee, 2018). Data-driven approaches, particularly machine learning methods, have been widely used for predicting train delays at target stations by mapping the input to the output without explicitly modelling train-event dependency structures to capture traffic flow dynamics (Spanninger et al., 2022).

Considering the practical application of the train delay prediction model in providing actionable information, Papers 4 and 5 propose multi-output framework designs focusing on the train delay predictions for multiple downstream stations at arbitrary times. Additionally, to formulate real-time predictions in response to real-time railway traffic conditions, Papers 4 and 5 introduce the conditional multi-output regression setting by incorporating the corresponding sets of training datasets based on the train's current location into the prediction model. The detailed discussions on the location-conditioned concept and multi-output frameworks can be found in Sections 8.3.1 and 8.3.2, respectively.

In Paper 4, location-conditioned concepts and multi-output frameworks are adopted together with tree-based algorithms, including LightGBM, GBR and RFR, due to their proven superior performance (Barbour et al., 2018; Kecman & Goverde, 2015), whereas linear regression serves as the baseline model for comparison purposes. Paper 5 adopts linear regression with a direct multi-output framework for comparability with SUR (discussed in detail in Section 5.2.2 Seemingly unrelated regression). To improve model accuracy, Paper 5 introduces two error adjustment strategies, which are elaborated upon in Section 8.3.3. The following subsection provides detailed information on tree-based models and linear regression.

5.3.1 Light gradient boosting machine

LightGBM is a tree-based machine learning algorithm that sequentially trains ensembles of decision trees by fitting negative loss gradients. The algorithm employs gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). GOSS focuses on training data instances with larger gradients for greater information gain and eliminates a large number of data instances with small gradients. On the other hand, EFB is a near-lossless feature selection method in LightGBM that reduces the number of features by grouping sparse, mutually exclusive features. Implementing LightGBM with GOSS and EFB accelerates training time while maintaining high accuracy levels.

5.3.2 Gradient boosting regression

GBR is an ensemble algorithm that sequentially adds several weak learner models based on the performance of the prior iteration's composite, culminating in a complex final model. During each iteration, a weak learner is built, and its training involves computing a gradient which represents the partial derivative of the loss function. This gradient facilitates adjustments to the model parameters by GBR, thereby reducing the error in the next round of training and gradually strengthening the model with each iteration. Ultimately, GBR learns by aggregating a weighted sum of all the weak learners.

5.3.3 Random forest regression

RFR is an ensemble of decision trees generally trained using the bagging method. The bagging method begins with the selection of the number of weak learners, denoted as N_c. Subsequently, bootstrap samples, N_c datasets, are generated through repetitive and independent sampling. In RFR, the selection splits are computed via a random subset rather than searching for the very best feature when splitting a node. This approach enhances tree diversity, trading a higher bias for a lower variance, thus yielding a better overall model. Once all the individual regression trees are trained, the final prediction is derived by averaging the predictions from all the regression trees.

5.3.4 Linear regression

Linear regression models are used to establish a relationship between one dependent variable (response) and two or more independent variables by fitting a linear equation to the data, as represented in Equation 5:

$$y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots \beta_p X_{i,p} + \varepsilon_i, \qquad i = 1, \cdots, n \tag{5}$$

where y is the dependent variable, X is the independent variable and ε is the disturbance or error term, considering the data set $\{y_i, x_{i,1}, \cdots, x_{i,p}\}_{i=1}^{n}$.

The term "linear" in multiple linear regression denotes the linearity of the model in the parameters, β_0, $\beta_1, \beta_2, \dots$, β_p , with each parameter multiplied by an independent or x variable, and the regression function being a sum of these "parameter times x variable" terms. Furthermore, ε is assumed to have a normal distribution with a mean of 0 and a constant variance of σ^2. This assumption implies that the variability of the response for fixed values of the independent variables is consistent regardless of the magnitude of the responses. Linear regression also assumes that the errors of the response variables are uncorrelated with each other.

5.4 Model evaluation

Both regression models and predictive analysis models undergo performance evaluation tests to assess their effectiveness in capturing relationships within the data and making accurate predictions. In Paper 2, the model's performance is evaluated using the coefficient of determination (R^2) and the AUC-ROC curve (Area Under the Curve of the Receiver Operating Characteristic). Meanwhile, Papers 3 and 4 assess model performance utilizing R2 and root-mean-square error (RMSE), while Papers 5 evaluates model accuracy using mean absolute error (MAE) and RMSE.

5.4.1 Area Under the Curve of the Receiver Operating Characteristic

A ROC curve is a probability curve depicting the true positive rate (TPR) against the false positive rate (FPR) that illustrates the performance of a classifier model like logistic regression. The TPR and FPR are calculated using Equations 6 and 7:

$$\text{TPR} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{FPR} = \frac{FP}{TP+FN} \tag{7}$$

where TP represents the number of true positives, and FN is the number of false negatives. TPR reflects the classifier's ability to identify all positive samples, while FPR indicates the ratio of negative instances incorrectly classified as positive. The AUC summarises the ROC curve by measuring a classifier’s capacity to distinguish between classes. A perfect classifier achieves an AUC-ROC of 1, while a purely random classifier is 0.5.

5.4.2 Coefficient of determination

R^2 is widely employed in predictive modelling of train delays to assess the goodness of fit of a model or the proportion of variance in the dependent variable explained by the independent variable (Pineda-Jaramillo et al., 2023). The model fits better if R^2 is close to 1.

$$R^2 = 1 - \frac{\sum_{k=1}^{N}(\hat{y}_k - y_k)^2}{\sum_{k=1}^{N}(y_k - \bar{y})} \tag{8}$$

where y_k and $\hat{y}_k$ represent the actual and predicted values, respectively; $\bar{y}$ is the mean of the dependent variable, and all variables are recorded in minutes.

5.4.3 Root-mean-square error and mean absolute error

RMSE and MAE are widely used metrics for assessing the performance of train delay prediction models in terms of prediction errors, as shown in Equations 9 and 10. These metrics are scale-dependent measures, meaning their scale depends on the scale of the data.

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|\hat{y}_k - y_k| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{y}_k - y_k)^2} \tag{10}$$

where y_k and $\hat{y}_k$ represent the actual and predicted values, recorded in minutes. The closer RMSE and MAE are to zero, the better the performance of the model.

6 Data preparation and case study

This chapter outlines the data preparation process, focusing particularly on the case study conducted in the thesis. It begins with a comprehensive overview of the datasets, detailing the variables derived from them. Subsequently, the chapter delineates the data preprocessing procedures undertaken to ensure data quality. The chapter ends by defining the data scope established for the case study.

6.1 Datasets

The thesis utilizes five distinct datasets, as indicated in Table 3. In the subsequent sections, the details of these datasets will be presented.

Table 3: Overview of data used

Paper	Datasets	Input variables	Output variables
1	Reviewed Literature	N/A	N/A
2	Train Operation Data	Train passes	Dwell time delays
3	Train Operation Data	Scheduled and actual dwell time, Scheduled and actual running time, train arrival delays at upstream stations, arrival delays of three trains before, historical mean for arrival delays of the train (hour, weekday, month), direction of travel, time of day, days of week	Arrival delays
	Weather data	Temperature, precipitation, snow depth, wind speed	
	Trackwork plan	Trackwork Period, trackwork restriction	
4,5	Train Operation Data	Actual dwell time, actual departure headway, scheduled running time, scheduled dwell time, scheduled headway, arrival delay, departure delay	Running times
6	Reviewed Literature	N/A	N/A

6.1.1 Reviewed literature

The data for Papers 1 and 6 was extracted from literature identified through the systematic literature review conducted in March 2023 for Paper 1 and January 2024 for Paper 6. The search was performed on the Web of Science and Scopus databases for English academic journals and conference papers, targeting English academic journals and conference papers with no restrictions on publication years. The process began with a literature search using keywords related to railway transport, including "train" and "rail*”, specifically in article titles to prevent confusion with unrelated terms like "training" and "train validation". This was followed by terms emphasising train delay prediction such as “delay," "forecasting" and "prediction," and those emphasising data-driven approaches such as "data-driven," "machine learning," "regression," "artificial intelligence," "deep learning," "neural network" and "statistical regression" in the titles, abstracts and keywords. The results from both databases were combined, and duplicate results were eliminated. Subsequently, a full-text review was conducted, excluding articles without full-text access, purely qualitative studies, papers unrelated to train events from a timetable perspective, and those primarily based on mathematical perspectives, optimisation, simulation, or queuing theory. After this review, forward and backward snowballing strategies were applied to the remaining papers to capture additional relevant literature. Backward snowballing involves checking the reference lists of remaining papers, while forward snowballing focuses on identifying new papers and citing the remaining papers. A thorough assessment of titles, keywords, and abstracts, followed by a full-text review, resulted in the selection of 56 and 65 papers for analysis in Papers 1 and 6, respectively.

6.1.2 Train Operation Data

The train operation data used in this thesis is provided by Trafikverket and is recorded using the signalling system. It includes detailed information on the scheduled and actual arrival and departure times of different trains at each station along the designated train path in Sweden, with one-minute time precision. Additionally, the data contains specific details such as train identification number, train route, train type, and information about the infrastructure (single or double track).

In Papers 2–5, which primarily focus on train delays, the train operation data serves as the main source of data. Given the highly utilised railway network and the presence of heterogeneous traffic in Sweden, there is an increased dependency between trains. Interactions among trains can result in non-scheduled events such as extended running or dwell times due to disturbances, which propagate knock-on delays to other trains in the network. These intricate relationships can be effectively captured through the train operation data. More specifically, train event variables derived from the train operation data have a direct impact on train delays and are crucial predictors in the train delay prediction models to enhance their prediction performance. Thus, train event variables such as previous train delays, scheduled and actual running and dwell times, etc. are used in Papers 2–5.

In addition to the aforementioned train event variables that keep track of both scheduled and actual train movements, Paper 3 explores the effects of other variables on train delays, such as network and calendar variables, which are extracted from the train operation data. These variables help capture variations in train operation conditions at the respective locations and times. Therefore, quantifying their impacts is worthwhile in assisting policymakers to make informed decisions for enhancing the management of current train operations in Sweden.

6.1.3 Weather data

Weather data is retrieved from the open-access weather information provided by the Swedish Meteorological and Hydrological Institute (SMHI). This dataset encompasses historical meteorological observations, including average snow depth, average precipitation, average temperature and maximum wind speed across the entire geographic area of Sweden. These weather variables are recorded at various intervals; snow depth, temperature, and precipitation are recorded once a day at each weather station, while wind speed is recorded hourly.

In Paper 3, weather-related factors such as temperature, precipitation, wind speed and snow depth are included as explanatory variables to explore their impact on train delays and to identify the potential improvements that can be made to the accuracy of real-time train delay prediction models. Previous studies, such as the work done by Oneto et al. (2016), have demonstrated that integrating weather data into models can enhance accuracy by approximately 10%. Similarly, Huang, Wen, Fu, Lessan, et al. (2020) found that neglecting weather-related factors harms prediction model performance, as evidenced by an increase in the loss function upon their removal, even though weather-related factors are considered less important compared to train operation factors. Since weather stations are not located at train stations, the weather data in Paper 6 is linked to train operation data by matching the coordinates of the weather station to the nearest train station on a day-by-day basis.

6.1.4 Trackwork plan

As with the train operation data, the trackwork plan is provided by Trafikverket. It specifies information such as the trackwork identification number and the location of the planned activity, and indicates the starting and ending stations for the trackwork. This dataset also includes details about the timing of the trackwork, which is specified by week number, time, and day pattern. It is worth noting that this data is collected manually on a weekly basis by Trafikverket since there is no automated system for this type of data, thus making the collection process time-consuming.

Ivina et al. (2021) identified a correlation between trackwork activities and increased train delays. In response to this correlation, Paper 3 includes variables into the trackwork plan that capture infrastructure unavailability and restrictions on train movement due to interference between trackwork and train operation. More specifically, the period of trackwork and speed restrictions due to trackwork from the trackwork plan are utilised as explanatory variables in Paper 3. The trackwork plan specifies the location of trackwork by unique signal numbers on the track segment between two assigned stations. To identify the train passages occurring during scheduled trackwork, each unique train passage between two assigned stations is identified from the train operation data, and then the sequence of stations on the train route between two trackwork stations is determined.

6.2 Data preprocessing

Data quality is crucial for quantitative methods, particularly when seeking insights from the data and constructing intelligent algorithms for practical applications. Raw datasets are usually characterised by class imbalance, data heterogeneity, high skewness, privacy, irrelevant and redundant features, continuous data, collinearity among metrics and noise in the data. Various data preprocessing tasks become essential to organise input for prediction algorithms and ensure that models learn from unbiased and well-structured data. The data preprocessing process utilised in the thesis is illustrated in Figure 2, with each stage detailed in the subsequent subsections.

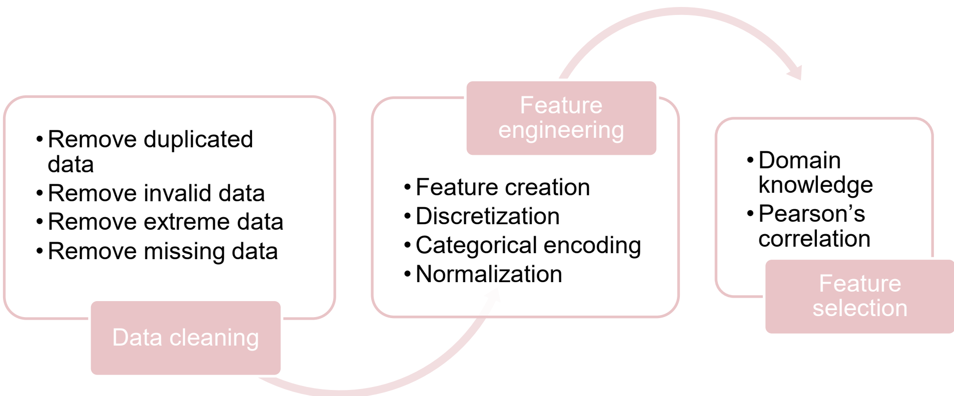


Figure 2: The data preprocessing process

6.2.1 Data cleaning

Inevitable faulty sampling caused by human or computer error results in missing data and noisy data, including duplicated data, invalid data and extreme data. Data cleaning, a process focused on detecting and rectifying (or eliminating) erroneous observations, becomes essential to manage missing values and address inconsistencies. The presence of irrelevant, noisy and unreliable data significantly affects model outcomes and knowledge discovery, thus posing challenges during the training phase. Two common approaches employed to handle noisy and missing data are the sample imputation technique or simply excluding observations with missing values. The exclusion of samples with missing values does not adversely affect the performance of the model if it constitutes a small ratio of the total datasets (Fenner, 2019). With sufficient amounts of data, Papers 2–5 can ignore trips with data completeness issues or data that contain missing observations or illogical values rather than employing imputation methods. Examples of errors encountered in Papers 2–5 include arrival times preceding the departure times of the previous station, duplicate train operation records and incomplete trip records.

6.2.2 Feature Engineering

Feature engineering is the process of extracting features from raw data and transforming them into formats suitable for modelling, thereby maximising the utility of the original dataset. The feature engineering techniques that are adopted in Papers 2–5 are as follows:

6.2.2.1 Feature creation

Feature creation involves developing new variables from existing data to enhance predictive models. New well-designed variables can often capture important information in a dataset more effectively than the original variables. As demonstrated in Papers 2–5, considering train arrival delays at the current station is crucial in train delay prediction. While train operation data typically provides arrival and departure times, focusing on delays at the current station is essential for predicting delays at subsequent stations. This variable can be derived from the train operation data by calculating the difference between actual arrival times and scheduled arrival times at the current station. Creating such a variable addresses a significant source of input for the train delay prediction model and contributes to more precise predictions for train delays at subsequent locations.

6.2.2.2 Discretization

Discretization involves logically grouping data values into bins, and it is applicable to both numerical and categorical data. It is important to note that discretization prevents overfitting at the expense of losing data granularity. Paper 2 employs discretization to address the non-linear relationship between weather variables and train delays. Økland and Olsson (2020) found that weather only negatively influences train punctuality when snowfall, precipitation and low temperatures hit certain thresholds. Following trial-and-error tests, temperatures are categorised into groups (e.g. normal, cold and extremely cold) based on their impact on delays. In Paper 3, the discretization technique replaces numerous values of continuous attributes (calendar, weather and maintenance variables) with a few interval labels, simplifying the original data and giving an easy-to-use, knowledge-level representation of mining results.

6.2.2.3 Categorical Encoding

Categorical encoding is the technique used to encode categorical features into numerical values, making them more comprehensible for algorithms. One hot encoding (OHE) is the categorical encoding technique selected for this thesis, wherein each category in a categorical variable is transformed into a binary feature. Each variable indicates whether a particular category is present (1) or absent (0) in the original variable. In Paper 2 for instance, binary variables are generated using OHE for dwell time delays, where deviations exceeding 0 minutes are considered delays requiring potential rescheduling, while delays of 0 minutes or less are less likely to impact negatively. Another example is that after discretizing the temperature variables into three distinct categories in Paper 3, OHE is applied to encode each category into binary variables—normal temperature, cold and extremely cold. This enables the algorithms to interpret these categories as binary features, making it easier for the prediction algorithm to generate meaningful diagnostic insights.

6.2.2.4 Normalization

Normalisation is the process of changing the values of numeric variables in the dataset to a common scale without distorting differences in the ranges of values. In the context of machine learning algorithms, especially those involving numerical optimisation, normalisation ensures that all input variables are on a standardised scale. This is essential to prevent variables with larger scales from dominating the learning process, which could lead to biassed model performance. Additionally, normalisation enhances the convergence of optimisation algorithms like gradient descent, thus contributing to faster and more consistent convergence. Normalisation improves the generalisation and interpretability of machine learning models by providing a standardised basis for comparing the contributions of different variables. Papers 2–5 adopted the common form of normalisation, standardisation, by centring the data around zero and scaling it to have a standard deviation of 1.

6.2.3 Feature selection

Feature selection involves the identification and retention of relevant features while discarding those that are redundant or irrelevant. As the saying goes, garbage in, garbage out. The prediction model is capable of learning efficiently only when the training data contains enough relevant features and not too many irrelevant ones. In Papers 2–5, initial feature selection relies on domain knowledge and common sense to determine a set of explanatory variables believed to have predictive value for incorporation into the models. Consequently, Pearson's correlation is then employed to identify variables with statistically significant correlations to the target variable, which is train arrival delay. This not only ensures the inclusion of influential input variables, but it also serves to detect multicollinearity; given the correlation that exists among the supposedly independent variables, it tends to hinder the ability of the model to generate an accurate prediction. In cases where there is high correlation between variables, the more significant variable is selected to avoid potential confounding effects and enhance the robustness of the analysis. This approach aligns with the overarching goal of optimising the model's performance and accuracy in train delay prediction.

6.3 Study areas and scope of data

Section 6.3 addresses the crucial considerations of study areas and the scope of data for each paper in the thesis. The cautious selection of study areas and the definition of data scope are vital components for conducting a comprehensive analysis of the subject under investigation since they provide a nuanced understanding of the subject and its complexities. In qualitative studies, the criteria used for selecting study scopes play a crucial role in shaping the depth and generalisability of the analysis, particularly in the context of data-driven train delay prediction models. Likewise, in quantitative studies, the identification of study areas and the scope of data are essential for case studies. This careful selection helps provide an accurate illustration of the application of theoretical concepts to practical scenarios, thus empowering researchers to derive insights, formulate interpretations and generate theories or hypotheses.

6.3.1 The global perspective in Papers 1 and 6

A global perspective was adopted in Papers 1 and 6 for the literature review when the literature search was unrestricted to the years and country of publication. Papers 1 and 6 encompasses all relevant literature to ensure the generalisability of the proposed framework. This approach is crucial since the intention of Paper 1 was to develop a generic, domain-specific framework that guides the design and evaluation of train delay prediction models, thereby accelerating the uptake of data-driven approaches to advanced practices in railways. Similarly, by considering the different model evaluation methods from existing literature, Paper 6 ensures the applicability of the proposed evaluation framework in diverse situations.

6.3.2 Swedish cases in Papers 2–5

Papers 2–5 focus on case studies conducted in Sweden. The railway system in Sweden is characterised by dense and heterogeneous train traffic with varying stopping patterns, acceleration and maximum speeds. More specifically, the mixed traffic in Sweden, which includes freight trains, passenger trains, high-speed trains, long-distance trains and service trains, shares the same infrastructure, causing an increase in dependency on trains. This leads to a sensitive system where it is hard to recover from delays, and the delays can easily propagate to other traffic once a disturbance or disruption occurs.

Figure 3: Capacity use in Sweden 2020 (Trafikverket, 2023)

One of the important train lines in Sweden is the Southern Main Line (SML), which runs southward between Stockholm and the third-largest urban area, Malmö, via Katrineholm, Norrköping and Mjölby. It is the main connection to Denmark and central Europe. The SML is important since it connects freight trains between two large freight yards in Hallsberg and Malmö. Figure 3 illustrates the capacity utilisation in Sweden, revealing that several lines in Sweden experience high capacity utilisation coupled with highly heterogeneous traffic, contributing to an overall delay-sensitive network with increased average train delays. Andersson (2014) highlighted that the most overutilized stretches on the SML are concentrated at the northern end near Stockholm. In parallel with the findings of Trafikanalys (2021), Andersson (2014) also finds that the punctuality of fast long-distance trains in Sweden operating on the SML is lower compared to that of other trains. While the punctuality for all other trains averages around 90%, fast long-distance trains range between 30 and70%. Due to the higher frequency of train services, punctuality during weekdays and rush hours tends to be notably lower than on weekends and during non-rush hours, respectively. This is important since train delay prediction tasks tend to be more challenging for trains with lower punctuality.

6.3.2.1 Stations irrespective of train line

Paper 2 focuses on exploring various types of passes for passenger trains in terms of dwell time delays. The dataset for Paper 2 includes data for passenger trains on double tracks in Sweden in 2014. To account for differences in travel behaviour between weekends and weekdays, observations for Saturday and Sunday were excluded. After filtering out trains with no scheduled or actual pass, the dataset for Paper 2 comprised a total of 403 000 observations. The case studies involved in Papers 2–5 are summarised in Table 4.

Table 4: Case study for the Paper 2–5

Paper	Date	Train type	Study area	Final number of observations
2	2014	Passenger train	All of Sweden	403 000
3	12/2016–12/2020	High-speed train	Copenhagen Central Station (KPH)–Stockholm Central Station (CST) CST–Sundsvall Central Station (SDC)	46 875
4	12/2016–12/2020	High-speed train	Hyllie Station–Norrköpings C Station	2 240–6 414
5	12/2016–12/2020	High-speed train	Hyllie Station–Linköping C Station	6 000

6.3.2.2 Stations along specific train line

Paper 3 examines train arrival delays on two high-speed train (HSR) lines: Copenhagen Central Station (KPH) to Stockholm Central Station (CST) and CST to Sundsvall Central Station (SDC). This represents four distinct train lines covering both directions on the two routes. To ensure consistency, five stations were chosen for each line to analyse the impact of various factors on train delays. For the CST to SDC line, these stations were Hudiksvall, Söderhamn, Gävle, Uppsala, and Arlanda, while Hässleholm, Alvesta, Nässjö, Linköping, and Norrköping were selected for the KPH to SDC line. The analysis focuses exclusively on HSRs, excluding other runs for further analysis. Despite efforts to include additional long-distance train lines, their limited dataset and statistical insignificance led to their exclusion in order to avoid potential biases. The data for Paper 3 covers the period from December 2016 to December 2020, which comprises 46 875 train records after pre-processing.

For Papers 4 and 5, the study focuses on the northbound direction of the SML. To maintain consistency and make it easier to compare results, nine stations were selected for Paper 3, and six of these were retained for Paper 4 within the same study area. The stations selected for Paper 3were Malmö C, Lund C, Hässleholm, Alvesta, Nässjö, Mjölby, Linköpings C, and Norrköping C, while Lund C, Hässleholm, Alvesta, Nässjö, Mjölby and Linköpings C were selected for Paper 4. The strategic selection of train stations along the SML is based on their frequent usage and adequate coverage of the long-distance trains passing through. This approach ensures that there is sufficient data available for a robust analysis, particularly considering the limited number of long-distance trains travelling in the network per day. The study period spans four years, from December 2016 to December 2020. To model train progression downstream closer to the final station, separate datasets were created with distinct sets of explanatory variables reflecting various current railway traffic conditions in time and space. For instance, since the study area has nine stations, eight separate datasets were prepared in Paper 3, whereas five separate datasets were prepared in Paper 4. These datasets vary in size, ranging from 2 240 to 6 414 observations depending on the completeness of the data from the current station to the final station.

7 Summary of Papers

This chapter presents a summary of each appended paper, outlining their respective purposes, methodologies, key results, and implications for practice. For a more in-depth understanding of each paper, please refer to the respective papers in the appendix.

7.1 Paper 1: A Review of Data-driven Approaches to Predict Train Delays

Paper 1 provides a comprehensive review of existing technical studies to present a unified framework for the development of domain-specific data-driven train delay prediction models. The framework disaggregates the complex prediction process into six key aspects. For the aspect of application scope, it categorizes data-driven train delay prediction models into long-term and short-term categories and notes a shift from understanding the impact of explanatory factors on train delays to real-world applications focusing on accuracy and robustness. The study suggests exploring dynamic multiple stations prediction for short-term train delay, which predicts train events at multiple stations at arbitrary times, as opposed to the limited application of the next station's train events prediction.

In terms of model inputs, feature selection commonly relies on domain knowledge and common sense, but alternative approaches like filter methods, wrapper methods and embedded methods offer less subjective options. Addressing common data quality issues such as noise, missing data and class imbalance through data pre-processing is crucial to ensure that train delay prediction models receive accurate information. Despite the challenges posed by imbalanced data, such as poor performance in predicting long delays, a comprehensive solution has yet to be found in train delay prediction studies. In terms of methodologies, there is a noticeable shift from statistical approaches to machine learning models for long-term prediction studies, especially given the ability of machine learning models to capture nonlinear relationships between variables. Short-term prediction emphasises accuracy, with a growing preference for neural networks and hybrid models. However, a complex model is not always the best solution; thus, the logic behind selecting a methodology for modelling train delay prediction models is also worth exploring.

In terms of output, predicting arrival delays is the most direct way of capturing disturbances in scheduled timetables, whereas predicting train process times (e.g. arrival time, departure time, dwell time and running time) and the impact of delays (e.g. the number of affected trains, total delayed time and total time of affected trains) provide important insight into potential timetable adjustment strategy and operational consequences, respectively. To ensure efficiency in the prediction model, attention should be given to predicting multiple-output variables. While accuracy is often emphasised in evaluation techniques, other evaluation aspects such as representational power, which ensures the fitting of the model, explainability, which justifies the use case of the model, and model validity, which assesses the degree to which the modelling framework assumption matches the characteristics of the problem, are equally vital in selecting the appropriate model.

7.2 Paper 2: The Effects of Train Passes on Dwell Time Delays in Sweden

In Paper 2, the primary objective was to increase the understanding of train passes and their influence on dwell time delays for trains in Sweden. This understanding is crucial for optimising timetable planning and train traffic management within the railway system. The focus was on improving the efficiency of trains passing each other, ultimately minimising delays and leading to a more punctual railway operation. An overview of various pass types in Sweden was provided first based on historical train operation data. The passes in Sweden can be broadly categorised into three types: cancel passes, scheduled passes and unscheduled passes. Notably, a remarkable 97% of the passes did not occur as scheduled, with the most common being cancelled passes (76%), followed by unscheduled passes (21%) and scheduled passes (3%). This indicates that timetables are difficult to realise with a high level of accuracy, which results in significantly fewer train passes in actual operations compared to the timetables. It also suggests dispatchers play a very active role in cancelling and rescheduling train passes, shifting them from one station to another.

A logistic regression model was employed to assess how different train passes impact train delays since the odds ratio derived from the logistic regression offers a direct comparison of different dispatching approaches, thus identifying which dispatching approach has the best odds of not causing delays. The results show that delays are less likely when passes are cancelled and more likely when passes are unscheduled when compared to scheduled passes. More specifically, the odds of dwell time delays for unscheduled passes were 2.6 times higher than scheduled passes, suggesting that timetables have little flexibility. On the other hand, the odds of dwell time delays for cancelled passes are 9.8 times less likely than for scheduled passes. This suggests that the cancellation of a train pass can be used as a strategy to mitigate the risk of delays during disturbances, and this is often done in practice.

However, in certain situations, passes cannot easily be cancelled, so they are shifted from one station to another, where they appear as unscheduled passes; as a result, this significantly increases the probability of a delay.

7.3 Paper 3: Analysing Factors Contributing to Real-time Train Arrival Delays using Seemingly Unrelated Regression Models

Paper 3 focuses on understanding the heterogeneous impact of various factors on train delays. To address this, an interpretable and statistically sound model known as the SUR model was employed. By considering the model's contemporaneous residual correlations resulting from the shared unobserved feature variables across the regression equations, the SUR model ensured a statistically more efficient estimation of regression coefficients. This is because at the trip level, missing residual terms such as rolling stock, tracks and driver styles are essentially the same. The modelling design incorporates the concept of real-time train arrival prediction, which treats the train delay prediction problem as a set of next-station arrival delay prediction models that are conditional on train location. High R2 values (between 0.8 and 0.9) and a low RMSE (3 minutes) confirmed that station-specific delay prediction models need to be developed to analyse the impacts of influencing factors on delays.

Paper 3 details the heterogeneous impact of a comprehensive set of factors on train delays and provides insights into the practical implications. Train operation variables (e.g. dwell time, running time, and delays from previous trains and upstream stations), as well as network variables (e.g. travelling direction), are found to have significant impacts on train arrival delays at downstream stations. Notably, the consecutive upstream station delay of the same train has a greater impact on the current station train delay than delays from further upstream stations. This is because the consecutive upstream stations are more reflective of current traffic conditions since the impacts of further upstream delays are absorbed by the consecutive upstream station delay. This underscores the importance of considering the nearest station information for effective real-time operations management.

Calendar variables such as peak hours and weekdays are identified as factors that can increase current station delays. This is attributed to the increased passenger volume during peak hours and the potential for congestion interference between trains due to the increased rail services on weekdays. Weather variables, including temperature, precipitation, snow depth and wind speed, exhibit varying effects on station arrival delays, possibly due to the fact that Sweden is a sizable country with huge variations in weather conditions. However, most weather variables show statistically insignificant impacts on train delays, suggesting that the railway transportation system in Sweden is less likely to be affected by weather. Maintenance variables, including periods of trackwork and temporary speed restrictions, can significantly increase train arrival delays. Maintenance activities caused train speed reductions and increased capacity utilisation, necessitating timetable rescheduling to minimise delays associated with maintenance work during low-traffic periods or at night.

7.4 Paper 4: Real-time Train Arrival Time Prediction at Multiple Stations and Arbitrary Times

Paper 4 focuses on the practical need for real-time train arrival predictions for multiple stations at arbitrary times, as opposed to one-step-ahead (next station) predictions that are operation-oriented and triggered only when a train arrives or departs from a specific station. This proposed approach aims to cater to passenger needs for predicting train arrival times at their concerned stations based on their current locations.

By using the multi-output regression and chained multi-output regression frameworks, the multi-output prediction problem is modelled as a regression problem that simultaneously predicts two or more numerical values. To leverage the most recent information, the multi-output regression models are trained conditionally on the current train location, incorporating the corresponding sets of factors from the training dataset into the model. The comparison results show that direct multi-output and chained multi-output regression models have comparable performance, thus demonstrating that taking into account previous stations' arrival time predictions has no significant impact on subsequent stations' predictions; it only leads to longer training time for chained multi-output regression models.

The superior performance of LightGBM compared to benchmark models such as GBR and RFR highlights its efficiency in predicting train delays. Further investigation into changes in prediction performance as the train travels along the route towards downstream stations reveals that Direct Multi-output LightGBM consistently improves prediction performance (except for the first three models). The bigger prediction errors at the Hässleholm, Alvesta, and Nässjö stations are primarily attributed to their high arrival time variability. Due to the availability of more real-time information for making predictions, the prediction performance at a station tends to improve as the train approaches it.

7.5 Paper 5: Real-time Train Arrival Time Prediction along the Swedish Southern Mainline

Paper 5 builds upon the work in Paper 4 and aims to enhance train arrival time prediction for multiple downstream stations at arbitrary times. The prediction problem in Paper 5 is also framed as a multi-output regression conditioned on the train's current location that leverages the most recent train operation information for optimal accuracy. In addition to the direct multi-output regression framework, Paper 5 explores the SUR framework, which considers correlated prediction errors across equations. Two iterative bias correction approaches are introduced—one-step before prediction error correction and upstream prediction error correction—to improve the performance of the prediction model by incorporating real-time observed information, specifically prediction errors at previous stations, along with predictors from historical data.

The findings show that the SUR approach does not have an obvious advantage over direct multiple output linear regression, suggesting that considering residual correlations in predictions does not enhance the prediction of train arrival times at subsequent stations, thereby confirming the independence of residual correlations in this prediction task. The superior performance of the direct multiple output linear regression model with upstream prediction error correction compared to models with one-step before prediction error correction and without bias correction modules underscores the importance of leveraging both historical and real-time information for improved model performance. These results affirm that iterative prediction error adjustment using all available information to allow the model to constantly adjust itself can improve the prediction effect.

The results also reveal that the prediction error increases as the distance to the destination increases since it requires the prediction of arrival times for a greater number of stations from the origin to the destination. This is expected because longer distances to the terminal destination station introduce more uncertainty and fluctuations in railway traffic conditions. Similar to Paper 4, the findings demonstrate that the prediction performance of the direct multiple output linear regression model at a station improves as the train moves closer to it. This is most likely due to the utilisation of more relevant real-time information that captures the actual dynamics of train traffic for more accurate predictions.

7.6 Paper 6: AP-GRIP Evaluation Framework for Data-driven Train Delay Prediction Models: Systematic Literature Review

Paper 6 aims to equip railway modellers and practitioners with the necessary tools to make a critical evaluation of prediction models and familiarise them with each important evaluation component prior to model deployment in the railway fields. To achieve this goal, Paper 6 presents a systematic literature review on data-driven train delay prediction models and introduces the novel AP-GRIP (Accuracy, Precision-Generalisability, Robustness, Interpretability, and Practicality) evaluation framework consisting of six key aspects.

Accuracy, crucial for assessing prediction model capability, involves measuring the similarity between observed and predicted values. Precision evaluates the dispersion of prediction error by capturing bias tendencies in specific methods, while generalisation ensures the model's transportability across diverse conditions. Robustness assesses model performance in unexpected situations, and interpretability provides insights into the predictions generated, thus enhancing the confidence of end-users. Practicability, an application-oriented aspect, assesses the task fulfilment capability of prediction models from an end-user perspective, recognizing that the tolerance for prediction errors varies based on the model's specific use case. To attain a thorough evaluation, the framework considers spatial, temporal and train-specific perspectives. It also recommends making benchmark model comparisons in addition to relying solely on an overall model performance perspective in order to identify the most effective train delay prediction models. After an evaluation is carried out using the framework, end-users are expected to possess an in-depth understanding of the prediction model's strengths and limitations. They will be able to determine when, where and how certain models perform well or poorly and thus offer well-informed suggestions for the continuous improvement of train delay prediction models.

8 Answering the research questions

As outlined in the introduction, the purpose of the thesis is to enhance the understanding of data-driven train delay prediction models. In this chapter, the findings of each paper are discussed in relation to the research questions posed in the thesis to fulfil its overarching aim.

8.1 Factors to consider when building a train delay prediction model

To facilitate the development of data-driven prediction models in the railway field, the framework proposed in Paper 1 systematically breaks down the train delay prediction process into six components: scope determination, model inputs, data quality, methodologies, model outputs, and evaluation techniques. The subsequent subsections provide detailed insights into each component. However, the part for feature variables under the model inputs and evaluation techniques subsections are explored and discussed in detail under RQ2 and RQ4, respectively. The components of train delay prediction models discussed in Section 8.1 are illustrated in Figure 4.

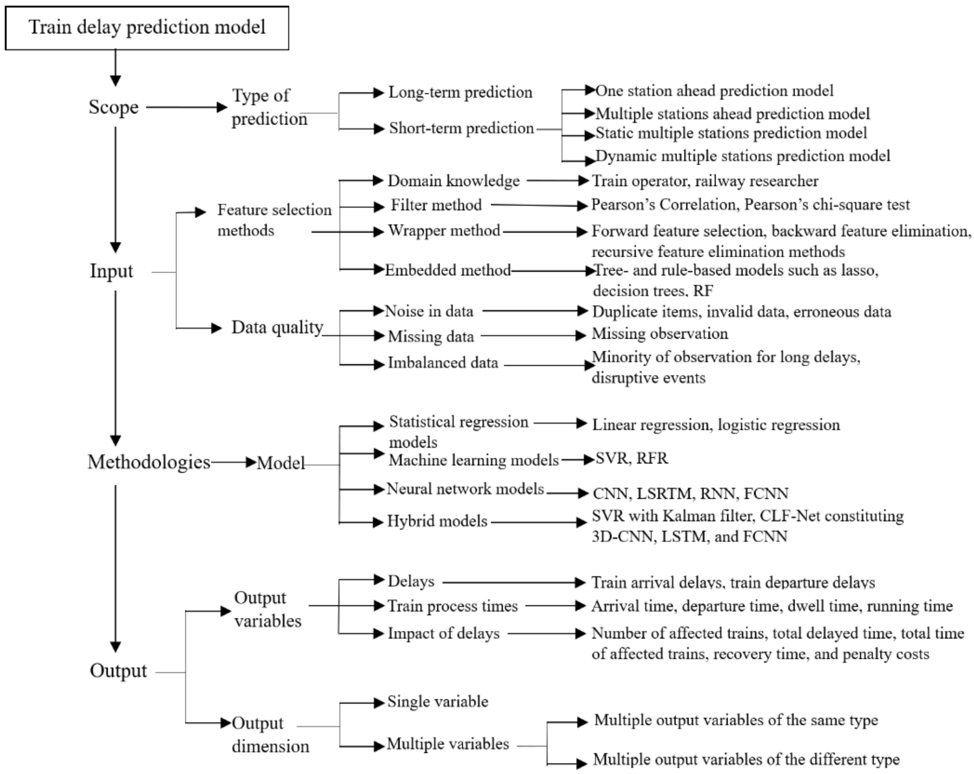


Figure 4: Components in existing train delay prediction models

8.1.1 Scope determination

Data-driven train delay prediction models can be categorised into long-term and short-term train delay prediction models based on the prediction horizons. Long-term train delay prediction models aim to understand the impact of various factors on railway system performance by utilising aggregated historical train operation data to predict delays several days or months in advance. On the other hand, short-term train delay prediction models, fed with real-time data, focus on generating accurate predictions for train delays. Spatially, short-term train delay prediction models are further classified into the following types: the one station ahead prediction model, multiple stations ahead prediction model, static multiple stations prediction model, and dynamic multiple stations prediction model. The one-station-ahead prediction model forecasts the train delay at the next station while the multi-station-ahead prediction model looks at train delays at multiple stations ahead. In contrast, both the static and dynamic multiple-station prediction models aim to predict train delays at all downstream stations. The static multiple stations prediction model provides a one-shot prediction without updates, while the dynamic multiple stations prediction model continually updates predictions as railway traffic information evolves.

8.1.2 Model Input

Feature selection involves the selection of important input variables to eliminate redundant or irrelevant features. The feature selection methods include the filter method, wrapper method, and embedded method. The filter method, such as Pearson's correlation, ensures the selection of input variables that have statistically significant relationships to the target variable. The wrapper method (e.g. forward feature selection, backward feature elimination, and recursive feature elimination methods) involves building similar prediction models but with different subsets of input variables; only those input variables that contribute to the best-performing model are selected. However, the wrapper method is not suitable when dealing with large datasets. The embedded or intrinsic method, also referred to as the feature selection process, is embedded in the predictive model, thus enabling the model to automatically select input variables that maximise the model's accuracy. For instance, feature importance is a built-in metric in a random forest model.

8.1.3 Data quality

Builders of train delay prediction models commonly encounter data quality issues, including class imbalance, missing data, and noise. Therefore, data pre-processing becomes crucial to prevent prediction models from being trained on erroneous data. Removal of missing and noisy observations is a common practice, particularly when the datasets are large enough. Imbalanced data, indicating data with a disproportionate observation ratio, remains a challenging issue in train delay prediction studies. For instance, the utilisation of imbalanced data or data with a long-tailed distribution leads to poor prediction performance for predicting long delays or requires a substantial amount of data for effective model training.

8.1.4 Methodologies

Statistical regression models such as linear regressions are mainly used to understand the variable impact on train delays and can also be used to predict train delays; however, their limitation lies in modelling complex and non-linear relationships. Supervised machine learning, particularly random forest regression, is widely applied in train delay prediction due to its ability to capture non-linear relationships and handle high-dimensional, noisy data for reliable and repeatable predictions. Machine learning, while less interpretable than statistical regression, excels at uncovering hidden knowledge from historical data. However, machine learning requires human-engineered spatiotemporal features to capture the spatial and temporal flow patterns of train operation. Conversely, the advantages of neural networks include automatic learning of spatiotemporal representations from raw data and the flexibility to integrate different architectures into hybrid models, which is efficient in the handling of heterogeneous and multi-attribute data in dynamic railway systems. Different neural network algorithms are effective in dealing with different types of data, such as LSTM or recurrent neural networks (RNN) for sequential data, convolution neural networks (CNNs) for spatial or image data, and fully connected neural networks (FCNN) for cross-sectional data. Hybrid models combining multiple algorithms enhance prediction robustness by leveraging uncorrelated prediction errors and reducing the risk of simultaneous failures.

8.1.5 Model outputs

In train delay prediction studies, output variables encompass parameters related to train delays, train process times and the impact of delays. Arrival delays are the most commonly used output in existing train delay prediction studies since they directly capture disturbance impacts. Accurate prediction of process times is crucial due to the sharing of infrastructure between trains, and trains exceeding the scheduled train process times hinder subsequent train operations. Consequently, the prediction of process times is important for timetable rescheduling and resolving conflicts between train paths. Predicting the parameters related to the impact of delays helps with real-time dispatching and operational decision-making. Despite many studies focusing on predicting one single aspect of train movements, neural networks algorithms and hybrid models can be employed to simultaneously predict multiple output variables, thus enriching the information relative to various train movement aspects.

8.2 Input variables improving the train delay prediction model

To determine effective input variables for train delay prediction models, the impacts of a comprehensive set of factors (as shown in Figure 5) on train arrival delays were examined. Utilising statistical regression coefficients, the relationships between these factors (indicated in bold and italics) and train delays were captured. Significance tests were applied to identify crucial variables to incorporate into the prediction model. The following subsections present the findings derived from the analysis.

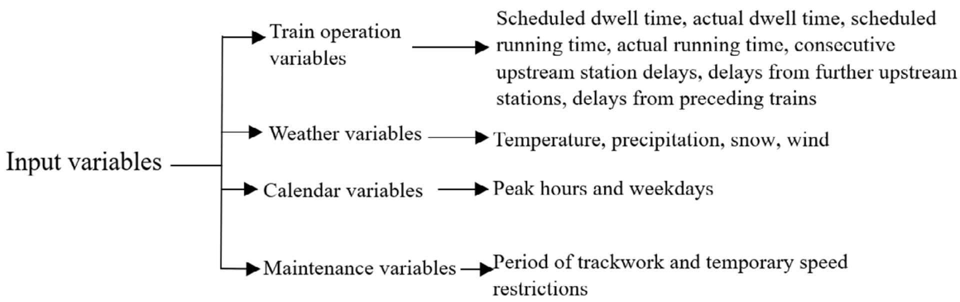


Figure 5: The factors examined in the thesis for their impact on train arrival delays

8.2.1 Operational variables

The ***scheduled dwell time*** exhibits a statistically significant and negative correlation with current station arrival delays, indicating that longer scheduled dwell times offer flexibility for delay recovery by absorbing train delays. Conversely, ***actual dwell time*** has a significant positive impact on current arrival delays, with passenger boarding and alighting activities being the primary contributors. It is worth noting that dispatching decisions by train operators, such as cancelling or allowing unscheduled train passes, also influence dwell times. The relationship between current arrival delays and the ***scheduled running time*** varies along the line, where longer scheduled running times at the first station led to earlier train arrivals, while subsequent stretches require more scheduled running times to absorb accumulated delays and provide flexibility for delay recovery. The ***actual running time*** is statistically significant and correlates negatively with current station arrival delays, indicating that a shorter actual running time following an overtaking operation leads to early train arrival, which results in trains queuing due to infrastructure capacity limitations. ***Consecutive upstream station delays*** show a statistically significant increase in relation to current station delays and possess the greatest potential to exacerbate the current station delays, thus outweighing ***delays from further upstream stations*** due to direct propagation to the current station. Current station delays can increase due to ***delays from preceding trains*** passing the same station due to the sharing of infrastructure between the current and preceding trains.

8.2.2 Weather variables

The effects of weather variables such as temperature, precipitation, snow depth and wind speed on current station arrival delays vary throughout the train lines. ***Cold and extremely cold temperatures*** reduced train delays at terminal stations, but interestingly, at middle stations, extremely cold temperatures increased train delays. The reduction in train delays at terminals is attributed to decreased train demand since passengers tend to cancel unnecessary trips during extreme temperatures. Conversely, the increased train delays at middle stations may be due to the train operations' vulnerability to extremely cold weather in this section. ***Precipitation*** was associated with increased train delays, likely due to the need for greater headway between trains and reduced speeds in wet conditions, all of which impacts the total travel time and causes network congestion delays. Limited alternative transportation options when it rains also contribute to increased train demand, causing further delays. The impact of ***snow and strong wind*** on train delays was generally insignificant, affirming the reliability of railway transport in snowy and windy conditions in Sweden. Notably, snow causes a reduction in train delays at the end of the train line, possibly due to changes in travel behaviour such as rescheduling trips to avoid traffic congestion by adjusting departure times or considering different routes.

8.2.3 Calendar variables

Calendar variables, such as peak hours and weekdays, have a significant correlation to increased train arrival delays. ***Morning peak*** traffic leads to additional train delays because of heavier passenger loads and passenger exchange activities such as late arrivals and clustered boarding, all of which leads to increased dwell delays. ***Afternoon peak*** traffic increases train delays along the first stretch of the line but results in earlier arrivals along the second stretch. This variation may be attributed to the flexibility of passengers in choosing when and where to travel during afternoon peak hours, thus causing passenger volume to fluctuate across different sections of the line. ***Weekday operations*** are more susceptible to train delays than weekend operations. The higher number of trains running during weekdays results in smaller headways between trains, creating a greater potential for congestion interference between trains.

8.2.4 Maintenance variables

Maintenance variables, specifically periods of trackwork and temporary speed restrictions, can significantly increase train arrival delays. ***Trackwork scheduled for both day and night*** exacerbates delays more than when ***trackwork is scheduled solely for day or night***, indicating that the greater the interference of scheduled trackwork with train movements, the longer the train delays. This is because scheduled trackwork causes capacity restrictions like track closures or single-track operations. Daytime trackwork results in more delays due to more train activity. A correlation exists between ***temporary speed restrictions due to trackwork*** and increased train delays along the entire the train line, except for the station at the beginning of the line, which possesses sufficient buffer time to accommodate such restrictions since it is less likely to be burdened with heavily accumulated delays from the previous station. For downstream stations, speed restrictions prolong travel time, leading to train punctuality issues. The unadjusted timetable based on initial speed limits depletes scheduled buffers, inevitably escalating delays across the network.

8.3 Approaches to enhance the train delay prediction model

This research question addresses the train delay problem by constructing a prediction model with robust predictive capabilities that are applicable in real-world scenarios. Location-conditioned concepts and error adjustment strategies are introduced to harness real-time information, thereby improving the model's prediction performance. In terms of practical application, the study proposes various multi-output frameworks designed to dynamically predict train delays for multiple downstream stations at arbitrary prediction times. The enhancements introduced for existing train delay prediction models are depicted in the Figure 6.

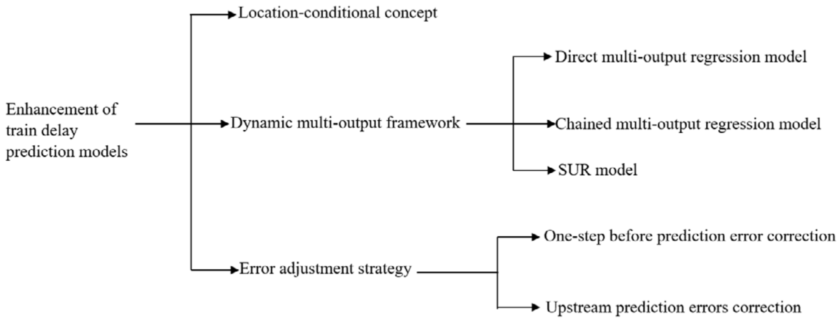


Figure 6: Enhancements for existing train delay prediction models

8.3.1 Location-conditioned concept

A location-conditioned concept is proposed to leverage the availability of the most up-to-date information for real-time prediction. Predictions are generated for future stations when a train is at the current station. In this approach, the current train location determines the inclusion of corresponding sets of factors from the training dataset into the prediction model, as defined by Equation 11. This location-conditioned concept focuses on the observability level of information, which signifies that only factors that can be observed in real-time and historically, such as initial delays, current train delays, and historical train delays, are accounted for when calculating the impact on the next train arrival delays. Taking into account operational information at the station closest to the target prediction station is especially important since the impact of the explanatory factors used for making predictions vary in time and space as trains progress towards their destination (Barbour et al., 2018). The following equation:

$$\hat{y} = f(X|i) \tag{11}$$

where $\hat{y} = (\hat{t}_{i+1}, \hat{t}_{i+2}, \ldots, \hat{t}_N)$, denotes the predicted train arrival delays at subsequent stations given current station *i*. X represents a set of predictor variables encompassing both historical and real-time explanatory factors.

8.3.2 Multi-output framework

In practical applications, passengers receive train arrival information for their particular stations via predictions based on all downstream stations rather than one station-ahead predictions only. Similarly, train dispatchers require a train delay prediction model that serves as a decision-support tool that can provide a comprehensive overview of the expected duration for a train to reach each intermediate station in order to complete the entire route. To address these needs, multi-output frameworks are employed to predict arrival times for multiple downstream stations at arbitrary times via inputs derived from the current station. More specifically, when the train is at station *i*, the model predicts arrival times for downstream stations (*i+1, ..., N*) concurrently. Upon the train's arrival at station *i+1*, the predictions are updated simultaneously for downstream stations (i+2, …, N). Figure 7 illustrates the multi-output framework for arrival time prediction. The thesis evaluates three multi-output frameworks: direct multi-output regression, chained multi-output regression, and SUR.

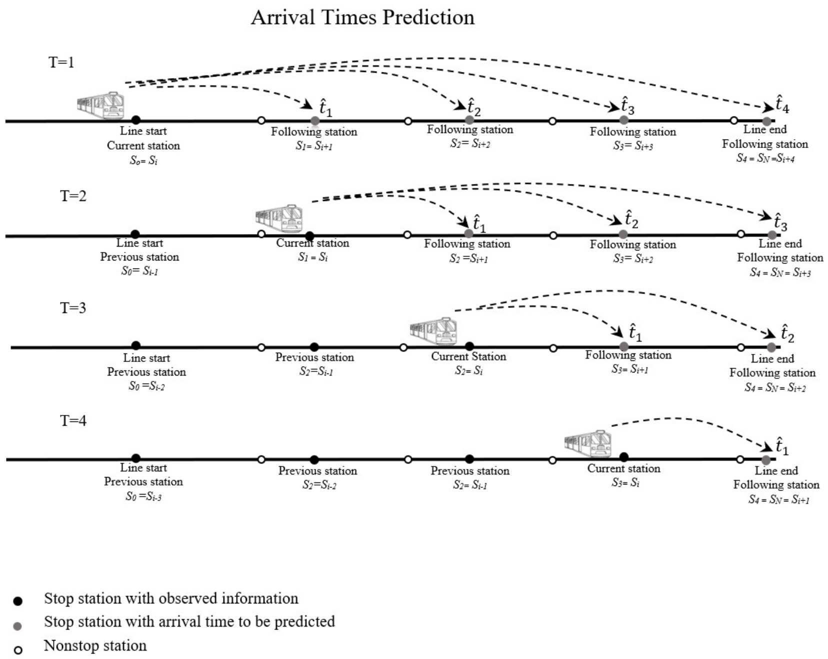


Figure 7: Conditional multi-outputs framework

The ***direct multi-output regression model*** tackles the regression problem by addressing each output independently by assuming their mutual independence. The ***chained multi-output regression model*** functions sequentially, where the first model predicts one output and subsequent models utilise the inputs and outputs from preceding models for subsequent predictions. The use of the ***SUR model*** involves two main stages. The first stage independently generates predictions, similar to the direct multi-output regression model. In the second stage, it considers correlated error terms across multiple regressions from the first stage. Based on the findings from Paper 3 and Paper 4, the direct multi-output regression models outperformed both the chained multi-output regression models and the SUR models, confirming that accounting for previous station arrival time predictions and correlated prediction residuals do not provide significant help in predicting subsequent stations' train arrival times.

8.3.3 Error adjustment strategies

Error adjustment strategies leverage real-time observed information, specifically prediction errors at current and previous stations, to enhance predictive performance. Prediction errors ($E_{i+1,i+1}$) are determined using Equation 12 when the actual train arrival time ($a_{i+1,i+1}$) is observed upon the train's arrival at the next station $i+1$. Two iterative correction approaches are implemented based on these prediction errors: one-step before prediction error correction and upstream prediction error correction.

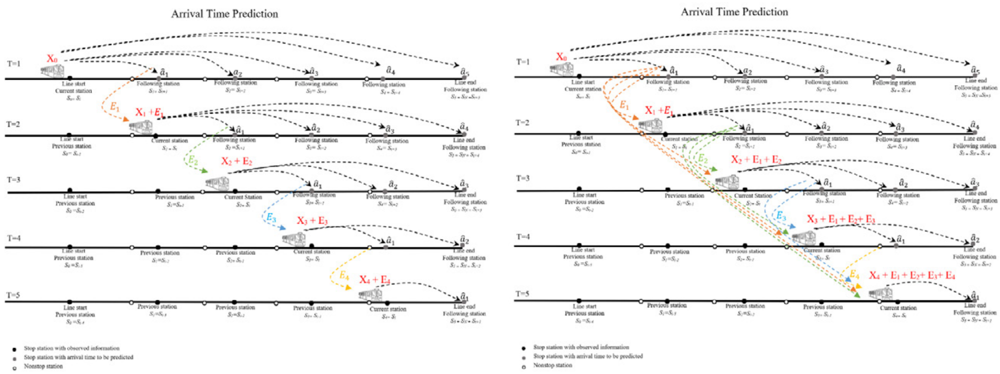


Figure 8: (a) One-step before prediction error correction; (b) Upstream prediction error correction

$$E_{i,i} = \hat{a}_{i,i-1} - a_{i,i} \tag{12}$$

where $E_{i,i}$ is the prediction error at station i when the train is at station i, $\hat{a}_{i,i-1}$ is the predicted arrival times of the train for station i when the train is at station i-1, and $a_{i,i}$ is the actual arrival time of the train for station i when the train is at station i.

In the ***one-step before prediction error correction*** approach, only the prediction error at the current station i where the train is currently located, $E_{i,i}$, is used as a predictor along with other predictors $X_{i,i}$ to predict train delays at subsequent stations. For instance, in Figure 8a, when the train arrives at station S_3 at T = 4, the actual train arrival times $a_{3,3}$ is observed and the prediction error $E_{3,3}$ is calculated. This error is then combined with $X_{3,3}$ to predict $(\hat{a}_{4,3}, \hat{a}_{5,3})$. In the ***upstream prediction error correction*** approach, all previously determined prediction errors and the prediction error at the current station, $(E_{1,1}, \dots, E_{i-1,i-1},\ E_{i,i})$, are used as predictors along with $X_{i,i}$. In Figure 8b for example, when the train arrives at the station S_3 at T = 4, the prediction error $E_{3,3}$ is calculated and used together with previously computed prediction errors $(E_{1,1},\ E_{2,2})$ and $X_{3,3}$ as predictors to forecast $\hat{a}_{4,3}, \hat{a}_{5,3}$. As evidenced in Paper 4, models with one-step before prediction error correction outperform those without any correction, thus highlighting the value of iterative prediction error adjustment using real-time information that enables the model to constantly adjust itself. Furthermore, models with upstream prediction error correction exhibit a slight improvement over those with one-step before correction, underscoring the significance of both historical and real-time information in enhancing model performance.

8.4 Evaluating train delay prediction models

To uncover the strengths and weaknesses of prediction models and consequently guide the decision-making processes related to the expansion, modification, or rejection of a prediction method, there is a need to evaluate the prediction model from six evaluation aspects (accuracy, precision, generalizability, robustness, interpretability and practicality) across multiple dimensions, including overall, spatial, temporal, and train-specific dimensions, as illustrated in Figure 9. It is worth noting that since the output variable for the train delay prediction models is typically a continuous variable such as train delays, the evaluation framework proposed is primarily designed for regression models rather than classification models.

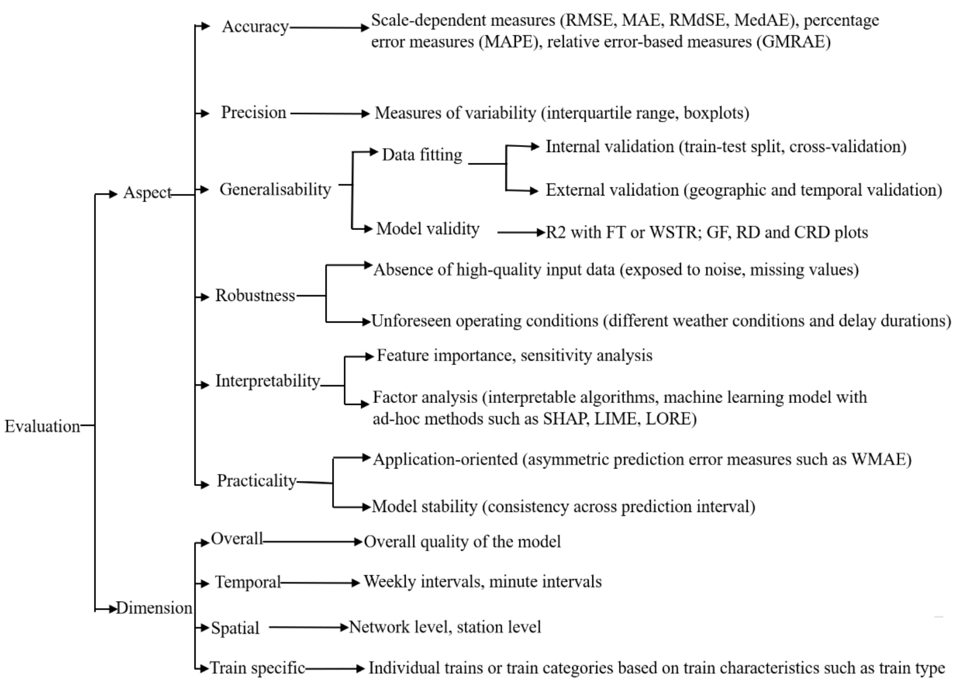


Figure 9: Measures for different evaluation aspects and dimensions

8.4.1 Evaluation aspect: Accuracy

The evaluation of model accuracy involves examining residual errors and gauging the closeness between observed and predicted values (Loague & Green, 1991). Accuracy measurements fall into three categories: scale-dependent, percentage error-based and relative error-based. Commonly employed scale-dependent measurements in train delay prediction studies, such as RMSE and MAE, are sensitive and easily interpretable, thus helping decision-makers comprehend the practical implications of prediction errors. Conversely, median error measurements such as root median squared error (RMdSE) offer robustness against outliers but are less informative. Percentage error measurements, including mean absolute percentage error (MAPE), are scale-independent and facilitate comparisons across studies. However, they encounter challenges with zero actual values, such as the potential for infinite values when actual values are zero, and skewed distributions when actual errors approach zero. Relative error-based measurements such as geometric mean relative absolute error (GMRAE) are scale-independent, resilient against outliers, and provide insights into model performance relative to benchmarks. However, they may lack precision in quantifying the adjustments needed in train operations to rectify delays. Given the emphasis on developing new models and the predominant use of simple baseline models in comparative analyses where all models use the same datasets and units of measurement, employing scale-dependent measures in train delay prediction studies remains a viable choice. Nevertheless, to provide a complete picture of model accuracy, it is essential to incorporate median error measures like RMdSE alongside the sensitive RMSE and MAE metrics.

8.4.2 Evaluation aspect: Precision

Precision, as defined by Walther and Moore (2005), refers to the statistical variance or the spread of data (Debanne, 2000). Measurements of variability, such as range, variance or standard deviation, can serve as precision measurements, with the interquartile range considered more robust due to its lower sensitivity to extreme values (Manikandan, 2011; Walther & Moore, 2005). While unscaled precision measurements such as the coefficient of variation (CV) have been advocated for other prediction domains, their application in train delay prediction is constrained by the prevalence of delays near zero, leading to highly sensitive minor mean fluctuations issues. Graphical methods, particularly boxplots, are commonly used to assess precision in train delay prediction studies. For example, they have been employed to compare the distribution of actual and predicted delays and visualise the distribution of prediction errors. Precision, which indicates the level of uncertainty in prediction errors, plays a crucial role in evaluating the reliability of predictions generated by predictive models, thus emphasizing the need for a comprehensive assessment that incorporates precision measurements (Du et al., 2011). Basic descriptive statistics such as the minimum, maximum, and mean of errors are commonly explored to provide a fundamental understanding of model precision, yet these metrics are sensitive and easily influenced by outliers. Therefore, more robust precision measures such as the interquartile range or boxplot of prediction errors should be incorporated into the precision evaluation process to reflect the level of uncertainty in the prediction errors.

8.4.3 Evaluation aspect: Generalisability

Generalisation, as defined by Bishop and Nasrabadi (2006), is the ability of a predictive model to make accurate predictions on previously unseen data. Challenges to generalisation include 1) data overfitting, where the model performs poorly with new data due to being closely tailored to the training data, and 2) a mismatch between the model's complexity and the characteristics of the data. Van Calster et al. (2023) highlight the importance of both internal and external validation to ensure that the models are not overfitted. Internal validation involves testing the model on data used during its development to ensure it remains valid under similar train operation conditions, while external validation assesses the model's performance on different data to verify the model's transportability across diverse train operation conditions that reflect real-world complexity. Internal validation involves techniques like train-test splits or cross-validation, while external validation can be achieved through temporal or geographic validation. Besides data overfitting, the model validity test is crucial in assessing the generalisation of model performance since it evaluates how well the model aligns with the problem's characteristics and assesses whether residuals exhibit a random pattern. Graphical methods, including observed versus predicted plots, residual distribution plots and cumulative error distribution plots, are commonly employed for this purpose. Washington (2020) proposed that various error specification tests should be considered for a comprehensive assessment of model fit, further enhancing the evaluation process. While external validation is not commonly practiced in train delay prediction studies, it is highly recommended before deploying prediction models to bridge the gap between model development and real-world implementation, ensuring that decision-related railway traffic management is not based on incorrect prediction models in realistic scenarios. Typically, the assessment of goodness-of-fit measures is limited to R^2, but it should be supplemented with another statistical test such as Akaike information criterion (AIC) or Schwarz's Bayesian Information Criterion (BIC), especially for evaluating nonlinear models.

8.4.4 Evaluation aspect: Interpretability

Interpretability, which focuses on users' comprehension of results, is crucial in prediction models for building trust and comprehending the impact of inputs on outputs. Doshi-Velez et al. (2017) propose two key elements for achieving interpretability: 1) providing human-interpretable information about factors and their weights in the decision-making process, often achieved through feature importance analysis in train delay prediction studies, and 2) answering counterfactual questions, which explore the impact of altering inputs on outputs for more comprehensive understanding. This can be easily achieved with the use of interpretable algorithms such as linear regression, logistic regression and decision trees. The simplicity of these models enables users to understand how these algorithms learn from input data and converge towards a solution. Interpretable algorithms have limited performance on high-dimensional data, whereas complex algorithms offer better performance but are less understandable (Mohseni et al., 2021). To address this trade-off, ad hoc methods such as the Shapley additive explanations (SHAP) framework, local interpretable model-agnostic explanation (LIME), anchors, local rule-based explanations (LORE) and model agnostic supervised local explanations (MAPLE) have been developed to provide explanations, particularly for complex models. For a more comprehensive understanding of the underlying mechanisms of prediction models in train delay studies, there is a need to transition towards type (2) explanation is crucial. The adoption of ad-hoc methods is recommended, as they enable the exploitation of advanced prediction models without compromising predictive accuracy. Furthermore, there is a need to encourage further exploration and integration of ad-hoc methods beyond the SHAP framework for enhanced model interpretability, as their utilisation in train delay prediction is still limited and their benefits are not fully realised.

8.4.5 Evaluation aspect: Robustness

Robustness is defined as a model's ability to operate correctly even in the presence of invalid inputs or challenging environmental conditions (IEEE, 1990). Despite its significance, especially in real-time systems, robustness testing, which focuses on unexpected events rather than normal system functioning, is often overlooked. The evaluation of robustness is essential to ensure that train delay prediction models can operate correctly even with suboptimal-quality input data, which includes various perturbations such as noise, missing values, measurement errors, data drift and processing errors. While the robustness of train delay prediction models in the presence of low-quality input has not been extensively studied, research in other fields has actively explored this area (e.g. Koçak et al. (2023); Sharma et al. (2019)). Moreover, the robustness of train delay prediction models should also be evaluated under exceptional or unforeseen operating conditions, such as different weather conditions and delay durations, including extreme delays. This verification process would confirm the model's reliability in providing estimates even during non-recurring events. Shahrokni and Feldt (2013) underscore the importance of evaluating models in realistic settings using datasets with perturbations representative of real-world scenarios in order to ensure their practical viability and applicability in industrial contexts. To avoid train delay prediction models from being solely academic contributions without real-world industrial use, it is crucial to evaluate them using datasets that include perturbations representative of real-world scenarios or exceptional circumstances, particularly disruptions leading to extreme delays.

8.4.6 Evaluation aspect: Practicability

Practicability evaluates the ability of a prediction model to fulfil its intended task from the perspective of end users. The tolerance for prediction errors varies based on the use case of the model. For example, passengers may perceive the overestimation of a delay unfavourably since it could lead to missed connections. Mathematically, practicability is measured using asymmetric prediction error measurements which penalise prediction errors that are unfavourable to end users. In real-time applications, the stability of predictions assesses the consistency of predictions at each interval along the train's journey, ensuring that end users trust the predictions' reliability. Passengers may consider predictions unreliable if they experience significant fluctuations in the information provided, which underscores the importance of stability in maintaining user trust. Recognising the significance of proactive actions based on ahead-of-time information, some studies (Meng et al., 2022; Taleongpong et al., 2020) have examined train delay prediction models at various prediction intervals to assess prediction uncertainty and better equip both operators and passengers for potential adjustments to travel plans. End users' need is often overlooked in the current study, and it is essential to address their perspective by employing asymmetric prediction error measures, such as weighted mean absolute error (WMAE), where heavier weights are assigned to the end-user unfavourable side of prediction errors, thus providing a more user-centric evaluation of model performance.

8.4.7 Evaluation dimensions

To thoroughly understand the prediction patterns of train delay prediction models in realistic operating scenarios, it is crucial to evaluate their performance across various dimensions while taking into account the six aspects mentioned earlier. Overall performance evaluation, the most commonly presented dimension in existing literature, provides a fundamental understanding of how the modelling method performs for the given prediction task, thus offering an overview of the model's quality by aggregating all observations in terms of prediction error measurements. While overall performance assessment is essential during the initial stage of model comparison, detailed evaluations become necessary to uncover underlying performance patterns and identify circumstances in which models excel or fall short. The evaluations must consider spatial and temporal dimensions since train delay prediction is a spatiotemporal problem providing train movement information across different stations and time periods (Zhang et al., 2021). The granularity of evaluation varies based on the purpose; spatial evaluation can range from the network level to the station level, while temporal evaluation can vary from weekly intervals to minute intervals. Another dimension is the train-specific dimension, which involves the analysis of model performance for individual trains or train categories. This analysis involves aggregating data based on train characteristics such as frequency, type, priority, empty rolling stock movements and maintenance fleet movements to distinguish which characteristics make certain train categories more predictable or prone to errors; this helps guide work to improve the model. For example, freight train operations, which are known for their greater variability, pose a challenge to most prediction models (Andersson, 2014). Since a model with good overall performance may underperform in specific situations, the multi-dimensional evaluation approach should be included in the evaluation process to provide a comprehensive understanding of a model's performance and aid in identifying areas for intervention to enhance predictive capabilities.

9 Synthesis

This section synthesizes the findings from the four research questions to address the overarching aim of the thesis, which is to advance the understanding of data-driven train delay prediction models. The data-driven train delay prediction model is discussed across five key components, as shown in Figure 10, which is a modified version of the data-driven train delay prediction framework introduced by Tiong et al. (2023).

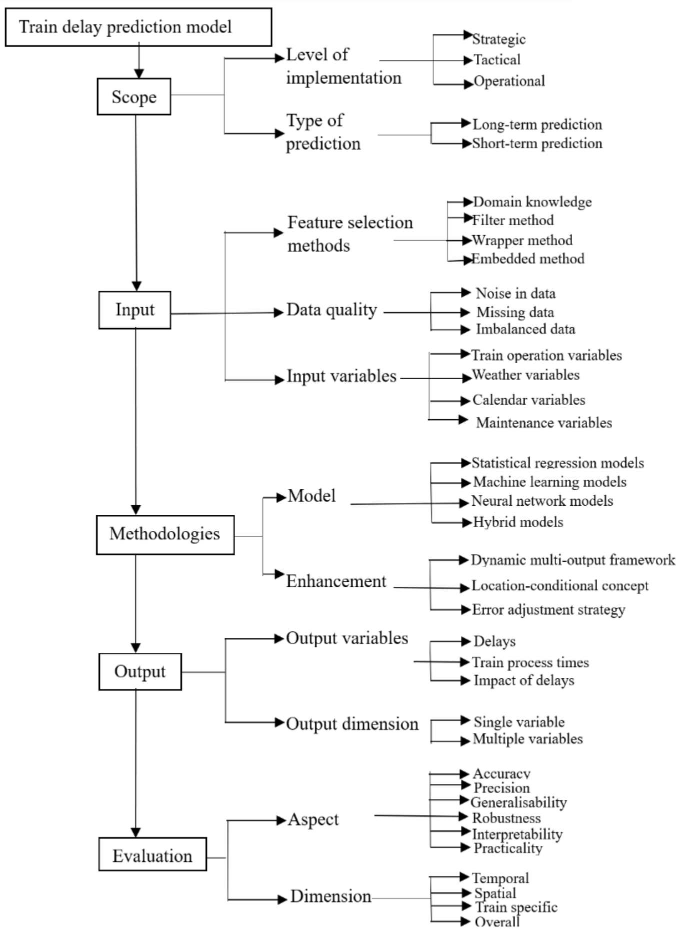


Figure 10: Train delay prediction framework. Source: Adapted from Tiong et al. (2023)

9.1 Scope

In terms of scope, **long-term train delay prediction models** are well-suited for **strategic and tactical-level** train traffic planning, where planning initiatives are implemented several years or days in advance. By examining the relationship between infrastructure investments or facility improvements and their impact on train events, long-term models aid in selecting the most effective investment plan within budget constraints. They reveal the heterogeneous impact of various factors, including rolling stock, crews and margin time, on train events, thus allowing operators to adjust timetables accordingly when anticipating disruptions or disturbances.

On the other hand, **short-term train delay prediction models** serve as intelligent decision-support systems for not only train dispatchers during **operational** train traffic management, but also for passenger information systems. By utilising the most updated information, these models provide insights into near-future train delays based on current operational circumstances. They enable train dispatchers to accurately anticipate conflicts, facilitate making adjustments to resolve conflicts, and enhance operational efficiency. Similarly, providing real-time information regarding expected train delays benefits passengers by assisting them in trip planning and identifying alternative connections.

9.2 Input

In terms of input, there are now many less subjective **feature selection methods** available, including the filter method (e.g. Pearson's correlation), the wrapper method (e.g. forward feature selection, backward feature elimination, and recursive feature elimination methods) and the embedded method (e.g. RFR), all of which complement researchers' knowledge in the domain when selecting input variables to maximise a model's prediction accuracy. To ensure the **quality of input data** and prevent prediction models from being trained on erroneous data, data preprocessing steps such as the elimination of missing and noisy data, as well as extreme data, are necessary, particularly when the removal of such data does not have a significant impact, especially in cases where the datasets are large enough. Due to the nature of imbalanced data or data with long-tail distributions, which can lead to poor performance in predicting long delays or require significantly more data to train predictive models for long train delay predictions, careful preprocessing is essential to address these challenges and improve model performance.

Input variables derived from train operation data, such as train delays at preceding stations, running times and dwell times, have the most significant impact on train delays and ultimately influence the performance of train delay prediction models. However, incorporating input variables from other datasets, such as weather-related and maintenance-related variables, can further enhance the model's responsiveness in diverse circumstances. Despite the fact that their contribution might not be as significant as that of train operation data, their inclusion adds a layer of adaptability to the prediction model and allows it to effectively adapt to different environmental or operational conditions. Spatial and temporal considerations are crucial in bolstering the predictive capabilities of the model. For instance, providing the model with the most recent observations from the nearest station increases its accuracy. This is evident in the findings of the thesis, where the impact of input variables from consecutive upstream station delays significantly influences the current station's train delay. In contrast, the influence of delays from further upstream stations is marginal, likely due to the absorption of these effects by the consecutive upstream station delay.

9.3 Methodologies

In term of methodologies, one drawback of existing studies is their overemphasis on the development of more advanced and sophisticated data-driven train delay prediction models. This trend is evident in the evolution of data-driven train delay prediction approaches, which have shifted from statistical approaches (Gorman, 2009) to conventional machine learning (Li, Wen, et al., 2020; Taleongpong et al., 2020; Wang & Zhang, 2019), to neural networks (Oneto et al., 2017, 2018; Wen et al., 2019) and now to hybrid-based models (Huang, Wen, Fu, Lessan, et al., 2020; Huang, Wen, Fu, Peng, et al., 2020; Lulli et al., 2018).

However, there is a growing need to address the challenges limiting their practical application. In this regard, the thesis introduces **dynamic multi-output train delay prediction models** since passengers and train operators need information not only for the next station but also for the downstream stations that will affect their designated journey. By continuously updating input data from current stations when a train reaches a given station, the framework can predict arrival times for multiple downstream stations at arbitrary times.

To further enhance the accuracy of model predictions, a **location-conditional concept and an error adjustment strategy** can be implemented. The **location-conditional concept** leverages certain characteristics of the Markov property by incorporating input based on the train’s current location. This concept recognizes that delays at any station are directly influenced by the delay status of the immediately preceding station. However, it offers greater flexibility than the Markov property since it does not require the computation of probability distributions, thus allowing the direct incorporation of raw train operation information into the model. Moreover, the location-conditional concept makes it possible to use past information as input, such as delays from the previous two stations, thus recognising the potential usefulness of past information for prediction rather than disregarding it entirely. **Error adjustment strategies** leverage both real-time and historical observed information to enhance predictive performance. Upon the train's arrival at a station, the actual arrival time is observed and the predictions for the current station made when the train was at all previous stations can be verified. Prediction errors can then be computed and incorporated as input variables for the prediction model. This iterative prediction error adjustment strategy using real-time information allows the model to constantly adapt and refine itself.

9.4 Output

In terms of output, the prediction of **arrival delays** stands out as the most common output for train delay prediction models, offering direct insights into disruptions within timetables. For timetable planning and adjustment, as well as for resolving conflicts between train paths and providing reliable passenger information, other output variables can also be considered, including variables related to **train process times** such as arrival time, departure time, dwell time and running time. To gain a thorough understanding of the consequences of delays on operations and to support informed decision-making in real-time train dispatching, the output variables can be variables quantifying the **impacts of delays**, such as the number of affected trains, total delayed time, total time of affected trains, recovery time and penalty costs. It is worth nothing that separate models are often employed for predicting different types of output variables, thus customising the models to **specific output types** and optimising their accuracy and effectiveness. However, in some studies, neural network algorithms and hybrid models were employed to simultaneously predict **multiple output variables of the same type**.

9.5 Evaluation

In terms of evaluation, the drawback of current data-driven train delay prediction studies is their overemphasis on the overall accuracy of the prediction models. To address this limitation and ensure a more comprehensive evaluation of model performance before deployment, the thesis introduces the **AP-GRIP evaluation framework.** This framework encompasses six evaluation aspects: accuracy, precision, generalisation, robustness, interpretability and practicality, all of which are assessed across four evaluation dimensions, which are the spatial, temporal, train-specific and overall dimensions.

In term of **accuracy**, scale-dependent measures such as RMSE are strongly preferred due to the comparisons made between models built using the same dataset with the same unit measurements. To gauge the **precision** of predictions and capture the level of uncertainty in prediction errors, measuring variabilities such as range, variance or standard deviation, along with graphical methods, particularly boxplots, can be employed. Apart from internal and external validation to check whether the model is overfitted or underfitted, model **generalisability** also includes assessing whether residuals exhibit a random pattern rather than a systematic trend by using statistical tests like R^2 with AIC or BIC. Model **robustness** involves evaluating model performance when input variables are subjected to noise and missing values, as well as assessing the model's ability to exhibit acceptable behaviour under exceptional operating conditions such as extreme delays and adverse weather conditions. Besides measuring a feature's weight of importance, model **interpretability** can delve deeper into assessing the impact of altering inputs on the output by using interpretable algorithms or employing complex machine learning models with the help of ad hoc methods. Evaluation in terms of **practicality** involves implementing customised penalties based on end users' tolerance for prediction errors as well as assessing the fluctuation in the predictions at each prediction interval along the train's journey.

To investigate underlying performance patterns comprehensively, the six aspects can be evaluated across **different stations, time periods and train-specific dimensions** while taking into account different levels of data granularity. This approach goes beyond evaluating overall model performance and includes spatial evaluations from the network level to the station level, temporal evaluations ranging from weekly intervals to minute intervals, and train-specific dimension analyses by aggregating the data based on train characteristics like train frequency, type (e.g. commuter, regional, high-speed), priority, empty rolling stock movements and maintenance fleet movements.

10 Conclusion

Due to extensive use and the inherent heterogeneity of train traffic in Sweden, the interdependence of train activities has significantly increased the sensitivity of the train system so that the delay of one train can easily propagate to others. Repeated experiences with delays can lead passengers to perceive that train transportation is unreliable, thus generating negative word-of-mouth publicity and inevitably deteriorating the image of railway transport. In response to these challenges, the development of train delay prediction models has emerged as a crucial research area to enhance the operational efficiency and capacity utilisation of the railway system. Despite the abundance of academic studies utilising diverse data-driven approaches for train delay prediction, there remains a lack of clarity regarding the specific requirements for developing such models. Thus, the goal of this thesis is to better clarify the development process of data-driven train delay prediction models.

Existing literature was reviewed for the thesis in order to synthesize the progress made in various aspects of data-driven train delay prediction model development and present technical solutions for several crucial model components, such as input variables, data quality, modelling techniques, and performance evaluation methods. The examination of existing approaches revealed various technical options in different modelling steps and highlighted the limitations of current modelling techniques as well as pitfalls in the practical application of train delay prediction models. Additionally, the thesis also explores the spatiotemporal relationship among factors such as train operations, network conditions, weather, maintenance, and calendar variables and their impact on train arrival delay. This analysis is vital for identifying the useful explanatory variables to be incorporated into real-time prediction models. By uncovering the relationships between these factors and train delays throughout the train line, the research findings also provide valuable insights for train planners so that they can implement proactive measures based on the train's location, thus minimising the negative impact of potential factors causing train delays.

The thesis also focuses on innovative approaches that leverage real-time and historical observed information to enhance train delay prediction model performance. The key methodological improvements contributed by the thesis include (1) a dynamic multi-output modelling framework generating predictions for all downstream stations at arbitrary times by incorporating spatial-temporal representations from the train operation data, and (2) adaptive error adjustment strategies that continuously correct the prediction based on observed train traffic information. In contrast to existing studies, which heavily emphasise prediction accuracy, the thesis introduces a comprehensive evaluation framework covering various aspects and dimensions crucial for model assessment. Analysing model performance across these aspects and dimensions provides a thorough understanding of predictive capabilities by identifying when, where, and why certain models excel or perform poorly under different circumstances.

In summary, the thesis synthesises current practices in data-driven train delay prediction studies while introducing innovative approaches to enhance model performance. It identifies limitations in existing evaluation processes and introduces a framework to address these gaps. The purpose of the thesis is not to give a full description of the train delay prediction model development process, but rather to offer a tool to consider before attempting to develop data-driven prediction algorithms. The thesis helps make this growing field of research more transparent, with the ultimate goal of accelerating the adoption of data-driven approaches in the railway research community.

10.1 Contribution

This section presents the main contributions of the research presented in this thesis. As outlined in the previous section, the research focused on studying, extending the knowledge of and improving train delay prediction models across various facets of the data-driven model development process, as detailed in the following section.

10.1.1 Enhancing the long-term train delay prediction model

The long-term train delay prediction model proposed in this study can be used to develop advisory strategies in train traffic management, especially in response to unexpected events such as accidents or adverse weather conditions. The model studies the heterogeneous impact of various factors on train events, allowing train operators to adjust their plans accordingly when expecting to encounter different scenarios several days or even months in advance. This gives train operators sufficient time to formulate effective train management plans. By understanding the factors causing train delays, more precisely which part of the train journey is largely affected and where the disturbance or disruption is expected to increase the most, effective dispatching strategies that can be implemented in practice to deal with the potential disturbance and disruption, thus avoiding inefficient infrastructure investment. In other words, the long-term train delay prediction model provides robust theoretical support for train operators as they make decisions related to train schedule planning, adjustment, and infrastructure investments.

10.1.2 Enhancing the short-term train delay prediction model

The short-term train arrival time prediction model proposed in the study plays a crucial role in enhancing passenger information systems by providing timely and accurate information on train arrival times at multiple stations, regardless of the train's current location. As the proposed model takes into account the latest train operation information when making predictions, it can generate timely information on service variations and station/train congestion. With the help of the model, passengers on trains or waiting on platforms are more likely to find train services with acceptable travel times to their destinations and thus board their preferred trains successfully despite having different destinations with different purposes (such as transfers or connecting to other modes). By predicting train arrival times at all downstream stations at any time regardless of where the train is, the proposed model assists passengers in setting realistic expectations in terms of the required travel times for their journey, thus mitigating the secondary effects caused by disruptions to their activities, such as work and study, and ultimately improving their travel experiences.

Furthermore, the proposed short-term train delay prediction model can help train dispatchers accurately anticipate conflicts and potential delays, allowing for proactive conflict resolution and optimal traffic planning. Unlike existing technologies that provide exact current information (such as the exact positioning of trains, their current speed, etc.) without any insight into how to recover, the proposed model offers future-oriented awareness. By providing train dispatchers with information about future train delays based on the current train operation situation, the proposed model can serve as an intelligent decision-support system to increase the dispatcher's ability to efficiently control train traffic. With the support of the short-term prediction model, the train dispatcher can foresee possible disruptions well in advance and implement timely compensatory actions to prevent delays from propagating throughout the network.

10.1.3 Uncovering the model development process

The proposed three-stage train delay prediction framework can accelerate the adoption of data-driven approaches when developing predictive models for decision support in the railway field. The framework outlines a systematic workflow comprising design concepts, modelling and evaluation while addressing six crucial aspects: scope, model inputs, data quality, methodologies, model outputs and evaluation techniques. Rather than offering a comprehensive description of the entire train delay prediction process, the framework serves as a thoughtful tool for researchers embarking on model development. The study emphasises a strong technical focus on streamlining the model development process and identifying up-to-date solutions for the technical problems researchers may potentially encounter, thus providing references for solving technical challenges in each aspect. It ensures that railway researchers can develop data-driven train delay prediction models correctly and efficiently. Simultaneously, the proposed framework establishes a robust foundation for future studies in order to develop more intricate real-time delay prediction algorithms.

10.1.4 Improving the model selection process

The proposed model performance evaluation framework provides a thorough and extensive procedure for selecting the most suitable train delay prediction model. The comprehensive assessment reveals the inadequate prediction patterns of the prediction models and provides an in-depth understanding of their predictive capabilities in diverse circumstances, thus assisting the researcher in determining the suitability of models for specific prediction tasks. The proposed AP-GRIP evaluation framework can serve as guidance in the decision-making processes related to the expansion, modification or rejection of specific prediction methods. By examining the models from various aspects and in multiple dimensions, the strengths and weaknesses of prediction models were revealed, thereby enabling researchers to make well-informed recommendations for continuously improving and refining train delay prediction models.

10.2 Recommendations to put into practice

Advances in technology have facilitated the collection, management and dissemination of data from railway transportation networks, resulting in a surge of academic studies utilising diverse data-driven approaches to model train traffic characteristics and produce train delay prediction in a diverse variety of settings. My work, which has been a step towards bringing data-driven approaches to the railway field, addresses aspects such as timetable planning, train operation management and passenger information systems.

10.2.1 Timetable planning

The results obtained from the developed long-term train delay prediction model offer more comprehensive theoretical support for train operators in various decision-making processes, such as proposing new policies, planning train timetables, managing real-time train operations and making infrastructure investments. Instead of relying on train operators' experience, which is very subjective, the results derived from the developed model quantitatively support the decision-making of train operators by improving train punctuality. For example, insights gained from the models revealed weaknesses in the current timetable and provided an accurate picture of the railway networks. More specifically, the higher punctuality of southbound trains compared to northbound trains along the Swedish Southern Mainline confirmed that inserting more margin shortly after the stations where the most delays occurred can effectively diminish train delays caused by various disturbances, thus improving the chances of quick recovery. This allows operators to understand and quantify the global effects of local operational choices and make informed decisions based on the characteristics of the train operation conditions, given that dispatchers are aware of the possible consequences of different control actions.

10.2.2 Train management systems

Implementing a data-driven approach for real-time prediction applications can enhance the support provided by current train management systems in Sweden. The existing dispatching work relies heavily on experience, leading to uncertainties and inconsistencies in decision-making when faced with similar circumstances. The highly utilised railway network with heterogeneous traffic increases the dependency between trains and the sensitivity of the network, leading to increased complexity when rescheduling during real-time train management. Rush-hour periods and congested stations, characterised by increased train services and reduced allocated buffer times, highlight the challenges encountered by dispatchers when trying to solve rescheduling problems based solely on past experience. In contrast, these stations and periods with heavy train service provide abundant data for data-driven train delay prediction models, thus enhancing their reliability in predicting train delays and generating feasible real-time solutions. However, the full automation of real-time traffic management by intelligent algorithms is not yet the optimal solution, especially for areas with light train service. Thus, the data-driven train delay prediction model can serve as an efficient decision support tool to accurately quantify the effects of various dispatching measures. It will also enable train dispatchers to focus on identifying implementable control actions and efficient replanning strategies such as rerouting and adding train services to optimise resource utilisation. However, the dispatcher's expertise and experience are still needed to monitor and adjust the generated predictions in order to successfully produce feasible schedules and accurate real-time information for train operations.

10.2.3 Passenger information system

The current passenger information system in Sweden relies on manual forecasts generated based on the scheduled timetable and which are updated by a dedicated staff taking into account current train delay times. The manual forecast does not involve any additional decision-making beyond the existing timetable information. In reality, various dispatching strategies are employed to handle perturbed or disrupted train traffic, such as rerouting, cancelling, rescheduling and resequencing trains. The research identifies the limitations of the manual approach used in the current passenger information system due to its tendency to concentrate on previous train delay times while ignoring extreme train traffic behaviours, and its inability to predict random events. The study suggests replacing the manual approach with data-driven methods known for revealing hidden knowledge through data learning. However, trust from the railway sector is crucial to accelerate advances in data-driven train delay prediction models that will eventually benefit the train user. Timely information about service adjustments is important for passengers that need to re-plan their journey effectively since it provides them with more choices than they have now for reacting to the delay information, and it reduces the possibility of being denied boarding, particularly during rush hours.

10.3 Limitations

As with most research studies, the thesis has its limitations. This section examines the limitations of each paper that constitutes the thesis. Given that Paper 1 derived technical options exclusively from existing data-driven train delay prediction studies, it might overlook the latest techniques that could serve as effective solutions to the key problems encountered in model development. This is due to the interdisciplinary nature of data-driven train delay prediction studies, and researchers in the computer science domain in particular are often at the forefront of identifying novel and efficient solutions to modelling challenges compared to those in the railway domains. Consequently, the approach adopted by Paper 1, which focused on previous railway domain studies, may hinder the transfer of knowledge and innovative solutions from the computer science domain to the railway domain, potentially slowing the adoption of practices in the railway field.

Gorman (2009) identified three types of train interactions crucial for predicting delays—meets, passes, and overtakes—where both overtaken and passing trains typically incur delays. However, a limitation of Paper 2 is its exclusive consideration of the impact of train passes on delays, thus restricting the exploration of valuable inputs from a broader spectrum of train interactions. While Paper 6 introduces the AP-GRIP evaluation framework, it also faces limitations when demonstrating and validating the applicability of the framework in practice due to the insufficient amount of data available derived from various train delay prediction models. Nevertheless, conducting a comprehensive case study is essential to providing valuable insights into how the analysis can be conducted and the challenges encountered when applying the proposed evaluation framework.

Papers 3–5 share a limitation in their case studies since they both focus on one to four lines in Sweden rather than extensive railway networks, prompting concerns about the model's applicability to unexplored sections of the network and different train lines. This signifies that the models may not be generalisable since they are trained with insufficient data, thus rendering them unable to capture train delays at stations or lines with varying frequencies of operation. Papers 4–5 have an additional limitation, which is that their train delay prediction models only incorporate train operation variables identified as very significant in Paper 2, omitting other factors like weather-related variables that are significant but less impactful on train delays. This modelling choice may result in non-responsive prediction models that struggle to accurately predict train delays in scenarios involving unexpected incidents such as accidents and adverse weather conditions. This is due to the construction of responsive predictive algorithms that require comprehensive variables reflecting complex traffic conditions, including unforeseen events.

10.4 Future research

Data-driven train delay prediction is constantly developing and requires concurrent research efforts focused on critical assessments of existing prediction models. It also requires in-depth investigation into comparatively under-researched prediction issues and the identification of weaknesses in current data-driven models, all of which can provide valuable insights and promising directions for future train delay prediction research. In light of this, several suggestions for further research are outlined below.

In the thesis, interpretable algorithms like logistic regression and SUR models are adopted to gain deeper insights into the factors influencing train arrival delays. While machine learning models are often more accurate than traditional statistical models, they are criticised for being “black boxes” due to a lack of transparency and interpretability. Thus, an interesting direction for further research involves exploring explainable AI approaches and utilising ad hoc methods such as SHAP to analyse machine learning outputs. Explainable AI approaches serve to identify significant factors and quantify their effects on train delays as well as address the interpretability challenge of machine learning methods. This area is noteworthy since interpretable algorithms, while having less accuracy, can be complemented by ad hoc methods without sacrificing the prediction performance of complex models. Despite numerous studies in the area of black-box model interpretability such as LIME (Ribeiro et al., 2016) and LORE (Guidotti et al., 2018), their adoption in the train delay prediction field remains limited.

The train delay prediction models proposed in this thesis (models in Papers 4 and 5) focus on delivering accurate predictions of future train delays under typical train traffic conditions by leveraging historical and real-time train operation data. Future research could explore the integration of data from multiple sources, including the causes of disruptions, influences on infrastructure, and passenger flows, especially for predicting train delays under abnormal conditions such as adverse weather and special events. Additionally, exploring online prediction using streaming data presents a compelling direction for research since challenges like missing data, noisy data and delayed observations are inevitable in practice. Furthermore, while the reliability of the proposed train delay prediction model is validated at the line level, future research can seek to enhance the proposed dynamic multi-output train delay prediction models to make predictions about very large railway networks with diverse operating characteristics.

The primary input for most existing train delay prediction models relies on the train events observed in train operation data, thus facilitating the prediction of daily train delays. However, the potential limitation to this lies in the fact that current and historical train events may not adequately represent future occurrences. One potential avenue for future research could involve developing methods to estimate synthetic train events that represent future events not yet observed in the train operation data. Combining synthetic train events with those derived from existing data into the train delay prediction models could enhance the robustness of the prediction generated for diverse train events to improve the decision-making ability of railway dispatchers in unexpected cases. This approach draws inspiration from the exploration of synthetic data in other research fields, such as the utilization of deep generative models (DGM), generative adversarial networks (GAN), and variational autoencoders (VAE) to comprehend system-level travel behaviour (Kim and Bansal (2023).

References

Adele, S., Tréfond-Alexandre, S., Dionisio, C., & Hoyau, P. A. (2019). Exploring the behavior of suburban train users in the event of disruptions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 344–362. https://doi.org/https://doi.org/10.1016/j.trf.2019.08.009

Andersson, E., Peterson, A., & Törnquist Krasemann, J. (2011, February 16-18). *Robustness in Swedish railway traffic timetables* 4th International Seminar on Railway Operations Modelling and Analysis, Sapienza-University of Rome. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A539662&dswid=-4552

Andersson, E. V. (2014). *Assessment of robustness in railway traffic timetables* [Linköping University Electronic Press]. http://dx.doi.org/10.3384/lic.diva-103676

Bao, X., Li, Y., Li, J., Shi, R., & Ding, X. (2021). Prediction of train arrival delay using hybrid ELM-PSO approach. *Journal of Advanced Transportation*, *2021*, 1–15. https://doi.org/https://doi.org/10.1155/2021/7763126

Barbour, W., Martinez Mori, J. C., Kuppa, S., & Work, D. B. (2018). Prediction of arrival times of freight traffic on US railroads using support vector regression. *Transportation Research Part C: Emerging Technologies*, *93*, 211–227. https://doi.org/https://doi.org/10.1016/j.trc.2018.05.019

Barron, A., Melo, P. C., Cohen, J. M., & Anderson, R. J. (2013). Passenger-focused management approach to measurement of train delay impacts. *Transportation Research Record*, *2351*(1), 46–53. https://doi.org/https://doi.org/10.3141/2351-06

Bešinovic, N. (2017). *Integrated capacity assessment and timetabling models for dense railway networks* [Delft University]. https://pure.tudelft.nl/ws/portalfiles/portal/19195941/NBesinovic_PhD_thesis.pdf

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer. https://link.springer.com/book/9780387310732

Blayac, T., & Stéphan, M. (2021). Are retrospective rail punctuality indicators useful? Evidence from users perceptions. *Transportation Research Part A: Policy and Practice*, *146*(3), 193–213. https://doi.org/https://doi.org/10.1016/j.tra.2021.01.013

Börjesson, M., & Eliasson, J. (2011). On the use of “average delay” as a measure of train reliability. *Transportation Research Part A: Policy and Practice*, *45*(3), 171–184. https://doi.org/https://doi.org/10.1016/j.tra.2010.12.002

Brazil, W., White, A., Nogal, M., Caulfield, B., O'Connor, A., & Morton, C. (2017). Weather and rail delays: Analysis of metropolitan rail in Dublin. *Journal of Transport Geography*, *59*, 69–76. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2017.01.008

Brons, M. R. E., & Rietveld, P. (2008). *Rail mode, access mode and station choice: The impact of travel time unreliability*. https://core.ac.uk/download/pdf/16378496.pdf

Carey, M., & Kwieciński, A. (1994). Stochastic approximation to the effects of headways on knock-on delays of trains. *Transportation Research Part B: Methodological*, *28*(4), 215–267. https://doi.org/https://doi.org/10.1016/0191-2615(94)90001-9

Ceder, A., & Hassold, S. (2015). Applied analysis for improving rail-network operations. *Journal of Rail Transport Planning & Management*, *5*(2), 50–63. https://doi.org/https://doi.org/10.1016/j.jrtpm.2015.06.001

Chiou, Y.-C., Jou, R.-C., & Yang, C.-H. (2015). Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice*, *78*, 161–177. https://doi.org/https://doi.org/10.1016/j.tra.2015.05.016

Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster. https://books.google.se/books?hl=en&lr=&id=mjVKEAAAQBAJ&oi=fnd&pg=PR9&dq=Deep+Learning+with+Python,+Second+Edition&ots=Ag7TAEXA-b&sig=KeK4xnfoL6AOrdoH3AF3zKMfAPY&redir_esc=y#v=onepage&q=Deep%20Learning%20with%20Python%2C%20Second%20Edition&f=false

Corman, F. (2010). *Real-time railway traffic management: Dispatching in complex, large and busy railway networks* [Delft University of Technology]. https://repository.tudelft.nl/islandora/search/contributor%3A%22Hansen%2C%20I.A.%20%28promotor%29%22

Corman, F., & Kecman, P. (2018). Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, *95*, 599–615. https://doi.org/https://doi.org/10.1016/j.trc.2018.08.003

D'Ariano, A. (2008). *Improving real-time train dispatching: models, algorithms and applications* [Delft University]. https://www.researchgate.net/profile/Andrea-Dariano/publication/27344535_Improving_real-time_train_dispatching_models_algorithms_and_applications/links/0deec521371a6572b9000000/Improving-real-time-train-dispatching-models-algorithms-and-applications.pdf

Daamen, W., Goverde, R. M. P., & Hansen, I. A. (2008). Non-discriminatory automatic registration of knock-on train delays. *Networks and Spatial Economics*, *9*(1), 47–61. https://doi.org/https://doi.org/10.1007/s11067-008-9087-2

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, *29*(3), 510–522. https://doi.org/https://doi.org/10.1016/j.ijforecast.2012.09.002

Debanne, S. M. (2000). The planning of clinical studies: Bias and precision. *Gastrointestinal Endoscopy*, *52*(6), 821–822. https://doi.org/https://doi.org/10.1067/mge.2000.110757

Dennis, H., Leo, G. K., Ramon, M. L., & Michiel, J. C. M. V. (2005). Operations Research in passenger railway transportation. *Statistica Neerlandica*, *59*(4), 467–497. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=797950

Dingler, M. H., Lai, Y. C., & Barkan, C. P. (2009). Impact of train type heterogeneity on single-track railway capacity. *Transportation Research Record*, *2117*(1), 41–49. https://doi.org/https://doi.org/10.3141/2117-06

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S. J., O'Brien, D., Shieber, S., Waldo, J., Weinberger, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. *SSRN Electronic Journal*. https://doi.org/https://doi.org/10.2139/ssrn.3064761

Du, N., Budescu, D. V., Shelly, M. K., & Omer, T. C. (2011). The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, *114*(2), 179–189. https://doi.org/https://doi.org/10.1016/j.obhdp.2010.10.005

Fenner, M. (2019). *Machine learning with Python for everyone.* Addison-Wesley Professional.

Givoni, M., Brand, C., & Watkiss, P. (2009). Are railways climate friendly? *Built Environment*, *35*(1), 70–86. https://doi.org/https://doi.org/10.2148/benv.35.1.70

Gorman, M. F. (2009). Statistical estimation of railroad congestion delay. *Transportation Research Part E: Logistics and Transportation Review*, *45*(3), 446–456. https://doi.org/https://doi.org/10.1016/j.tre.2008.08.004

Goverde, R. M., & Hansen, I. A. (2013). *Performance indicators for railway timetables.* IEEE International Conference on intelligent rail transportation proceedings.

Goverde, R. M. P. (2005). *Punctuality of railway operation and timetable stability analysis* [Delft University of Technology]. https://repository.tudelft.nl/islandora/search/contributor%3A%22Hansen%2C%20I.A.%20%28promotor%29%22

Goverde, R. M. P. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological*, *41*(2), 179–201. https://doi.org/https://doi.org/10.1016/j.trb.2006.02.003

Grandhi, B. S., Chaniotakis, E., Thomann, S., Laube, F., & Antoniou, C. (2021). An estimation framework to quantify railway disruption parameters. *IET Intelligent Transport Systems*, *15*(10), 1256–1268. https://doi.org/https://doi.org/10.1049/itr2.12095

Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F., & Giannotti, F. (2018). *Local rule-based explanations of black box decision systems.* ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.1805.10820

Guirao, B., Garcia-Pastor, A., & Lopez-Lambas, M. E. (2016). The importance of service quality attributes in public transportation: Narrowing the gap between scientific

research and practitioners' needs. *Transport Policy 49*, 68–77. https://doi.org/https://doi.org/10.1016/j.tranpol.2016.04.003

Harris, N. G., Mjøsund, C. S., & Haugland, H. (2013). Improving railway performance in Norway. *Journal of Rail Transport Planning & Management*, *3*(4), 172–180. https://doi.org/https://doi.org/10.1016/j.jrtpm.2014.02.002

Higgins, A., & Kozan, E. (1998). Modeling train delays in urban networks. *Transportation Science*, *32*(4), 346–357. https://www.jstor.org/stable/25768833?seq=1

Hofman, M., Madsen, L., Jespersen Groth, J. C., J, & Larsen, J. (2006). *Robustness and recovery in train scheduling-a case study from DSB S-tog a/s.* 6th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'06). https://drops.dagstuhl.de/storage/01oasics/oasics-vol005-atmos2006/OASIcs.ATMOS.2006.687/OASIcs.ATMOS.2006.687.pdf

Holmgren, J. (2020). The effect of public transport quality on car ownership – A source of wider benefits? *Research in Transportation Economics*, *83*, 100957. https://doi.org/https://doi.org/10.1016/j.retrec.2020.100957

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/https://doi.org/10.1177/1049732305276687

Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., Peng, Q., & Xu, X. (2020). Modeling train operation as sequences: A study of delay prediction with operation and weather data. *Transportation Research Part E: Logistics and Transportation Review*, *141*(102022). https://doi.org/https://doi.org/10.1016/j.tre.2020.102022

Huang, P., Wen, C., Fu, L., Peng, Q., & Tang, Y. (2020). A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems. *Information Sciences*, *516*, 234–253. https://doi.org/https://doi.org/10.1016/j.ins.2019.12.053

Huisman, T., & Boucherie, R. J. (2001). Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B: Methodological*, *35*(3), 271–292. https://doi.org/https://doi.org/10.1016/S0191-2615(99)00051-X

IEEE. (1990). *IEEE standard glossary of software engineering terminology.* https://doi.org/https://doi.org/10.1109/ieeestd.1990.101064

Ivina, D., Palmqvist, C. W., Olsson, N., & Winslott Hiselius, L. (2021). *Train delays due to trackwork in Sweden.* 9th International Conference on Railway Operations Modelling and Analysis (ICROMA)–RailBeijing., https://www.k2centrum.se/sites/default/files/fields/field_uppladdad_rapport/59_train_delays_due_to_trackworks_in_sweden_002.pdf

JBS. (2021). *TTT – Tillsammans för Tåg i Tid Årssammanfattning 2021* https://bransch.trafikverket.se/contentassets/de2780dd12d847a6a5bae5c5f74907db/ttt_arssammanfattning_2021.pdf

Jiang, S., Persson, C., & Akesson, J. (2019, October 27-30). *Punctuality prediction: Combined probability approach and random forest modelling with railway delay statistics in Sweden* 2019 IEEE Intelligent Transportation Systems Conference (ITSC), http://dx.doi.org/10.1109/itsc.2019.8916892

Jovanović, P., Kecman, P., Bojović, N., & Mandić, D. (2017). Optimal allocation of buffer times to increase train schedule robustness. *European Journal of Operational Research*, *256*(1), 45–54. https://doi.org/https://doi.org/10.1016/j.ejor.2016.05.013

Kecman, P. (2014). *Models for predictive railway traffic management.* [Delft University]. https://www.researchgate.net/publication/268277780_Models_for_Predictive_Railway_Traffic_Management

Kecman, P., & Goverde, R. M. P. (2015). Predictive modelling of running and dwell times in railway traffic. *Public Transport*, *7*(3), 295–319. https://doi.org/https://doi.org/10.1007/s12469-015-0106-7

Kim, E. J., & Bansal, P. (2023). A deep generative model for feasible and diverse population synthesis. *Transportation Research Part C: Emerging Technologies*, *148*(104053). https://doi.org/https://doi.org/10.1016/j.trc.2023.104053

Koçak, B., Cuocolo, R., Dos Santos, D. P., Stanzione, A., & Ugga, L. (2023). Must-have qualities of clinical research on artificial intelligence and machine learning. *Balkan Medical Journal*, *40*(1), 3. https://pubmed.ncbi.nlm.nih.gov/36578657/

Kuipers, R. A., Palmqvist, C.-W., Olsson, N. O. E., & Winslott Hiselius, L. (2021). The passenger's influence on dwell times at station platforms: A literature review. *Transport Reviews*, *41*(6), 721–741. https://doi.org/https://doi.org/10.1080/01441647.2021.1887960

Laifa, H., khcherif, R., & Ben Ghezalaa, H. H. (2021). Train delay prediction in Tunisian railway through LightGBM model. *Procedia Computer Science*, *192*, 981–990. https://doi.org/https://doi.org/10.1016/j.procs.2021.08.101

Li, D., Daamen, W., & Goverde, R. M. P. (2016). Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station. *Journal of Advanced Transportation*, *50*(5), 877–896. https://doi.org/https://doi.org/10.1002/atr.1380

Li, D., Goverde, R. M., Daamen, W., & He, H. (2014). *Train dwell time distributions at short stop stations*. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC*)*, 3410-3415. https://doi.org/https://doi.org/10.1109/ITSC.2014.6958076

Li, Z., Huang, P., Wen, C., Tang, Y., & Jiang, X. (2020). Predictive models for influence of primary delays using high-speed train operation records. *Journal of Forecasting*, *39*(8), 1198–1212. https://doi.org/https://doi.org/10.1002/for.2685

Li, Z., Wen, C., Hu, R., Xu, C., Huang, P., & Jiang, X. (2020). Near-term train delay prediction in the Dutch railways network. *International Journal of Rail Transportation*, *9*(6), 520-539. https://doi.org/https://doi.org/10.1080/23248378.2020.1843194

Loague, K., & Green, R. E. (1991). Statistical and graphical methods for evaluating solute transport models: Overview and application. *Journal of Contaminant Hydrology*, *7*(1-2), 51–73. https://doi.org/https://doi.org/10.1016/0169-7722(91)90038-3

Lovett, A. H., Dick, C. T., & Barkan, C. P. (2015). *Determining freight train delay costs on railroad lines in North America*. Proceedings of Rail Tokyo.

Lulli, A., Oneto, L., Canepa, R., Petralli, S., & Anguita, D. (2018, October 1-3). *Large-Scale Railway Networks Train Movements: A Dynamic, Interpretable, and Robust Hybrid Data Analytics System.* 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), http://dx.doi.org/10.1109/dsaa.2018.00048

Luo, J., Huang, P., & Peng, Q. (2022). A multi-output deep learning model based on Bayesian optimization for sequential train delays prediction. *International Journal of Rail Transportation*, *11*(5), 705–731. https://doi.org/https://doi.org/10.1080/23248378.2022.2094484

Manikandan, S. (2011). Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*, *2*(4), 315. https://www.proquest.com/docview/898546209?pq-origsite=gscholar&fromopenview=true

Marković, N., Milinković, S., Tikhonov, K. S., & Schonfeld, P. (2015). Analyzing passenger train arrival delays with support vector regression. *Transportation Research Part C: Emerging Technologies*, *56*, 251–262. https://doi.org/https://doi.org/10.1016/j.trc.2015.04.004

Medeossi, G., Longo, G., & de Fabris, S. (2011). A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management*, *1*(1), 1–13. https://doi.org/https://doi.org/10.1016/j.jrtpm.2011.07.001

Meng, M., Toan, T. D., Wong, Y. D., & Lam, S. H. (2022). Short-term travel-time prediction using support vector machine and nearest neighbor method. *Transportation Research Record: Journal of the Transportation Research Board*, *2676*(6), 353–365. https://doi.org/https://doi.org/10.1177/03611981221074371

Milinković, S., Marković, M., Vesković, S., Ivić, M., & Pavlović, N. (2013). A fuzzy Petri net model to estimate train delays. *Simulation Modelling Practice and Theory*, *33*, 144–157. https://doi.org/https://doi.org/10.1016/j.simpat.2012.12.005

Milliet de Faverges, M., Russolillo, G., Picouleau, C., Merabet, B., & Houzel, B. (2018, November 4-7). *Estimating Long-Term Delay Risk with Generalized Linear Models.* 2018 21st International Conference on Intelligent Transportation Systems (ITSC), http://dx.doi.org/10.1109/itsc.2018.8569507

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems*, *11*(3-4), 1–45. https://doi.org/https://doi.org/10.1145/3387166

Monsuur, F., Enoch, M., Quddus, M., & Meek, S. (2021). Modelling the impact of rail delays on passenger satisfaction. *Transportation Research Part A: Policy and Practice*, *152*, 19–35. https://doi.org/https://doi.org/10.1016/j.tra.2021.08.002

Mou, W., Cheng, Z., & Wen, C. (2019, June 17-20). *Predictive model of train delays in a railway system.* 8th International Conference on Railway Operations Modelling Analysis (RailNorrköping), Norrköping, Sweden. https://ep.liu.se/ecp/069/059/ecp19069059.pdf

Nabian, M. A., Alemazkoor, N., & Meidani, H. (2019). Predicting near-term train schedule performance and delay using bi-level random forests. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(5), 564–573. https://doi.org/https://doi.org/10.1177/0361198119840339

Nair, R., Hoang, T. L., Laumanns, M., Chen, B., Cogill, R., Szabó, J., & Walter, T. (2019). An ensemble prediction model for train delays. *Transportation Research Part C: Emerging Technologies*, *104*, 196-209. https://doi.org/https://doi.org/10.1016/j.trc.2019.04.026

Nie, L., & Hansen, I. A. (2005). System analysis of train operations and track occupancy at railway stations. *European Journal of Transport and Infrastructure Research* *5*(1). https://journals.open.tudelft.nl/ejtir/article/view/4331

Økland, A., & Olsson, N. O. E. (2020). Punctuality development and delay explanation factors on Norwegian railways in the period 2005–2014. *Public Transport*, *13*(1), 127–161. https://doi.org/https://doi.org/10.1007/s12469-020-00236-y

Olsson, N. O., & Haugland, H. (2004). Influencing factors on train punctuality—results from some Norwegian studies. *Transport Policy*, *11*(4), 387-397. https://doi.org/https://doi.org/10.1016/j.tranpol.2004.07.001

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., & Anguita, D. (2016, October 17-19). *Advanced analytics for train delay prediction systems by including exogenous weather data.* 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montréal, Canada. http://dx.doi.org/10.1109/dsaa.2016.57

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., & Anguita, D. (2017). Dynamic delay predictions for large-scale railway networks: Deep and shallow extreme learning machines tuned via thresholdout. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *47*(10), 2754–2767. https://doi.org/https://doi.org/10.1109/tsmc.2017.2693209

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., & Anguita, D. (2018). Train delay prediction systems: A big data analytics perspective. *Big Data Research*, *11*, 54–64. https://doi.org/https://doi.org/10.1016/j.bdr.2017.05.002

Ongsulee, P., Chotchaung, V., Bamrungsi, E., & Rodcheewit, T. (2018, November 21-23). *Big data, predictive analytics and machine learning.* 2018 16th international conference on ICT and knowledge engineering (ICT&KE), Bangkok, Thailand. http://dx.doi.org/10.1109/itsc.2013.6728330

Palmqvist, C. W., O.E. Olsson, N., & Winslott Hiselius, L. (2020). Some influencing factors for passenger train punctuality in Sweden. *International Journal of Prognostics and Health Management*, *8*(3). https://doi.org/https://doi.org/10.36001/ijphm.2017.v8i3.2649

Palmqvist, C. W. (2019). *Delays and Timetabling for Passenger Trains* [Lund University]. Sweden. https://lucris.lub.lu.se/ws/portalfiles/portal/70626078/Carl_William_Palmqvist_web.pdf

Parbo, J., Nielsen, O. A., & Prato, C. G. (2016). Passenger perspectives in railway timetabling: A literature review. *Transport Reviews*, *36*(4), 500–526. https://doi.org/https://doi.org/10.1080/01441647.2015.1113574

Pedersen, T., Nygreen, T., & Lindfeldt, A. (2018). *Analysis of temporal factors influencing minimum dwell time distributions* WIT Transactions on the Built Environment, https://books.google.se/books?hl=en&lr=&id=Ttd1DwAAQBAJ&oi=fnd&pg=PA447&dq=Analysis+of+temporal+factors+influencing+minimum+dwell+time+distributions&ots=7uo-oFltce&sig=w5DN0CI6PU8TwPDkEfJL7eQCGEI&redir_esc=y#v=onepage&q=Analysis%20of%20temporal%20factors%20influencing%20minimum%20dwell%20time%20distributions&f=false

Peterson, A. (2012). Towards a robust traffic timetable for the Swedish Southern Mainline. *Proceedings of the computers in railways XIII: Computer system design and operation in the railway and other transit systems*. WIT Press

Pineda-Jaramillo, J., Bigi, F., Bosi, T., Viti, F., & D'ariano, A. (2023). Short-term arrival delay time prediction in freight rail operations using data-driven models. *IEEE Access*, *11*, 46966-46978. https://doi.org/https://doi.org/10.1109/access.2023.3275022

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?". *Proceedings of the 22nd ACM SIGKDD nternational Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery. http://dx.doi.org/10.1145/2939672.2939778

Sarstedt, M., & Mooi, E. (2014). *A Concise Guide to Market Research*. Springer. https://link.springer.com/book/10.1007/978-3-642-12541-6

Scheepmaker, G. M. Pudney, P. J., Albrecht, A. R., Goverde, R. M., & Howlett, P. G. (2020). Optimal running time supplement distribution in train schedules for energy-efficient train control. *Journal of Rail Transport Planning & Management*, *14*(100180). https://doi.org/https://doi.org/10.1016/j.jrtpm.2020.100180

Shahrokni, A., & Feldt, R. (2013). A systematic review of software robustness. *Information and Software Technology*, *55*(1), 1–17. https://doi.org/https://doi.org/10.1016/j.infsof.2012.06.002

Sharma, S., Henderson, J., & Ghosh, J. (2019). Certifai: Counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models. ArXiv. https://doi.org/https://doi.org/10.1145/3375627.3375812

Shi, R., Xu, X., Li, J., & Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Applied Soft Computing*, *109*, 107538. https://doi.org/https://doi.org/10.1016/j.asoc.2021.107538

Spanninger, T., Trivella, A., Büchel, B., & Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning & Management*, *22*(100312). https://doi.org/https://doi.org/10.1016/j.jrtpm.2022.100312

Swanborn, P. (2010). Case study research: What, why and how? *Balkan Medical Journal*, 1-192. https://doi.org/https://doi.org/10.4135/9781526485168

Taleongpong, P., Hu, S., Jiang, Z., Wu, C., Popo-Ola, S., & Han, K. (2020). Machine learning techniques to predict reactionary delays and other associated key performance indicators on British railway network. *Journal of Intelligent Transportation Systems*, *26*(3), 311–329. https://doi.org/https://doi.org/10.1080/15472450.2020.1858822

Tiong, K. Y., Ma, Z., & Palmqvist, C. W. (2023). A review of data-driven approaches to predict train delays. *Transportation Research Part C: Emerging Technologies*, *148*(104027). https://doi.org/https://doi.org/10.1016/j.trc.2023.104027

Trafikanalys. (2018a). *Bantrafik 2017*. https://www.trafa.se/globalassets/statistik/bantrafik/bantrafik/2017/statistikblad-bantrafik-2017.pdf

Trafikanalys. (2018b). *Punktlighet på järnväg 2017*. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2017/statistikblad-punktlighet-pa-jarnvag-2017.pdf
Trafikanalys. (2020). *Bantrafik 2019*. https://www.trafa.se/globalassets/statistik/bantrafik/bantrafik/2019/statistikblad-bantrafik-2019.pdf
Trafikanalys. (2021). *Punktlighet på järnväg 2020*. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2020/statistikblad-punktlighet-pa-jarnvag-2020.pdf
Trafikanalys. (2022a). *Bantrafik 2021*. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2022/punktlighet-pa-jarnvag-2021.pdf
Trafikanalys. (2022b). *Punktlighet på järnväg 2021*. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2022/punktlighet-pa-jarnvag-2021.pdf
Trafikanalys. (2023a). *Bantrafik 2022*. https://www.trafa.se/globalassets/statistik/bantrafik/bantrafik/2022/bantrafik-2022.pdf
Trafikanalys. (2023b). *Punktlighet på järnväg 2022*. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2023/punktlighet-pa-jarnvag-2022.pdf
Trafikverket. (2017). *Riktlinjer Täthet Mellan Tåg*. https://bransch.trafikverket.se/contentassets/45256f843f5c4292899ce5c63622c213/tathet-mellan-tag-tagplan-2021-v1.0.pdf
Trafikverket. (2023). *The capacity situation in 2022*. https://bransch.trafikverket.se/for-dig-i-branschen/jarnvag/Kapacitet/
UIC. (2004). *UIC code 406*.
Van Calster, B., Steyerberg, E. W., Wynants, L., & van Smeden, M. (2023). There is no such thing as a validated prediction model. *BMC Medicine*, *21*(1), Article 70. https://doi.org/https://doi.org/10.1186/s12916-023-02779-w
van Loon, R., Rietveld, P., & Brons, M. (2011). Travel-time reliability impacts on railway passenger demand: A revealed preference analysis. *Journal of Transport Geography*, *19*(4), 917–925. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2010.11.009
Veiseth, M., Olsson, N., & Saetermo, I. A. F. (2007, 2007/08/17). Infrastructure's influence on rail punctuality. *Urban Transport XIII: Urban Transport and the Environment in the 21st Century*. WIT Press. http://dx.doi.org/10.2495/ut070451
Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, *24*(5), 533–557. https://doi.org/https://doi.org/10.1080/0144164042000195072
Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, *43*, 3–19. https://doi.org/https://doi.org/10.1016/j.trc.2014.01.005

Vromans, M. J., Dekker, R., & Kroon, L. G. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research*, *172*(2), 647-665. https://doi.org/https://doi.org/10.1016/j.ejor.2004.10.010

Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, *28*(6), 815–829. https://doi.org/https://doi.org/10.1111/j.2005.0906-7590.04112.x

Wang, P., & Zhang, Q.-P. (2019). Train delay analysis and prediction based on big data fusion. *Transportation Safety and Environment*, *1*(1), 79–88. https://doi.org/https://doi.org/10.1093/tse/tdy001

Wang, Y., Wen, C., & Huang, P. (2021). Predicting the effectiveness of supplement time on delay recoveries: A support vector regression approach. *International Journal of Rail Transportation*, *10*(3), 375–392. https://doi.org/https://doi.org/10.1080/23248378.2021.1937355

Washington, S. Karlaftis, M. G., Mannering, F. & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. CRC press.

Weber, K., Zangl, M., Zahid, M. U., & Holzner, M. (2022). *Environmental Impact Evaluation of a European High Speed Railway Network along the 'European Silk Road'*. Vienna Institute for International Economic Studies (wiiw). https://wiiw.ac.at/environmental-impact-evaluation-of-a-european-high-speed-railway-network-along-the-european-silk-road-dlp-5837.pdf

Wen, C., Mou, W., Huang, P., & Li, Z. (2019). A predictive model of train delays on a railway line. *Journal of Forecasting*, *39*(3), 470–488. https://doi.org/https://doi.org/10.1002/for.2639

Yaghini, M., Khoshraftar, M. M., & Seyedabadi, M. (2012). Railway passenger train delay prediction via neural network model. *Journal of Advanced Transportation*, *47*(3), 355-368. https://doi.org/https://doi.org/10.1002/atr.193

Yuan, J., & Hansen, I. A. (2008). Closed form expression of optimal buffer times between scheduled trains at railway bottlenecks. *Proceedings of the 2008 11th international IEEE conference on intelligent transportation systems*. IEEE. https://ieeexplore.ieee.org/abstract/document/4732539/

Zakeri, G., & Olsson, N. O. E. (2017). Investigation of punctuality of local trains – The case of Oslo area. *Transportation Research Procedia*, *27*, 373–379. https://doi.org/https://doi.org/10.1016/j.trpro.2017.12.080

Zhang, D., Peng, Y., Zhang, Y., Wu, D., Wang, H., & Zhang, H. (2021). Train time delay prediction for high-speed train dispatching based on spatio-temporal graph convolutional network. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2434–2444. https://doi.org/https://doi.org/10.1109/tits.2021.3097064

Zhuang, H., Feng, L., Wen, C., Peng, Q., & Tang, Q. (2016). High-speed railway train timetable conflict prediction based on fuzzy temporal knowledge reasoning. *Engineering*, *2*(3), 366–373. https://doi.org/https://doi.org/10.1016/j.eng.2016.03.019