Deceleration behaviour of commuter heavy rail vehicles

Stochasticity, Causality, Correlation and Impact on Infrastructure Occupation

T.A. Grincell

Delft University of Technology

TU Delft

RICARDO

Deceleration behaviour of commuter heavy rail vehicles

by

T.A. Grincell

in partial fulfilment of the requirements for the degree of

Master of Science
in Civil Engineering - Transport & Planning

at the Delft University of Technology,
to be defended publicly on Thursday 16 May 2019 at 16:00.

Student number:	1504282	
Project duration:	December, 2017 – May, 2019	
Thesis committee:	Chairman:	Prof. dr. R.M.P. Goverde Delft University of Technology Faculty Civil Engineering Dep. Transport & Planning
	External Supervisor:	dr. ir. V.L. Markine Delft University of Technology Faculty Civil Engineering Dep. Road & Railway Engineering
	Daily Supervisor University:	dr. ir. N. Bešinović Delft University of Technology Faculty Civil Engineering Dep. Transport & Planning
	Daily Supervisor Company:	J.A.A. Welvaarts Ricardo Rail Netherlands

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

RICARDO

Preface

Here before you lies the research thesis that forms the partial fulfilment of the requirements for the degree of Master of Science in Civil Engineering, specialising in Transport & Planning, at the Delft University of Technology. The research has been carried out between December 2017 and May 2019 in cooperation with Ricardo Rail Netherlands, who have provided me with a workplace, knowledge from expert colleagues, contacts in the dutch rail industry and my daily supervision.

While high frequency train runs and optimising infrastructure occupation are increasingly popular topics used in the further development of heavy rail public transport, the focus so far has been on the improvement of the vehicle's acceleration and reducing the headway between vehicles. Investigating the deceleration behaviour of heavy rail vehicle, has opened my eyes to the diversity in deceleration behaviour, has given me insights to the possibilities for optimisations in railway systems and has taught me a lot about data acquisition, processing and analysis. I have always been interested in public transport systems and truly believe that Heavy and Light Rail systems are the future regarding mass passenger transportation.

I would like to thank all my colleagues at Ricardo Rail, the department 'Prestatie & Innovatie' (PI) within NS and my student colleagues at the university for your insights, interesting discussions and well needed conversational distractions.

Further, I would like to thank Ramon Lentink for being my contact and sponsor within NS. His efforts in guiding the process of data acquisition within NS, help in arranging the preliminary dataset used in the pilot and his interest in my research were greatly appreciated. I would like to thank Wilco Tielman for being my contact within ProRail. As well as for his support in helping me understand the requested data, his help checking the quality or availability of the data when those questions arose after observing the data myself, and his interest in my research and analysis approach while sharing some extra insights his research and projects.

Of course, I would like to thank my entire committee and my family. I have greatly appreciated the committee's valuable insights, discussion and guidance given, along with your patience in dealing with my stubborn moments. To my family who have had seemingly endless patience with me during the entirety of my study career and who have supported me unconditionally.

And lastly, I would like to thank you, the reader, for your interest in the subject matter and your curiosity towards this research. I hope you find it an interesting and enjoyable read.

T.A. Grincell
Delft, May 2019

Summary

When travelling between stations, the observant passenger may have noticed that no two train runs are the same, caused by your departure or arrival at stations at different times or the varying length of time for getting from A to B. The most obvious causality from these differences are delays or restrictive track signalling. However, even when the conditions appear to be very similar and no obvious restricted track signalling is present, these differences between the train runs persist.

The Dutch Train Operator Company (TOC) and Infrastructure Manager (IM), being NS and ProRail respectively, have generally observed there to be a spread in the running times between all stations. While investigations and research have been done regarding the acceleration or maximum velocity of the heavy rail commuter vehicles, of which some tested the stochastic nature of their general operational behaviour, the deceleration behaviour of these vehicles were at most described in a more general sense while either reconstructing the entire speed profiles of rail vehicles in available scientific papers (Bešinović et al., 2013a,b,c; Medeossi et al., 2011) or investigated as a byproduct of other researches regarding technology and safety. Current timetabling tools or simulation software account for a single defined deceleration behaviour based on static conditions and only consider the braking rate as a constant coefficient, describing a sort of (near) binary state of on/off in brake application, of which some allow for variation of the simulated or estimated braking rate, rather than allowing for a dynamic or non-uniform development of the braking rate and different deceleration behaviour to be applied to account for different driver mentality and behaviour. This, however, assumes nominal vehicle characteristics and operating conditions (i.e. conditions relevant for train runs, based on stated track / vehicle design specifications and vehicle characteristics) and a uniform behaviour from all train drivers. Considering the human aspect and widely differing external conditions (e.g. weather, timeliness of train runs), how valid would this assumption be when reviewing the realised deceleration behaviour? This has further raised the question to what extent the impact is of this stochastic behaviour on infrastructure occupation, especially around stations.

This research has the objective to empirically determine the perceived stochastic nature of the deceleration regimes and to determine the impact it has on infrastructure occupation within the network corridor, which was approached by posing the research question:

How does the stochastic behaviour, observed in the deceleration approach of a planned station stop, impact the infrastructure occupation in the network corridor between stations?

Answering this research question is achieved by the development of a data-driven reconstruction model to estimate the speed profiles of realised train runs that elaborates on the deceleration regimes in a more dynamic and generalised (i.e. able to be applied to any rail network corridor) manner to provide a more detailed description of the realised deceleration behaviour. This research introduces a conceptual framework for the reconstruction of the deceleration behaviour of heavy rail vehicles. The general structure and concept approach are inspired by Bešinović et al. (2013a,b,c) and Medeossi et al. (2011).

The proposed framework differentiates from its inspiration by it's interest and focus on the deceleration behaviour within the proposed reconstruction model, aptly named Deceleration Reconstruction (DR) model, instead of reconstructing complete velocity profiles. This is done through the implementation of sub-regimes (i.e. smaller driving regimes describing one of four driving states of a heavy rail vehicle) and deceleration regime profiles defined by a combination of said sub-regimes. These sub-regimes are subject to the rail sector defined differential equation, describing Newton's second law of motion, with the vehicle resistance expressed by the Davis equation (Davis, 1926). This equation is further expanded

in this research to accommodate the testing of non-uniform braking behaviour by writing the braking rate as a function of either velocity or time, inspired by Maurya and Bokare (2012). The framework incorporates a statistical analysis because, besides the realised train trajectories and corresponding parameters, the statistical distributions of these parameters and related performance indicators (PI) are considered the end products for this research.

Furthermore, this research introduces the concept of 'Data Fusion' for all the realised location tracking data to be used in the DR model. The datasets used from the different sources (i.e. the single source of train describer data and 1 to 2 sources of GPS data) are formatted, corrected and derivable data is created to unify, expand and fill in the gaps for the provided data sources. After the aligning the different location tracking data sources, they are fused in which the strengths of all the different sources are inherited while their individual weaknesses are compensated. Once fused, the dataset of individual train runs are filtered to discard the undesired, hindered or invalid train runs before feeding the DR model with the fused location tracking data of the selected realised train runs.

The DR model has several layers of nested for loops in its programming, testing each predefined deceleration regime profile and braking variant for each station approach in the corridor of a single realised train run. Each combination of deceleration regime profile and braking variant has their own specific β-vector (i.e. a list or sequence of variable parameters describing the vehicle's behaviour within the DR model) for the optimisation algorithm, whose structure is developed automatically in the dataset and coefficient preparations preceding the actual optimisation of the minimisation problem. The minimisation problem is defined as minimising the sum of absolute errors of the surface areas under the speed profile between each pair of data points , with the surface area under the realised speed profile defined by the average velocity multiplied by the distance between the two data points and the area under the estimated speed profile defined by a double integral over distance of the velocity differential equation describing the estimated behaviour for a specific sub-regime. The optimisation algorithm used on this non-linear minimisation problem, is a Genetic Algorithm (GA) method. For this research, a customised GA is developed, which is dubbed "Elitism with Randomised Population Migration and Diminishing Mutation". This is an algorithm which retains a small pool of best fitting solution vectors for both the next generation and for developing the 'offspring' vectors through cross-over and mutation with a diminishing mutation range, while refreshing and maintaining the general population size through the migration of a group of completely randomised vectors.

To validate the analysis methodology, a case study was developed over the corridor 's Hertogenbosch - Utrecht (Ht-Ut) for the (inter-)regional commuter line series 6900 during the year 2017 for the rolling stock type Sprinter Light Train (SLT). Both location tracking sources (i.e. enriched GPS (MTPS) and train describer data(TROTS)), infrastructure data, signalling data and the network timetable for this train series was provided by Dutch IM ProRail and Dutch TOC NS. The nominal vehicle characteristics, used as references in the performance indicators (e.g. running times, deceleration loss times, braking rates, track section occupation duration) and as guides and bounds in the optimisation algorithm, were provided by TU Delft and Ricardo Rail Nederland.

The results of the analyses done in this research, have empirically shown the existence of the stochastic deceleration behaviour through the stochasticity of the performance indicators and has shown the existence of a strong and linear correlation between the already known spread in realised running times and that of deceleration loss times (Figure 1). The results have also shown the impact that this stochasticity has on infrastructure occupation, which in this research is simplified and described by track section occupation duration. It has shown performance impacts, relative to the minimum running profile, of up to 1.5 to 2 times the occupation duration and inversely dropping the infrastructure occupation to anywhere from two thirds to half of the potential track section capacity (Figure 2a).

The DR model has provided more detail to the estimated deceleration behaviour and has shown that finding the most prominent realised deceleration regime behaviour is possible

with this analysis methodology and the currently available data sources. For example, it has shown the realised data for station Geldermalsen best fits the deceleration regime profiles with two braking regimes with constant braking coefficient. It has also shown that describing the braking rate as a non-uniform function is a very plausible alternative to the simplified approach of defining the braking rate as a constant coefficient and merits further investigation, as many of the best fitting profiles returned with a non-uniform braking variant. Observing the estimated braking rates of the best fitted profiles and variants for the tested train runs, the distributions show that there is a significant gap between the expected braking rate of 0.8 m/s^2 and the estimated braking rates.In the majority of the estimated deceleration regimes, a more conservative behaviour is observed relative to the expected deceleration behaviour.In case of station Geldermalsen in Figure 2b, this means a mean of estimated braking rates ranging from approx. 0.45 to 0.65 m/s^2.

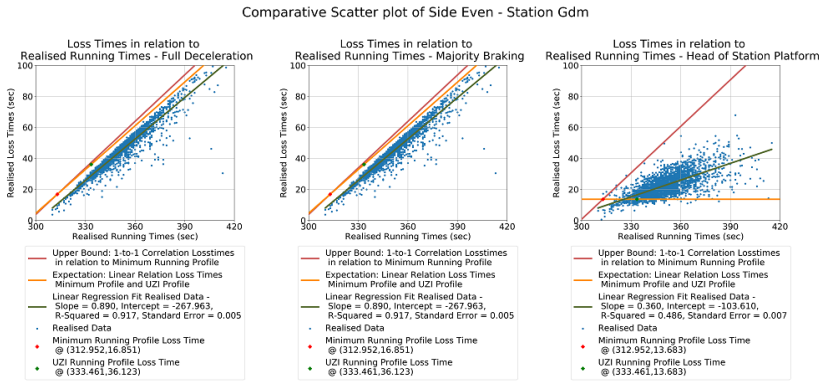
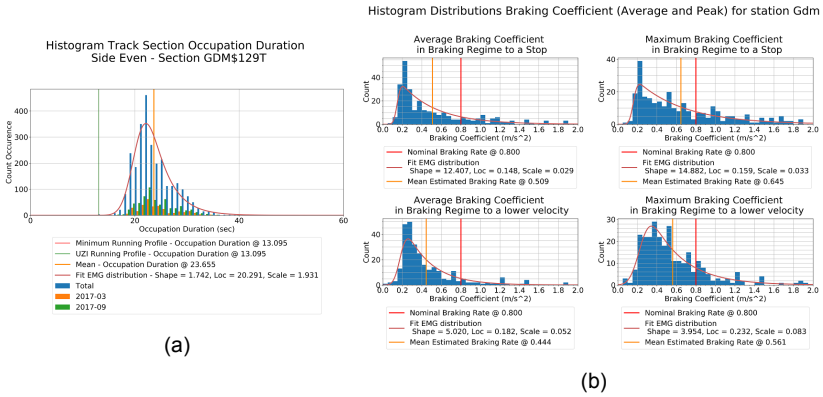


Figure 1: Scatter plot correlation realised running times versus deceleration loss times - station Geldermalsen (Gdm)



(a) (b)

Figure 2: Example of the results from empirical analysis and DR model:
(a) Histogram of the realised occupation durations for track section GDM$129T (see full size Figure K.1a)
(b) Histogram plots of the estimated braking rates for station Geldermalsen (Gdm) (see full size Figure H.1)

These results can be used to improve scheduling tools and microscopic models, by using improved estimations of braking rates and deceleration regimes, and by providing a more refined and dynamic approach of the deceleration regime within these systems. The 'Data Fusion' process has shown great potential in reinforcing the strengths and compensating for the weaknesses inherent to the used data sources, and providing robustness in data availability while providing a more refined location tracking within track sections, which merits further development of this concept. Further investigation is required to fully understand the causality behind the deceleration behaviour in more specific detail, such as the effects of station platform design or open-track signalling placement, in order to account for it within scheduling tools or to provide an environment to train operators to adhere to the expected deceleration regime. Further research is required to provide a more definitive answer to the implemented braking regime and its non-uniformity the use of more refined data, different data sources and further improvements to the DR model.

Glossary

Term	Description
ATB-EG	Automatische Trein Beveiliging - Eerste Generatie, The Dutch national ATP applied to the main heavy rail network.
ATB-Vv	Automatische Trein Beveiliging - Verbeterde Versie,
ATO	Automatic Train Operation
Assistive Force	Opposite to Resistive Forces, Applies its force in the intended travel direction.
BDA	Big Data Analysis
Big Data	Data characterised by the 6 V's, namely Volume, Variety, Velocity, Veracity, Value and Variability.
Coasting	A driving regime in which both the tractive effort and the brakes of the rail vehicle are turned off, decelerating purely on the resistance efforts.
Cruising	A driving regime in which the tractive effort is adjusted to match the resistance efforts, allowing the vehicle to run at a constant velocity.
Cycle Time	Duration or period over which the scheduling pattern of a cyclic/periodic timetable is developed.
DAS	Driver Advisory System
Deceleration Regime	A regime using sub-regimes, describing the smaller regimes in which coasting, cruising and braking are applied within the same intention of deceleration.
Driving Regime	A control operation applied to the vehicle with regards to train movement.
Epoch Time	A method of counting time in Unix and later other computer systems, expressed as an 32-bit integer, counting the seconds since Epoch, which corresponds with the date 01-01-1970 at 00:00:00. A float or decimal can be used to express time more precisely in terms of microseconds.
GPS	Global Positioning System, Usually referring to the general implementation of Satelite-based Location Tracking.
IC	InterCity, Trains that provide a more national level rail connection
IM	Infrastructure Manager

Jerk	Rate of change of acceleration / deceleration. Second order derivative of velocity. Usually described as m/s^3, but in case of velocity dependency, such as Equation (2.9) – d_2, as $1/s$.
NS	Nederlandse Spoorwegen, The Dutch National TOC
Nominal	Expected or designed value within expected margin of error
ProRail	The Dutch National IM
RMSE	Root Mean Squared Error
RSSLI	Rolling Stock, Speed Limit and Infrastructure, refers to the infrastructure model used in this research.
SGMm	Stadsgewestelijk Materieel gemoderniseerd, the retrofitted and modernised version of the older rolling stock type used on the Dutch rail network.
SLT	Sprinter Light Train, A commuter heavy rail vehicle or Sprinter type train.
Sprinter	Trains that provide a more (inter-)regional level rail connection
Sub-regime	Smaller driving regimes within the 2 of the 4 general driving regimes characterised by an active attempt to alter the current velocity (i.e. Acceleration and Braking), with a nuanced difference between neighbouring sub-regimes of the same type(e.g. two braking sub-regimes with different braking coefficients).
TNV	TreinNummerVolgsysteem, Train Number Following system, the predecessor to the TROTS system
TOC	Train Operating Company
Timetable Day	The day of a timetable in the Dutch railway sector run from 2am to 2am the next day (02:00 - 02:00), unlike the regular day beginning and end at 12am (00:00 - 00:00).
Train Describer Data	The data gathered from infrastructure sensors, recording track section occupation and release times, train description steps, signalling states (stop/go), switch states (left/right).
Train Event Recorder Data	The data of events recorded from within the rolling stock, measuring train positions, velocities, traction and braking applications.
UZI	Universeel Zuinig rijden Idee, translates to Universal Efficient driving Idea and is an energy efficiency plan, describing a driving regime in which the driver is expected to apply maximum acceleration, minimal braking and applies coasting for as long as the time allowances would allow.

List of Tables

List of Figures

Contents

1

Introduction

When travelling between stations, the observant passenger may have noticed that no two train runs are the same, caused by your departure or arrival at stations at different times or the varying length of time for getting from A to B. The most obvious causality from these differences are delays or restrictive track signalling. However, even when the conditions appear to be very similar and no obvious restricted track signalling is present, these differences between the train runs persist.

This research investigates the stochastic deceleration behaviour of commuter heavy rail vehicles, in what stochastic form (i.e. distribution shape and parameters) this is perceived and what the impact of it on infrastructure occupation. Even the assumptions, such as the braking regime consists of a single and constant braking rate and deceleration approach to a station consists of a single braking regime and sometimes preceded by a coasting regime, are challenged in this research and are to be tested for their validity. This research develops a reconstruction model based on realised traffic data to obtain and analyse the realised deceleration behaviour in a data driven approach.

1.1. Situation Definition

The Dutch Train Operator Company (TOC) and Infrastructure Manager (IM), being NS and ProRail respectively, have generally observed there to be a spread in the running times between all stations. While investigations and research have been done regarding the acceleration or maximum velocity of the heavy rail commuter vehicles, of which some tested the stochastic nature of their general operational behaviour, the deceleration behaviour of these vehicles were at most described in a more general sense while either reconstructing the entire speed profiles of rail vehicles in available scientific papers (Bešinović et al., 2013a,b,c; Medeossi et al., 2011) or investigated as a byproduct of other researches regarding technology and safety. Current timetabling tools or simulation software account for a single defined deceleration behaviour based on static conditions and only consider the braking rate as a constant coefficient, describing a sort of (near) binary state of on/off in brake application, of which some allow for variation of the simulated or estimated braking rate, rather than allowing for a dynamic or non-uniform development of the braking rate and different deceleration behaviour to be applied to account for different driver mentality and behaviour. This, however, assumes nominal vehicle characteristics and operating conditions (i.e. conditions relevant for train runs, based on stated track / vehicle design specifications and vehicle characteristics) and a uniform behaviour from all train drivers. Considering the human aspect and widely differing external conditions (e.g. weather, timeliness of train runs), how valid would this assumption be when reviewing the realised deceleration behaviour? This has further raised the question to what extent the impact is of this stochastic behaviour on infrastructure occupation, especially around stations.

1.2. Objective

The primary objective of this research is to empirically determine the impact of the perceived stochastic nature of the deceleration driving regimes, as presented in the work of Bešinović et al. (2013a,b,c), Medeossi et al. (2011) and Tielman (2015), has on the infrastructure occupation.

To evaluate the realised deceleration behaviour, a model has to be developed to run the analysis. This model has its focus on the deceleration regimes, with a more detailed analysis approach to determine the existence of different sub-regimes. Sub-regimes, in this research, are defined as smaller regimes within the general driving regimes, with a nuanced difference between neighbouring sub-regimes of the same type(e.g. two braking sub-regimes with different braking coefficients).

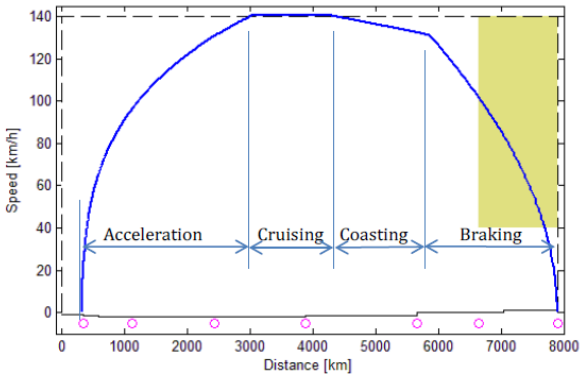


Figure 1.1: A plot with a general velocity profile with the four main operational driving regimes Bešinović et al. (2013c)

The secondary objectives that arise are the causality and correlations related to the stochastic nature of the deceleration behaviour and what the sensitivity/impact of these influential factors is.

These objectives are set to investigate the existence and causality of the stochastic behaviour in the deceleration driving regime of a commuter heavy rail vehicle, the impact these influential factors have on the stochasticity and the impact this stochastic behaviour has on the network through means of infrastructure occupation.

As a result, a better understanding is expected to be reached on the deceleration behaviour of heavy rail vehicles and their drivers. This will provide some initial findings to help improve the relation between timetabling and the stochastic nature of the vehicle's deceleration behaviour, and to introduce the development and implementation of a different approach in describing the vehicle's deceleration behaviour when calculating running times for timetable development.

1.3. Research Questions

To aid the focus of the research, the approach of the research will be to answer a main question that will cover the extent of this research, with several supporting smaller questions that are more specific on the different aspects of the main question and will lead to a more detailed and in-depth answer.

Main Research Question

To understand the role of a driver's deceleration behaviour in current climate of concerns for the Infrastructure Manager(IM) and Train Operating Company (TOC), the impact of the stochasticity on the infrastructure occupation will be investigated. This leads to the following research questions:

How does the stochastic behaviour, observed in the deceleration approach of a planned station stop, impact the infrastructure occupation in the network corridor between stations?

Research Sub Questions

The research sub questions in support of answering the main question are as follows:

1. *What vehicle parameters are required to fully describe the behaviour in a deceleration approach of a planned stop?*
2. *How would the collection and processing of the required data for the behaviour analysis be defined?*
3. *How would the data-driven reconstruction model be defined to allow for the fusion of different database sources and to analyse the realised data to reconstruct the deceleration behaviour of vehicles in detail?*
4. *What are the distribution types and stochastic values of the realised deceleration behaviour of commuter heavy rail vehicles?*
5. *Could any sub-regimes be determined in the realised deceleration behaviour?*
6. *How do the parameter distributions of the realised deceleration behaviour compare to the nominal parameters of the expected deceleration behaviour?*
7. *What are the causalities for the stochastic variations found in a vehicle's deceleration behaviour?*

1.4. Research Approach

In order to achieve the objective of this research and to provide a platform to answer the stated research questions, an approach is made and summarised by:

- Present a data-driven, reconstruction model that will reconstruct the realised velocity profiles and estimate the realised deceleration regimes in more detail through means of implementing sub-regimes.
- Present a means of combining and processing both vehicle-based and infrastructure-based location tracking datasets.
- Empirically and visually provide evidence regarding the stochasticity of the deceleration driving regime in commuter heavy rail vehicles.
- Empirically and visually provide evidence regarding the causality and its impact and sensitivity, affecting the stochastic distributions of the realised deceleration behaviour.

The research approach and analysis methodology is elaborated on in Chapter 4.

1.5. Scope & Case Study

The scope for this research is limited by the selection and boundaries set below in order to keep the complexity of this research manageable and keep the focus of the research on its objectives.

Driving Regimes and Operational Scenario

The focus of this research is on the deceleration behaviour of commuter heavy rail vehicles. There are several scenarios in which any rail vehicle would decelerate. This could range from decreasing velocity due to (dynamic) speed limits, delay propagation from a preceding train or having to come to a complete stop for either a red rail signal or entering a planned stop at a station platform.

In this research, the choice was made to focus on the planned stops at a station platform from conflict-free train runs. This was chosen to eliminate driver behaviours constrained by traffic conflicts and to observe the driving behaviour under normal operational conditions. This would include the delayed train runs that are not constrained in their deceleration approaches.

The deceleration of a vehicle has many variables that are suspected to influence the actual physics of decelerating a long and heavy vehicle to a stop. What makes this operation more complex in comparison to the other driving regimes, is that it has its target location as a fixed point, with a relative small margin of difference, at the end of the motion instead of having the target/goal of the motion be a floating point, like with acceleration regimes. In addition, the deceleration target point is out of sight until the last few seconds in the station approach, leaving very little room for adjusting in an over-assertive station approach.

However, why examine the entire deceleration approach and not just the braking regime? The braking regime and braking coefficient, in considering driving behaviour, loses context if only a part of the deceleration behaviour is taken into account. One station approach might have very short braking times and high braking coefficients, but could still take a long time to complete its approach into a station, leading to higher block reservation times and possibly lose a lot of potential infrastructure occupation. While another approach might not be as assertive in their braking and taking longer in their braking times, but have a shorter overall approach time and most likely be more efficient for the infrastructure occupation near the switches near the station platforms. Therefore, to maintain the context with these station approaches, the entire deceleration behaviour leading to the planned station stop are considered.

Rolling Stock

For this research, the scope will focus on the commuter heavy rail vehicles (e.g. SLT, SGMm) as those are expected to yield the more significant impact from this research due to a higher density in station stops (i.e. more planned station stops per corridor compared to intercity or shorter distance between station stops). These higher stop rates will provide more stops over a relatively short corridor in which we can guarantee the same driver operating the rail vehicle. This would allow the testing for correlations while assuming similar driver mentality. This will help eliminate contamination in determining correlation between other aspects, such as different station/platform designs.

Another consequence of the higher stopping rates is that the commuter trains have a lower average operational velocity and longer station-station running times as well as a more pronounced (i.e. compounded) effect of the possible stochastic nature on the vehicle operational behaviour, causing the infrastructure occupation to be impacted relatively more severe than intercity trains.

The decision to narrow the choice of the rolling stock down to a single vehicle type, is to eliminate different vehicle characteristics caused by the design of the vehicle and to focus

the correlations and statistical results on driver behaviour and external influences (e.g. weather, vehicle or track conditions, vehicle delay state).

For the analysis the nominal vehicle characteristics will be used to define the default vehicle resistance values and coefficients. There are, however, not considered fixed values due to the intra-vehicle-type mechanical stochasticity (i.e. differences between vehicles from the same vehicle type).

Infrastructure

To create a clear distinction between external influential factors (e.g. weather, vehicle or track conditions, vehicle delay state) and personal driver characteristics, this research has chosen to focus its investigation a single corridor between two major cities with plenty of stations and different platform designs. With this the assumption can be made, with a high degree of certainty, that there were no driver changes along the corridor and have plenty of station stops to understand the correlation between different stations and differentiate it from the correlation of driver deceleration behaviour. This corridor will also have a high intensity of rail vehicles. This is to test the stochasticity against a lean supplemented train schedule and what the impact is on infrastructure occupation.

Time Period

The selection of the time period is considered a balance between computational time and for the statistical analysis to provide statistical significant results on the deceleration behaviour and any correlations drawn in this research. The decision is made to run the main analysis of a time window of an entire timetable year.

Realised Data

Due to the heritage of the proposed analysis model, the realised data is used from both vehicle (i.e. Vehicle GPS: Location-Time, Velocity) and infrastructure (i.e. Train Describer Data: track section occupation and release times, section signalling) for the analysis. The analysis method for this research and the required data will be discussed in Chapter 4 of this research.

Case Study

To further specify on the different aspects mentioned above in defining the scope for this research, the following can be summarised on the previous scope aspects to define the case study for this research:

Driving Regimes and Operational Scenario:

The deceleration regimes Coasting and Braking were examined to provide the complete context of the deceleration behaviour for the planned station stop. Only the conflict-free train runs are used for the analysis.

Rolling Stock - SLT:

The vehicle type chosen for the case study was the Sprinter Light Train or SLT for the assumed more modern on-board control and logging system. For the default vehicle parameters, the nominal values of the vehicle characteristics (e.g. resistance coefficients) for an SLT-10 were used.

Infrastructure - ”Ut - Ht”:

A few different corridors were examined, such as ”The Hague - Rotterdam” and ”Amsterdam - Utrecht”. In the end, the decision was made to use the ”A2 South Corridor”/”Utrecht(Ut) - ’s Hertogenbosch(Ht)”.

Time Period - timetable year 2017:

The timetable year of 2017 was selected, as it was the most recent available data source to eliminate any contamination of recently implemented operational or structural changes. This was also the year in which trials with high frequency train traffic was tested on the chosen corridor. (NS, 2016; ProRail, 2016, 2018; SpoorPro, 2018) The SLT Sprinter trains were not part of the current roll-out of the high frequency train schedule, but it is assumed that, due to the interaction on mixed traffic corridor, the commuter rail vehicles have been affecting and were affected by the high frequency timetable. These test days make for an interesting insight in the effects of implementation of High Frequency timetabling on the driver’s deceleration behaviour and the effects of stochastic behaviour on the effectivity and sensitivity of High Frequency timetabling.

Train Run Series - Series 6900:

When the considerations of rolling stock and infrastructure were being made, the idea of using a specific train run series arose to provide a more consistent dataset. Train run series 6900 of the timetable year 2017 seemed to align with the chosen rolling stock(i.e. the entire train run series was run with SLT vehicles) and the demanded aspects on the case study infrastructure, helping solidify the infrastructure choice.

2

Literature Study

The summarised reviews of different literature sources are grouped by their intent or influence in this research.

2.1. Literature Reconstruction Model & Operational Behaviour

This section holds the summary of the different mainstream approaches in reconstructing the realised velocity profiles and therefore train operator and vehicle behaviour. This is then followed with the literature used to support the analysis methodology of this research.

2.1.1. Basics of modelling train movements

Before discussing the works, research and papers related to the reconstruction model, a short description of the two main mathematical approaches used in calculating a vehicle's velocity is given, along with a couple relevant pros and cons.

Kinematics

One of the main mathematical approaches is through the use of kinematics. The related equations are are all based on four variables(i.e. velocity(split into initial and end), time, distance, acceleration) and require the use of three to calculate a fourth. Albrecht et al. (2006, 2010a,b) used this method to describe the vehicle motion, with parameters calibrated by track occupation data, which was collected through the use of train describer systems (Daamen et al., 2009; Goverde and Meng, 2011; Kecman and Goverde, 2012a,b). The general kinematic equations used, are:

$$\begin{aligned} s &= v_i t + 0.5at^2 \\ v_f &= v_i + at \\ v_f^2 &= v_i^2 + 2as \\ s &= 0.5(v_i + v_f)t \end{aligned} \tag{2.1}$$

In which s is distance, t is time, a is acceleration or deceleration when negative and v_i and v_f describe the vehicle's velocity in its respectively initial and final state.

Dynamics

The other main mathematical approach is through the use of dynamics, in particular the use of Newton's dynamic motion equations, expressed as Equation (2.2). The equation is a (partial) differential equation with coefficients and variables reflecting the different forces acting upon and reacting to an entity in the form of deformation or motion, with the train being modelled as a mass point. This assumption is widely accepted and used in previous research and practice. (Hansen and Pachl, 2008) This dynamics equation can be formally expressed as Equation (2.3).

$$\begin{aligned} p &= p(t) = mv(t) \\ F &= \frac{dp}{dt} \\ F &= m\frac{dv}{dt} \end{aligned} \tag{2.2}$$

$$f_t(v) - r(v) = f_s(v) = m\frac{dv}{dt} \tag{2.3}$$

With p defined as the vehicle's momentum, F as the resultant force of the vehicle, v as vehicle velocity, t as time, m as vehicle mass(with rotational mass factor implicitly accounted for), $f_t(v)$ as velocity dependant tractive force, $r(v)$ velocity dependent resistive force and $f_s(v)$ being the surplus force exerted on or by the rail vehicle. The resistive force $r(v)$ can be further broken down into the rail vehicle resistive forces $r_v(v)$, which is usually based on the Davis Equation (Davis, 1926), and the resistive forces related to the track geometry, gradient (r_g) and curvature (r_c) (Trani, 2018). Each of these functions can be further described by the following equations:

$$\begin{aligned} f_t(v) &= \min(c_0 + c_1 v; c_2/v) \quad \text{(Assumed)} \\ r_v(v) &= r_0 + r_1 v + r_2 v^2 \\ r_c &= 0.01\frac{k}{rad_c} \\ r_g &= \frac{g \cdot grad}{1000} \end{aligned} \tag{2.4}$$

With v describing the rail vehicle velocity, g describing gravitational acceleration (9.81 m/s^2), k being a dimensionless variable describing the impact of the curvature radius (rad_c, expressed in m) on the curvature resistive force (r_c) and $grad$ (expressed as proportion) describing the longitudinal gradient of the track section. Lastly, c_0, c_1 and c_2 describe the rail vehicle's tractive coefficients and r_0, r_1 and r_2 describe the rail vehicle's resistive coefficients according to the applied Davis Equation. Furthermore, the tractive and resistive force equations have already been adjusted to be mass-specific for the purpose of this research, either explicitly through mass variables or implicitly through the equation coefficients, and are therefore expressed as their related acceleration.

Kinematic vs Dynamics

Using the kinematic equations assumes a constant acceleration or deceleration during the driving (sub-)regime or considered track section. The equations are also trajectory dependent and can therefore calibration not be used for a different train run even if the rolling stock is considered identical in vehicle characteristics. Furthermore, kinematic equations only implicitly account for the different forces, making it harder to distinguish between the different aspects of the vehicle's motion(e.g. driving behaviour, infrastructure/vehicle design characteristics, weather/track/vehicle conditions). The vehicle traction and resistance forces, as seen in Equation (2.4) are also, for the most part, velocity dependent and

will therefore change with differences in velocity, making it inherently variable.

Dynamics Equation allows differentiation of acceleration / resistance coefficients. The dynamics equation and its coefficients are trajectory independent by accounting for velocity differences and making the coefficients generalised for both vehicle and infrastructure characteristics. Therefore previous calibration of the coefficients can be used on different train runs with similar vehicle and infrastructure characteristics. Due to these differences, the proposed reconstruction model described in this research, just like the research elaborated on below, uses the dynamics equation Equation (2.2) adjusted to describe the vehicle velocity in distance instead of time, as seen later in Equation (2.6).

2.1.2. Using stochastic blocking times to improve timetable planning. – (Medeossi et al., 2011)

Medeossi et al. (2011) developed a method of generating and evaluation of blocking timetables with the running times based on stochastic behaviour of heavy rail vehicles, as other research (Albrecht et al., 2006; Goverde, 2005) had done. This method was calibrated with realised train runs in the corridor Trieste - Udine in Northern Italy.

Due to the inaccessibility/unavailability of high quality track describer data or archived digital train recorder data from the Italian IM and TOC, Medeossi et al. (2011) used the GPS location tracking of the vehicles and then linked and adjusted them onto the underlying infrastructure of the corridor. He was able to achieve usuable results and developed his stochastic approach to blocking schedules.

Medeossi et al. (2011) used the kinematic motion equations, in which he opted for the use of performance coefficients and fixed, nominal vehicle characteristics. Medeossi et al. (2011) had to therefore define 'driving regime'-specific functions, instead of applying a generalised function for velocity profile description.

The minimisation problem defined for the optimisation of the model's estimation, applied a mean squared error calculation on the velocity component of the different GPS points with the model's estimated velocity, as seen in Equation (2.5). The optimisation method used to solve the minisation problem was applying an simulated annealing algorithm.

$$\begin{aligned} &\text{Minimise} \quad \sum_{i=1}^{N} (v_{GPS}(i) - v_{calc}(i))^2 \\ &\text{Subject to:} \quad v_c = f(p_a, \ldots, p_{br}, R, F, m, BWP) \end{aligned} \tag{2.5}$$

The equation used to calculate the velocity, was 'a conventional motion equation' and variable performance parameters and pre-determined (or 'deterministic') parameters related to vehicle, infrastructure and timetable. (i.e. resistance R, tractive effort F, mass m and braked weight percentage (BWP))

The differences with the later research of Bešinović et al. (2013b), was the use of a kinematic motion function with fixed, nominal vehicle characteristics and a general performance factor, compared to a generalised differential equation describing the velocity difference over distance with bounded variable vehicle characteristics. This meant 4 different 'driving regime'-specific functions. This, along with the general performance coefficients, made it an inflexible system, as it limited the degrees of freedom in the regime by fixing the individual vehicle characteristics.

This research, however, has provided insight in the existence of stochastic driving behaviour and has shown it to be possible to calculate an estimation of the realised driving behaviour from GPS location tracking data.

2.1.3. Simulation Based Optimisation Model – (Bešinović et al., 2013a,b,c)

Bešinović et al. (2013b) introduced the framework of using a simulation based model with realised train describer data(in this case track occupation and release times linked to vehicle ID) to accurately describe the vehicle coefficients for timetable implementation. In Bešinović et al. (2013a) the model was further refined through a sensitivity analysis, concluding with Bešinović et al. (2013c) validating the simulations with realised vehicle GPS data. These papers produced 'Chapter 7 - Calibration of Train Speed Profiles' of the PhD research by Bešinović (2017) and had the purpose of creating a reliable method of determining running times and velocity profiles for accurate estimation and timetable development in dense railway networks. The proposed simulation model was based on the Newtonian Dynamics equations of motion. In particular, the dynamics equation used in the model was written as a differential equation of velocity over distance, as described in Equation (2.6).

$$\frac{dv}{ds} = \frac{f_t(v) - r_v(v) - b - r_g - r_c}{v} \tag{2.6}$$

In which b is the braking coefficient, v the vehicle velocity, gradient resistance r_g, curvature resistance r_c and the tractive effort ($f_t(v)$) and vehicle resistance ($r_v(v)$) described by the following equations:

$$\begin{aligned} f_t(v) &= \min(c_0 + c_1 v; c_2/v) \\ r_v(v) &= r_0 + r_1 v + r_2 v^2 \end{aligned} \tag{2.7}$$

In Bešinović et al. (2013b), the sensitivity analysis of the parameters showed that c_1 and r_1 have a negligible variance and could therefore be assumed fixed values. They set c_1 to zero, due to the insignificant relevance, and r_1 to the default value of the rolling stock characteristics, provided by the TOC.

The optimisation problem as proposed by Bešinović et al. (2013b) was the minimisation of the sum of absolute differences between observed and simulated running times between track sections and can be seen in Equation (2.8). It was minimised with an heuristic solving algorithm, namely a Genetic Algorithm.

$$\begin{aligned} \text{Minimise} \quad & \sum_{i \in N} |t_i{}^{observed} - t_i{}^{simulated}| \\ \text{Subject to:} \quad & \frac{dv}{ds} = \frac{f_t(v) - r_v(v) - b - r_g - r_c}{v} \\ & \frac{dt}{ds} = \frac{1}{v} \\ & c_i \in [c_i^{lb}, c_i^{ub}], \text{for } i = 0, 2 \\ & r_i \in [r_i^{lb}, r_i^{ub}], \text{for } i = 0, 2 \\ & b_{stop} \in [b_{stop}^{lb}, b_{stop}^{ub}] \\ & \theta_{cruising}(s) \in [\theta_{cruising}^{lb}(s), \theta_{cruising}^{ub}(s)] \\ & \theta_{coasting}(s) \in [\theta_{coasting}^{lb}(s), \theta_{coasting}^{ub}(s)] \\ & v(0) = v_0 = 0, v(N) = v_{end} = 0 \end{aligned} \tag{2.8}$$

2.1.4. An Agent-based Approach to Simulating Train Driver Behaviour – (Tielman, 2015)

Tielman (2015) developed a methodology and optimisation algorithm in order to improve on the validity of the microscopic simulation model FRISO, used by ProRail, through the use of agent based train operators. These agents that he developed and trained through machine learning, concluded with an improved result in FRISO simulating train operator behaviour, when comparing the agent based model and the FRISO base model to realised traffic data.

However it was not perfect as some comparisons resulted in still showing a noticable, but improved, deviation in results between realised and calculated behaviour. Much like the functions used by Medeossi et al. (2011), Tielman was limited by the 'linear' equations used in ProRail's FRISO and was therefore limited to constant vehicle parameters per driving regime (e.g. the use of single and constant braking and acceleration coefficients within the same driving regime). This was dealt with by modelling the 'final approach to a planned stop or a red signal' into two different braking regimes and if necessary filled in a coasting / cruising regime. (Tielman, 2015, p.49)

When developing his agents, he investigated the Driving time Until Emergency braking curve (DUE). This was so that the simulated trains would not start their braking regime too early. The gamma distribution was drawn as a reference for his random draw of initial DUE used, as the initial DUE was deemed stochastic and independent from the initial vehicle velocity.

There were other correlations drawn (e.g. between the initial velocity and initial distance between start and stop of a deceleration approach to a stop) that would further provide conclusions, that would imply that stochasticity could be found even in the on-set times and duration of different driving regimes with different scenarios and signalling.

2.1.5. Study of Deceleration Behaviour of Different Vehicle Types – (Maurya and Bokare, 2012)

Due to the limited research material publicly available on the deceleration behaviour of rail vehicles, an exploration into different transportation or traffic research fields was deemed necessary. The research fields covering road vehicles hold significantly more research on deceleration behaviour, with the paper of Maurya and Bokare (2012) particularly interesting and relevant for this research.

This research focused on determining the different deceleration behaviour for distinctly different vehicle types(i.e. Trucks, cars, motorised two and three-wheeler). It differentiated itself from the majority of past research by examining the deceleration behaviour in a free-flow, heterogeneous highway by asking all road users (i.e. cars, trucks and other motorised vehicles) to drive at their maximum (desired) velocity to a complete stop in the shortest amount of time instead of restricting themselves to only cars and trucks in a homogeneous traffic state, and. The deceleration of all road users was recorded using GPS trackers.

It was determined that the different vehicle types held significantly different deceleration behaviour. Vehicles with a higher initial velocity had a longer deceleration time and distance, while experiencing higher maximum and mean deceleration rates during there deceleration approach. All vehicle types during their deceleration approach, had a deceleration rate that initially increased, attained maximum deceleration and then decreased their deceleration rate again, once the vehicle velocity dropped below a critical velocity.

For all, except the cars, a dual regime model was developed to describe the deceleration rate of the observed deceleration behaviour, with the regime above the critical velocity (Regime I, Equation (2.9) – d_1) being described by an inverse exponential function and the regime below the critical velocity (Regime II, Equation (2.9) – d_2) being described as a linear function. The deceleration rate of the observed cars was described as a velocity dependent, second order polynomial (Equation (2.9) – d_c). For validation, the research used them to

interpolate and fit the functions based on the observed deceleration-velocity data and applied a 'Paired T-test' to defend their null hypothesis as described in Equation (2.9) – μ. This null hypothesis was tested for a 95% confidence interval($\alpha = 0.05$) and it was determined that the null hypothesis could not be rejected.

$$\begin{aligned} d_1 &= k_1 \cdot e^{-k_2 v} \\ d_2 &= \alpha + \beta v \\ d_c &= -k_3 v^2 + k_4 v + k_5 \\ \mu &= \mu_o - \mu_m = 0 \end{aligned} \tag{2.9}$$

$k_1 \dots k_5$ = model parameters
α = minimum deceleration (m/s^2)
β = Jerk, velocity dependant $(1/s)$
v = vehicle velocity (m/s)
μ_o = mean deceleration calculated from observed (m/s^2)
μ_m = mean deceleration obtained from model (m/s^2)

It is not quite sure how strong the parallels are between the respective research fields of road and rail vehicles. However, it is interesting to see that, even in research about the driving behaviour of road users, deceleration rates are observed to be non-uniform. Any assumptions of behavioural parallels will have to be thoroughly tested. For instance, the defence or assumption that the deceleration rates are velocity dependent, felt short and superficial. This will also mean an uncertainty in regards to the 'critical velocity' that was observed in the deceleration rate of the road users and if this would translate to the deceleration behaviour of heavy rail vehicle operators.

2.2. Literature Data Processing & Statistical Analysis

Considering the data driven approach of this research, some key research literature had to be reviewed, with regards to Data Mining and Data Processing. These papers had a significant impact on the current method of mining and pre-processing of the data by the Dutch IM and TOC. The purpose of this section had a less direct influence on this research, compared to the previous papers. This will, however, function more as background information and appreciation for the technological advancements made, allowing research like this MSc Thesis research to be feasible.

Ghofrani et al. (2018) reviewed the implementation of Big Data and the importance of this development within the different branches of the railway sector. It reviewed and summarised the different research papers, using Big Data for their analysis or creating mining/processing methods/tools for their Big Data Analysis (BDA). The review was further elaborated by categorising them by Railway Transportation Systems(RTS) area (i.e. Safety, Operations, Maintenance), level, model type (i.e. Descriptive, Prescriptive, Predictive) and other big data techniques.

In the RTS area of Operations, Intelligent Rail Transportation Systems (IRTS) has delivered several innovative technologies to IMs and TOCs in an effort to help them make optimised and efficient decisions. The application of BDA in IRTS has improved timetabling, simulation models and prediction models by allowing faster and more data processing. This has provided more means to evaluate, validate or train the analysis tools and make better decisions faster by providing more and relevant information.

Current and modern automatic signalling systems are based on the track-clear detection. This includes the train describer systems that keep track of train movements through their train description (e.g. train identification number). This forms the core of the automatic

route setting and centralised traffic management. However, it wasn't until the start of the 21st century that the logs were archived for railway operations analysis. Before that moment, these logs were only stored for a few days as support for investigations of possible accidents. Considered one of the forerunners in processing train describer data for analysis, Goverde, Daamen and Hansen, with their work presented in Daamen et al. (2009); Goverde (2005); Goverde et al. (2008), developed a mining tool on the older TreinNummer-Volgsysteem(TNV) for the Dutch IM, presented as the tool named TNV-Conflict. A newer analysis methodology was developed and it was the spiritual off-spring of the previous mining tool. Now developed to mine and process the data from the Dutch IM's newer train describer system, TROTS (Kecman and Goverde, 2012a,b). This is the technology that would allow for the BDA in the Railway Operations to exist in the Dutch Railway industry and this, or a derivative, is what runs behind the scenes when any Data Acquisition Request is made.

Collecting, processing and transforming these data sources, which are very well described by Ghofrani et al. (2018), shown in Table 2.1, in combination with timetabling data for descriptive analysis of train delays and timetable improvements, are considered the first applications of BDA in railway operations. Through the concept of process mining, with domain knowledge, researchers were able to process train describer records, retrieve train positions at the level of track section occupation. (Daamen et al., 2009; Kecman and Goverde, 2012a) Also route conflicts could be determined from these logs, from which secondary delays due to unscheduled braking could be derived. (Daamen et al., 2009; Goverde et al., 2008)

Big Data Sources	Typical contents
Train Describer Data	Track occupation and release times, train description steps, signal states(stop/go), switch states (left/right)
Traffic control delay data	Delays at stations or other timetable points
GPS data	Train positions
Train event recorder data	Train positions and velocities, traction, brake applications
Traffic control incident registration data	Begin and end of disruptions, failing elements
Timetable data	Arrival and departure times, train routes, stops
Ticket sales data	Tickets available
Automatic Fare Collection data (smart card data)	Passenger check-in and check-out times
Website data	Timetables, recommended travels and prices, train delays, disruption location and times, online ticket sales

Table 2.1: Data Sources described by Ghofrani et al. (2018) – Table 3

The train describer data enabled data-driven predictive models, using historical and real-time data, with methods such as robust regression, regression trees and random forests (Kecman and Goverde, 2015a), and event graph models for online track conflict analysis and train delay predictions (Hansen et al., 2010; Kecman and Goverde, 2015a). Another application of train position data has been to estimate train characteristics using graphical tools on train event recorder data (De Fabris et al., 2008), Simulated Annealing on vehicle GPS data (Medeossi et al., 2011), and a Genetic Algorithm on track occupation data (Bešinović et al., 2013b).

De Fabris et al. (2008) used the train event recorder data to derive realised train characteristics. This data-set consists of not only the vehicle's location, but also its velocity, traction and braking applications. The issue with this type of data is the decentralised and temporary storage, making it difficult to use for analysis on a larger scale. Another popu-

lar means of vehicle-based vehicle tracking data, is using the vehicle's GPS tracking data forthe same purpose. (Medeossi et al., 2011) This would provide with less information and a lower tracking accuracy, but is deemed more accessible for researchers, due to policy and centralised availability.

2.3. Current Practice Running time calculations & Nominal Values

The other side of the coin in trying to understand realised vehicle and operator behaviour, is to understand what the assumed/expected/nominal vehicle and operator behaviour is and how these variables are derived or founded. The current practice by the Dutch TOC and IM in regard to braking regimes, is primarily that the braking regimes apply a constant braking coefficient. This assumes that braking is applied instantly by the operator and a negligible amount of time is spent on the build up or release for the braking force applied by the vehicle's braking systems. This allows for a simpler solution and faster computation at the cost of accurately describing real operator behaviour.

Another current practice applied by the TOC and IM in regard to designing the network timetable, is to calculate the minimum running time between stations based on maximum acceleration allowed (i.e. limited by either traction in the wheel-rail interface, available engine power or passenger comfort), running at the traction's speed limit and braking just in time with constant and nominally defined deceleration rate (i.e. defined per vehicle type based on vehicle tests, SLT vehicles allow for a nominal deceleration rate of 0.8 m/s^2 during a normal braking regime) to reach the intended stop location. This running time is then extended with time allowances initially defined by a fixed percentage of the minimum running time and later further extended with time allowances based on the mean delay time between stations found in the recent historical data. The first measure of time allowances is meant to add robustness to the network timetable and allow for energy efficiency when running time surplus is expected under performance plans, such as Universeel Zuinig rijden Idee (UZI). The second measure of time allowances is to adjust for either the initial error in calculation of the minimum running time or the underestimation of the fixed percentage time allowance compared to realised train runs and is to further increase robustness of the network and to decrease the percentage of delays.

This method of minimum running time calculations based on maximum acceleration and nominal constant braking, and the time allowance calculations based on a fixed percentage of the minimum running time and the realised mean delay, does not (or partially) allow for the existence of stochasticity based on the variance in the human operator's judgement and/or behaviour. For the more assertive operational behaviour or more on-time train runs with sufficiently surplus in time allowances, this is not an actual problem as the time allowance surplus can be turned into a more energy efficient train run by applying a coasting regime, in which a train draws minimal power by turning off the traction motors and allowing the vehicle to roll for a significant distance.

The problem of this method of running time calculations arises with the "top half" of the stochastic distribution in regards to the delays found in the historical data of the realised running times. This is because the root cause of the stochasticity is not addressed in either a solution to the variance in operator's driving behaviour or in the accounting for stochastic behaviour in the running time and network timetable calculations.

2.4. State of the Art - Scientific Gap and Inspiration

The scientific gap found while reviewing scientific papers, was that the simulation and reconstruction models presented by Bešinović et al. (2013a,a,b); Medeossi et al. (2011); Tielman (2015) mainly focused on a single source of 'Time-Location' data (i.e. Train describer data or vehicle GPS data) and therefore inherited the benefits and downsides of these data sources. When it came to the modelling of the deceleration regime, their models were also limited to a constant and uniform braking coefficient or resistance coefficients. Tielman

(2015) tried to solve it by braking the deceleration regime into smaller regimes, but was within these smaller regimes still limited to FRISO's 'linear' functions.

While reviewing other fields of study for papers on deceleration behaviour, Maurya and Bokare (2012) provided an interesting view in describing the deceleration behaviour of different road users and examined the non-uniformity on it. Testing the non-uniformity of the deceleration regime in the railway system is something that has not been attempted. However, the manner in which the non-uniformity was handled in Maurya and Bokare (2012) was symptomatic by describing the realised braking behaviour in a velocity-deceleration plot and fitting model parameters to that dimension. No real vehicle characteristics or driver behaviour could be linked to these coefficients.

So the scientific gap, as seen per this research, is the combining or fusing of the different data sources in order to enrich each other by providing a finer mesh-grid of 'Time-Location' points while tagging more detailed information of the infrastructure and signalling states. Also, no known attempt has been made to describe the deceleration rate of heavy rail vehicle as a non-uniform motion. Furthermore, none of the above mentioned papers have drawn correlation between the realised deceleration behaviour and external influences (i.e. weather, light) or railway system conditions (i.e. timeliness vehicle, track conditions, vehicle characteristics, line of sight, station platform-signal lay out) These papers, however have given many of the inspirations applied or tested within this MSc thesis research.

Bešinović et al. (2013a,b,c) provided the inspiration for the proposed analysis methodology and vehicle velocity profile reconstruction, and the incentive to investigate the stochasticity of the vehicle's deceleration approach to a planned (station) stop through the use of track section 'Time-Location', infrastructure and track side signalling data.

Medeossi et al. (2011) provided support for the inspiration found for the analysis methodology and the research approach to use 'Time-Location' data to reconstruct the vehicle velocity profiles and analyse the stochasticity in the deceleration behaviour. This research specifically added the inspiration to use the vehicle's gps data to provide a richer 'Time-Location' dataset, as well as to investigate the stochastic variation of the realised blocking times of track sections, especially near stations, to provide a means of quantifying the performance impact of the infrastructure occupation.

Tielman (2015) provided not only more evidence and support of the existence of a stochastic nature in the rail vehicle's operational cycle (i.e. accelerate, cruise, coast, brake, repeat) through means of stochastic vehicle parameters (e.g. braking coefficient), but also a stochastic variance in the on-set times of the different driving regimes. Furthermore it provided proof of the existence of different patterns in the deceleration approach to a planned stop by defining smaller driving regimes within the deceleration approach, creating a further differentiation in describing the rail vehicle's deceleration approach. This provided support for the creation of the driving sub-regimes and the evaluation of different deceleration approach profiles.

Maurya and Bokare (2012) provided the inspiration for the attempt to describe the rail vehicle's deceleration rate as a non-uniform variable within the (sub-) regimes, even after defining distinct sub-regimes within a rail vehicle's deceleration approach. Some of the deceleration rate functions initially appeared to describe certain rail vehicle deceleration approaches better than the widely used constant deceleration rate. However, since the defence of the velocity dependency felt short (Maurya and Bokare, 2012, p.260), both velocity dependent deceleration rate functions, as well as time dependent alternatives, were tested for their fitness.

3

Performance Indicators

This chapter will describe the performance indicators and what these indicators signify and describe. Furthermore, this chapter will elaborate on the considerations made for using them and the methods and tools required to provide results and scientific support for these performance indicators. The performance indicators and their respective (mathematical) formulation are summarised in Table 3.1 at the end of this chapter.

3.1. Background Indicators

To validate or quantify the performance of anything, parameters or indicators have to be defined, to provide a form of measure. The focus of this research is on the stochastic nature of the operator's deceleration behaviour and the impact it has on the realised infrastructure occupation.

To say anything about the stochastic nature of the vehicle's deceleration behaviour, the behaviour is broken down into several performance indicators that describe the deceleration behaviour. This could be vehicle related (e.g. Braking coefficients, Regime Duration, Regime Profile) or time related (e.g. Running Times, Loss Times, DUEs) performance indicators, based on realised data and provide the stochastic distribution of the indicators.

When it comes to the impact the stochasticity has on the network, several track-based performance indicators are available. This could be a percentage or ratio of time a track section is utilised, the more general infrastructure occupation at corridor level or track throughput. However, since the nature of this behaviour is assumed stochastic, the stochasticity of the performance has to be taken into account to provide a complete picture. This would show the track potential, limits or sensitivity to the variation of vehicle behaviour.

This will, however, mean that the performance indicators used in this research to quantify the performance of the case study, are mostly either be vehicle or time related and most likely have a stochastic nature. The performance indicators used in this research can be grouped into Vehicle related, Time related and Track related.

3.2. Vehicle related PI

To accurately describe the performance of the deceleration behaviour of the realised train runs in relation to the scheduled and expected/nominal train runs, some of the vehicle related aspects have to be taken into account and can be used as a performance indicator. The vehicle related indicators used in this research are the deceleration rate, sub-regime composition (or deceleration regime profile), transition hardness between two sub-regimes and the vehicle resistance coefficients. All these are defined, described and explained below.

Braking Rate

Name:	Braking Rate
Maths Formulation:	Variable Notation, see Section 4.4.3 for mathematical formulation of the braking rates.
Unit:	m/s^2
Description:	A constant value or non-uniform variable described as a function. The non-uniformity can be defined either as a Dual-Step, Triple-Step or a second order polynomial.

The braking rate profile describes the braking rate within a braking (sub-)regime. The questions to be asked here are: *"Is the braking rate uniform?"* and *"What profile can be drawn from the realised data?"*. The braking rate will be defined as the rate of change of velocity in the temporal dimension and therefore defined as m/s^2. Any rate of change of the deceleration, known as Jerk, will be defined as rate of change in the temporal dimension and therefore defined as m/s^3. This will, however, not imply any assumption on the dependency of a non-uniform braking rate.

The deceleration, or the coefficients of a non-uniform deceleration function, will be calculated from realised tracking data through the use of the reconstruction model, at an accuracy of 2 decimals. The non-uniform deceleration function will form to a second order polynomial and/or be described by a Dual Step or Triple Step approach(i.e. Braking Sub-regime described by further breaking down into multiple sections within a single intended motion.). This will tested by evaluating the goodness for fit for several pre-defined profile concepts. The evaluation will be described as a value of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). In case of multiple runs or series, the distribution of these values will be given, to demonstrate the likelihood of this profile being either vehicle-type-specific, dependant on the operational conditions or train operator related.

Deceleration Regime Profile

Name:	Deceleration Regime Profile
Maths Formulation:	e.g. Br-Co-Br, Co-Br-Cr-Br (numerically translated 0-1-0 and 1-0-2-0 resp.)
Description:	Sequence of sub-regimes in a deceleration regime, defined by their respective abbreviation or a number of the sub-regimes (i.e. cruising as Cr or 2, coasting as Co or 1, and braking as Br or 0).

The deceleration regime profile describes the pattern and duration of the different sub-regimes used in the vehicle's deceleration approach. This leads to the following questions on the profile composition: *"What sub-regimes are the deceleration regimes comprised of?"* and *"What are the likely profiles that can be drawn from analysis?"*.

This will be calculated with the reconstruction model by testing the goodness of fit of several pre-defined regime profiles. This will be described as a value of R-squared or RMSE. In case of multiple runs or series, the distribution of these values will be given, to demonstrate the likelihood of this profile being station-specific, dependant on the operational conditions or train operator related.

Vehicle Resistance

Name:	Vehicle Resistance Coefficients
Maths Formulation:	$r_v = r_0 + r_1 v + r_2 v^2$
Unit:	N/kg, N/kg, $(Ns)/(kgm)$, $(Ns^2)/(kgm^2)$ – resp. unit for r_v, r_0, r_1, r_2
Description:	Mass-specific vehicle characteristic coefficients related to the Davis Equation.

Just like Bešinović et al. (2013b) discussed the vehicle resistance coefficients in the development of their reconstruction model, these variables will also be examined in the development of this research's reconstruction model. This will be partly to evaluate the reconstruction model in producing reasonable coefficients and partly to assess the quality of the nominal vehicle coefficients in comparison to the "real" coefficients.

3.3. Time Related PI

To relate the deceleration behaviour to vehicle and track performance, the use of time related performance indicators is generally used. In this research, the differences in realised running times between stations and the deceleration loss times were used as they describe the performance on the network the clearest.

Realised Running Times

Name:	Realised Running Time
Maths Formulation:	$t_{Realised}$
Unit:	s
Description:	The difference between realised running times and scheduled running times. Indicator for usage of time allowances by the deceleration regime.

When considering the time related indicators while evaluating the performance of the realised vehicle operations, the key performance indicator and most direct measure of performance is the vehicle realised running time between stations. This is lead by the questions: *"What is the realised running time?"* and *"How do the realised running times compare to the scheduled times?"*.

By comparing the difference in realised running times to the calculated reference running times, a performance impact can be assigned. To understand if this performance impact was due to the usage of the time allowances or due to the train operator's trust in the track conditions, a comparison to the realised and scheduled departure times could be made to create that distinction.

Deceleration Loss Times

Name:	Deceleration loss time
Maths Formulation:	$t_{Deceleration} - t_{Maximum\_Velocity}$
Unit:	s
Description:	Loss time expressed as the time lost due to deceleration compared to the time required to traverse the same distance at the track section's speed limit.

The Dutch rail sector has a time related performance indicator, known as a loss time. This is described as the time 'lost' due to an regime implementation, compared to the time required for a vehicle to traverse the same distance at the track section's static maximum speed limit.

In this research, the specific loss time of interest are the loss times over the deceleration regime. Considering that the deceleration behaviour is stochastic in nature, the distance that the loss times are measured over is variable. However, the measure of performance is in the difference between the two running times (of the driving regime and the constant operational velocity) and shorter distances by other deceleration regimes will have no to negligible loss times over the distances not measured compared to larger distances. Therefore, this will still allow for a good comparison and performance measuring tool. To understand the performance differences between realised runs and the scheduled behaviour, deceleration loss times of both the minimum running times and the running times with time allowances will be calculated per station and used to compare the performance between realisation and scheduling.

3.4. Track Related PI

In relation to what the impact of the stochastic deceleration behaviour has on the infrastructure, a performance indicator related to the infrastructure is required. To evaluate the performance impact of the realised deceleration regimes by rail vehicles on the infrastructure, the infrastructure occupation will be described at track section level.

Considering the available data will cover only a portion of the complete corridor usage, the realised infrastructure occupation and the scheduled infrastructure occupation will consist only of the current selection of rolling stock. To understand the performance impact, the realised infrastructure occupation and scheduled infrastructure occupation will be compared to each other.

Track Section Occupation Duration

Name:	Track Section Occupation Duration
Maths Formulation:	$t_{Release} - t_{Occupation}$
Unit:	s
Description:	The duration of a vehicle traversing a track section, measured from the first axle entering the track section ($t_{Occupation}$) until the last axle exiting the track section ($t_{Release}$).

When measuring the infrastructure performance at a microscopic (i.e. track section) level, it allows for the localisation of infrastructure capacity bottlenecks and sensitivity analysis on the track sections. This is accompanied by the following questions during the investigation: *"What are the infrastructure occupation rates at the track section levels?"*, *"Which track sections form the capacity bottleneck for the corridor?"* and *"How severe is the impact of the realised stochastic deceleration behaviour at these critical track sections?"*.

In this research, the infrastructure occupation was simplified to the track section's occupation duration, due to several unknown factors such as the out-of-scope train runs that use the track section, the other unknown time variables in the scheduling blocking time (e.g. switch/release time, sight/reaction time, setup time). Only the times of track section occupancy and release are empirically measured by the train describer data, as discussed in Section 4.3 with regards to this research's used data pre-processing.

3.5. Stochastic Distribution & Nominal Value Comparison

The expectation of the current scheduling tools and models, with regards to the heavy rail vehicles, is for the train runs to run invariable and consistent. The realisation is that this is not the case considering the human operator and external influences. However, on the Dutch rail network, the IM and TOC plan for time allowances with the train runs to increase the flexibility and robustness of the network to compensate for this variability. This is still mostly deterministic and either assigned ad hoc to train runs or in advance by a standard measure of time allowances (e.g. this is 5% of the minimum running times on the Dutch rail network before rounding to the whole minute). This is all considering the nominal values of all aspects of the railway system coefficients, parameters and characteristics.

Furthermore, there is also a maximum period of time a track section is allowed to be reserved and occupied within a cycle time (i.e. duration/period over which the scheduling pattern of a cyclic/periodic timetable is developed). All these allowances and margins allow for minor fluctuations, but what are the consequences of using that aspect of the railway system and how often does this happen? So this raises the question: *"How do the realised train runs compare to the scheduled expectations?"*.

Due to the stochastic nature of the realised train runs, all nominal values and estimated time allowances will be used to test for the performance of these realised train runs in reference of the scheduled and nominal/expected behaviour by the IM and TOC. The distribution of the realised performances will be fitted to theoretical distributions in order to investigate if the probability distributions are consistent and could be used to make predictions of future performances.

Indicator Relation	Indicator Name	Math Formulation	Indicator Unit
Vehicle	Braking Rate	$b = const.$ or $non-unif.$	m/s^2
	Deceleration Regime	*Co-Br-Cr-Br* or numerical *1-0-2-0*	seq. of regime abbrev. or numerical
	Vehicle Resistance	$r_v = r_0 + r_1 v + r_2 v^2$	N/kg, N/kg, $(Ns)/(kgm)$, $(Ns^2)/(kgm^2)$
Time	Realised Running Time	$t_{Realised}$	s
	Deceleration Loss Time	$t_{Deceleration} - t_{Maximum\_Velocity}$	s
Track	Infrastructure Occupation	$t_{Release} - t_{Occupation}$	s

Table 3.1: Summary Table Indicators

4

Conceptual Framework & Deceleration Reconstruction Model

When analysing the realised deceleration behaviour of train operators, large datasets are required to accurately describe the different deceleration behaviours found in the realised data between the different types of operators and even larger datasets are required to analyse the stochastic nature of the realised deceleration behaviour. This analysis methodology is known as a data driven approach, as it uses and combines multiple data sources of which the datasets can be classed as Big Data as the data used is over a large number of days (i.e. 365 days) and train runs (i.e. approx. 70 runs per day), collecting a large number of datapoints per train run (i.e. 1 datapoint per 10 seconds and 1 datapoint per track section in the corridor per train run) with each data point holding several pieces of information(e.g. train ID, rolling stock ID, location, velocity, time, track section ID, signalling). As a means to assist with the analysis of the data driven approach, a conceptual framework (i.e. a support system or structure around a research's analysis methodology or research concepts) and reconstruction model is proposed to analyse the stochastic nature of the realised deceleration behaviour and test the impact of the stochasticity on the infrastructure occupation in comparison to the deterministic expected driving behaviour and scheduled infrastructure occupation.

This chapter discusses the proposed conceptual framework, the required data sources in this research, the data processing required prior to the reconstruction model and further elaborates on the details of the reconstruction model, such as the mathematical theory, structure, implementation, purpose and scope of the framework and reconstruction model, as well as the desired output products of the analysis methodology.

4.1. Conceptual Framework

This research introduces a conceptual framework for the reconstruction of the deceleration behaviour of heavy rail vehicles. The general structure and concept approach are inspired by Bešinović et al. (2013a,b,c) and Medeossi et al. (2011), in the sense of using location tracking data from GPS and infrastructure and their analysis application to reconstruct the realised driving behaviour.

The proposed framework differentiates from its inspiration by it's interest and focus on the deceleration behaviour within the proposed reconstruction model, aptly named Deceleration Reconstruction (DR) model, instead of reconstructing complete velocity

profiles. It instead elaborates in more detail what occurs in the deceleration regime (i.e. the operational action taken in order to decrease the vehicle velocity to the desired velocity.), which is explained in detail in Section 4.4. The proposed framework is further expanded to incorporate the statistical analysis. This is because it is not just the realised train trajectories and corresponding parameters, but also the statistical distributions of these parameters and related performance indicators (PI) that are the considered end products for this research. This framework is visualised in Figure 4.1 and has several stages it passes through, namely the Pre-processing of the data used by the DR model, the Deceleration Reconstruction Analysis with the DR model and concludes with Post-processing of the modelling results.

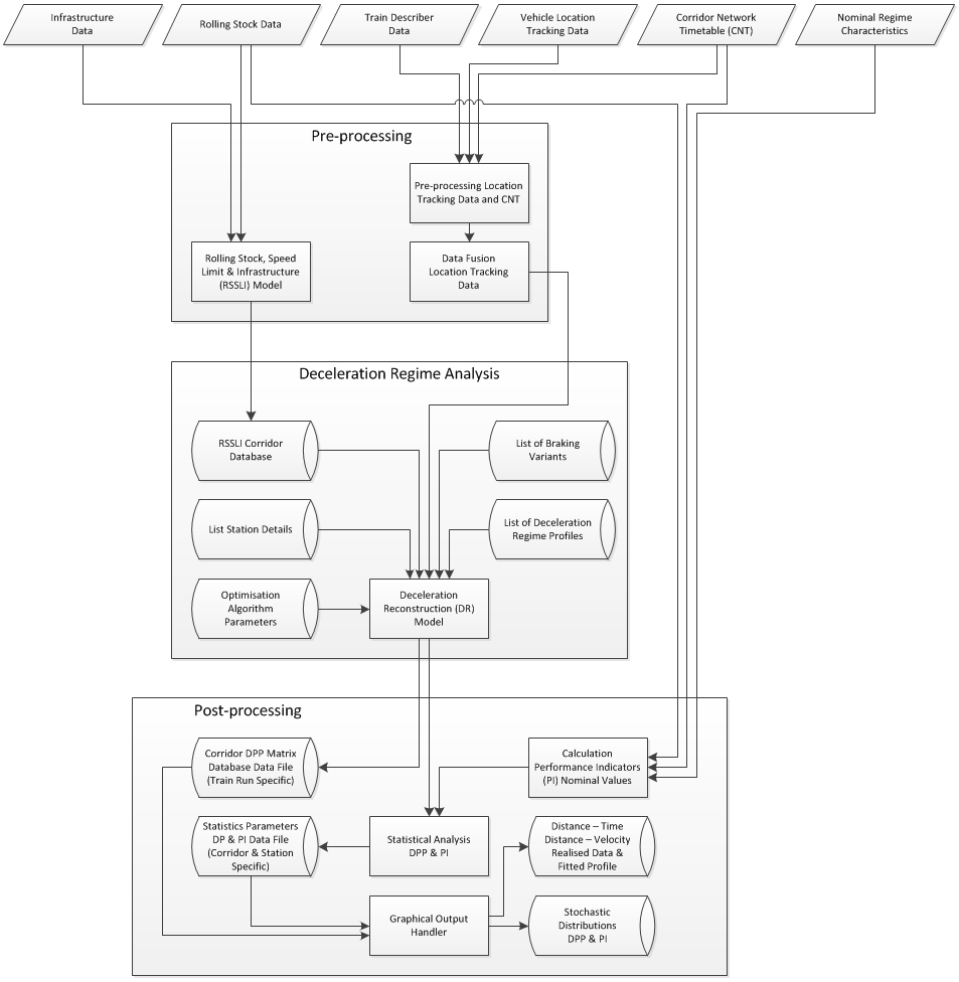


Figure 4.1: Process Structure for the conceptual framework used in this research, describing the data flow and process logic behind the analysis.

Pre-processing

It starts with the intake of the datasets from their respective sources and pre-processes the data to their desired formats and extracts the desired data per datapoint of their respective sources. Since this framework analyses a single specific corridor and a specific rolling stock type, while comparing multiple train runs over a large window in time, it therefore is sensible to map and model the infrastructure and Rolling Stock once in a single set-up process, namely the ”Rolling Stock Speed Limit & Infrastructure (RSSLI) model”. The static models are generated with infrastructure data and used to develop the corridor’s infrastructure model. The nominal rolling stock parameters are stored similar to the infrastructure in the RSSLI model. This is further explained in Section 4.3.1.

Some specific rolling stock related data (e.g. rolling stock composition) are used along side the network timetable in the data pre-processing, in order to extract the desired GPS datapoints from the raw gps data and to further enrich it with more detailed information. This type of processing is useful in case the TOC or IM cannot provide a processed version of vehicle location tracking data. In case of available processed vehicle location tracking data, the ”Pre-processing Location Tracking Data and CNT” process is used to validate the data points with the corridor network timetable (CNT). This is to catch and filter out the data points that are inaccurately linked to a train run or was separated from the rest of the train run due to changing traffic days. This process consists of data formatting and data filtering, found in Section 4.3.1 and Section 4.3.2.

A further part of the data pre-processing is the (re-)calculation/validation of the distances provided in the location tracking data sources. This is done to correct or adjust for errors or gaps in the data, shortly after reformatting the raw data sources. The distance calculation were done to each location tracking data sources separately before continuing on with the data fusion process to minimise the magnitude of error and to provide the data alignment the right means to align the different data sources with. This is further explained in Section 4.3.3.

The last pre-processing step ”Data Fusion Location Tracking Data” is where the fusion of the different location tracking data occurs. To allow for the fusing of the different sources and sometimes even different groups with these sources (i.e. combined rolling stock compositions such as ”SLT-4 SLT-6” or ”SLT-6 SLT-6” are two rolling stock physically coupled and have two GPS sensors active during the train run), all sources will need to be aligned to each other (see Section 4.3.4). During the data fusion, the different data sources are combined and rearranged by their distance along the train path. From each data source, the data of interest is migrated over with each datapoint. Since the location and velocity of the vehicle are of particular interest for this research, extra care is given in the calculation and preservation of the location tracking data’s vehicle velocity data. Each data sources is given their own data entry point in the fused location tracking dataset and the mean velocity is used, calculated at each data point in the fused dataset. This process is elaborated on in Section 4.3.5.

Deceleration Reconstruction Analysis

From the generated fused and corridor aligned location tracking dataset files, a train run is selected to load a train run specific, fused and corridor aligned location tracking dataset to computer memory for analysis. Along with RSSLI model dataset, it is used for analysis in the Deceleration Reconstruction (DR) model(see Section 4.4). The DR model uses the location tracking data to estimate the realised deceleration regimes applied during the station approaches over the corridor. It challenges the assumption of a singular driver action or regime by introducing sub-regimes (i.e. smaller regimes within a larger regime of a single operational intent). It further challenges the assumption that, when applying the brakes, a uniform (i.e. constant) braking rate is generated that effectively resembled a binary on/off mechanics, compared to a more natural build up and release of the braking application. These challenged assumptions are tested by evaluating realised location tracking data for different deceleration regime profiles (see Section 4.4.5 and different braking variants (see Section 4.4.3).

With infrastructure data from the RSSLI model, the DR model extracts realised deceleration sub-regime profile (i.e. a sequence of sub-regimes as described in Section 4.4.4, lengths and duration of each sub-regime) and the detail coefficients, such as realised resistance coefficients, the braking rates (i.e. both as uniform variable as non-uniform function and coefficients as described in Section 4.4.3), and the performance indicators described in Chapter 3 to be used in the statistical analysis. The reconstruction model is elaborated on in Section 4.4.

Post-processing

After all the location tracking data has been processed and the desired information and parameters are obtained, this data gets stored in train run specific data files to be accessed for statistical and comparative analysis. Their graphical representation are stored along with it to provide a visual aid and means of inspecting the quality of the specific train run and the analysis done on the train runs.

Using the specific train run data files, the last process is to run the statistical analysis to obtain the desired stochastic distribution details and generate the graphical output of the specific deceleration regime parameters and the research defined performance indicators (Chapter 3). To understand the performance impact and differences of the realised train runs, the expected performance of the scheduled train runs with their nominal parameters (i.e. braking rate, resistance forces) and the TOC's/IM's current practice of running time calculations, are calculated. (see Section 4.5).

These nominal performance parameters are used in the statistical analysis for comparison to the realised train runs, to showcase the differences in driver behaviour during the deceleration regime, the related performance impact of these differences and the causality/sensitivity of the influential factor (e.g. seasonality, on-timeness of the rail vehicles).

4.2. Data

For this research's data-driven analysis, six different datasets are used. Infrastructure Data and Rolling Stock Data are used to build a base model of the train path's infrastructure and static restrictions and the nominal characteristics of the rolling stock to accurately describe the vehicle's location during the reconstruction of the realised deceleration regimes. Rolling Stock data, Train Describer data, Vehicle-sided Location Tracking data and the Corridor Network Timetable are used to process and align the vehicle-sided with the infrastructure-sided vehicle location tracking data and enrich it with the signalling. The Corridor Network Timetable and the Nominal Regime Characteristics are used to describe the scheduled train runs and expected implemented deceleration regimes. The details of the used datasets are elaborated on in their subsections below and summarised in Table B.1, found in the Appendix.

4.2.1. Infrastructure Data

The infrastructural data required for the analysis and the reconstruction model, is to map the corridor to accurately describe the vehicle's location and behaviour and to understand the infrastructural influence on the implementation of the driving behaviour. This requires the track section ID, lengths, GPS Location, type of section(i.e. switch/'open track'/level crossing), relation to neighbouring section, static speed limits, curvature and gradients. This information for the infrastructural data is gathered from Dutch IM's rail infrastructure specialised GIS system InfraAtlas and processed in InfraMonitor.

Using the track section IDs, the relation between the infrastructure and the location tracking data from train describer data is made possible, while the infrastructure GPS location helps to align the vehicle location tracking data to the infrastructure. The length of track section are used to calculate the average vehicle velocity from the track section occupation time in the train describer dataset. The static speed limits, the type of track section and it's relation to the neighbouring sections help identify the track section's restrictions and the possible reasons for them. The track section's curvature and gradient are related to the effects of the infrastructure acting on the realised driving behaviour through the resistive forces r_c and r_g, of which the latter could be considered an assistive force (i.e. external force working in the same direction as the intended action) when accounting for track sections with a negative gradient (i.e. going downhill). For the signalling data, data required for this research describes their GPS location, type of signal(i.e. Controlled or Automated) and their relation to the track sections (or in case near stations, relation to stopping platforms) in or-

der to understand the control points in the infrastructure in relation to the realised driving behaviour. Out of both (i.e. Infrastructure and Signalling) data sources, the relevant information for this research is extracted and formatted according to this research's format.

4.2.2. Rolling Stock Data

The rolling stock data required for the reconstruction model, are the nominal vehicle characteristics, such as vehicle mass (including normal loading and rotating mass factor), vehicle resistance coefficients related to the Davis Equation (Trani, 2018) and nominal braking coefficients. This information is obtained from TU Delft and Ricardo Rail Netherlands and its values are derived from the performance testing Ricardo Rail Netherlands did in the certification process of the SLT rolling stock type. This data is required in order to provide the reconstruction model a starting vector to optimise from and provides the statistical analysis the values required to calculate the expected performance from nominal and scheduled parameters in order to compare the differences to the realised deceleration behaviour.

4.2.3. Train Describer Data

Train describer data is infrastructure-sided train tracking data, based on train numbers and messages received from the signalling and interlocking systems. The Dutch IM's train describer system, TROTS, logs the trains at the level of track sections. The TROTS data log holds relevant data (i.e. track section occupation time, track section release times, train IDs, track section ID, track section lengths, average vehicle velocity through the track section) which provide a means of train location tracking from the infrastructure sensors, along with the relevant signalling aspects to provide the system requested driving behaviour through dynamic speed limit management.

The required data from the TROTS logs for the reconstruction model, are track section ID, train run/rolling stock ID, track section occupation time, track section release time, current section signalling aspects, following section signalling aspects, track section lengths and average vehicle velocity in the track section. The signalling aspects of the current and following sections are used to understand the nature of the dynamic restrictions and the expected regime implementation. These TROTS data logs are obtained from the data analysis department of the Dutch IM ProRail, called Prestatie Analyse Bureau (PAB).

4.2.4. Vehicle-sided Location Tracking Data

Vehicle-sided location tracking data provides a different means of vehicle tracking and has the benefit of providing the actual vehicle location instead of an estimate between two known points as with the train describer data. This would provide a more accurate or finer tracking of the vehicles inside a track section. Tracking a vehicle standalone (i.e. without any external system for tracking/registering) is usually done through on-board GPS sensor tracking. Data gathered from these GPS sensors consists of the rolling stock ID, a time-stamp expressed in Unix/Epoch time(i.e. number of seconds since 01-01-1970 00:00), GPS longitude and latitude, and sometimes even the momentary vehicle velocity measured by the GPS sensor. The polling rate of these sensors is currently 0.1 Hz (i.e. 1 data point per 10 seconds).

The IM and TOC have a processed/enriched version of these GPS datasets, called MTPS, linking the GPS data to the realised train run, taken from the network timetable, aligning it to the travelled track section and corresponding signalling information. Using the enriched GPS datasets, would save an extra pre-processing step in the analysis. However, the conceptual framework of the data analysis does account for this pre-processing process in case of absence of the enriched dataset or failing of the desired data quality. The test pilot before the analysis of this research's case study, using the raw GPS data, has shown that the quality and accuracy of the GPS will bear little to no impact on the accuracy of the analysis results.This was confirmed by the data specialist at ProRail who mentioned

that the Kalman filter was only applied sparingly in the extreme cases of noisy datapoints, which was not necessary in the provided datasets of this research's case study.

The required data from the vehicle-sided location tracking data for the analysis, are the train ID, rolling stock ID, timestamp of the datapoint, GPS Latitude, GPS Longitude. The GPS Vehicle momentary velocity is mainly used for extra validation of the accuracy of the analysis and the selection of the estimated deceleration sub-regime composition. Considering the reconstruction method of the DR model and the quality of GPS data, the track section ID and the GPS coordinate adjustments are not required as input data, as they can be acquire from the train describer data or offer minor accuracy improvements respectively. The enriched vehicle-sided location tracking data, MTPS, will only provide benefit in reducing the effort and processing time of this research's analysis.

4.2.5. Corridor Network Timetable

The network timetable for the case study's corridor, is required for the analysis to provide the scheduled and expected train run times for the performance comparison to the realised train runs. The network timetable also provides the times of the realised train runs at these timetabling points, called "Dienstregelpunten" in the Dutch railway sector. Of these actualised times, the on-timeness of the realised train run can be derived to analyse the behavioural differences in applying the deceleration regime between the on-time and the delayed train runs. The network timetable is further used in linking the raw rolling stock GPS data to the corresponding realised train run data by comparing the rolling stock IDs and Epoch times. For the enriched vehicle-sided tracking data (MTPS), the network timetable is used to check and filter the data used in the actual realised train run and remove any activity registered outside of the realised train run.

For every timetable entry, the relevant data stored and required for the analysis, are the timetable day (i.e. timetable days in the Dutch railway sector run from 2am to 2am the next day, instead of 12am to 12am of a regular day), train number, the origin timetable point, the origin's action type, the origin's planned execution time, the origin's realised execution time, the destination timetable point, the destination's action type, the destination's planned execution time, the destination's realised execution time, the rolling stock ID, the rolling stock type and the rolling stock position. The network timetable is obtained from the Dutch TOC NS and the Dutch IM ProRail.

4.2.6. Nominal Regime Characteristics

To describe the expected deceleration regime during the scheduled running times, the nominal regime characteristics need to be defined to reconstruct the expected velocity profile between the stations in the corridor and to derive the expected system performance from this. For this research to describe the expected velocity profiles, the parameters for the expected acceleration, braking, time allowances, applied driving regime plans and performance plans (e.g. Universeel Zuinig rijden Idee (UZI)) are required. The theory and details of driving performance plans are elaborated on in the thesis work of Scheepmaker (2013). The strategies "Time-Optimal" and "UZI" are used from his research in order to compare the current Dutch TOC and IM practices and expected behaviour to the realised behaviour and its performance with regards to the proposed performance indicators. The "Time-Optimal" driving strategy holds the driving behaviour of a train run that minimises the running time through maximum acceleration to maximum velocity, maintaining the maximum velocity until braking at the exact time with a nominal braking rate to come to a stop at the desired location. The "UZI" performance strategy holds a driving behaviour for a train run that uses coasting in order to consume the remaining time allowances of a scheduled running time, given a running time of the minimum running time with a time allowance of roughly 5 percent of the minimum running time, which in turn reduces the mechanical energy usage.

4.3. Pre-processing

The forms of data pre-processing considered in this research are the formatting and filtering of the different data types, the creation of derivable data points (e.g. cumulative distance, Unix/Epoch timestamp from date-time text) prior to the analysis and more importantly the alignment of the location tracking data (i.e. vehicle GPS data and train describer data) to each other and the infrastructure train path over the corridor. Through this alignment, the fusion of the different location tracking datasets is made possible, along with the applied signalling aspects.

4.3.1. Data Formatting

All of the gathered data needs to be formatted in a way that would clean up the relevant data while removing the redundant or irrelevant data points, fixing minor errors in derivable data and filling in the minor gaps in datasets. This is to keep the data handling in the analysis as efficient as possible, while allowing the data association of the input data to be readable, clear and uniform. All raw datasets required as input data have some form of formatting done to them. Most of it is done in storage in a different file format, formatting date-time structures, renaming of columns headers or adjusting decimal points of numeric values. However, the most extensive formatting would be the development of the RSSLI-model, because of the differences in track section naming, and sometimes even grouping of track sections, between the infrastructure data from InfraAtlas and the train describer data from TROTS or any other traffic control related programs. This is done with the use of a translation table provided by ProRail and it is done to allow a more one-to-one access of the relevant infrastructure data in the analysis and reconstruction model in particular. The same is done to the rolling stock data to gather the relevant rolling stock data into a single dataset for easier access by the model. The RSSLI-model process is still considered data formatting, as it is renaming, rearranging and regrouping of the raw data, rather than being truly transformative.

4.3.2. Data Filtering

Along with formatting the datasets, the raw data has to be filtered. The data filtering is focused on cleaning up the location tracking data. This involves checking if every datapoint, for a given train run provided in the tracking data, actually belonged to the scheduled train run and not falsely related to data points gathered during a maintenance session of the same time table day or due network rescheduling for example.

To validate if the train runs and their station stops are useful for the research, two filtering approaches are used. With the signalling information provided in the train describer data, these realised train runs are further filtered on any dynamic speed restrictions (e.g. 'Yellow8', 'Yellow' or 'Red' signalling). This is to eliminate the hindered train runs and provide the unhindered runs to evaluate the deceleration regime by drivers restrained by nothing but the static infrastructure restrictions in their station approach. The second filtering approach is the "V-Box" method, in which train runs between stations is invalidated when the vehicle velocity within a specific section of train path drops below a set value. This is to eliminate train runs that hold erroneous velocity/tracking data or to eliminate train runs that limited their velocity for other reasons that were not caught by filtering on a given dynamic signalling.

Using these filtering approaches before the reconstruction model might filter out realised train run's station stops with valid and unrestrained deceleration regimes. However, it is assumed that correct filtering on signalling aspects (i.e. using the right combination of signalling aspects for filtering) or correct box filtering (i.e. removal of erroneous data points by selecting the right "V-box") before the reconstruction model will provide very few losses of such train runs and the gains made in computation time with the DR model not having to process train runs that are discarded, outweigh said losses in train runs.

4.3.3. Creation Derivable Data

To reduce the complexity of the reconstruction model and the analysis computation, some derivable data is calculated prior to the DR model. The derivable data generated from the input datasets are, in the case of the train describer data, the Unix or Epoch timestamp and the cumulative distance travelled on the realised train path. The Epoch timestamp is generated to make the comparison of time values of tracking data easier, since it represents time as an integer or decimal number instead of as a string in an agreed upon format. The cumulative distance is to allow for easier mapping and aligning the location tracking with the other sources. It also allows for easier plotting of the data to create a means to see the realised deceleration behaviour. The zero-point(start) of the train path in the infrastructure data is adjusted to a chosen fixed GPS point and used as a common anchor point to align the start of the infrastructure side data with the start of the vehicle-sided location tracking data.

In the case of the vehicle location tracking data, being either GPS as used in the pilot of the analysis methodology or MTPS as used in the case study (Section 4.2.4), the derivable data that is generated, are the distance, cumulative distance and, if not present, the Epoch timestamp. As stated in Section 4.2.4 in case of raw GPS data, it will first have to be formatted, filtered and aligned with the network timetable. To derive the relative distance and cumulative distance of the estimated train path from the GPS coordinates, the individual data points have their relative distance to their neighbouring coordinates calculated by the Vincenty's Inverse Formula. This is an iterative method that assumes the more realistic shape of the Earth as an oblate spheroid, as seen in Vincenty (2018). This provides an accuracy of less than 1 mm, compared to the 0.5% of the Haversine Formula, with negligible performance loss and is therefore used over other GPS distance calculation methods.

The first GPS coordinate in the series of datapoints in the location tracking of a realised train run, has its relative distance, and with that also its cumulative distance, calculated from the earlier mentioned chosen fixed GPS point that is used as a common anchor point for alignment later. The cumulative distance for all the following datapoints is a simple addition of its relative distance to the previous datapoint and that previous datapoints' cumulative distance. For dealing with the fact that the actual train path is not always a direct straight line between points, the assumption is made that the error between the two points(e.g. due to track curvature) is minimal due to the relatively short distances between the location tracking points.

4.3.4. Data Alignment

Before being able to fuse the different sources of location tracking data, each will have to be aligned to allow for a correct and smooth fusion. This is generally done when generating the derivable data as expressed in the subsection before. However, one final alignment needs to be made, which is the off-set of the GPS sensor to the leading axle of the (combined) rolling stock. This is handled firstly by comparing the epoch times of the vehicle location tracking data with the epoch time found in the train describer data for an exact match and the off-set is derived of the difference in distance along the train path. If no epoch time stamp is matched, the epoch time of the first track section is selected which falls between the smallest possible interval of epoch time in the vehicle location track data. The off-set is derived of the difference between the distance from the train describer data and a linear approximation of distance from the vehicle location tracking data at the same epoch time of the train describer data. The fact is known, that a linear interpolation between two GPS points at a ten second interval would lead to an error in determining the off-set. However, the reasoning and assumption is that the differences are minimal due to expected low velocity when exiting the station and moving into the first track section. The assumption is that the error would fall under an order of magnitude comparable to the GPS sensor's accuracy. An option to minimise the error of this interpolation, is to register the GPS sensor data at a higher polling rate of 1 Hz, decreasing the differences in time, distance and velocity.

4.3.5. Data Fusion

An integral part of this research is the fusion of the both infrastructure-sided (i.e. train describer data) and vehicle-sided (i.e. Vehicle GPS/MTPS) location tracking data. This allows for the enrichment of the GPS data with infrastructure-sided information such as current signalling, while creating a finer meshgrid of location tracking compared to solely using train describer data, and allowing better tracking of the vehicle within the track sections. After the (re)creation of the derivable datapoints and data alignment, both location tracking datasets share a common GPS anchor point, which is used as the common zero-point of all realised train paths. While fusing the tracking data, the data is aligned on both the epoch timestamp and cumulative distance along the train path. The datapoints get a key added to them, describing their source dataset (i.e. GPS/TROTS/etc.) and their cumulative distance(i.e. distance along the train path) remains in relation to their neighbouring datapoints from the same sources and the calculated alignment off-set. Their relative distance is recalculated between datapoints of all the different sources. This is to allow for both the fusion benefits while maintaining the ability to distinguish between the different tracking sources.

With the fusion and alignment, each data point of each datasource is enriched with the related datapoints of different sources, based on association and comparison. For instance, GPS datapoints get enriched with the track section their occupying along with the current and following dynamic signalling, while train describer data gets enriched with the estimated vehicle velocity through interpolation of two GPS datapoints and their GPS velocity data. As expressed in Section 4.3.4 when determining the off-set of the GPS sensor relative to the rolling stock's 'head', the margin of error should be low enough to allow this without creating problems in the analysis. This was assumed at low velocities at the beginning of the train run to minimise the margin of error for the GPS-sensor off-set. The velocity interpolation at higher velocities and bigger velocity differences would lead to higher errors, but for the use of velocity interpolation in the reconstruction model the error is assumed comparable to the measurements taken directly from the GPS sensors. When examining the example speed profile for a realised train run in Figure 4.2, the results of the data alignment appear to align nicely and the effects of the data fusion on the speed profile can be seen.

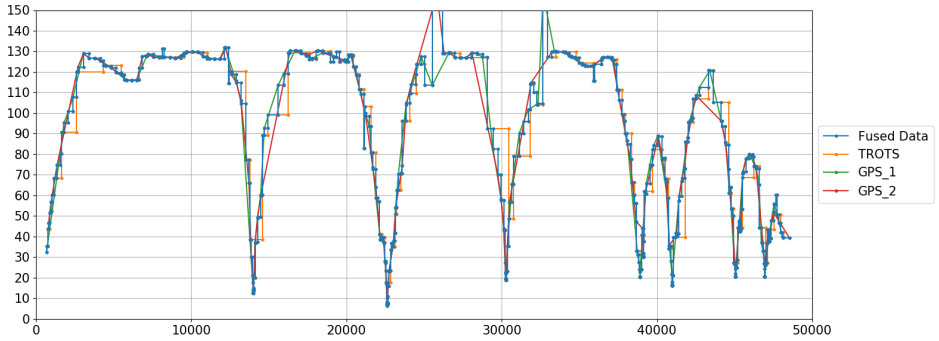


Figure 4.2: Example plot of realised location tracking data speed profile, for both the fused location tracking data as for the individual sources of location tracking. These sources in this example are the train describer data provided by the Dutch TROTS system and two GPS data sources from a single sensor on each individual rolling stock within the presented rolling stock combination.

For the deceleration reconstruction analysis, the process of data fusion allows for the velocity data from each respective source to continue to exist as separate entries in the fused dataset, which then fills in the gaps at the other data points from other sources through backfill and linear interpolation. This is then complemented with the mean velocity of all these velocity entries in the fused dataset. This mean velocity helps even out the extreme shifts in velocity due to minor tracking errors from GPS, rounding errors in the train describer data or calculation rounding errors from the linear interpolation described above.

4.4. Deceleration Reconstruction Model

This research proposes two different methods in deceleration reconstruction and are referred to as Point-to-Point Differential Backtracking and Velocity Difference Backtracking. The first method was developed to estimate the driving regime between two of location tracking datapoints and based on the retrieved differential parameters determine if the vehicle was in its deceleration regime for the station stop and its likely driving regime(e.g. coasting or braking). The second method was developed to streamline the analysis process by comparing the velocities between the datapoints to determine the start and end of the deceleration regime and used pre-defined deceleration regime profiles to determine the profile of the realised deceleration regime.

Both methods use the same set of mathematical equations to reconstruct the vehicle movement and characteristics. They rely on the dynamics equations of Newton's Law of Motion. The dynamics equations, as used in this research, are further explained in Section 4.4.1. Before reconstructing the deceleration regime, the defining characteristics for the four main driving regimes are determined relating to the dynamics equations, which are described in Section 4.4.2. Both DR model methods are explained in Section 4.4.4. In both model methods, the optimisation problem is defined as a minimisation of the difference in surface area underneath the speed-profile between the averaged velocities from the realised location tracking data and the estimated vehicle velocity from the DR model. To solve this optimisation problem, a genetic algorithm is used. Both the optimisation problem and the solving algorithm are further explained in Section 4.4.6.

4.4.1. Dynamics equations for the Deceleration Reconstruction Model

The analysis of the train movement relies on the dynamics equations of Newton's Law of Motion. These are formally described as general Equation (2.2) and rail-specific Equation (2.3). Since this research focuses on the deceleration behaviour of a commuter heavy rail vehicle, the dynamics equation used in this research is rewritten to focus on the resistance and braking equations.

The resistance equations due to vehicle and infrastructure are still part of the differential equations. The vehicle related resistance equation is based on the Davis equation. Also the gradient and curvature resistance functions are considered in the differential equation, even though the gradients and curvatures (especially the latter) in the Netherlands cause minimal restrictions, due to either low gradients or large curvature radii, and therefore have a minimal expected impact(visually represented in Figure 4.3). All these equations are considered mass-specific. This is to determine the resistance forces in terms of an acceleration. This is done implicitly in the case of the vehicle related resistance force r_v in the coefficients r_0, r_1 and r_2. For the gradient resistance force r_g, it is done explicitly through the removal of the mass variable m in the equation. For the provided curvature resistance force r_c, it was already set as a mass-specific equation, of which their units are set to N/kg or rewritten as an acceleration m/s^2.

$$\begin{aligned} r_v(v) &= r_0 + (r_1 * v) + (r_2 * v^2) \\ r_g &= \frac{9.81 * grad_{long}}{1000} \\ r_c &= \frac{0.01 * k}{Rad_c} \end{aligned} \tag{4.1}$$

Of which r_0, r_1 and r_2 (units resp. N/kg,$(N*s)/(m*kg)$ and $(N*s^2)/(m^2*kg)$) are vehicle resistance coefficients linked to the Davis Equation (Trani, 2018), v expressing the vehicle velocity in m/s, $grad_{long}$ the track gradient (in ‰) in longitudinal or driving direction, Rad_c the curvature radius in m and k a dimensionless coefficient, usually bound between 500 and 1200, describing the influence of the curvature on the resistance forces the vehicle is subjected to. With these specific equations, Equation (2.6), used in Bešinović et al.

(2013a,b,c), is rewritten as Equation (4.2), with b being the variable for braking rate. In case of non-uniformity the differential equation is expanded to define the braking rate as a time or velocity dependent function. Details of the braking rate description are elaborated in Section 4.4.3.

$$\frac{dv}{ds} = \frac{-r_v(v) - r_g - r_c - b}{v} \tag{4.2}$$

4.4.2. Driving Regime Characteristics

To correctly identify the realised driving (sub-) regime, the identifying characteristics of those driving regimes need to be determined that can (in combination) unambiguously describe the driving regime being applied. Without access to systems such as on-board train event recorder data, a few agreements and checks are needed to be set on the identifying characteristics to determine these (sub-)regimes. All four of the main driving regimes are mentioned, even though this research is only focuses on the related to the deceleration regime of a heavy rail vehicle. This is in order to make a clear identification of the analysed realised data and to provide the complete overview on driving regime characteristics.

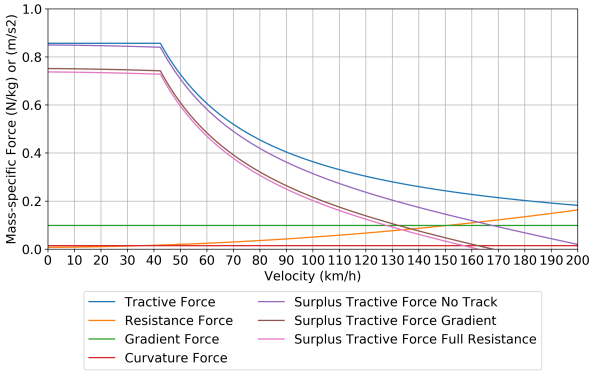


Figure 4.3: Velocity dependent mass-specific forces based on nominal coefficients. Shown is the effect of individual forces and different combination on the surplus force exerted by the vehicle, either expressed as N/kg or m/s2.

Acceleration

Since the acceleration regime is considered out of scope for this research, the traction force was removed from the velocity-distance differential, as seen in Equation (4.2). However, while identifying the different sub-regimes from realised data, the reconstruction model has a possibility of running into these acceleration driving regimes. To identify these regimes, the braking rate b is permitted to hold negative value to compensate for the lack of traction components in the dynamics differential and is used to indicate such regime. Therefore, when initially identifying driving sub-regimes or for when the GPS velocity data quality is not reliable, the identifying characteristics used are $b < 0$ and $dv/ds > 0$. In case of reliable quality velocity data from the GPS sensors, the driving (sub-)regimes are determined through essentially comparing $v_i > v_{i+1}$.

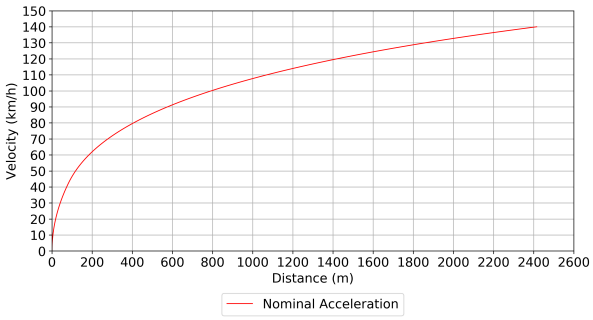


Figure 4.4: Acceleration Regime highlighted, shown with nominal acceleration and resistance coefficients. Defined by $v_i < v_{i+1}$

Cruising

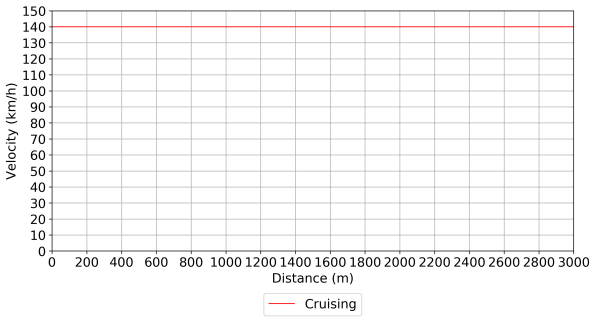


Figure 4.5: Cruising Regime highlighted, defined by the condition $v_start = v_end$

Even though the cruising regime is not part of the deceleration regime when it's preceding the deceleration regime, it is however part of the deceleration regime when the cruising sub-regime has two neighbouring sub-regimes being either coasting or braking. The definition of cruising is such that the vehicle's velocity is constant and therefore have $dv/ds = 0$. To unambiguously describe the cruising regime, a further distinction needs to be made between a vehicle cruising and a vehicle being stationary, which technically has a rate of change in velocity of zero. Therefore, to be considered a cruising regime, it will have to hold to the condition that the vehicle velocity is considered non-zero ($v > 0$). Since realised data will most likely not provide an exact match to the condition $dv/ds = 0$, a margin of error will have to be applied to compensate any inaccuracies in GPS velocity measurements.

Coasting

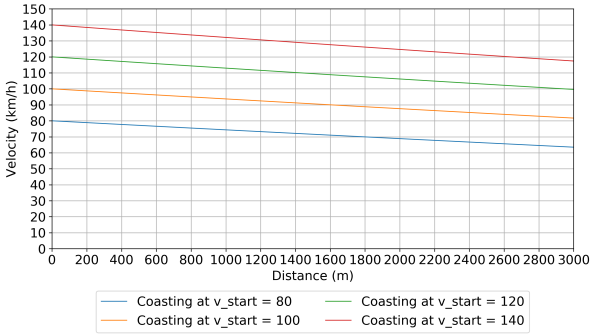


Figure 4.6: Coasting Regime highlighted, Nominal resistance coefficients shown for different starting velocities. Defined by $v_i > v_{i+1}$ and $b = 0$

The coasting regime is considered part of the deceleration regime applied by the train operator and is a regime in which the traction engines are turned off and no braking is applied. This leads to the identifying characteristic condition $b = 0$ for the coasting regime. However, this on it's own is not unambiguous enough, because this condition is present in the cruising regimes and at vehicle standstill. To make it unambiguous, a second regime characteristic condition is required, being $dv/ds < 0$. This would cover the eliminations of the cruising regime ($dv/ds = 0$ and $b < 0$) and vehicle standstill ($dv/ds = 0$ and $b = 0$). Due to the exactness of the equality conditions and the inaccuracy of measuring the realised data, a consideration needs to be made to give those equality conditions a margin of error. Due to the sensitivity of these conditions obtained from the velocity-distance differential at extremely low velocities and the margin of error applied, there is a possibility that the model would falsely identify a low velocity(e.g. walking pace) coasting regime as a vehicle standstill. Therefore a third condition is added to the defining regime characteristics, being $v > 0$, which is less sensitive than the other two conditions and eliminates the false identification of vehicle standstills.

Braking

The braking regime is the last regime to be considered part of the deceleration regime. It's primary characteristic condition is that the braking rate should be larger than zero ($b > 0$) or $b >$ margin of error in case of realised train runs to make sure its not due to a more resistive vehicle coefficient, creating a false identification for a braking regime instead of a coasting regime. To seal the unambiguous identification of the braking regime, the longitudinal gradient of the infrastructure has to be taken into account and this is done through the characteristic condition $dv/ds < 0$. This is to cover for the false identification of a braking regime when the brakes could actually be applied on a steep gradient to maintain a certain velocity and should therefore be classified as a cruising regime.

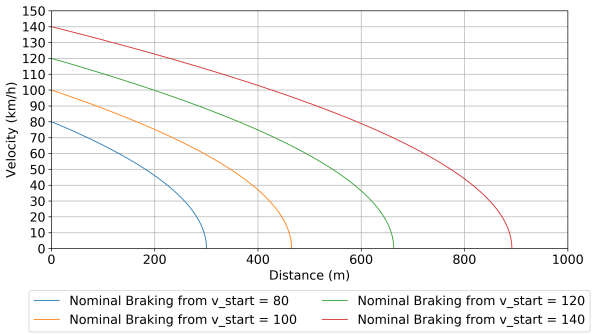


Figure 4.7: Braking Regime highlighted, Nominal resistance and braking coefficients shown for different starting velocities. Defined by $v_i > v_{i+1}$ and $b > 0$

The identification of the deceleration related regimes of coasting and braking described above are for when the quality of the vehicle velocity obtained through the GPS sensor is not considered reliable enough. In the case of the reliability being present, as is in this research, the deceleration related regimes can be identified through essentially comparing $v_i > v_{i+1}$. This condition will not make the distinction between coasting or braking. However, for identifying the deceleration regime distance this distinction is not necessary, as it is followed by a fitting of a defined sub-regime composition profile.

4.4.3. Braking Rate Variants

Due to the review of Maurya and Bokare (2012) and the initial testing of the realised train run data, a concept was developed that would challenge the idea that the train operators always apply a singular and constant braking rate in their deceleration regime into a planned stop at a station. The manner in which the braking rate is applied, is defined in this research as either uniform defined as a constant coefficient or as a non-uniform defined variable, either described by a Dual/Triple Step method or as a second order polynomial function (either time or velocity dependent). Table 4.1 shows the different methods of describing the braking rate tested in this research, with Figure 4.8 providing the visual comparison of the different braking variants.

Table 4.1: Deceleration Regime - Braking Rate Variants

Braking Rate Name	Uniformity Type	Mathematical Type	Mathematical Formulation
Constant	Uniform	Constant Variable	const.
Dual Step	Non-Uniform	Piece-wise Function	$b_{stop} = \begin{cases} b_{apply}(t) & 0 \leq t < t_1 \\ b_{release}(t) & t_1 \leq t \leq t_2 \end{cases}$
Triple Step	Non-Uniform	Piece-wise Function	$b_{stop} = \begin{cases} b_{apply}(t) & 0 \leq t < t_1 \\ b_{constant} & t_1 \leq t < t_2 \\ b_{release}(t) & t_2 \leq t \leq t_3 \end{cases}$
Time Dependent Polynomial	Non-Uniform	Second Order Polynomial Function	$b(t) = k_0 + k_1 * t + k_2 * t^2$
Velocity Dependent Polynomial	Non-Uniform	Second Order Polynomial Function	$b(v) = k_0 + k_1 * v + k_2 * v^2$

In the Dual Step and the Triple Step, the b_{apply} is the braking rate when applying the brakes to the desired braking rate and the $b_{release}$ in the case of releasing the braking. The b_{apply} and $b_{release}$ are either defined by an (inverted) second order polynomial or linear function. The $b_{constant}$ in Triple Step implies that for a length of time the desired braking rate is applied at a constant rate. In Maurya and Bokare (2012), the second order polynomial was defined as velocity dependent. However, to test the validity of this dependency, a time dependent second order polynomial is added as a likely non-uniform description of the train operator's behaviour when applying the brakes.

Each of the described braking variant takes different variables to express their respective braking rates and as such use a β-vector (i.e. a list or sequence of variable parameters describing the vehicle's behaviour within the DR model) for the optimisation method, tailored for that specific variant. These braking variants also used a tailored differential equation, as expressed in Equation (4.3), in which the vehicle's mass was accounted for separately to allow the vehicle coefficients not to become too small, which in turn decreases the impact of floating point errors. The second order polynomial braking functions are rewritten from the standard form of $b = k_0 + k_1 * x + k_2 * x^2$ to the vertex form $b = k_a * (x - k_h)^2 + k_b$ with variables k_a and k_h further defining them in a relation of k_b and Δx, namely $k_a = (-4 * k_b)/(\Delta x^2)$ and $k_h = 0.5 * \Delta t$ or $k_h = v_{end} + 0.5 * \Delta v$, with x being either the dimension time (t) or velocity (v) that the braking variant depends on.

This is done to provide a consistent behaviour of the polynomial braking variants applied to a braking regime and to produce verifiable markers defining the braking regime in space-time and velocity, while eliminating an extra element in the β-vector for the optimisation algorithm to account for. For the piece-wise functions 'Dual-Step' and 'Triple-Step', a linear function is used to define the brake application and release, defined as $b = l_0 + l_1 * t$ with the coefficients l_0 and l_1 expressed in terms of time difference Δt, switching moment t_{switch} and maximum braking rate k_b for the same consistent behaviour and verifiability as the second order polynomials.

$$
\begin{aligned}
\frac{dv}{ds} &= \frac{-(\frac{(r_0+(r_1*v)+(r_2*v^2))}{m} + b)}{v} \\
\frac{dv}{ds} &= \frac{-(\frac{(r_0+(r_1*v)+(r_2*v^2))}{m} + (k_a*(v-k_h)^2 + k_b))}{v} \\
\frac{dv}{dt} &= \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + (k_a*(t-k_h)^2 + k_b) \\
\frac{dv}{dt} &= \begin{cases} \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + (l_{0_0} + l_{1_0}*t) & 0 \le t \le t_{switch} \\ \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + (l_{0_1} + l_{1_1}*t) & t_{switch} \le t \le \Delta t \end{cases} \\
\frac{dv}{dt} &= \begin{cases} \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + (l_{0_0} + l_{1_0}*t) & 0 \le t \le t_{switch_0} \\ \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + k_b & t_{switch_0} \le t \le t_{switch_1} \\ \frac{-((r_0+(r_1*v)+(r_2*v^2))}{m} + (l_{0_1} + l_{1_1}*t) & t_{switch_1} \le t \le \Delta t \end{cases}
\end{aligned}
\tag{4.3}
$$

Other alterations made, are firstly to calculate the differentials for the time-dependent braking regimes in the time dimension, instead of the distance dimension, in order to simplify the differential equation with two dependent variables to a single dependency differential equation. This is done by removing the division by velocity and spanning the braking regime over a timespan instead of a distance, as seen in Equation (4.3). To reunite these braking regimes with the rest of the deceleration regime, the distance points are derived from the velocity curve and a chosen time step over which the braking regime is sampled at. Secondly, the resistance functions for the gradient r_g and curvature r_c were removed. The effects assumed to be negligible due to the low gradients and the large radius curvatures over the chosen case study corridor.

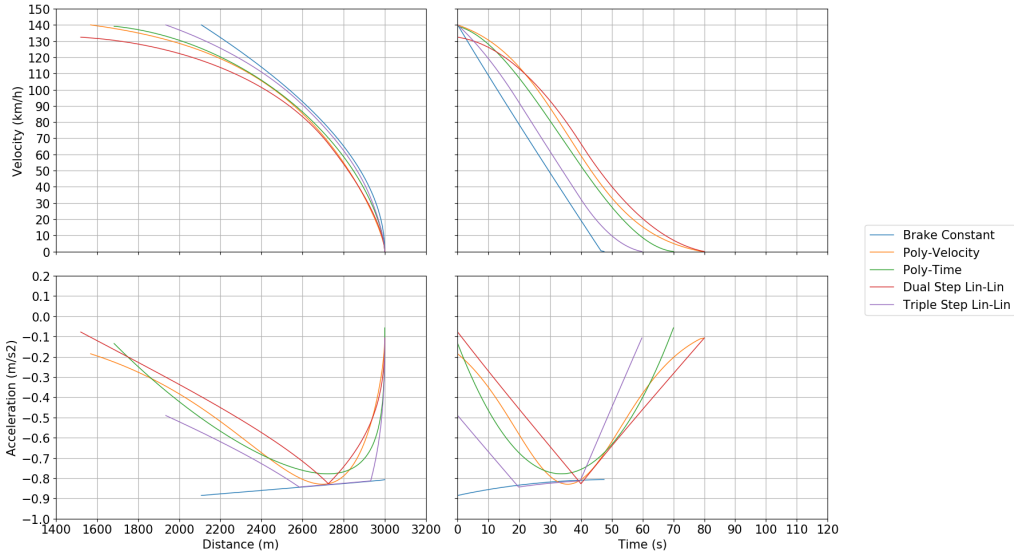


Figure 4.8: Example plot of all proposed braking variants sampled to resemble constant braking coefficient deceleration in the distance domain and the effects when comparing the variants in the time domain side by side.

When comparing the different braking variants in Figure 4.8, looking at the typical speed profile domain, Distance, the presented examples appear very similar in shape. However, when the domain is shifted to that of Time, the presented examples show their clear and distinct differences in their deceleration with a significant difference in their time required to come to a stop. Note that all variants reach close to the same maximum deceleration rate, but how the difference in brake application has an impact on the speed profiles in both distance and time.

Constant Braking Coefficient

The only uniform braking variant and the variant that is used in the development of running times and speed profiles for use in current network timetable designs and simulations. This approach perceives the braking application in a very binary on/off manner. Note that the deceleration curve is constant in the time domain and is only curved due to the velocity dependent vehicle resistance, looking at the highlighted red line in Figure D.2.

Second order polynomial (velocity dependent)

The first group of non-uniform braking variants to be examined is the second order polynomial braking variants, similar to Maurya and Bokare (2012). In that paper they made a case for the braking behaviour to be considered velocity dependent, in which drivers seemed to increase their braking rate in continuously smaller steps, until they reached a critical velocity after which they slowly started to release their brakes and decreased their braking rates in increasingly larger steps. Note that in this research, this braking variant (see Figure D.3) is considered symmetrical in approach with the critical velocity being halfway the intended velocity delta during the braking application.

Second order polynomial (time dependent)

The second polynomial variant added to this research to test for alternative dependencies. As distance at higher velocities is harder to judge accurately, in this research the alternative second order polynomial is considered time dependent as it is as easily observed as the speedometer.

The equation describing this braking variant is identical to that of the velocity dependent second order polynomial, except for using the variable time and delta time to adhere to. In Figure D.4, note the similarities between both second order polynomials with the exception of the (lack of) 'flare' at the start and end of the braking regime, the wider vertex nose at the maximum deceleration rate and the slight skewness of the velocity dependent variant.

Dual-Step & Triple-Step Braking

This group of braking variants is developed based on the two stage behaviour Maurya and Bokare (2012) observed in the braking behaviour of truck drivers. As both trucks and rail vehicles are big vehicles with a large mass, it was deemed a reasonable assumption to expect comparable behaviour between trucks drivers and rail vehicle operators. This research shows preference for testing the time dependent Dual-Step variants as it is assumed that the time component is easier to observe for the driver and would adjust according to time rather than vehicle velocity. The selection for either one of the dependency types was mainly to minimise the testing of near identical braking applications.

This research tests the following combinations of brake application and release: "Linear - Linear"(see Figure D.5 and Figure D.6). The combinations are limited to linear braking application and release, but could be expanded with an second order polynomial brake application and an inverted second order polynomial brake release. The theory behind the option of the polynomial combinations, is that the closer a train operator gets to the desired braking rate, the more he tries to fine-tune towards his goal, using increasingly smaller adjustments.

To allow for partial (non-)uniform braking behaviour, the Triple-Step approach is introduced. This approach is identical to the Dual-Step approach with the difference of a third step between the brake application and brake release. In this research, this step is considered to be the use of a constant braking coefficient with the theory that the operator might maintain his desired braking rate for a while over longer braking regime or larger desired velocity delta's.

4.4.4. Determining Deceleration Regime

Two methods were used for the analysis to determine the deceleration regime of the realised train runs. Both reconstruct the deceleration regime by backtracking from vehicle standstill, assuming this was an intended station stop, until the conditions are met from which the deceleration regimes are considered to have started on. The reason for backtracking over the realised data, compared to following the train over its realised train run, are due to the integration problem and differential equation. When solving the integration problem presented with Equation (4.2) , only the vehicle velocity v_{end} at the end of the deceleration regime to a complete stop can be used as a boundary condition, as it is the only velocity that can empirically and accurately be determined.

With the boundary condition $v_{end} = 0$ and the differential equation having solely negative coefficients, the logical solution approach is to backtrack from the last point in the deceleration regime and calculate and optimise the rest according to this fixed boundary condition. Shown in Figure 4.9 are the process structures related to the two reconstruction methods that are used in this research and elaborated on further down in this section. Figure 4.9a visualises the more complex, iterative and computational heavy approach of Point-to-Point Differential Backtracking and grouping of the identified driving regimes between the tracking datapoints. This method is useful for directly identifying and grouping without prior knowledge of possible deceleration regimes. The more streamlined approach, visualised in Figure 4.9b, has fewer iterative steps as it does not need to build its own deceleration regime profile, due to prior knowledge and pre-defining of the deceleration regime profiles. This method was developed to reduce the heavy computation time experienced in the pilot analysis, which used the first backtracking method.

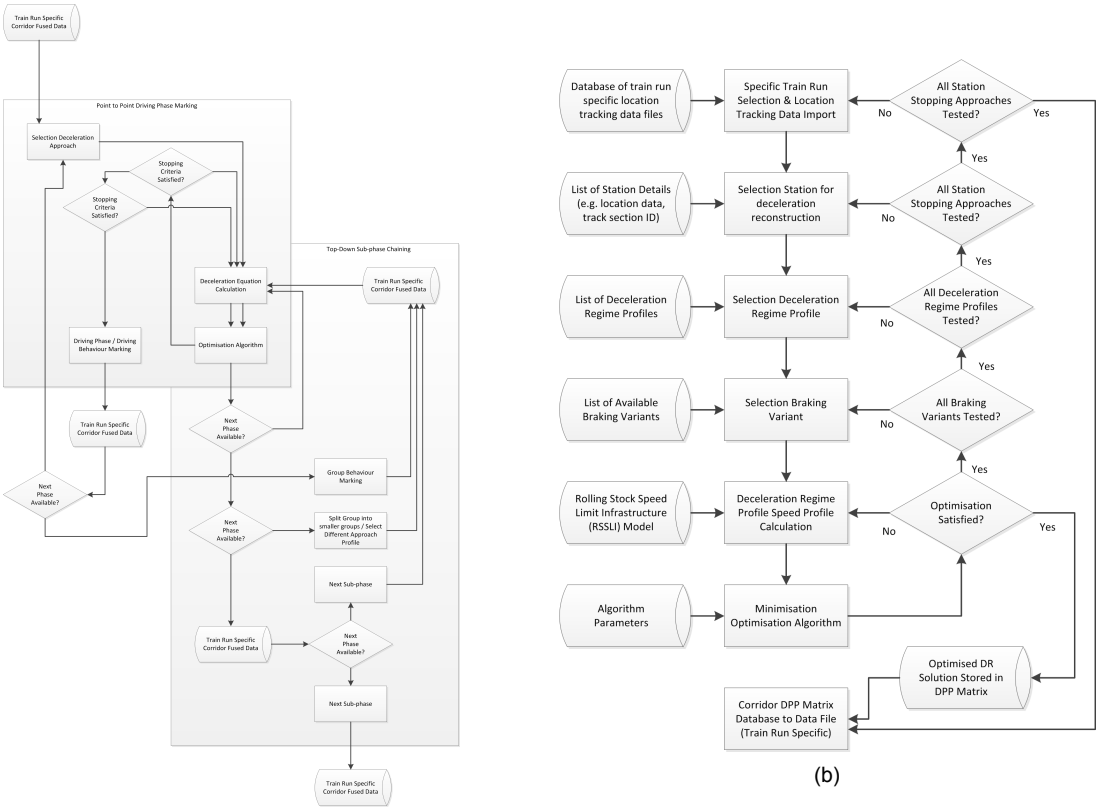


Figure 4.9: Deceleration Reconstruction Models displaying both reconstruction methods tested.
(a): Point-to-Point Differential Backtracking, comparing differential parameters to determine driving sub-regime and grouping points reevaluating differential parameters. For larger diagram see Figure C.2
(b): Velocity Difference Backtracking, comparing difference velocity and deceleration regime profiles. For larger view see Figure C.3

Sub-Regimes

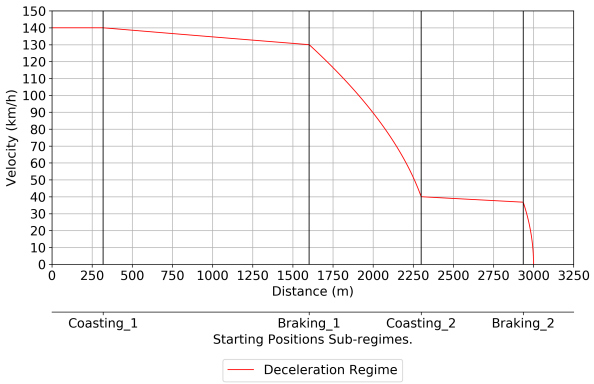


Figure 4.10: A deceleration regime presented with several sub-regimes implemented and their starting points marked. All these (sub-) regimes belong to the same singular intent of decelerating to a stop.

Before continuing with the elaboration of the two reconstruction methods, the term sub-regime has to be explained in detail and visualised(see Figure 4.10) to fully understand the reconstruction methods described further below and the deceleration regime profiles derived from these methods. During the analysis pilot and while using the ”Point-to-Point Differential Backtracking” method, the definition of sub-regime is the distinction made between neighbouring regimes identified in the reconstruction method as being the same driving regime but having distinctly different regime characteristics. They are named sub-regimes, as they are smaller driving regimes that otherwise would have been considered a single, larger regimes(i.e. two braking regimes with different braking coefficients instead of combined into a singular regime with a single and averaged braking coefficient).

However, with the introduction of the non-uniformity of the braking coefficient, these cases get covered under the ’Dual-Step’,’Triple-Step’ or polynomial defined braking functions, with the exception of an illogical application of the braking coefficient in an oscillating fashion (i.e. hard-soft-hard-soft application of the brakes). Due to the introduction of non-uniform braking and deceleration regime profiles, the definition of the sub-regime is altered to describe all of the smaller driving regimes applied within a singular operational intent (e.g. all the regimes applied within the same intent to decelerate to a complete stop at a station), which can be seen in Figure 4.10. This last definition is used in the reconstruction method ”Velocity Difference Backtracking”.

Method 1 - Point-to-Point Differential Backtracking

Without prior knowledge of the deceleration regime composition profiles, the identification of the estimated driving regime between the datapoints and the grouping of similar neighbouring regimes are used to determine the realised deceleration regime and would simultaneously allow for the identification of the different driving regimes within the deceleration regime and, if present, the distinction between similar driving regimes with different coefficients. For that purpose, the dynamics differential equation (Equation (4.2)), its integration and the minimisation problem (Equation (4.6)) are used to determine the deceleration regime and vehicle characteristics. From that, the driving regime is identified and a decision made upon the outcome on if the deceleration regime found in its entirety. The process steps in determining each sub-regime within the entire deceleration regime is for this method as follows:

1. Backtracking from vehicle standstill.
2. Point-to-Point sub-regime identification through obtaining regime characteristics from the dynamics equation.
3. Chaining and group testing sub-regimes for consolidation.

The implementation of this method for this research, as shown in detail in Figure 4.9a, is to take the available data of a specific train run and start the backtracking in the given datapoints to search for a point near the stations at which the vehicle velocity is near zero, allowing for some fluctuations of the GPS sensor tracking error. From that point, it will then backtrack point by point, while calculating the differential equation,through integration and minimisation optimisation, with a uniform braking coefficient, as described in

Equation (4.2), and determine the corresponding coefficients. From these coefficients, the realised driving regimes are estimated, based on the characteristic conditions described in Section 4.4.2.

This cycle continues until the reconstruction model estimates the driving regime to be acceleration, at which it breaks of the cycle and returns the current built-up chain of differential coefficients as the deceleration regime. It continues its search locating a point at which the vehicle is deemed to be at a standstill and continues this cycle until it runs out of datapoints to process within this trainrun. After estimating the driving regimes corresponding to the datapoints, the reconstruction method attempts to group similar driving regimes by grouping the datapoints of neighbouring sub-regimes and recalculating the differential coefficients, allowing non-uniform braking variants as described in Section 4.4.3 , and evaluates the grouped values to their individual coefficients for consistency and identification of neighbouring driving regimes with distinctly different differential coefficients. After finalising the grouping of the datapoints, the reconstructed deceleration regimes for that specific train run are stored in a database for later use in the statistical analysis. This larger iterative cycle continues until it runs out of train runs to process from the given datasets.

This method has the benefit of not relying on the velocity data gathered by the vehicle GPS or infrastructure sensors, or uses any prior knowledge on deceleration regimes over a given corridor. However, this method proved to be too heavy in computation time for its purpose, due to the large number of iterations and cycles. By applying this method on a sampleset of location tracking data, insights are gained about the possible deceleration regimes being applied. Possible deceleration regime profiles can be pre-defined from this knowledge to apply a more streamlined method of reconstructing the realised deceleration regime from location tracking data, leading into the second method: ”Velocity Difference Backtracking”.

Method 2 - Velocity Difference Backtracking

With computational efficiency in mind, reevaluating the deceleration regime analysis with the gained knowledge on the realised deceleration regime profiles lead to the development of a reconstruction method comparing the vehicle velocity per data point. The momentary vehicle velocity would be preferred, when available and deemed reliable. However, testing has shown that the use of the average vehicle velocity, calculated from the difference in distance and time, worked as a great alternative.

While backtracking from standstill, this method essentially compares $v_i \geq v_{i+1}$, to determine the location tracking datapoint defining the start of the deceleration regime. This start point would either be defined as a cruise or coast sub-regime to allow for the reconstruction model to identify the cruising sub-regimes between two braking sub-regimes in the realised deceleration regime. Similar to the first reconstruction method, this comparative backtracking stops when it estimates the regime between the datapoints to be acceleration. However, as a difference to the first method, the second method only returns the datapoints estimated to be the starting point and end point of the deceleration regime.

The analysis then continues to analyse the determined distance to find the best fit of the pre-defined deceleration regime profiles, using gained knowledge from the first method, to the realised datapoints in terms of sub-regime composition and braking rate variants. This reconstruction method’s process cycle can summarised in three steps:

1. Backtracking from vehicle standstill.
2. Point-to-Point sub-regime identification through $v_i \geq v_{i+1}$ comparison.
3. Deceleration regime fitting sub-regime composition and deceleration rate variants.

The implementation of this second method during this research, as shown in detail in Figure 4.9b, is to take the available data of a specific train run and apply the same backtracking as described in the first method. The difference between the methods lies in the comparative conditions that determine the starting point of a deceleration regime. Instead of

calculating the differential coefficients between every datapoint, in this method it is simplified to a singular comparison between datapoints of the condition $v_i \geq v_{i+1}$. This loses the benefit of knowing where and how long each sub-regime lasted and loses the identification of the driving regime estimated. This alteration, however, allows for the removal of a significant number of iterations. Due to the loss of the duration, location and identification of the estimated sub-regimes, knowledge is required to pre-define profiles of the deceleration regime sub-regime composition.

With the deceleration profiles pre-defined and the datapoints determined that estimate the starting point and end point of the deceleration regime, all the pre-defined profiles are evaluated in the deceleration reconstruction model, returning all the coefficients required to describe the profiles (Section 4.4.5) and their goodness of fit (i.e. a value describing how well the fitted function aligns with the evaluated datapoints). All the profile coefficients and their goodness of fit are stored along with all the other deceleration regimes evaluated within that specific train run, for manual validation and for statistical and comparative analyses at a later stage.

For this research, the sub-regime composition profiles are pre-defined manually, determined from the analysis pilot using the first reconstruction method and assessing the graphical outputs. The deceleration rate variants are introduced after the development of the first reconstruction method and the literature review of Maurya and Bokare (2012).

4.4.5. Deceleration sub-regime composition

The handling of multiple deceleration regimes in a single train run or corridor, is covered in two different ways depending on the used DR analysis method. When using the ”Point-to-Point Differential Backtracking” method, the model determines the realised driving regimes at a Point-to-Point sub-regime level at first. To handle multiple cycles of driving regimes, used in a realised train run, it forms a list or chain of driving regimes with independent coefficients.

When this analysis method is used for the DR model, these lists are then scanned to determine possible groupings of neighbouring similar sub-regimes and then recalculated as a single driving regime to validate it as a whole. This method focuses on the vehicle and regime characteristics of the coasting and braking regimes. The separate analysis of both regimes leads in some cases to where both regimes are identified, due to the single data point gap smoothing applied in the grouping phase of the analysis for sub-regime chains and the single data point ’edge’ smoothing for singular registered sub-regimes. This is due to the sensitivity of differential coefficients and the slight fluctuation of the realised vehicle velocity measured by the GPS sensor.

The ”Velocity Difference Backtracking” method uses the insights gathered from the analysis pilot to determine the more likely used deceleration regime profiles to use them as predefined sub-regime composition profiles to fit to the realised train run data. The sequences within these profiles are given by the abbreviations of the driving regimes coasting (Co), cruising (Cr) and braking (Br). In general, the profiles observed within the analysis pilot are listed in Table 4.2 and grouped in braking regime families, named after the number of braking regime found in these deceleration regime profiles (e.g. ’2-Br’ grouping deceleration regime profiles with two braking regimes). Profiles prefixed with UZI, describe profiles with a leading coasting regime instead of the implicitly defined cruising regime. The ’Single’ or ’Double’ refers to the number of coasting/cruising regimes between braking regimes, observed as a ’stepped’ deceleration in the speed profiles.

Table 4.2: Sub-regime composition profiles of different deceleration regimes.

Name	Sequence	Family	Optimisation Vector
Singular Braking	Br	1-Br	$\beta = [r_0, r_1, r_2, S_0, \Delta v_0, b_0]$
UZI Deceleration	Co-Br		$\beta = [r_0, r_1, r_2, S_0, S_1, \Delta v_0, b_0]$
Single Coasting Step	Br-Co-Br	2-Br	$\beta = [r_0, r_1, r_2, S_0, S_1, S_2, \Delta v_0, \Delta v_1, b_0, b_1]$
UZI Single Coasting Step	Co-Br-Co-Br		$\beta = [r_0, r_1, r_2, S_0, \dots, S_3, \Delta v_0, \Delta v_1, b_0, b_1]$
Single Cruising Step	Br-Cr-Br		$\beta = [r_0, r_1, r_2, S_0, S_1, S_2, \Delta v_0, \Delta v_1, b_0, b_1]$
UZI Single Cruising Step	Co-Br-Cr-Br		$\beta = [r_0, r_1, r_2, S_0, \dots, S_3, \Delta v_0, \Delta v_1, b_0, b_1]$
Double Coasting Step	Br-Co-Br-Co-Br	3-Br	$\beta = [r_0, r_1, r_2, S_0, S_1, \dots, S_4, \Delta v_0, \Delta v_1, \Delta v_2, b_0, b_1, b_2]$
UZI Double Coasting Step	Co-Br-Co-Br-Co-Br		$\beta = [r_0, r_1, r_2, S_0, S_1, \dots, S_5, \Delta v_0, \Delta v_1, \Delta v_2, b_0, b_1, b_2]$
Double Cruising Step	Br-Cr-Br-Cr-Br		$\beta = [r_0, r_1, r_2, S_0, S_1, \dots, S_4, \Delta v_0, \Delta v_1, \Delta v_2, b_0, b_1, b_2]$
UZI Double Cruising Step	Co-Br-Cr-Br-Cr-Br		$\beta = [r_0, r_1, r_2, S_0, S_1, \dots, S_5, \Delta v_0, \Delta v_1, \Delta v_2, b_0, b_1, b_2]$

No profiles with neighbouring braking sub-regimes (i.e. …-Br-Br-…) are identified as likely sub-regime composition profile. This is due to the simultaneous evaluation of the (non-)uniformity of the realised braking rate applied in the deceleration regime, which implements Dual-Step and Triple-Step braking variants. The β-vector for the optimisation method, expressed in Table 4.2, represent the vectors used for the braking variant 'Constant Coefficient'. For the non-uniform braking variants, b_i is replaced with k_{b_i} and for the time-dependent braking variants the vectors are expanded with variable Δt_i. In case of the piece-wise braking variants, the β-vector is expanded with the variable defining the switching moment t_{switch_i}.

4.4.6. Optimisation Problem & Algorithm

The optimisation problem in the reconstruction model of the realised deceleration regime, is an error minimisation problem. While Bešinović et al. (2013a,b,c) and Medeossi et al. (2011) used a minimisation of the error in time or velocity respectively, this research has the benefit of enjoying strengths of both data sources (i.e. Vehicle GPS Data and Train Describer Data) and therefore can use both the vehicle velocity as distance-time data to minimise the error of the reconstruction calculation. For the deceleration reconstruction, the sum of the integral X is used, defined as a surface area underneath the velocity curve, describing the speed profile, over distance.

$$X_{n'} = \int_{s_n}^{s_{n+1}} \int_{s_n}^{s_{n+1}} \frac{dv}{ds}\, ds\, ds \tag{4.4}$$

$$X_{n'_{avg}} = \frac{s_{n+1} - s_n}{t_{n+1} - t_n} * (s_{n+1} - s_n) \tag{4.5}$$

Equation (4.4) is used to calculate the surface area under the speed profile of the reconstructed deceleration regime. For the realised data, it was simplified to the average velocity between the two data points, calculated over the difference in distance, leading to Equation (4.5). The time component t is expressed with an accuracy of whole seconds, The optimisation problem is defined as follows:

$$\begin{aligned}
&\text{Minimise} && \sum_{n'} |X_{n'} - X_{n'_{avg}}| \\
&\text{Subject to:} && \frac{dv}{ds} = \text{see Equation (4.3)} \\
& && v_{end}(0) = 0 \\
& && v_{start}(n+1) = v_{end}(n) \\
& && S_i \in [S_i^{lb}, S_i^{ub}] \text{ for } i = 0 \dots 5 \\
& && b_k \in [b^{lb}, b^{ub}], \text{ for } k = 0 \dots 2 \\
& && \delta v_k \in [\delta v_k^{lb}, \delta v_k^{ub}] \text{ for } k = 0 \dots 2 \\
& && r_m \in [r_m^{lb}, r_m^{ub}], \text{ for } m = 0, 2 \\
& && \sum_k dv_k \leq v_{speedlimit} \\
& && \sum_i S_i \leq S_{deceleration}
\end{aligned} \tag{4.6}$$

Note that there is a distinction between the optimisation variable 'Distance Sub-regime' being uppercase S and regular description for distance being lowercase s. The variables b, r and δv describe the braking rates, resistance coefficients and the velocity differences respectively and the indexes i, k and m relate respectively to the index of the sub-regime and braking regime in the tested deceleration regime and to which coefficient of the Davis equation (Equation (4.1)) is referred to. Lastly, index n refers to the index of data points used in the optimisation process, with N being the count of data points and index n' referring to the space between data points n and $n+1$. The one-side bounded optimisation leaves the other end open to errors in the momentary vehicle velocities measured by the GPS sensors. With the reliability assessment of the GPS velocity data, the realised momentary vehicle velocities are used to validate the deceleration regime reconstruction. This validation is done with a Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), as seen in Equation (4.7).

$$\begin{aligned}
\text{MAE} &= \frac{\sum_{n=0}^{N} \left(|v_{sim_n} - v_{real_n}|\right)}{N} \\
\text{RMSE} &= \sqrt{\frac{\sum_{n=0}^{N} \left(v_{sim_n}^2 - v_{real_n}^2\right)}{N}}
\end{aligned} \tag{4.7}$$

The minimisation problem is solved using a Genetic Algorithm (GA), which is a solving algorithm that inherits its properties from the evolutionary theory in nature and "Survival of the Fittest". (Goldberg, 1989; Holland, 1992; Mitchell, 1996) There are different variations in evolutionary algorithms. The one used in this research starts with a population of different solution vectors, of which some are entered based on nominal values or boundary constraints and the rest randomly drawn from a uniform distribution between the vector variables' boundaries. Their "fitness" is tested by calculating the answer to the equations in the minimisation problem and a selection is made to form a pool of solutions (i.e. parents) used to produce new solutions by mixing the vector variables and mutate the variables slightly (i.e. breed offspring through genetic crossover and mutation). The new population pool of solutions consist firstly of the parents and their offspring, which is then resupplied with new randomised solutions to maintain the original population pool size of

solutions(i.e. new blood or immigration). This process is continued over several generation until either the fittest solution is good enough or until a specified number of generations is reached and the fittest solution is taken from the last generation population pool of solutions. The algorithm's parameters and settings are discussed in Chapter 5.

4.5. Post-Processing

The post-processing of the results after analysing the realised train runs with the deceleration reconstruction model, involve statistical and comparative analysis of the data inline with the performance indicators used in this research. This is done by calculating the performance indicators, as described in Chapter 3, per individual run and furthermore determines the distribution per given parameter of the given performance indicator. From that, the statistical distribution type and coefficients are estimated and compared to the nominal performance indicator values.

The comparative analysis takes the realised train runs and applies some differentiating characteristics (e.g. seasonality, delay/timeliness) that are suspected influences on the applied deceleration regime. Of these divisions, the statistical distributions are determined through the statistical analysis process and any significant correlations are drawn from this. This again will have the nominal performance indicators or behaviour characteristics present as reference.

Finally, the analysis results of both the reconstruction model and statistical model will be stored for reference and further use in refining the comparative analysis. Also a graphical representation of both the realised train runs, fitted deceleration regime profiles for the spatial data and the statistical distributions and their statistical coefficients of the resulting performance indicators and coefficients will be generated. This to provide a clear overview of the results and makes it easier to draw any conclusions.

4.6. Chapter Summary

This chapter describes the analysis methodology and the development approach of a data-driven reconstruction model to estimate the speed profiles of realised train runs that elaborates on the deceleration regimes in a more dynamic and generalised (i.e. able to be applied to any rail network corridor) manner to provide a more detailed description of the realised deceleration behaviour. This research introduces a conceptual framework for the reconstruction of the deceleration behaviour of heavy rail vehicles. The general structure and concept approach are inspired by Bešinović et al. (2013a,b,c) and Medeossi et al. (2011).

The proposed framework differentiates from its inspiration by its interest and focus on the deceleration behaviour within the proposed reconstruction model, aptly named Deceleration Reconstruction (DR) model, instead of reconstructing complete velocity profiles. This is done through the implementation of sub-regimes (i.e. smaller driving regimes describing one of four driving states of a heavy rail vehicle) and deceleration regime profiles defined by a combination of said sub-regimes (see Section 4.4.4).

These sub-regimes are subject to the rail sector defined differential equation, describing Newton's second law of motion, with the vehicle resistance expressed by the Davis equation (Davis, 1926). This equation is further expanded in this research to accommodate the testing of non-uniform braking behaviour by writing the braking rate as a function of either velocity or time, inspired by Maurya and Bokare (2012) and described in Section 4.4.3. The framework incorporates a statistical and comparative analysis because, besides the realised train trajectories and corresponding parameters, the statistical distributions of these parameters and related performance indicators (PI) and their interrelations are considered the end products for this research. This is described and elaborated in Section 4.5.

Furthermore, this research introduces the concept of 'Data Fusion' (Section 4.3.5) for all the realised location tracking data to be used in the DR model (Sections 4.2.3 and 4.2.4). The datasets used from the different sources (i.e. the single source of train describer data and 1 to 2 sources of GPS data) are formatted, corrected and derivable data is created to unify, expand and fill in the gaps for the provided data sources (Sections 4.3.1 to 4.3.3). After the aligning the different location tracking data sources (Section 4.3.4), they are fused in which the strengths of all the different sources are inherited while their individual weaknesses are compensated. Once fused, the dataset of individual train runs are filtered to discard the undesired, hindered or invalid train runs before feeding the DR model with the fused location tracking data of the selected realised train runs.

The DR model, as elaborated on in Section 4.4, has several layers of nested for loops in its programming, testing each pre-defined deceleration regime profile and braking variant for each station approach in the corridor of a single realised train run. Each combination of deceleration regime profile and braking variant has their own specific β-vector (i.e. a list or sequence of variable parameters describing the vehicle's behaviour within the DR model) for the optimisation algorithm, whose structure is developed automatically in the dataset and coefficient preparations preceding the optimisation of the minimisation problem.

The minimisation problem is defined as minimising the sum of absolute errors of the surface areas under the speed profile between each pair of data points , with the surface area under the realised speed profile defined by the average velocity multiplied by the distance between the two data points and the area under the estimated speed profile defined by a double integral over distance of the velocity differential equation describing the estimated behaviour for a specific sub-regime (Section 4.4.6).

The optimisation algorithm used on this non-linear minimisation problem, is a Genetic Algorithm (GA) method. For this research, a customised GA is developed, which is dubbed "Elitism with Randomised Population Migration and Diminishing Mutation". This is an algorithm which retains a small pool of best fitting solution vectors for both the next generation and for developing the 'offspring' vectors through cross-over and mutation with a diminishing mutation range, while refreshing and maintaining the general population size through the migration of a group of completely randomised vectors.

5

Model & Statistical Analysis

This chapter will discuss the findings regarding the model's performance and development, the results from the empirical, statistical and comparative analysis on the fused datasets of location tracking data from aprox. 13 300 realised train runs and track section infrastructure occupation of the track sections over the entire corridor, and the results from applying the DR model on 280 realised train runs pertaining to a case study set out in Section 1.5.

The scope of the case study is further reduced to focus on 3 of the 8 stations in the corridor "Ht - Ut" and for two characteristic weeks instead of the entire year of 2017. This is done for performance reasons as discussed in Section 5.1.2. The stations examined are Geldermalsen (Gdm, a large station with a speed restricted switching area), Houten (Htn, a small station with a short distanced, open-track station approach) and Zaltbommel (Zbm, a small station with a long distanced, open-track station approach). For the analysis with the DR model, the train runs of both the last week of March and the first week of September are tested. For the time-related tests (i.e. running times, deceleration loss times and track section occupation duration), the selection of train runs is kept to the entire year as these tests were not as CPU-intense.

The analyses applied on the DR model, are a performance analysis to test the model's analysis speed and result accuracy, a statistical analysis on the performance indicators described in Chapter 3 resulting from the DR model or corridor's network timetable, and a comparative analysis is applied to investigate the relations between the different performance indicators.

5.1. Model Analysis

When implementing the DR model and optimisation method on the chosen case study, some interesting results and consequences were found with regards to the model implementation, data processing and modelling performance, and the effects this had on the case study scope, statistical and comparative analysis. Below will discuss the details of the application of the model and the concessions made based on the modelling performance.

5.1.1. Model Setup

Preceding the details of the developed DR model, a quick introduction of the used software and hardware is given to provide context to the discussion about the model's development. After which the setup of the model and data processing is discussed in detail.

The programming language used to build and implement the DR model, process the (test/pilot) data and run a statistical and comparative analysis over the model results, was Python 3.6 (v3.6.5) with a standard suite analysis libraries, provided by Anaconda3. The

libraries used, were NumPy (v1.15), Numba (v0.39.0), SciPy (v1.1.0), Pandas (v0.23.4) and Matplotlib (v2.2.3).

The hardware used to test the performance on the pilot data, at the university was a system with an Intel e5-1620v2 with 8Gb RAM DDR3 and Operating System (OS) Windows 7, at home an AMD 2200G with 8Gb 2400MHz DDR4 RAM and OS Windows 10, and lastly at NS a system with an Intel Xeon e5-1650v4 with 16Gb 2400MHz DDR4 RAM and OS Windows 8.1 installed. The data usage of the case study presented in this research was completely run locally on the system provided by NS.

Code Structure

The process structure of the framework, DR model with Velocity Difference Backtracking, the minimisation problem and the Genetic Algorithm as described in Chapter 4, were used in the development of the Python programming code used in this research. The different processes of the conceptual framework had their own individual python file with all the related main functions and supporting local settings and functions, which was imported, ran and controlled from the main python file. This main file held all the input parameters, common variables and file path structures, along with a 'switchboard' of Boolean variables, using the 'True' or 'False' statement to turn parts of the main control file on or off. This control file allowed for the analysis to be broken up into smaller running sessions. The 'communication' between the different processes happened through the partially processed/analysed data stored after process was used on the data. This allowed for a 'staged' analysis and writing process, by not having to restart the analysis entirely at the beginning each time an alteration, adjustment or unit test was done on a code segment. By breaking the coding into their distinct processes and roles, it helped with maintaining clearer view of the process structure at a higher abstraction level, while making the processes and their internal functions easier to navigate.

Data Pre-processing

The data pre-processing for the case study is done as described in Section 4.3 with the exception of the omission of the data filtering based on signalling profiles, deciding to rely completely on the V-box filtering. This was due to the complexity of how the signalling filter would be implemented and minimal benefit due to the filtering redundancy. For the Even running side of the corridor, this meant the following list of parameters (resp. Distance Start (m), Distance End (m), Velocity Threshold (km/h)): [5000,12000,80], [16500,20000,90], [20000,21000,60], [21000,21800,30], [26000,29000,80], [29000,29600,50], [33000,37000,90], [39600,40200,50], [42000,44400,50], [45600,46500,50], [47400,47600,30].

For the data alignment, both the vehicle MTPS data and the infrastructure TROTS and InfraAtlas data were given a common GPS anchor point to align to. These points were chosen near the back end of the station platform of the first station in the corridor run. For the stations 's Hertogenbosch (Ht), Utrecht Centraal (Ut) and Geldermalsen (Gdm), the following GPS points were used respectively: [51.6874554, 5.2923015], [52.0899235,5.1084159], [51.8807017,5.2723727,22300]. The MTPS data was aligned by calculating the distance of the first GPS point available per each of the rolling stock combination. The infrastructure was slightly less direct, as there were no GPS coordinates available of the track sections. Given the InfraAtlas infrastructure data, every infrastructure part has a known relative position to the first infrastructure entry of the trainpath. Among those infrastructure parts, are the train signals of which a list of GPS coordinates was available. Therefore the distance of the first signal was calculated to the GPS anchor point and from there every other part of the infrastructure data was aligned. The alignment of the GPS sensor to the head of the train and to the infrastructure was done as described in Section 4.3.4.

Results Pre-processing Realised Data

When observing the results of the data fusion, it shows that the intended result of combining multiple location tracking sources is achieved. Some refinements are left to be made in vehicle location estimation and data source alignment, but this will largely come from using a more consistent and more frequently sampled vehicle-sided location tracking data and the knowledge of the relative position of the GPS sensor within the vehicle. Other refinements can be made around the station when the vehicle is nearing a complete stop. The relatively small errors in the GPS tracking start to have a larger effect at such low velocities seen close to a vehicle's stopping moment. Another issues with the inherent errors from GPS tracking are that a vehicle is never perfectly still in the GPS data, making it difficult to determine the exact stopping location for the DR model to use the data or for the calculations of the performance indicators. These effects mentioned can be seen in the speed profile plots over the corridor as large steps in both time and the velocity that is averaged over a longer distance between two known points. The other effect is that in the speed profile, the vehicle velocity never reaches zero, but only ever so often gets very close to it when the data point density of the vehicle-sided location tracking data is high and consistent enough. The effects of using the vehicle's GPS to pinpoint the deceleration behaviour inside the station can be seen in the distance-time plot, in which a vertical 'bouncing' is observed near the station stops. As the speed profile of the realised data was more important to this research, the data points were aligned and sorted first on distance followed by time index, rather than vice versa. Both manners of sorting, however, would have shown a similar 'bouncing' in the coordinates of the data points. This margin of error in the GPS data points and 'bouncing' of the data could be resolved by implementing a more finer data mesh-grid, infrastructure-sided tracking method within the station.

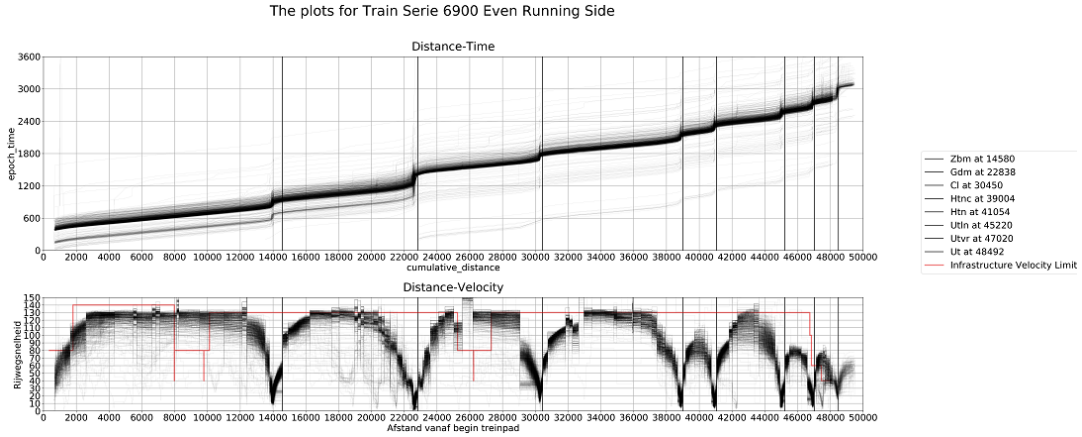


Figure 5.1: Unfiltered Fused Realised Data - Even Running Side for the entire month of September 2017 (# train runs: 1046, stations marked in legend)

After the fusion, a filtering of the train runs was required as can be seen in Figure 5.1 with some visibly disrupted station approaches between 18 000 and 22 000 m and around 29 000 m. The 'V-Box' filtering was implemented as a simpler approach for filtering out the train runs who had a hindered station approach, as well as to filter out the invalid train runs of which the data failed to align at the designated stations, either station 's Hertogenbosch (Ht) or Geldermalsen (Gdm) for the shorter runs at night, or register correctly (e.g. interrupted train runs with partial records). The results of the 'V-box' filtering (Section 4.3.2) applied, can be seen in Figure 5.2 at the aforementioned distance markers, which has cleaned up the dataset while still maintaining a large enough count of train runs for any statistical significance.

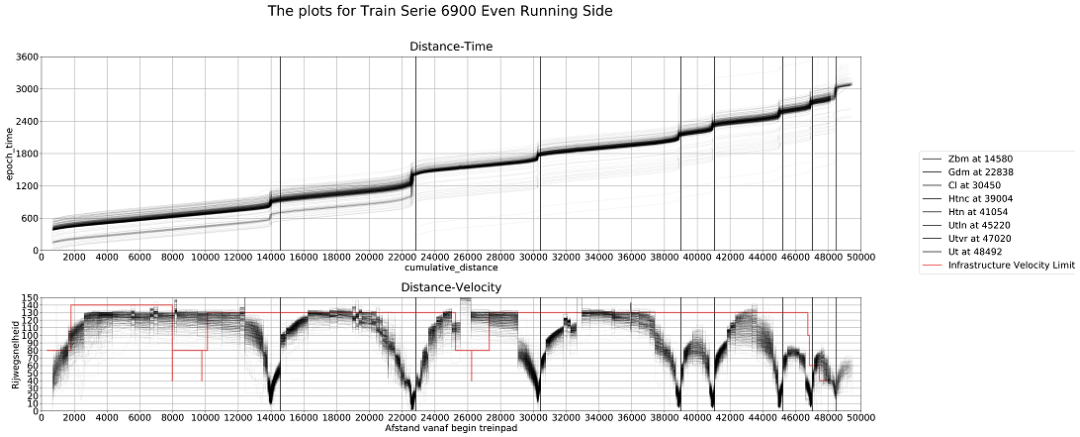


Figure 5.2: Filtered Fused Realised Data - Even Running Side for the entire month of September 2017 (# train runs: 645, stations marked in legend)

Deceleration Reconstruction Model

As for the coding of the DR model and the minimisation problem: for each of the braking variants, a different function was written to house the calculation of the deceleration curves and the sum of absolute errors. This was because the braking regime would take different variables to express the braking rates and as such used a β-vector tailored for that specific variant. These braking variants also used a tailored differential equation, as expressed in Equation (4.3).

This was done to provide a consistent behaviour of the braking regime and verifiable markers to the braking regime in space-time and velocity, while eliminating an extra element in the β-vector for the optimisation algorithm to account for. For the piece-wise functions, a linear function was used to define the brake application and release, as seen in Section 4.4.3, for the same consistent behaviour and verifiability as the second order polynomials.

Optimisation Algorithm

The optimisation algorithm used to solve the minimisation problem of the DR model, as defined in Section 4.4.6. The population pool size of vectors was set to 32 vectors. The parent/mating pool size was set to 8 vectors, while the offspring pool size was set to 16 vectors. Due to more offspring than parents, 1 parent would have 2 mates to provide both off springs (i.e. Parent 1 would breed with Parent 2 and 3, Parent 2 would breed with Parent 3 and 4, etc.). This was done to eliminate identical crossovers and to provide more diverse crossovers rather than an opposite cross-over combination compared to the first offspring. To further randomise the crossover process, a random number was drawn from a uniform distribution between 0 and 1 per vector variable for each offspring and a 50% chance was given to each parent to pass on their value of said vector variable. A similar randomness was present in the mutation phase of the pool of offspring vectors to determine if a vector variable should or should not mutate. To help converge the fittest selection, a mutation range restriction is added that will linearly narrow to a specified lower bound as the generation count progresses towards the limit of the number of generations allowed, defined as (1 - (restriction factor * (generation count / total number of generations))).

After the crossover and mutation, both the parent vectors and the offspring vectors were pooled together and resupplied with randomised vectors to keep a population size of 32 vectors. The solution evolution was given two stopping criteria, which was a limit on the number of population generations (40 generations) and a stop when the fittest vector was good enough (sum of absolute error less than 10% of the deceleration distance). This was keep a check on the optimisation time and reduces the time spent on unnecessary generations if a suitable vector is found early on.

This GA variant is dubbed 'Elitism with Randomised Population Migration and Diminishing Mutation' Genetic Algorithm. In the table below is an overview given of the details of the settings used in the optimisation algorithm and the boundaries given for the vector variables and mutations used in solving the minimisation problem.

Table 5.1: Overview Algorithm Details and Minimisation Problem Bounds

Name	Value	Name	Bounds [Lower, Upper]	Mutation Bounds [Lower, Upper]
Population Size	32	$r_0\ (N)$	[2154.15, 2632.85]	[-10.0, 10.0]
Parent Pool Size	8	$r_1\ (Ns/m)$	[77.04, 77.04]	[0.0, 0.0]
Offspring Pool Size	16	$r_2\ (Ns^2/m^2)$	[12.8304, 15.6816]	[-1.0, 1.0]
Generation Count	40	$S_i\ (m)$	[100, inf]	[-100.0, 100.0]
Restriction Factor	0.75	$\Delta v_i\ (m/s)$	[2.5, inf]	[-20.0, 20.0]
Mutation Threshold	0.75	$\Delta t_i\ (s)$	[5.0, 200.0]	[-20.0, 20.0]
Fitness Threshold	0.1	$t_{switch}\ (s)$	[1.0, 200.0]	[-10.0,10.0]
		$b_i,\ kb_i\ (m/s^2)$	[0.05, 2.0]	[-1.0, 1.0]

Nominal Coefficient Values

To use as a reference vector in the DR model and as a reference for the performance indicators, the minimum running time profile and the UZI-profile based on the 'recovery time'-supplemented and rounded running time, were used and these were based on the nominal values of the slowest train combination used on this service line (Series 6900). However, when comparing the different configurations of SLT rolling stock, it was observed that the performance of the variants were close to identical. Therefore the more common configuration, the SLT-10, was selected and its nominal values were used. These values were provided by TU Delft, but generated by Lloyd's Register or now known as Ricardo Rail. For the traction components, the c0 was provided as 297000 N. The other traction components c1 and c2 were not provided but c1 was set to zero as this was tested by Bešinović et al. (2013a,b,c) and c2 was derived from the traction curves and tables provided with the nominal dataset and observing the critical velocity at 10 m/s and therefore deriving the c2 component as 10 * 297000 Nm/s. The nominal vehicle resistance coefficients r0, r1, r2 were provided as resp. 2393.5 N, 77.04 Ns/m and 14.256 Ns^2/m^2. The nominal braking rate was provided as 0.8 m/s^2. The rolling stock configuration's mass and length were provided to be 343400 kg and 169.9 m. The mass used for the rolling stock configuration included a normal loading (i.e. assumed equally distributed) and a rolling mass factor (i.e. compensating for kinetic energy lost in rotational forces).

The recovery-time supplement used was set to 5 % of the minimum running time between stations and rounded to the nearest minute. Extra time supplements were added to the stations of 18 seconds, with the exception of station Gdm and Ut getting 60 seconds supplemented to their scheduled running time, as they are considered large stations with a priority on punctuality. Further variables given a value for a reference vector in the DR model, are the sub-regime lengths S_i, Desired Velocity Reduction Δv, Braking Regime Duration Δt, Piece-wise Function Switch Moment t_{switch} and non-uniform maximum braking rate k_b, respectively 200.0 m, 10.0 m/s, 40.0 s, 20.0 s and 0.8 m/s^2.

5.1.2. Model Performance

When developing the model and analysis code in Python, some performance testing was required as the full range of stations, deceleration regime profiles and braking variants took a long time to complete the analysis of a single train run over the case study corridor. The same test run was used for the performance testing and code profiling, implementing the cProfile library, native to Python. The analysis of the test run took approximately 7500, 1700, 1600 seconds to complete the single run with the full range of stations, deceleration regime profiles and braking variants. The first result was deemed indicative to the state of the system rather than the coding, but the latter two results seemed consistent even after rerunning the performance test per system. The 1600-1700 seconds or 26.667-28.333 minutes per train run leads to an unfeasible analysis duration if all the train runs for the entire year of 2017 are considered. Granted, the DR Analysis code is a single-processed(one work unit/thread or sequential) programming code, but this was due to the application of multiprocessing at a higher level in which the different train runs over the case study corridor were analysed in parallel of each other.

Selection Train Stations

Even when leveraging the multiprocessing library in Python, a decision had to be made to narrow the selection of train runs, stations, deceleration profiles and braking variants to test for this part of the research for the sake of time management. The selection of the train runs was decided upon through the availability of the data, quality of the data, the number of train runs and seasonality conditions. The selection of train runs settled on all the train runs in the Even ('s Hertogenbosch to Utrecht Centraal) direction for the last week of March (24-03-2017 til 30-03-2017) and the second week of September (06-09-2017 til 10-09-2017). This was done to allow for seasonality and different weather conditions to be considered, with the warm and dry weather conditions in the selected week of March and the wet weather conditions in September, while keeping a large enough dataset to hold a statistical significance.

In regards to the selection of stations, deceleration regime profiles and brake variants, the selection was made to further reduce the analysis time per train run. The stations selected were the long distanced open-track station Zaltbommel (Zbm), the large station with a speed-restricted switch area Geldermalsen (Gdm) and the short distanced open-track station Houten (Htn) in order to test the deceleration behaviour in different situations. The selection of deceleration regime profiles was done for each station individually and by observing the realised data for the several months, a selection of likely profiles was made. The deceleration regime profiles selected are, for Zbm the single braking regime profiles ('Br' and 'Co-Br'), for Gdm the dual and triple braking regime profiles (from 'Br-Co-Br' til 'Co-Br-Cr-Br-Cr-Br') and for Htn the single and dual braking regime profiles (from 'Br' til 'Co-Br-Cr-Br') in the DR analysis. For the brake variants, originally the dual and triple step piece-wise functions would have had sub-versions tested with different braking application and release, other than a linear increase or decrease of the braking coefficient. However, due to the minor differences compared to a linear application and release, the decision was made to reduce both dual step and triple step brake variants to just one of their sub-versions. This reduced the total number of brake variants from 11 to 5 variants, keeping the constant coefficient, second order polynomial velocity-dependent and time-dependent.

Error Quality

Besides calculation speed being a modelling performance measure, the accuracy of the solutions and number of generations required to do so is an important aspect. As mentioned previously, the GA optimisation has two stopping criteria, one being the fitness threshold and the other being the number of generations in an optimisation attempt. The fitness threshold was set to 0.1 or 10% of deceleration distance, as this would result in a solution fitness with an normalised sum of absolute errors (norm. SAE) of 0.1 m/s or lower. A

lower normalised SAE was deemed unnecessary for this research to be able to identify the realised deceleration behaviour and would start to fall in the realm of the GPS's margin of error. A ratio was used instead an absolute threshold as each deceleration run at different stations would yield significantly different SAE's and therefore an absolute threshold would yield different levels of accuracy for each deceleration reconstruction. The noisiness and data point density of the realised location tracking data and the randomness of the GA optimisation meant that this threshold often was not the stopping criteria of the GA optimisation. The main stopping criteria was the number of permitted generations, which was set to 40 generations. This stopping criteria was fine tuned from performance testing with the pilot data test run, striking the balance between calculation speed and solution fitness. The fitness convergence for this GA optimisation and minimisation problem meant that 40 generations were required to consistently get a solution fitness with a norm. SAE of under 1.0 m/s. Anything less than 40 would lead the solution fitness to be more dependent on the GA's randomness and a relatively minor gain in calculation speed. To further improve on the solution fitness consistently found after 40 generations, the number of permitted generations would have to increase significantly, causing the calculation time to increase to a duration unsuitable for this research. After 40 generations of a population pool of 32 vectors, up to 968 vectors are tested. If more parallel processing power would be available, further code optimisations are made or the data usage intent would shift to a smaller dataset, the number of permitted generations could be increased accordingly to provide the GA optimisation a higher chance at finding the true optimal solution or at least reach a solution much closer to the optimal.

In terms of the quality with regards to the best fitting curves on the realised data, plots are made to visualise the distribution of normalised Sum of Absolute Error (Norm. SAE) for the best fitting deceleration curves in Figure 5.3 and to visualise the differences between the best fitting curves with two other best fitting alternatives for 5 of the 280 fitted train runs in Figure 5.4 (others in Figures E.2 to E.6). Observing the distribution of norm. SAE for the best fitting deceleration regimes at each station(Figure 5.3), several things become clear about the chosen number of optimisation generations. For all the stations tested, the mean values are below desired norm. SAE threshold of 1.0 m/s. However, when observing the distributions, a portion of the best fitting curves is still above the 1.0 m/s threshold. This shows that the 40 generation limit used might not have been enough, even though the majority has managed to find a optimisation solution under this threshold.

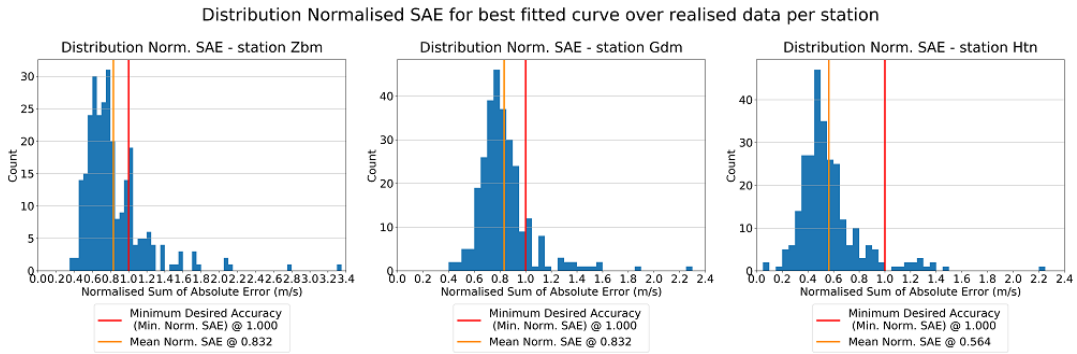


Figure 5.3: Histogram Normalised Sum of Absolute Error
Best fitting deceleration regime curve on the realised running data, given per station.

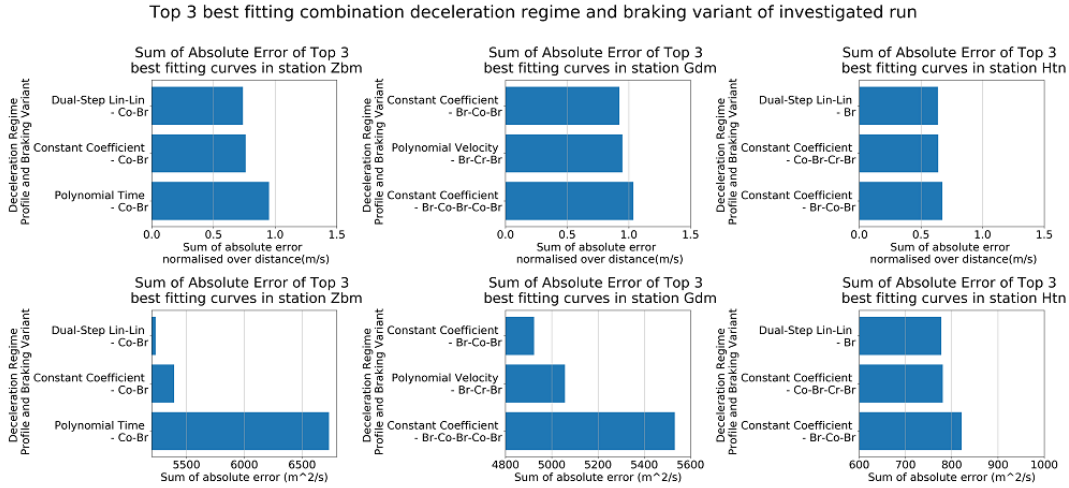
This could be due to the noisiness or big step in the realised data, leading to a larger error value as the optimum solution. However, considering the number of possible β-vector combination with the given vector variable bounds and the maximum number of possible tested vectors, it is safe to assume that a part of these error values being too big is due to the randomness of the GA optimisation. More finer stepped, less noisy realised data would help alleviate some of the error value, but future research will need to either improve the optimisation algorithm or extend the number of allowed generations to more than the 40 generations allowed in this research. Considering that the majority of the results are below

the error threshold and investigating the fitting of the deceleration regimes of a handful of train runs (see Figure 5.4b)

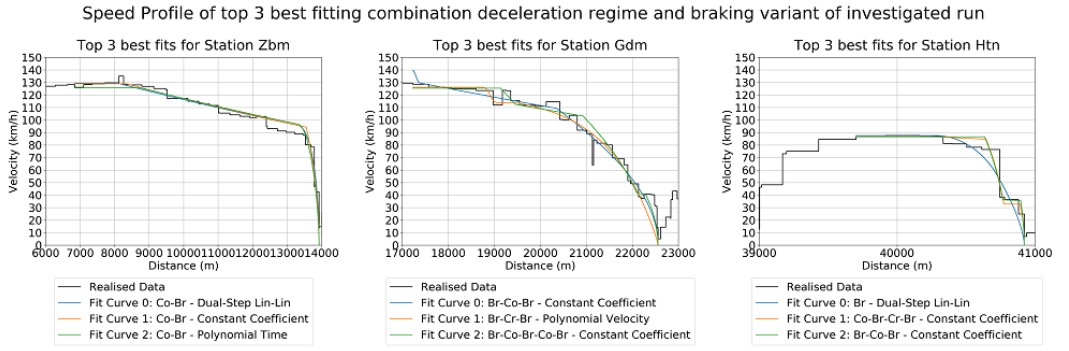
Deceleration Curve Fitness

Investigating the fitness of the selected deceleration regime curves relative to the two closest fitting alternatives, both the (norm.) SAE and the speed profiles are highlighted to observe their fitness to the realised data and each other at each of the selected stations (Gdm, Htn, Zbm). While the SAE seen in Figure 5.4a seem to contain a large error and wildly differ between stations, this is due to the long deceleration distance which also varies from station to station.

Therefore to make the results comparable between stations and make the errors more relatable, the SAE is normalised over the complete deceleration distance. With the normalised SAE, it is apparent that the fitness of the deceleration regime curves is assumed to be quite accurate, which is reinforced when observing the speed profiles of these deceleration regimes fitted over the realised data, in Figure 5.4b.



(a)



(b)

Figure 5.4: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes

Elaborating on the relative fitness between the different deceleration regimes, it becomes apparent that the braking variants have a very minor impact in altering the speed profile curves with the current level of point density and quality of the data sampled. The biggest impact to be found in fitting the data is the estimated deceleration regime. This is assumed when observing the fitness value and shape of the speed profile, as for example in the run

presented in Figure 5.4 or Figure E.3. However, even the selected deceleration regime will show some peculiarities in the optimisation, as seen in Figure 5.4, in which the fitted curve 0 at station Gdm is estimated to be a 'Br-Co-Br Constant Coefficient', but shows an unexpected short braking regime just at the start of the deceleration regime. This second braking regime was expected, but at the other end of the deceleration regime describing the braking behaviour from 40 to 0 km/h.

Peculiarities like these unexpected locations of sub-regime transitions, are most likely the result of the randomness getting close to a local minimum early on in the optimisation generations. Allowing more optimisation generations or expanding the randomised 'Immigration' pool-size, could have allowed for the randomness to get a more reasonable solution that could have provided a better fit. Another approach would be to restrict the band of the sub-regimes to search in an expected range. Restricting where the sub-regimes are allowed to exist would increase the chance of a more sensible optimisation solution. However, it would also restrict the possibilities of different deceleration behaviours to be analysed in the realised deceleration approach into the same station and therefore limit the generalisation of the DR model. This will need to be further investigated in future research or tailored to a station specifically through fine-tuning and observation of the realised data.

5.2. Results & Statistical Analysis

The results, from the DR model output and the other analyses from this research, with regards to the performance indicators (i.e. realised running times, deceleration loss times, track section occupation duration, vehicle braking rate, as described in Chapter 3) are discussed in this section and provide interesting insights to aid in answering the research questions. In this section the results are discussed per aspect of this research, along with an evaluation of their stochastic nature.

5.2.1. Deceleration Regime Composition & Braking Variants

While visualising the results of the DR model, some interesting observations and conclusions can be made from investigating the overview plots of the best fitting deceleration regime curves at the three selected stations (i.e. Gdm, Htn, Zbm). In the overview plots, three bar plots are presented that hold the count of best fitting deceleration regime curves per braking variants, per deceleration regime profiles and the unique combination of both braking variants and deceleration regime profiles.

Station Geldermalsen (Gdm)

Observing in the realised location tracking data over the entire year (Figure 5.2), all of the realised deceleration approaches towards the large, speed restricted station of Geldermalsen (Gdm) fit within either the '2-Br' or '3-Br' deceleration profile family (see Table 4.2), for which the DR model's deceleration profile selection has accounted for. When observing the distribution of the best fitting deceleration regime in Figure 5.5, it is a near even split between '2-Br' and '3-Br' families, with a relatively even distribution between individual deceleration regimes, leading to believe that the different versions of '2-Br' or '3-Br' profiles (e.g. 'Br-Cr-Br' versus 'Br-Co-Br' or 'Br-Cr-Br' vs 'Co-Br-Cr-Br') are too nuanced to distinguish themselves during the fitting. It is a different matter when observing the distribution between the best fitting braking variants, as here the 'Constant Coefficient' variant has been used approx. 130 times with a tied second place for 'Dual-Step Lin-Lin' and 'Polynomial Velocity' with each half of that count.

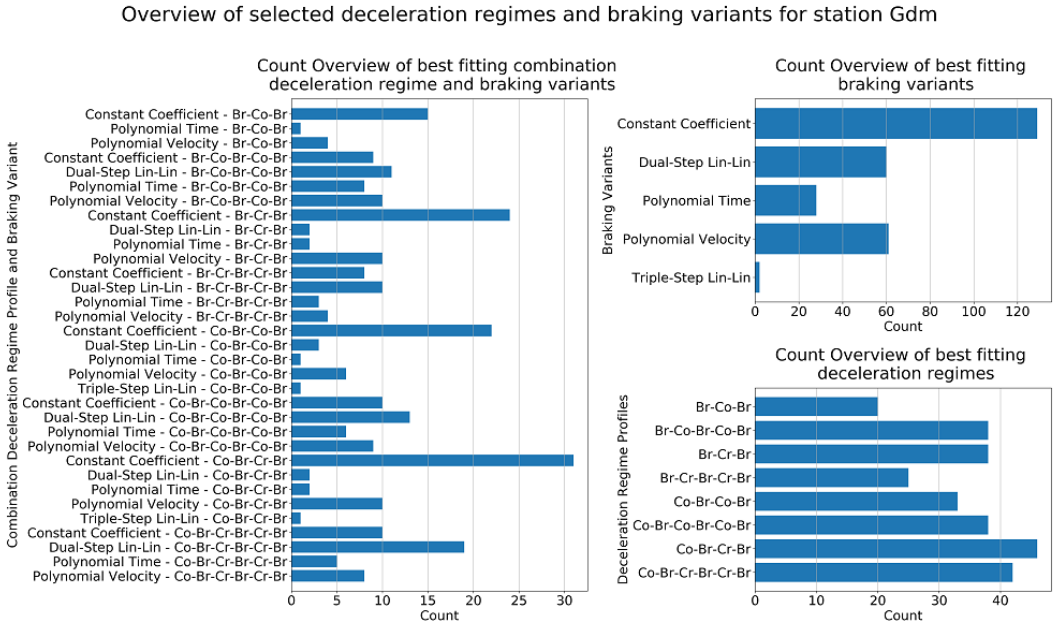


Figure 5.5: Overview Best Fitting Deceleration Regimes and Braking Variants - Gdm

The final braking regime into station Gdm is at approximately 35-40 km/h, which could lead to a too short of a braking regime distance to observe a good distinct braking behaviour. However, the first braking regime in the case of the '2-Br' deceleration regime profiles is present over a larger velocity delta, which should allow the distinct braking behaviour to manifest more clearly. This could be the reason why just over a half of all the counts at station Gdm went to both polynomial and the dual-step variants. However, when observing the unique combination of deceleration profile and braking variant in Figure 5.5, 4 of the top 5 most counted fitted regime profiles are held by the combinations with a 'Constant Coefficient' variant and a version of the '2-Br' deceleration regime profiles. The other combination in the top 5 is the 'Dual-Step Lin-Lin' braking variant and '3-Br' deceleration regime profile combination and only comes in at number 4.

The conclusion can be made that for station Gdm, the majority of the deceleration profiles adhere to the two braking regimes using a constant braking rate. The braking behaviour, however, would require more in-depth research with finer data, as the velocity delta along with the current mesh grid of the realised are suspected to be the reason for the inability to fully express the nuances at those distances and data point densities.

Station Houten (Htn)

With regards to the short-distanced, open-track, small station Houten (Htn), at most two braking regimes could be observed from the realised train runs and therefore the DR model tested for both the '1-Br' and '2-Br' deceleration regime families (see Table 4.2). However, this could have been misidentified, due to the influence of the train describer data making big jumps of average velocity in the speed profiles (i.e. large velocity difference over a relative short distance between relatively large track sections) of the fused location tracking data of the realised train runs, as seen in Figure 5.2. The fitting of the '2-Br' deceleration regime profiles and the relatively low lower bound of the sub-regime distance (i.e. 100 m) could have led the results of the DR modelling to skew in favour of the '2-Br' deceleration regime family.

However, when checking the overview plot of station Htn in Figure 5.6, the single braking regime without preceding coast regime still has a distinct lead over the 2-Br profiles. Observing the distribution of count over the different braking variants, the 'Dual-Step Lin-Lin' takes the lead with approx. 120 counts and the 'Constant Coefficient' variant a sec-

ond place with nearly 80 counts. The selection of analysed deceleration profiles should not have influenced the fitting of the braking variants, or if at all it would be in favour of the 'Constant Coefficient' as the shorter braking regimes would have a hard time distinguishing the nuances of the braking behaviour.

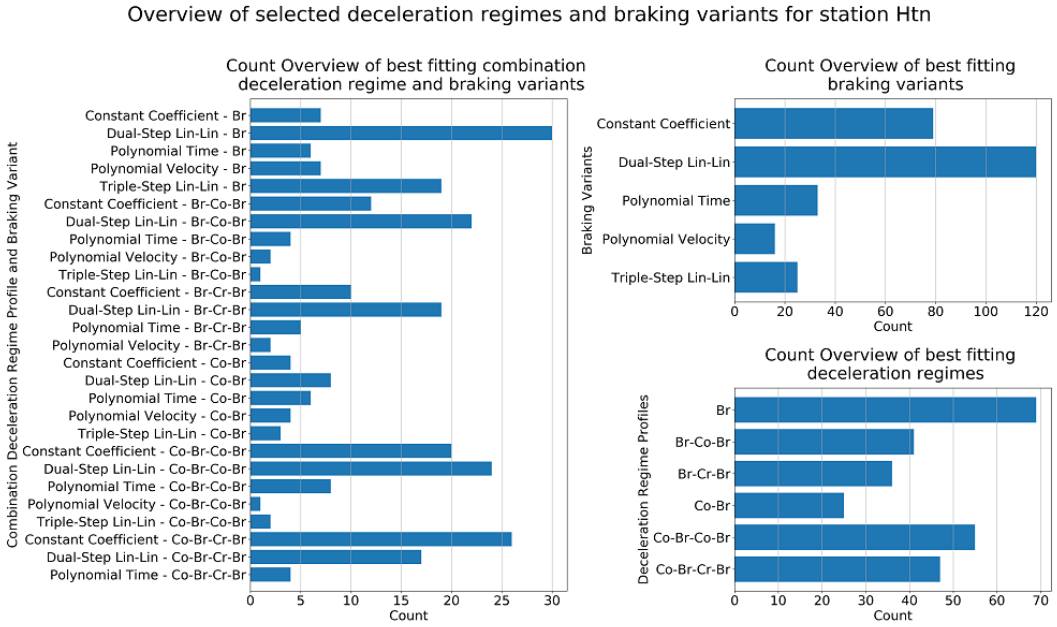


Figure 5.6: Overview Best Fitting Deceleration Regimes and Braking Variants - Htn

With the combination of deceleration profile and braking variants, the 2-Br paint a clear picture with the majority of counts for that particular deceleration profile going to the 'Constant Coefficient' and the 'Dual-Step Lin-Lin' braking variants. With the single braking regime profiles, the braking variants are more distributed in the case of 'Co-Br' and in case of 'Br' the two leading braking variants by far are the 'Dual-Step Lin-Lin' and the 'Triple-Step Lin-Lin'. Even though the maximum velocity in the train run to Htn reaches 100 km/h, it was enough to allow for more distinction in the nuances of the braking behaviour. The conclusion that can be drawn for station Htn is that linear braking application and release becomes more distinct as the braking regimes span over a larger velocity delta (Example fitted speed profile in Figure 5.4b).

Station Zaltbommel (Zbm)

Observing the fused location tracking data of the realised train runs for the long-distanced, open-track, small station Zaltbommel (Zbm) in Figure 5.2, only the deceleration regimes 'Br' and 'Co-Br' could be identified, both of which belong to the single braking regime family '1-Br' (see Table 4.2). Observing the overview of the fitted deceleration behaviour, two third of the deceleration profile counts are for deceleration profile with coasting. This seems logical, as Zbm is an open-track station with no speed restricted areas and is the first station stop in this corridor run. This means it is the least likely station on which its given running time supplement are used for delay recovery. As the train run to Zbm is long-distanced and open-tracked, the large velocity delta in the braking regimes allow for the nuances of the braking variants to distinguish in the DR model. The most popular braking variant for station Zbm, with over 110 counts, is the 'Polynomial Velocity' variant.

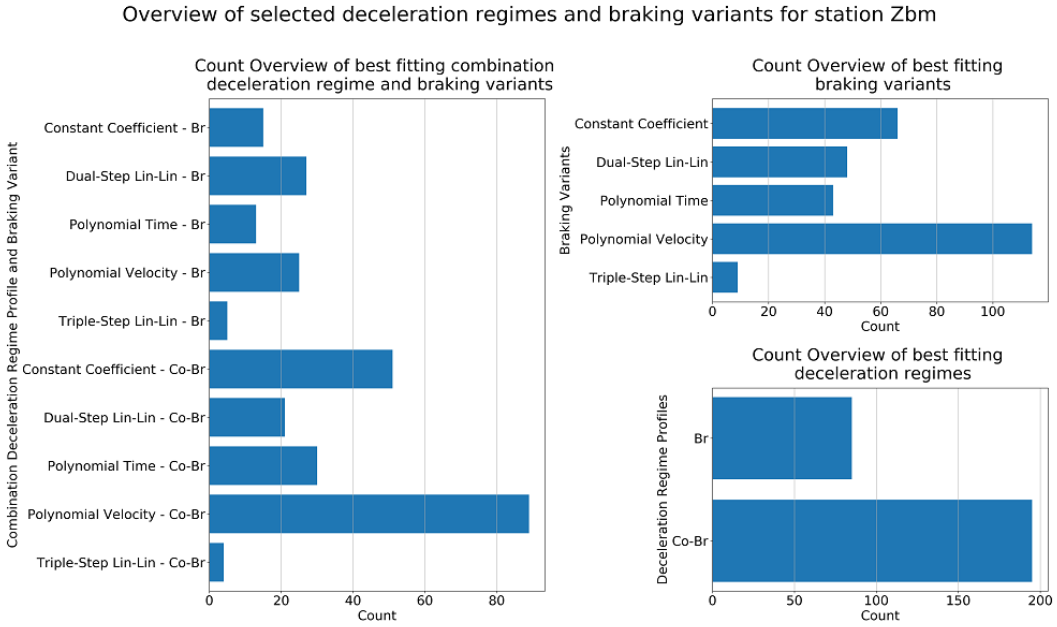


Figure 5.7: Overview Best Fitting Deceleration Regimes and Braking Variants - Zbm

The most popular combination is the 'Polynomial Velocity - Co-Br' combination, with a second place for the 'Constant Coefficient - Co-Br' combination, with respectively 90 and 50 counts to their name. The conclusion to be drawn from station Zbm is that when braking from 130 km/h, there are two popular driving behaviours, that being on to hold onto the convention of a constant braking rate and another in which the driver applies the braking rate in a smoother fashion.

5.2.2. Vehicle Coefficients

Analysing the vehicle resistance coefficients for validation, a couple of observations can be made of the validity and sensitivity of the estimated coefficients.

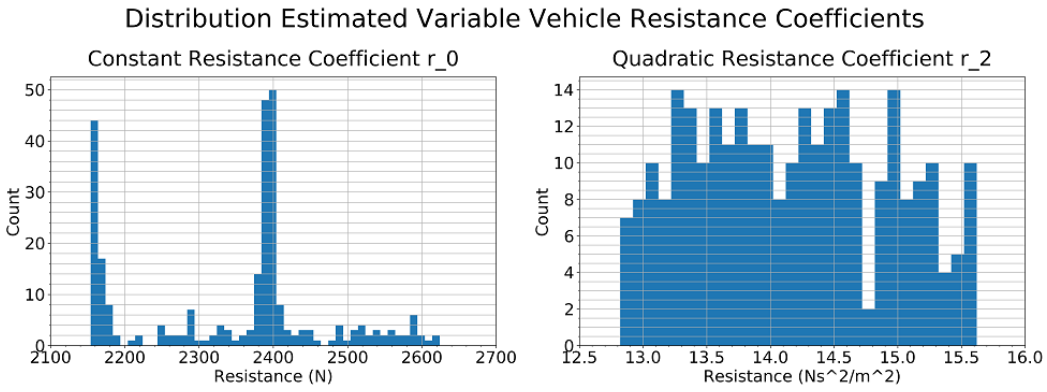


Figure 5.8: Histogram Vehicle Resistance Coefficients - Station Gdm

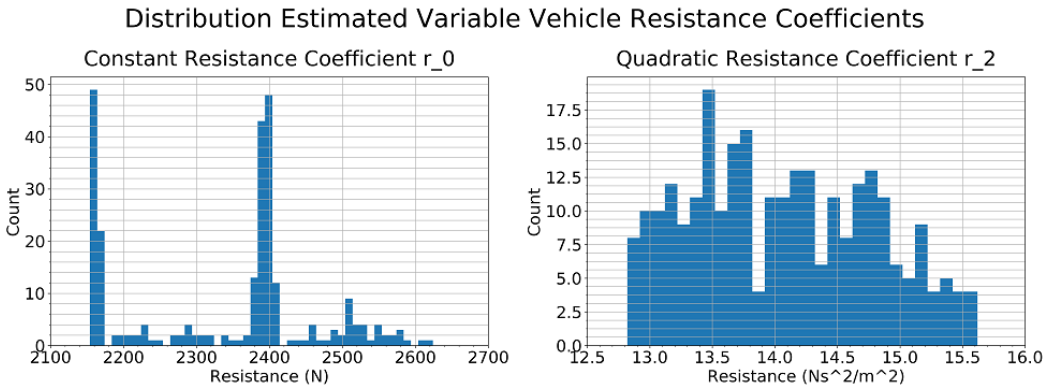


Figure 5.9: Histogram Vehicle Resistance Coefficients - Station Htn

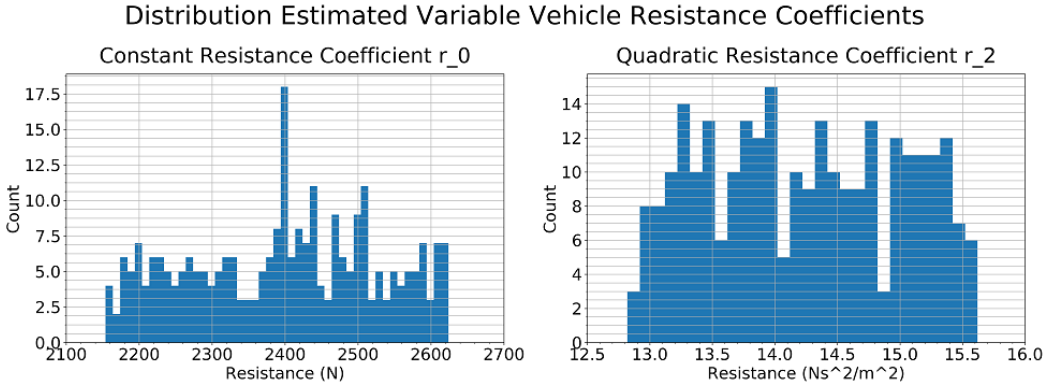


Figure 5.10: Histogram Vehicle Resistance Coefficients - Station Zbm

While the resistance coefficients at station Gdm and Htn (Figure 5.8 and Figure 5.9 respectively) show two distinct popular values near the lower bounds and the nominal value for r_0, station Zbm (Figure 5.10) shows a more distributed range with a less profound popular value near the nominal value. This shows that the given nominal value of r_0 is quite accurate and that the DR model hits the lower bounds to compensate for other parts of the deceleration curve not quite fitting correctly to the realised data. The distributions in all of the stations for r_2 show that this coefficient is sensitive variable, which makes sense considering the coefficient's size in relation to r_0 and the quadratic nature of this coefficient. Note that the resistance coefficients calculated were not maintained over the entire length of the corridor and were recalculated while analysing the individual deceleration regimes at each station. Partly out of practicality of the coded analysis methodology, but also partly due to these coefficients being affected by conditions such as axle load, wind resistance and rail type which can differ even between stations. As the vehicle resistance coefficients are not the main focus of this research, it was left at this. These results will need a more elaborate investigation in following studies.

5.2.3. Braking Rate

Analysing the results of the DR model with regard to braking rates, a distinction had to be made between the mean braking rate and the peak braking rate, as this research covers non-uniform braking variants. As in Bešinović et al. (2013a), a further distinction is made between braking rate to a stop and a braking rate to a lower velocity, when multiple braking regimes were present in the fitted deceleration profile. Observing these braking rates in Figure 5.11, the distributions of the estimated braking rates appear to adhere to a distinct Exponentially Modified Gaussian (EMG) distribution with a shape coefficient larger than 1 is present in all of the stations braking rate histograms (Station Gdm in Figure 5.11 and Stations Htn and Zbm in Figure H.2 and Figure H.3 respectively). The exception to that statement is the maximum braking rate to a lower velocity at station Htn in Figure H.2. This reinforces the assessment earlier about the '2-Br' profiles to misidentified.

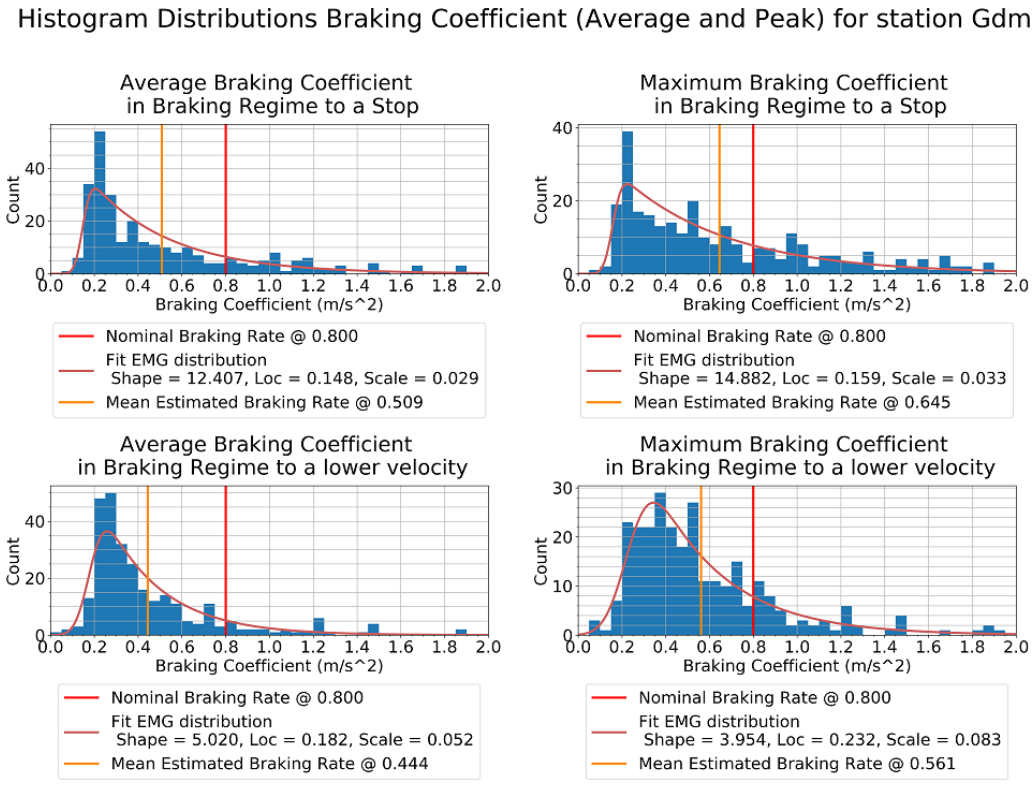


Figure 5.11: Distributions Mean and Peak Braking Rate for stop and lower velocity - Gdm. Vertical red line is the value of the nominal braking rate at 0.8 m/s^2.

These results shown in Figure 5.11 reinforce the research assumption of stochastic braking rates, as the distributions in braking rates would have been significantly more narrow for the distribution to be considered the result of a measurement error. The differences in the distributions between the distinction of averaged and maximum braking rates show that manner in which the nominal braking rate is defined can matter, especially when addressing non-uniform braking behaviour. The distribution of the maximum braking rate fitted by the DR model has a less pronounced peak near 0.2 m/s^2 with an overall higher estimated braking rate of 0.645 m/s^2 compared to 0.509 m/s^2 as seen by the shift in the line of the mean estimated braking rate. While the distributions of the maximum braking rates seem to diffused more towards the higher values, the distributions maintain very similar distribution shapes compared to their averaged braking rate counterpart with minor changes to the shape, location and scale coefficients.

As seen in Figure 5.11, the DR model estimated several braking rates up to 2 m/s^2, which is extremely high for an operational braking rate as seen by the nominal braking rate at 0.8

m/s^2. The DR model allowed the braking rate up to 2 m/s^2 in order not to hinder or influence the optimisation of the deceleration regime estimation. However, this is even high for an emergency braking rate and a more likely result of compensating for the larger velocity differences caused by the train describer data as seen in Figure 5.4b.

Comparing to the red line of the nominal braking rate of 0.8 m/s^2, Figure 5.11 shows clearly that the majority of the realised braking rates are significantly lower in regards to the mean braking rates estimated. This becomes apparent when observing the mean braking rate being lower in every distinction made in Figure 5.11. This difference does become less significant when observing the maximum braking rate estimated for station approach. Testing non-uniform braking rates with the DR model, allowed for another perspective in braking rate distributions and gives a possible explanation to why the calculated braking rates appear to be significantly lower than the nominal and expected braking rate. Even though this alleviates the severity of this disparity between the realised and expected braking behaviour, it still doesn't quite solve it and resulting distributions are still assumed stochastic braking behaviour of which the majority is still significantly conservative compared to the nominal braking rate.

5.2.4. Running Times

To answer the research question of the existence of stochastic deceleration behaviour in a station stop, the stochastic distribution of the realised running times has to be verified. This is again done for the three selected station Gdm, Htn and Zbm. To obtain these results, the network timetable was used instead of the DR model, as the model's focus lies on the deceleration part of the station-to-station train run. Of the results from the processing of the network timetable, both a histogram and boxplot are presented. From the histograms, the type/shape of the distribution can be estimated. The boxplots were made to make a better distinction between the different months to test seasonality, discussed in Section 5.3.4, and the total distribution of available train runs. Any correlation regarding the deceleration behaviour and realised running times, is drawn with the help of the comparative scatter plots, further discussed in Section 5.3.4. The boxplots also give a clear indication as to the outliers in the dataset of results. The boxplots have the advantage of showing the distribution of the sample set without being squashed by a larger data set, as compared to their respective histograms. From the histograms of the three stations (Gdm in Figure 5.12, Htn in Figure 5.13, Zbm in Figure 5.14), the distributions of the running times compare very similarly to a normal distribution, however the slight skewness of the bell shape would indicate a 'Gamma' distribution with shape coefficient $k > 1$ being the more likely fit.

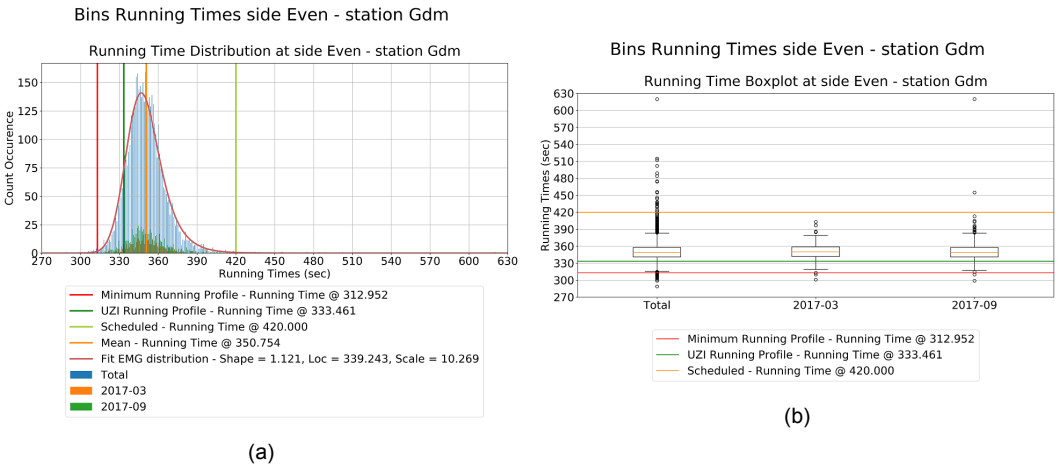


Figure 5.12: (a): Distribution Running Times - Even Running Side Gdm
(b): Boxplot Running Times - Even Running Side Gdm

Observing the histogram for Gdm in Figure 5.12a, nearly all train runs have a running time between the minimum running time and the scheduled running time, with the majority having a full minute to spare. On the contrary, the majority is slower than the UZI guide-book would suggest. This can be seen more clearly in the boxplot in Figure 5.12b, where the main body or box, representing the group between the lower and upper quartiles of the dataset, just above the green line of the UZI-profile running time. Meanwhile the whisker and cap (i.e. the vertical and horizontal line on the outer ends of the boxplot), representing the most extreme, non outlier data point, just under 40 *s* of the scheduled running time.

This behaviour is slightly different for station Htn in Figure 5.13, as the difference between UZI-profile running time and scheduled running time is smaller and still the distribution is shifted closer to the scheduled running time, seen in Figure 5.13a. In Figure 5.13b, the main body of the boxplot is pressed again the scheduled running time, with the caps and whiskers approximately over the scheduled running time. A clear cause for the relative difference in scheduled running time, is the method of calculating the scheduled running time. At station Gdm, the minimum running time was rounded up and further supplemented with 60 seconds, while Htn was rounded to the nearest 60 seconds and supplemented with only 18 seconds. Both rounding methods had the same effect of rounding up for these stations. However, the margin added to the running times for Gdm and Htn as a rounding and recovery supplements, was with approximately 107 and 29 *s* respectively.

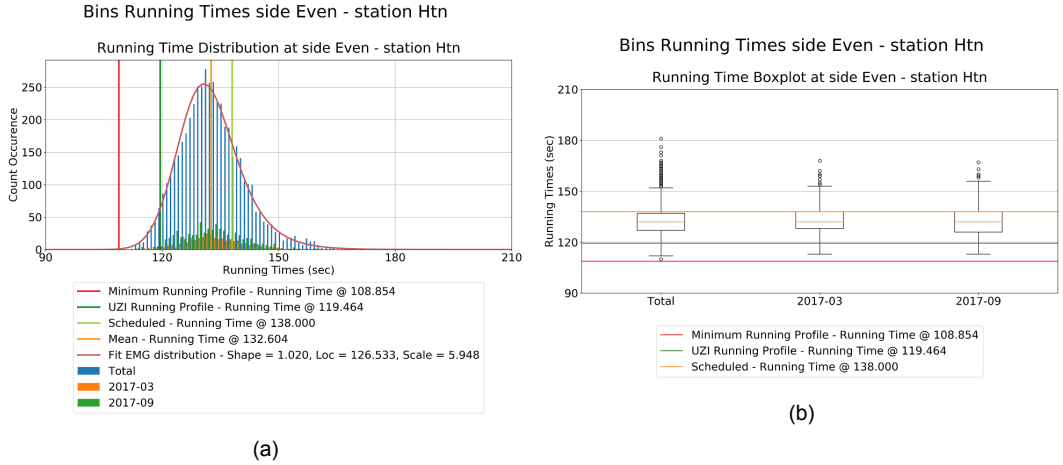


Figure 5.13: (a): Distribution Running Times - Even Running Side Htn
(b): Boxplot Running Times - Even Running Side Htn

In the development of the network timetable, the minimum running times are rounded a whole minute. However, depending on the station size, the methods of rounding are different. For the larger stations, the minimum running time is rounded up to the whole minute. However, for the smaller stations, these minimum running times are rounded to the nearest whole minute. On top of this different rounding method, there is a difference in the recovery time supplements given between large and small stations. Large stations are generally used for the delay recovery within the corridor, getting several minutes added, while the smaller stations will get a recovery time supplement of 18 *s* with a station dwell time of 42 *s*, due the departure times being rounded to the whole minute. This is due to the punctuality priority given to the larger stations with a leaner running profile to the smaller stations, in order to provide a significant recovery time buffer to adequately address any delays at large, important nodes within network.

The effects of the different rounding methods are more clearly seen in the comparison between minimum and scheduled running times for station Zbm. In here the minimum running time of 427.2 *s* was rounded down and supplemented with 18 *s* to provide a scheduled running time of 438 *s*. This leads to the peculiarity of the UZI-profile taking

significantly longer than the scheduled running time, as it underestimates the operational behaviour required to run the distance between stations within the scheduled running time. Furthermore, this skews the on-timeness of the train runs with the majority being 20-60 seconds late. This is masked with the 42 seconds dwell time at station Zbm, but this rounding down of the running times forces all trains to either or both run lean in their running times and dwell times or having to play catch-up from Zbm to Gdm. However, despite this leanness in the running times, the driving behaviour seems more relaxed, just from observing the distribution of the realised running times, compared to that of Gdm or Htn with a difference in upper and lower quartiles of 30-40 *s* at Zbm compared to the approximately 20 and 10 *s* for Gdm and Htn respectively.

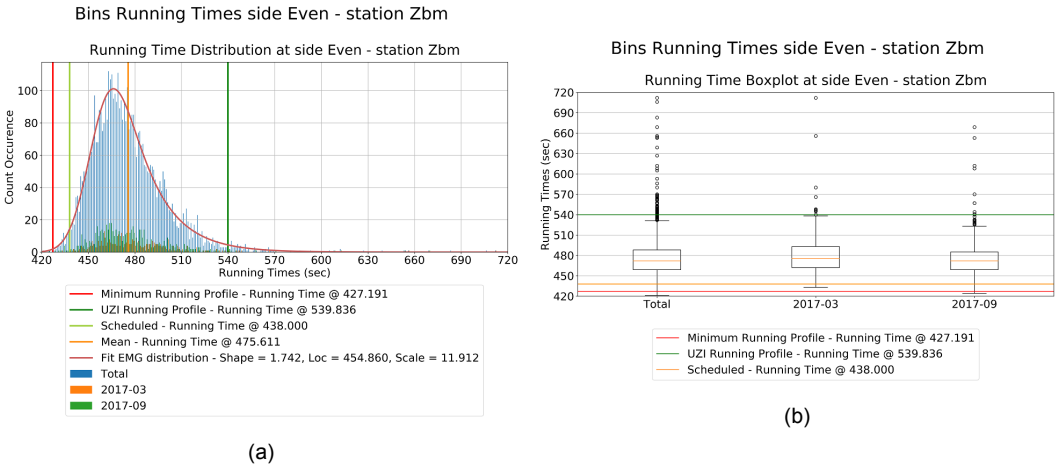


Figure 5.14: (a): Distribution Running Times - Even Running Side Zbm
(b): Boxplot Running Times - Even Running Side Zbm

The conclusions to be drawn from these stations running time distributions, is that all of the empirical distributions are likely to adhere to theoretical distributions such as a Gamma or Exponentially Modified Gaussian (EMG) distribution with shape coefficient greater than 1 and will try to adhere a running time between the running time defined by the UZI method and the scheduled running time if given the sufficient time supplements. The difference found with distributions of lean time supplemented runs, is the skew of the distribution leaning closer to the scheduled running time while still showing a similar spread in running times.

5.2.5. Deceleration Loss Times

While analysing the deceleration loss times of the realised train runs, it became apparent that calculations were needed from multiple distance points to describe and specify the effects of the deceleration behaviour on the loss times. The decision fell on three points in the deceleration distance to portrait the realised deceleration behaviour. These points were chosen to describe the full deceleration distance loss times, the loss times at the distance marker where the majority of the train runs were observed starting their final deceleration into the station, and lastly the shortest distance marker close to the head of the station platform at roughly 400-600 m before the estimated stopping point of the vehicles.

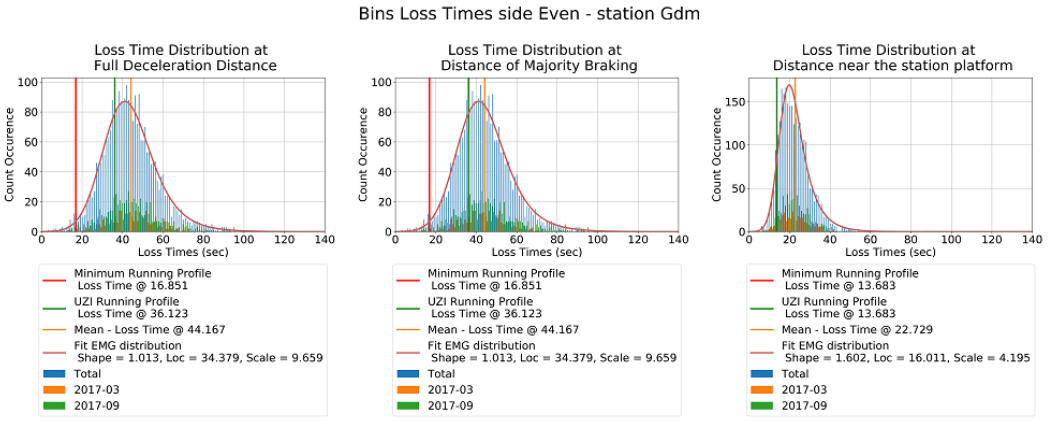
The three distance markers chosen for each of the stations of the corridor and these markers are selected to be the beginning or head of the relevant track section, as to eliminate the significant differences between the measured vehicle-sided location tracking data. The station stopping point is estimated as the realised stopping point could not be accurately measured from the infrastructure or could not be reliably measured by the vehicle-sided tracking data. As the rolling stock composition and the driver's ability to stop near the designated sign on the platform are considered consistent enough, it is assumed that any miscalculation leading from the wrongly identified stopping point would not significantly affect the relative position of the data to its other data points and would only shift along its axis in the relevant loss time plots as a single group without affecting distribution shapes or correlative results. The track sections used as distance markers for the stations Geldermalsen (Gdm), Houten (Htn) and Zaltbommel (Zbm), are presented in Table 5.2.

Table 5.2: Track Section ID's related to the distance markers used at their respective stations in describing loss times. Distance expressed in meters to assumed stopping point at station platform.

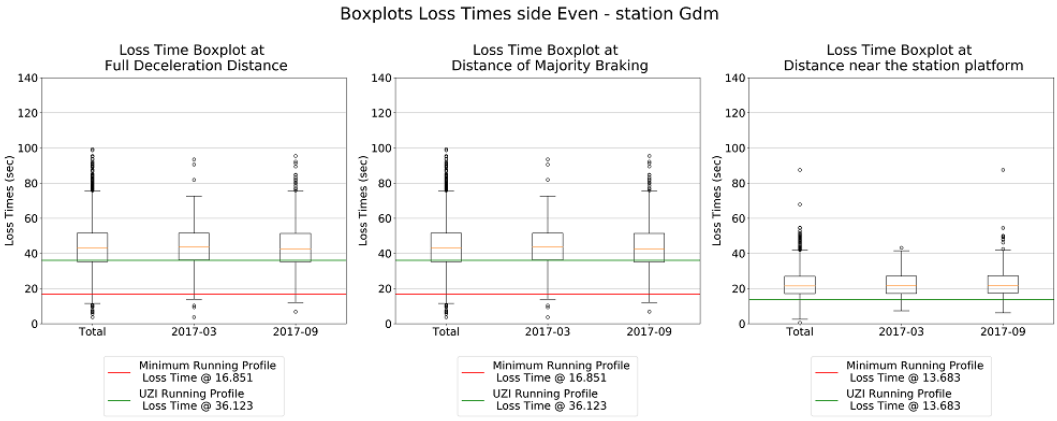
Station Name	Track Section ID Head Platform (m)	Track Section ID Majority Braking (m)	Track Section ID Full Deceleration (m)
Gdm	GDM$139T (821)	MTNA$692AT (5259)	MTNA$692AT (5259)
Htn	HTN$1844CT (486)	HTN$1865T (657)	HTN$1864AT (1288)
Zbm	OZBM$201BT (584)	OZBM$225BT (1994)	HDL$251AT (7269)
Cl	LEK$656T (1110)	LEK$656T (1110)	GDM$666AT (4410)

A fourth station Culemborg (Cl) is also represented in the table, as this station shows some interesting results in the comparative plot of realised running times and loss times, discussed in Section 5.3.3. For completeness, the track section IDs used as distance markers for station Culemborg are mentioned along with the investigated three stations, even though these distance markers are only used for the results in the comparative analysis in the aforementioned section.

Observing the histogram plots for the stations Gdm, Htn and Zbm (respectively Figure 5.15, Figure 5.16 and Figure 5.17), an Exponentially Modified Gaussian distribution with a shape coefficient greater than 1 is present, similar to that seen earlier in the running time distributions. All three stations, to some degree, have realised train runs that had a lower deceleration loss time compared to the calculated minimum running time based on the nominal vehicle coefficients. This could partly be addressed by drivers applying a more assertive deceleration behaviour and applying a braking rate that is higher than the nominal value of 0.8 m/s^2, which is considered the expected, normal operational braking behaviour. Even with an higher than nominal braking rate, it would only address a portion of the presented loss times. As mentioned earlier, the stopping point of the realised runs had to be pre-defined for a consistent and reliable reference. Therefore, it is most likely the case of the loss times being underestimated for a most of the train runs at these stations. However, the distribution shape is assumed to be valid for all of the distances at all of the stations, as their relative performance to other realised runs is assumed consistent.



(a)

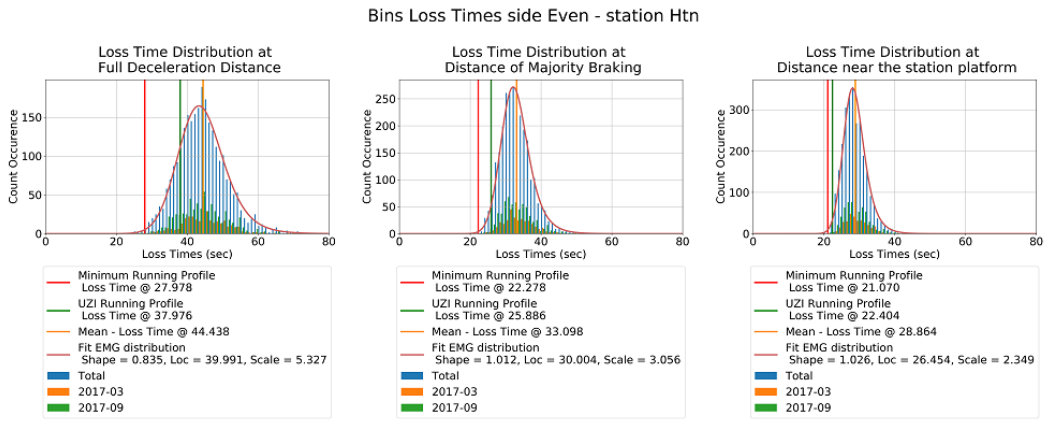


(b)

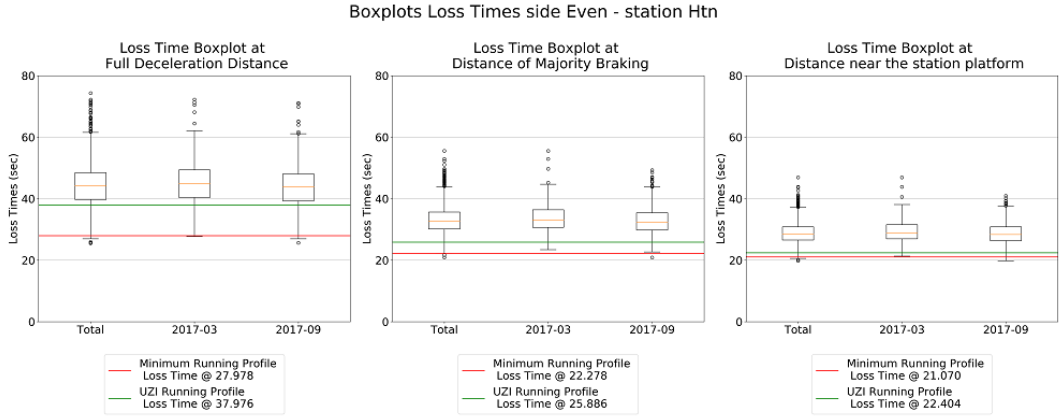
Figure 5.15: (a): Distribution Deceleration Loss Times - Even Running Side Gdm
(b): Boxplot Deceleration Loss Times - Even Running Side Gdm

When drawing the comparison between the different distances over which the loss times are calculated, it is unsurprising to see that the closer the vehicles get to their stopping point, the smaller the standard deviation appears to be. Interestingly though, it appears to be the case that the standard deviation also appears to decrease relative to the mean deceleration loss times. This would signify a more consistent performance closer to the end of the deceleration regime than at the point at which the majority of the train runs start their final part of their deceleration regime. Even though the loss times become more consistent nearing the station stopping point, a visibly apparent standard deviation remains. For the stations Gdm and Htn(Figure 5.15b and Figure 5.16b resp.), the whiskers (i.e. the most extreme non-outlier) reach a loss time of 1.5 to 2 times that of the loss times calculated for the minimum and UZI-profiled train run references. This would understandably lead to a considerable disparity in performance between the expected deceleration behaviour and the realised deceleration behaviour.

For station Gdm in Figure 5.15a, roughly 90% of the realised loss times are longer than the minimum running profile, while estimated from the histograms approximately 55-60% of the realised loss times were longer than the UZI-profiled loss time for the full deceleration distance. Observing the boxplots in Figure 5.15b, it becomes clear that the main body of the box is consistently describing a longer loss time than the UZI-profiled loss time for each of the distances calculated. This means in actuality that 75% of the realised deceleration regimes have a loss time longer than expected with the UZI-profile, with 50% up to 10-15 s slower.



(a)

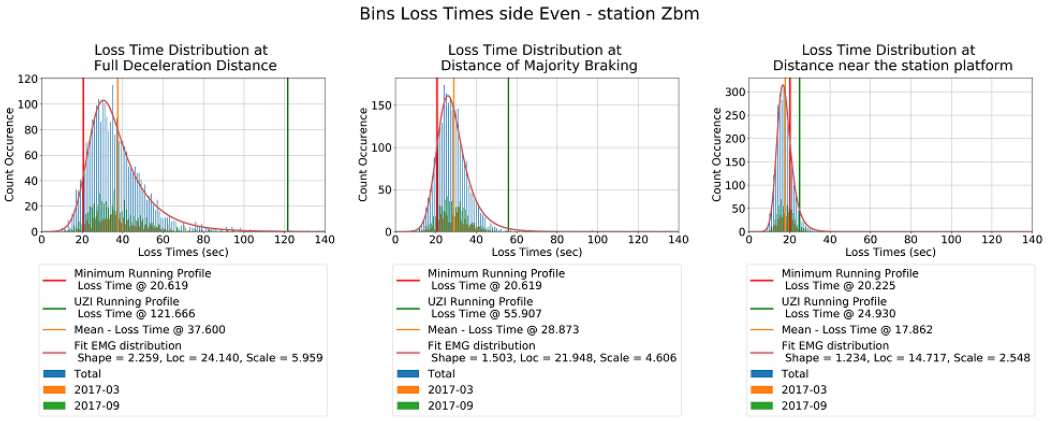


(b)

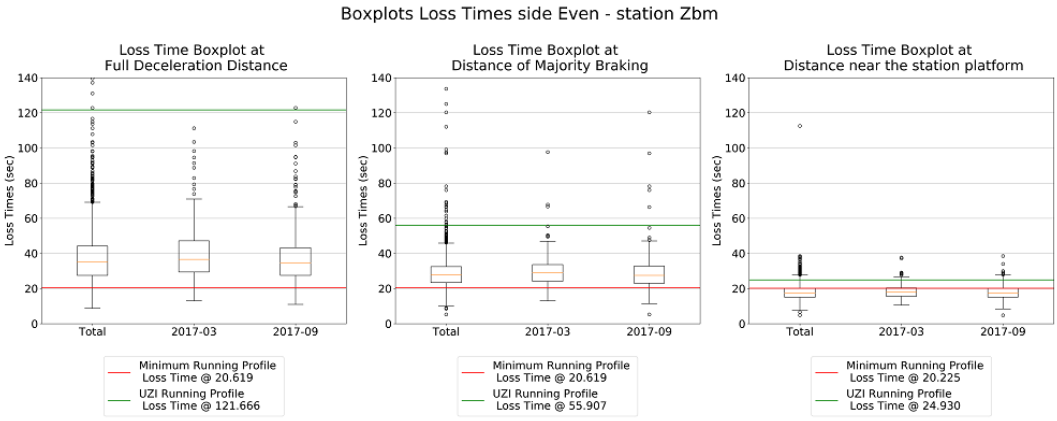
Figure 5.16: (a): Distribution Deceleration Loss Times - Even Running Side Htn
(b): Boxplot Deceleration Loss Times - Even Running Side Htn

In station Htn, the standard deviation is smaller in comparison to station Gdm and also shows the same behavioural consistency increasing the closer the realised runs got to their station stopping point. However, due to the leaner running time supplements (the previously mentioned 29 *s* over Gdm's 107 *s*), a more assertive UZI-profile is derived that allows for less differentiation in the expected driving behaviour. This skews the deceleration loss times at station Htn to be longer relative to the UZI-profile expected loss time.

However, when observing the calculated running times for the UZI-profile and the estimated scheduled time for stations Gdm and Htn (Figure 5.12, Figure 5.13), the majority of the realised running times falls between these two calculated times. Therefore, a fair assumption could be made for both stations that the large majority of the deceleration behaviour, and with that the deceleration loss times, will fall between the expected loss times of the UZI-profile and the remainder of the scheduled running time. Even though this means that the effects on arrival punctuality are limited, it also means that more time is lost on the deceleration regime than expected and that more time is spent occupying possibly critical track sections on a station approach, negatively impacting the infrastructure occupation.



(a)



(b)

Figure 5.17: (a): Distribution Deceleration Loss Times - Even Running Side Zbm
(b): Boxplot Deceleration Loss Times - Even Running Side Zbm

With station Zbm, however, the underestimated operational behaviour preceding to the arrival at this station leads to an exceedingly high deceleration loss time over the entire deceleration regime. This is most likely due to the underestimation of the resulting coasting time/distance or an overestimation of the given time supplements within the scheduled running time. As Figure 5.17 shows, the realised loss times are mostly distributed just above the calculated loss time related to a minimum running time profile. However, as the similar shaped distributions between running times and loss times show some correlation (further discussed in Section 5.3.3), it is assumed that a similar correlation can be made for the expected deceleration loss time based on the scheduled running time. As such and unlike with stations Gdm and Htn, the deceleration loss times from the realised runs will be longer than the expected, scheduled deceleration loss times. This would indicate a trade-off between punctuality and a more conservative deceleration behaviour in the approach to station Zbm, rather than between energy efficiency and punctuality.

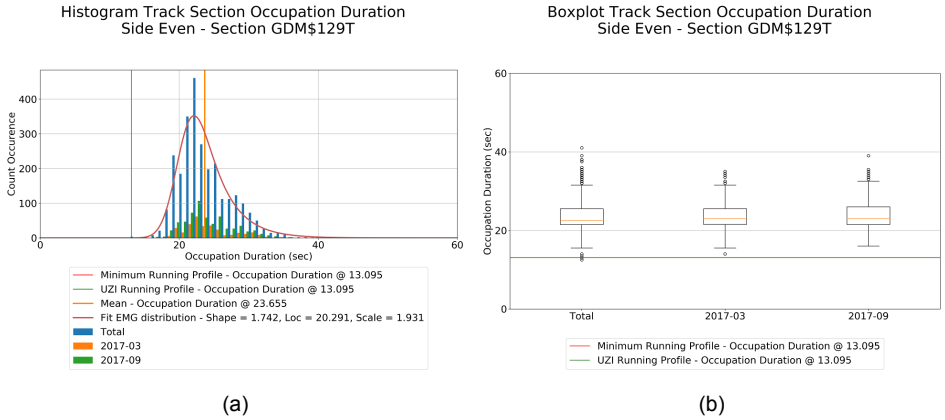
In regards to the loss time distribution of station Zbm, the aforementioned peculiarity of a lower realised deceleration loss time relative to the calculated loss time from a minimum running profile, is clearest seen here. This could partly be due to the more assertive braking rates applied due to the lean supplemented running times. However that would contradict the more conservative braking rates estimated from the DR model seen in Figure H.3 and the significant coasting distances seen in Figure 5.1. Therefore the most likely reason for the lower deceleration loss times would be the incorrectly estimated stopping point at the station platform.

The conclusion to be drawn from the distribution of realised deceleration loss times, is that

the deceleration loss times are not (strongly) correlated to the available time a driver might have left in their scheduled running times as the case of a lean supplemented running profile (seen approaching station Zbm) seems to show a trade-off between service punctuality and chosen deceleration behaviour.

5.2.6. Track Occupation

While calculating the track occupation duration for each track section of the given infrastructure data, the focused track sections as presented in this research, were the track sections just before the station platforms at critical points where the difference in track occupation duration could influence infrastructure occupation with their deceleration behaviour. This means the track sections in which a track switch is present for train overtaking or utilising other platforms with a shared track switch. This meant for station Gdm, Htn and Zbm the highlighting of the track sections GDM$129T (Figure 5.18), HTN$1844CT (Figure 5.19) and OZBM$201BT (Figure 5.20) respectively.



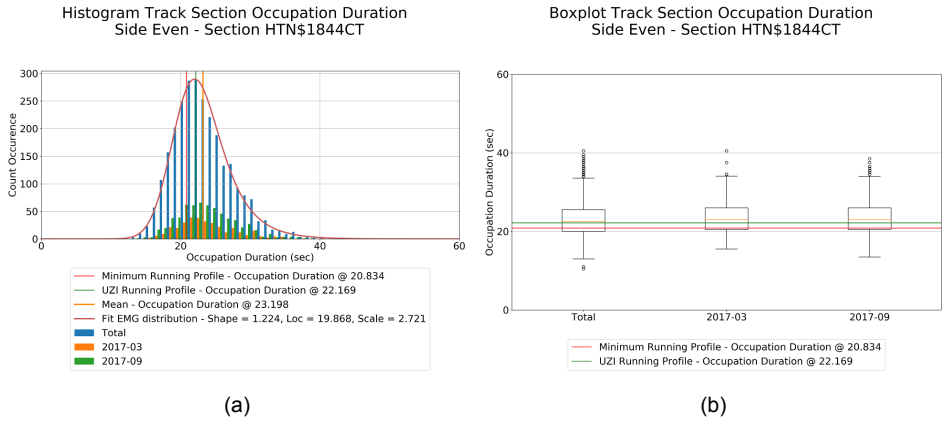
(a) (b)

Figure 5.18: (a): Distribution Track Section Occupation Duration - Even Running Side Gdm - Section GDM$129T
(b): Boxplot Track Section Occupation Duration - Even Running Side Gdm - Section GDM$129T

The first finding, while observing the three track sections, is the distribution shape and relative deviation significantly similar to the realised running times and realised deceleration loss times. This would reinforce the assumption of there being a strong correlation between the deceleration behaviour and the realised running times between stations. due to the strong similarities with the realised running times, deceleration loss times and the correlations made regarding seasonality, the assumption is made that this will be true for the correlations made regarding the deceleration behaviour and track section performance. The comparative analysis of the impact, that the deceleration behaviour has on track section occupation duration, will thereby be indirectly discussed in Section 5.3.4 through the comparative analysis of the realised running times and deceleration loss times.

At station Gdm (Figure 5.18), the majority of the track occupation duration falls between 20 and 40 s while the calculated track occupation was roughly 13 s. A part of the difference could be addressed with a disparity between the actual switching time of this track section and the assumed negligible switching time. However this would only off-set the calculated reference point. More likely, the disparity between the realised track occupation duration and the calculated one, would be a lower velocity through this track section due a more conservative station approach. This would likely be due to the numerous track switches near the station. Another possibility could be the fact that the platform of station Gdm is divided into an 'A' and a 'B' side, which would essentially reduce designated stopping platform (approx. 200m compared to Htn approx. 250 m). Lastly, the differences to their respective references could be due to the proximity of the platform's stopping point to the track section's signal. Both these platform differences could effectively reduce the margin of allowed error around the stopping point for Gdm, compared to stations like Htn and

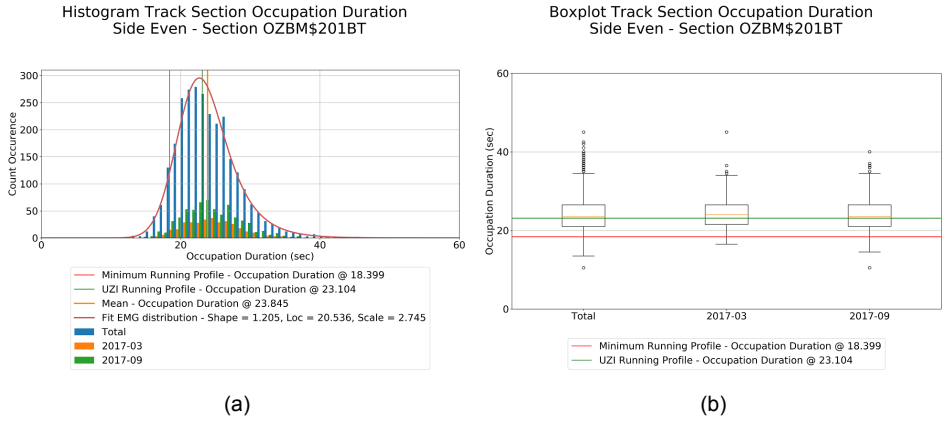
Cl. To confirm this assessment, further investigation to the relation between deceleration approach and the differences between station platform is required.



(a) (b)

Figure 5.19: (a): Distribution Track Section Occupation Duration - Even Running Side Htn - Section HTN$1844CT
(b): Boxplot Track Section Occupation Duration - Even Running Side Htn - Section HTN$1844CT

Observing the results for station Htn and Zbm, a significant portion of the realised distributions registers a shorter track occupation duration than the calculated track occupation for the minimum running time profile. The skew of the distribution could be due to having fewer track switches to navigate and open-track design, allowing for higher entry velocities and more assertive braking behaviour. However, the estimated braking rates for these station, derived from the DR model, are significantly more conservative than the nominal braking rate value for it to explain the results in Figure 5.19 or Figure 5.20. The more likely reason for the realised results would be a different configuration of rolling stock combination, more specifically the rolling stock's total length. In case of a single rolling stock configuration, such as an SLT-4 or SLT-6, compared to the assumed common rolling stock configuration SLT-10 (i.e. SLT-4 + SLT-6), this could (nearly) half the train length. Assuming similar velocities through the track sections, it would have a similar effect on the track occupation duration. This could explain the relative difference to the reference track occupation duration. However, this would still not explain the bottom outliers of approx. 10-11 s occupation duration. With the shortest configuration SLT possible (SLT-4) at 70 m and the length of HTN$1844CT being 240 m, it would still translate to an average velocity of 90-120 km/h, which is high for this track section and for the distance to the station's stopping point, as the realised train runs plotted in Figure 5.2 show an average velocity of 60-80 km/h. This peculiarity would require further investigation.



(a) (b)

Figure 5.20: (a): Distribution Track Section Occupation Duration - Even Running Side Zbm - Section OZBM$201BT
(b): Boxplot Track Section Occupation Duration - Even Running Side Zbm - Section OZBM$201BT

5.3. Comparative Analysis

For the comparative analysis, comparisons were drawn between aspects of realised data (i.e. departure delay, running time and deceleration loss time) and hypotheses on seasonality and deceleration behaviour were tested. These are discussed below and visualised through scatter plots, which in case of seasonality and deceleration behaviour (Section 5.3.4) makes use of colours to distinguish different groups.

5.3.1. Departure Delay vs Running Times

The comparison between the departure delay and the realised running times was made by calculating the realised running times from the realised departure and arrival times between stations and by calculating the delay (in *s*) from the difference in expected and realised departure times. Both the expected and the realised station (i.e. departure and arrival) times were extracted from the network timetable. The departure delay and the realised running times are presented in Figures 5.21 to 5.23, showing the departure delay with the corresponding realised running time in a scatter plot.

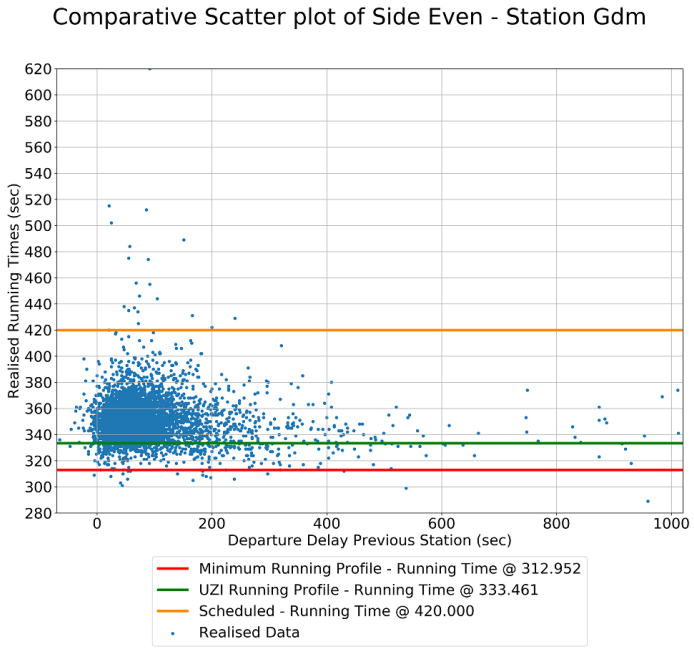


Figure 5.21: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Gdm

When observing the scatter plots for the three stations, no simple or straightforward correlation can initially be drawn. The centre of mass for the scattered data points, which appear to cover the majority of the data points, seems to be located on the delayed side of the train run realisation within the first 60-120 *s*. Each of the three station differ slightly in delay around which the points are massed. This could be explained with the dwell times on the previous stations and their ability to absorb any delays. For instance in Figure 5.23, the mass seems to be located between 0 and 60 *s* of departure delay, while Figure 5.21 shows a wider range of departure delays the realised train runs mass around. This particular difference can easily be explained due to the scheduled running times between Ht and Zbm causing the UZI-method to underestimate the required performance for this run. This in turn causes a significant distribution of realised running times between station Zbm and Gdm. However, this would only cover a 140 *s* of the departure delay at station Gdm, when observing the range of the majority of the running time distribution to station Zbm (Figure 5.26). The departure delay for the train run between station Zbm and Gdm appears to be a compounded delay, combining the previous departure delay with the previous run-

ning time distribution with only partial delay recovery within the realised running time. This could be the consequence of a lean scheduled running time, but would require further investigation.

There does not seem to be any direct interaction between the departure delay and the realised running times. The possible correlation to be observed, would be the formation of an upper and a lower bounds, which narrow towards a specific running time with an increasing departure delay. The upper bound would appear to follow a exponential curve with a form of $c * x^{-1}$, while the lower bound follows a linear or near linear curve with a very shallow sloping. While the shapes of these boundaries are open for interpretation, the consistency of the realised running times and thereby the consistency of the driver's behaviour seems to improve with the increasing departure delay. Interestingly, the consistency of the driving behaviour does not seem to approach or converge to the minimum running times, but to a running time that is more conservative. For station Gdm in Figure 5.21, these bounds seem to approach the running time of 333 s, which roughly coincides with the expected running time calculated based on the UZI-method.

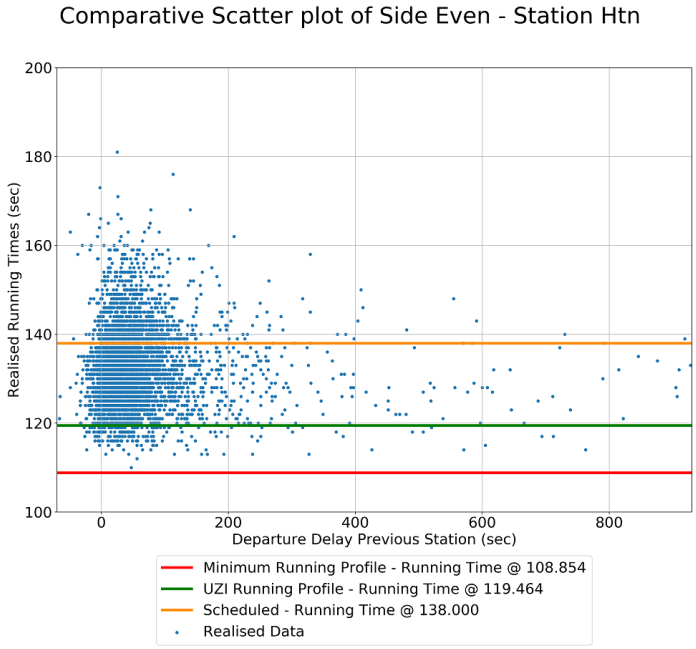


Figure 5.22: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Htn

For station Htn in Figure 5.22 the differences in running times are smaller, making it harder to estimate a possible convergence point in the realised running time. However, it seems the upper and lower bounds converge to a running time of 120-135 s This assumes that the operational behaviour towards station Htn has been fairly consistent and shows little dependency on the departure delay with a low convergence rate. The situation at station Zbm shows a significant spread in running times, but again a slow convergence rate, estimated to go towards a running time between 440 and 500 s. The big spread in running times show a wide variety in operational behaviour, becoming (slightly) more consistent with the increasing departure delay compared to the 100-120 s spread seen in the train runs with less than a minute departure delay.

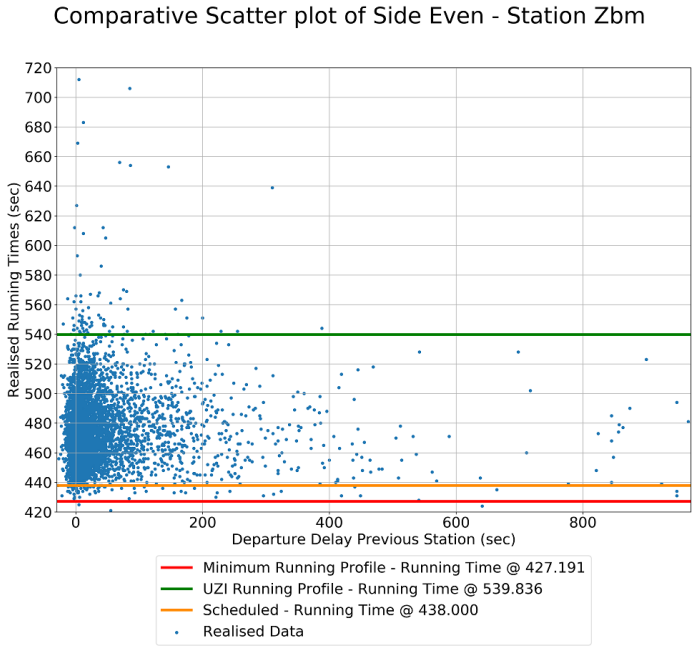


Figure 5.23: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Zbm

5.3.2. Departure Delay vs Deceleration Loss Times

Similar to the comparison scatter plots for the realised running times, the deceleration loss times are compared to the departure delays. Since its a comparison with the deceleration loss times, the three distance points are used as described in Table 5.2. Interestingly, a similar shaped group of data points are plotted for each station at each of selected distances.

Observing the results for station Gdm in Figure 5.24, a similar curved upper bound could be imagined, as seen in Figure 5.21. Except in this case, the curve would be more pronounced at the full deceleration distance compared to one found closer to the station stop. However when comparing the absolute size of the spread in both deceleration loss times and realised running times, the sizes are very similar and therefore the sharpness of the upper bound curve is just warped due to the scaling on the vertical axis.

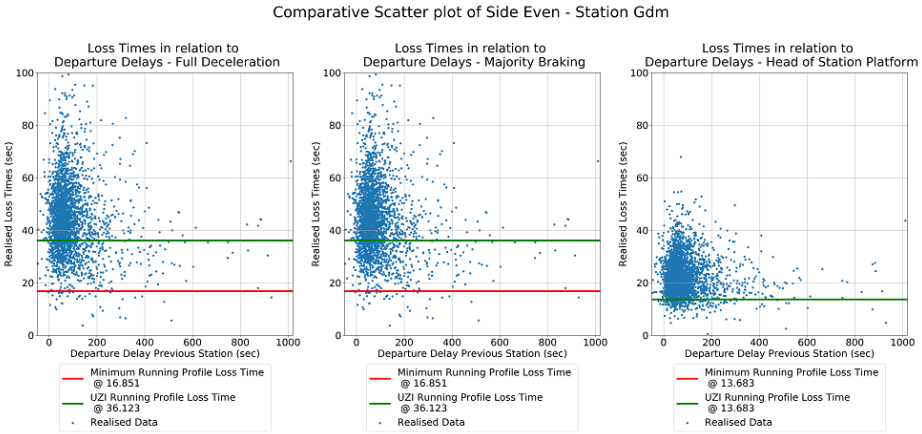


Figure 5.24: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Gdm

As seen in Figure 5.25 and Figure 5.26, the same observation of absolute values of the spread can be made. Station Zbm shows the similarly large distribution spread as seen in the realised running time (Figure 5.29) with loss times ranging from approx. 20 s up to 100

s for the majority of the spread and with outlier loss times as high as 140 *s*. The closer the loss times calculations are taken in relation to the station stop, the consistent the spread in the calculations become with the majority of the deceleration loss times at the head of the station platform being between 10 and 30 *s*. There is one exception outlier at approx. 110 *s*, but this is most likely a hindered train occupying the track section that has survived the data filtering process preceding this analysis.

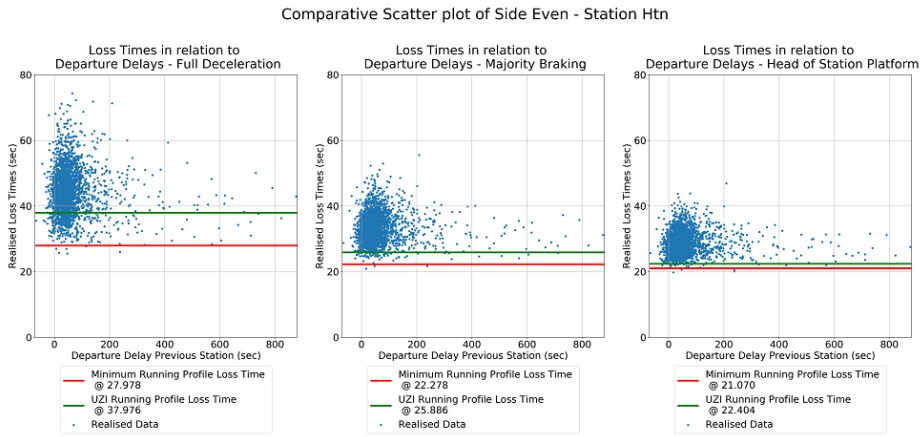


Figure 5.25: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Htn

For all three stations, a similar convergence of the loss times can be seen compared to the realised running times at their respective stations. All of the loss time scatter plot convergence points appear to be larger than calculated reference, leading to the further reinforcement of a more conservative deceleration behaviour consistently at every point in the deceleration regime. The exception to these convergence points, is at station Zbm, which is most likely another consequence of the aforementioned station stop alignment problem. These deceleration loss time results and their convergence points lead to the assumption that even under a time constraint the selected deceleration regime by the train operator is still more conservative than the behaviour expected and used in the current models and scheduling tools using nominal operational values.

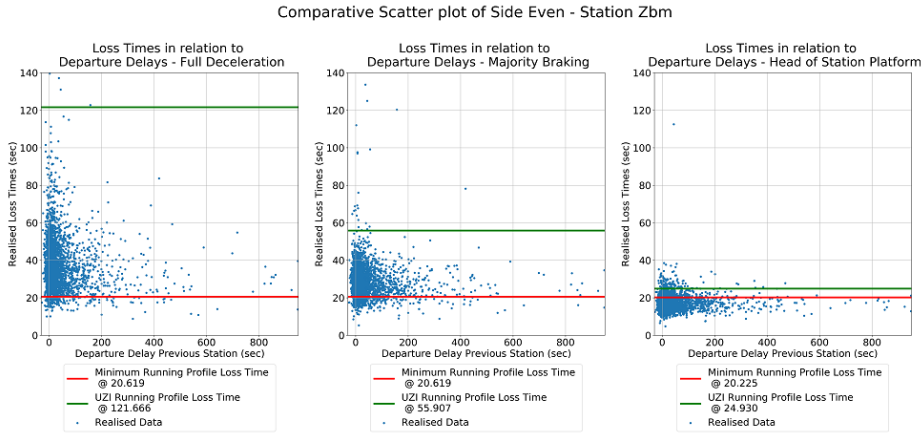


Figure 5.26: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Zbm

These results have shown that there is a convergence in deceleration loss times and with that assumed convergence of the deceleration behaviour, which is still seen to be performing less optimal than the scheduling tools and models were expecting. This finding needs further analysis and could lead to either an investigation analysing the underlying causality for the more conservative behaviour to address it, or lead to an alteration of the scheduling tools and models to account for a more conservative deceleration behaviour.

5.3.3. Running Times vs Deceleration Loss Times

As the scatter plots of the preceding comparative analyses (i.e. the realised running times and deceleration loss times against the departure delays) have shown strong similarities in the presented data, it led to the assumption of the existence of a strong correlation between the realised running times and the respective deceleration loss times. This led to this comparative analysis directly comparing the running times to the deceleration loss times, in order to understand the correlation between them. The following scatter plots show some interesting and visually strong correlations between the two datasets.

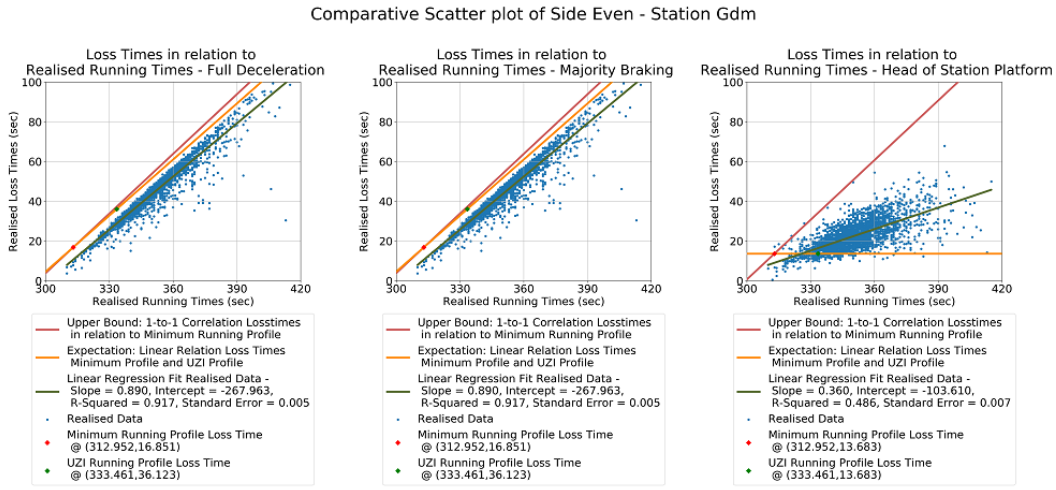


Figure 5.27: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Gdm

Within these presented scatter plots, two lines are plotted to be used as reference. The dark red line represents the upper bound at a 1-to-1 correlation between running times and loss times in relation to the calculated minimum running time (i.e. 1 *s* loss time for every 1 *s* running time with the origin at the calculated loss time and running time of the minimum running time profile). The orange (lighter) line represents the expectation, being the linear relation between the calculated minimum running time profile and the calculated UZI-profile (i.e. a straight line on which the calculated running time and loss time are found on for both the reference profiles). The dark green line seen in the scatter plots represents the linear regression fit over the data points presented.

In Figure 5.27, a very clear correlation can be seen in the plots for the full deceleration distance, with a near 1 to 1 correlation, which is reinforced by the slope coefficient being 0.890 for the fitted linear regression. However getting closer to the station stopping point, this direct correlation starts to diffuse by forming a significant spread that diverges with the increase in running time. With both the upper bound and expectation capturing the dataset, the slope coefficient is estimated at 0.360 for the fitted linear regression. The assumption is that the deceleration loss times at these distance markers only partially account for the differences in running times. The more spread grouping of the scatter plots nearer to the station stop would indicate a significant difference in deceleration behaviour during the final stages of the deceleration regime, leading to the measurable differences seen in deceleration loss times. This divergence is implied by the r-squared values decreasing from 0.917 to 0.486 for their respective linear regression fits.

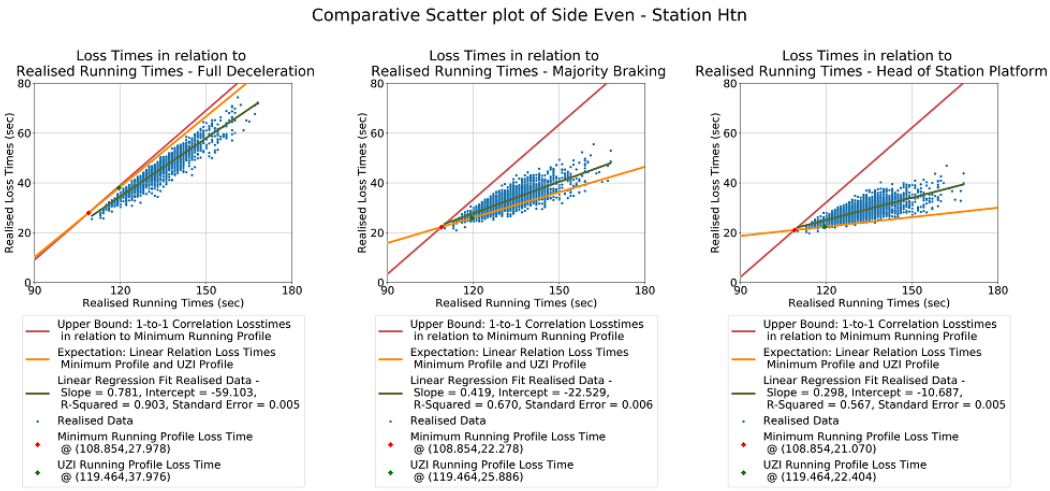


Figure 5.28: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Htn

In Figure 5.28 a gradual but clear progression is visible between the different distances. This would indicate that the cause for the difference in running times for this station can be found spread equally over the entire deceleration regime. The linear regression confirms the gradual change with a slope coefficient of 0.781, 0.419 and 0.298 respectively from left to right. The spread of the scatter plots, however, appears to remain relatively consistent between the three distances with the only difference between the three distances being the skew of the scatter, which appears to inline with the linear relation found between the upper bound and expected linear relation. A slight divergence of the spread is implied when observing the r-squared values gradually decreasing from 0.903 to 0.670 and 0.567 for their respective linear regression fits.

For station Zbm in Figure 5.29, the correlation between realised running times and deceleration loss times at the full deceleration distance is not as strongly defined as with stations Gdm or Htn. Between the three distances, a similar gradual transition as seen at station Htn can be found at station Zbm, with the slope coefficients of 0.417, 0.198 and 0.056 for the fitted linear regression. The scatter plots, however, show a more condensed point cloud when getting closer to the station stop which would indicate a more consistent deceleration performance and deceleration behaviour. This change in diffusion would indicate that a significant portion of the deceleration differences can be found in the earlier stages of the deceleration regime.

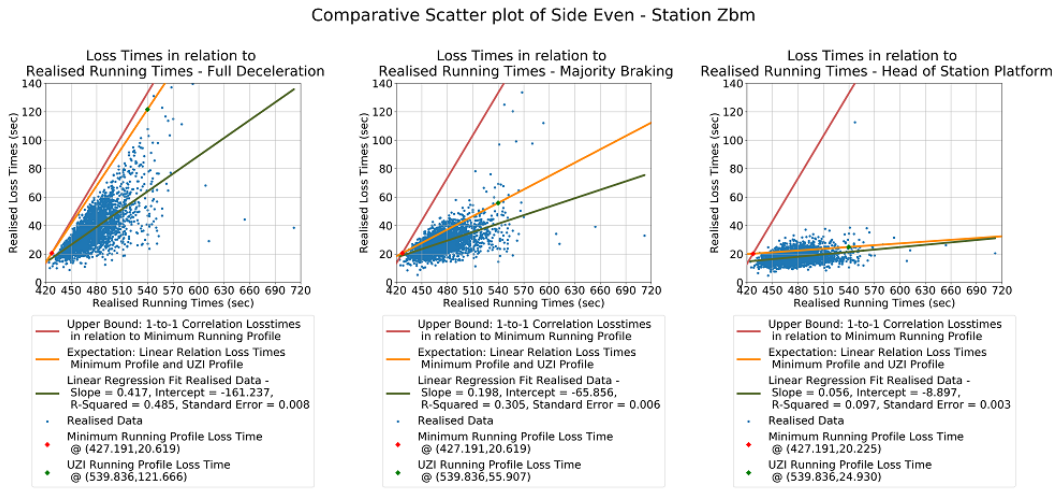


Figure 5.29: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Zbm

The station stopping point estimation, as previously mentioned, has led to an underestimated deceleration loss time (seen in Figure 5.27 and Figure 5.29). However, corrections to the stopping point would lead to the scatter plots only to shift up vertically, while remaining their relative position to each of the other realised datapoints. Any correlations drawn in regards to seasonality and deceleration regimes, are discussed in Section 5.3.4.

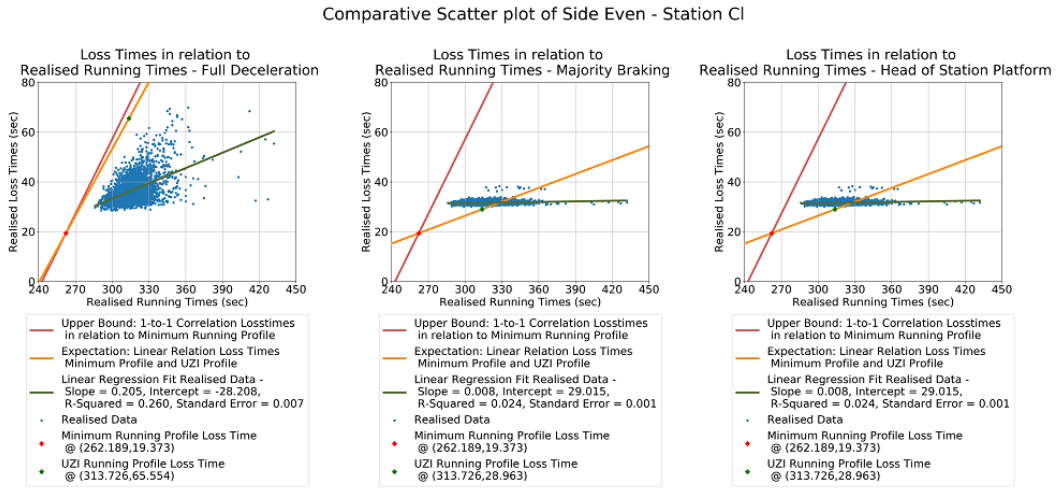


Figure 5.30: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Cl

During this particular comparative analysis, the relations were plotted for all the stations in the corridor direction 'Ht - Ut'. An interesting station worth discussing is the deceleration approach to station Culemborg (Cl), which was taken out of scope due to the DR model performance. While at a full deceleration distance the deceleration behaviour seems diffused with a linear regression estimating a slope coefficient of 0.205 and a r-squared value of 0.260, the deceleration behaviour near the station platform becomes consistent. Furthermore, the skew of the point cloud becomes interestingly flat, with a slope coefficient of 0.008 for the linear regression. This would indicate that for this particular case, no dependency can be found between the realised running times and the deceleration loss times. This would indicate that for station Cl the causality for the distribution of the realised running times can be found at the start of the deceleration regime.

5.3.4. Causality: Seasonality & Deceleration Behaviour

To find any causality to the results found and discussed in the previous sections, the seasonality was tested as the weather and seasonal conditions are assumed to have the bigger impact on a driver's driving behaviour and operational performance. This test of seasonality is done by highlighting two distinctly different months, March and September. To verify the findings, the monthly highlights were shifted to a month earlier, presenting the frosty February and the auspiciously summery August. Comparing both the boxplots and histograms, no significant difference was observed in both the loss times as the running times. Therefore it was assumed that the seasonality comparison between the months March and September, of which a week was used in the DR model, would be the better choice for the seasonality testing.

To be able to draw correlations between driver's estimated deceleration regime and the resulting performance, the same relation between departure delays, realised running times and deceleration loss times are plotted. However for the following plots, only the selection of realised train runs, of which the deceleration regimes are estimated with the DR model, is presented. Two other highlights are made in these scatter plots, distinguishing between the five tested braking variants and the different estimated deceleration regimes. All the highlighted scatter plots (i.e. seasonality, braking variants, deceleration regimes) are the same or similar to the plots of the previous sections describing the different relations, with the exception of the added colouring to identify the different groups and the train run selection in regards to the braking variant and deceleration regime highlight.

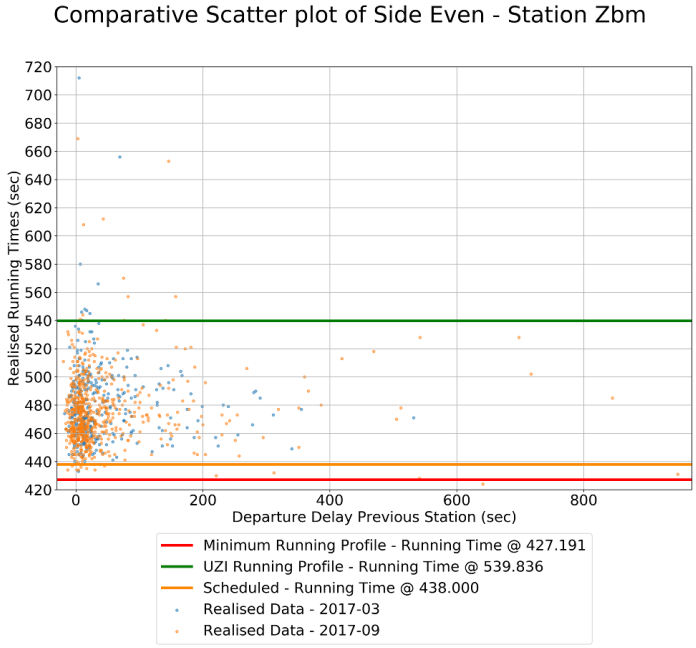


Figure 5.31: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Seasonality Highlight - Zbm

Observing the seasonality for station Zbm in Figure 5.31, no distinct groupings can be found between the different months highlighted, which reinforces the findings regarding the similarity in the histograms and boxplots, shown in Figure 5.14. This lack of grouping is not limited to station Zbm, as it can also be seen in the scatter plots for station Gdm (Figure L.4) and station Htn (Figure L.8).

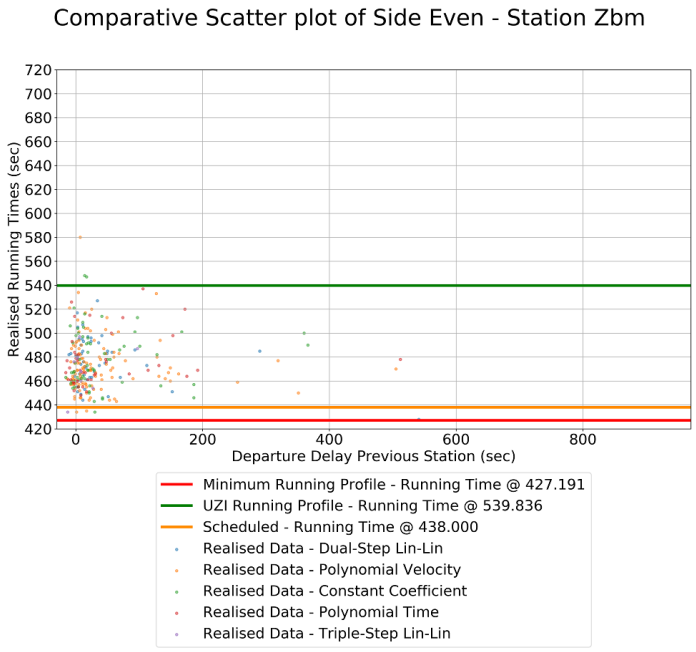


Figure 5.32: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Brake Variant Highlight - Zbm

The scatterplots for highlighting the braking variants (Figure 5.32) and the deceleration regimes (Figure 5.33) differ on the selected group of datapoints to match the train runs used in the DR model analysis. No clear groupings are seen in either of the highlighted scatter plots, relating either to the running time or departure delay. This leads to the conclusion that there is no correlation between the different braking variants or deceleration regimes and the realised running times. This conclusion can be extended on stations Gdm and Htn, as they show similar results in Figures L.5, L.6, L.9 and L.10.

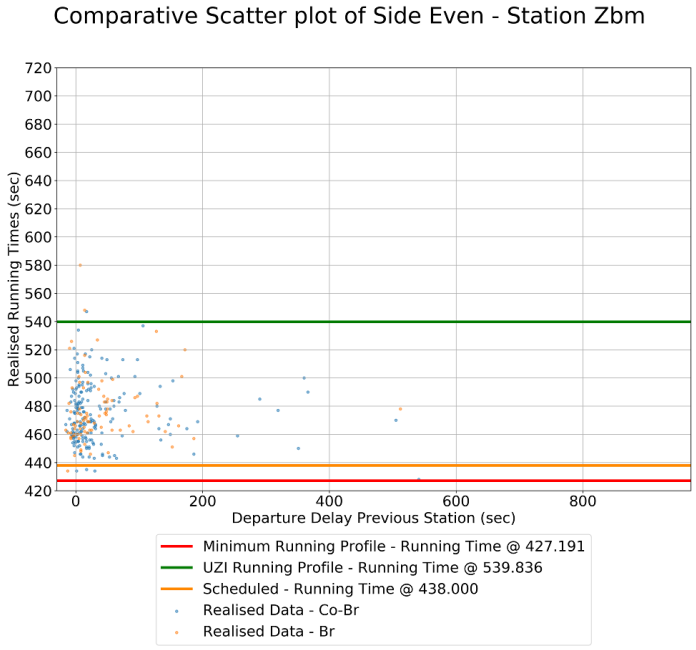


Figure 5.33: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Deceleration Regime Highlight- Zbm

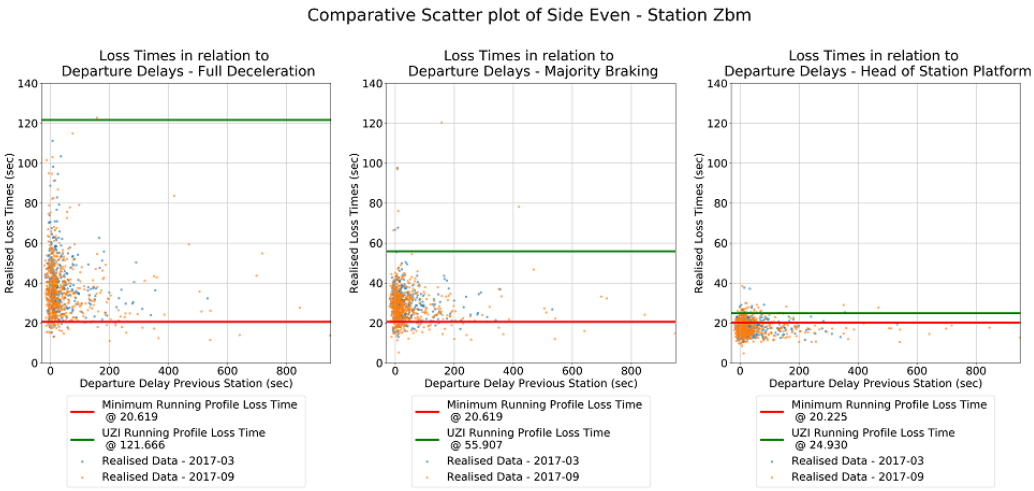


Figure 5.34: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Seasonality Highlight - Zbm

Expanding the same highlighting to the scatter plot relations departure delay versus deceleration loss times and running times versus deceleration loss times can be seen in Figures 5.34 and 5.35 for seasonality (for all three highlights and all three stations, see Figure L.18 till Figure L.42), which show a similar lack of correlation to the tested differentiation by the highlights. This further reinforces the conclusion of no correlations to be drawn in the tested relations for the tested highlights.

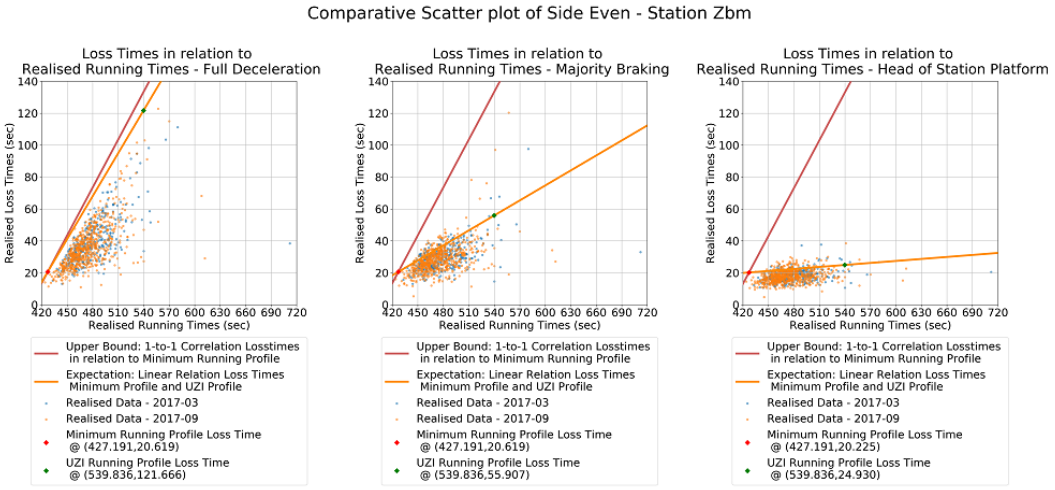


Figure 5.35: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Seasonality Highlight - Zbm

5.4. Conclusion Analysis Results & DR Model Validation

The DR model is heavy on resources which results in a long computation time. This could have also provided more logical and accurate results with some boundary tweaks, finer data quality or denser data point density. The current results however, are promising from an accuracy standpoint, but also from a model generalisation standpoint by dynamically defining the deceleration regimes and sub-regime lengths.

Analysis regarding non-uniform braking behaviour is promising as the DR model identifies a significant number of best fitting profiles to non-uniform variants, making them a very plausible alternative to the constant braking rate. However, this would require more consistent, more frequent and higher quality data, would require a more detailed research into the topic of non-uniformity and might require different data sources. More frequent time/distance sampling along with higher consistency for the current data sources would improve the solutions and would provide a better insight to the applied deceleration regime and brake variant.

The empirical distributions have shown clear stochastic behaviour with regards to realised running times, deceleration loss times and track section occupation duration. All are estimated to adhere to the Exponentially Modified Gaussian (EMG) distribution family with a shape coefficient of at least greater than 1. This provides empirical evidence to the existence of stochastic deceleration behaviour and the resulting impact on infrastructure occupation.

Comparative analysis has shown performance convergence in running time and loss time with increasing departure delay. It has visualised and quantified the correlation between the running times and the loss times, which show a strong linear correlation. Lastly, it has shown no direct correlation in performance between estimated driver behaviour or seasonality, with regards to running times, loss times and track occupation.

6

Conclusion & Recommendations

After applying the knowledge gained from the literature review in Chapter 2, testing the analysis methodology described in Chapter 4 and generating the results seen in Chapter 5, according to the performance indicators described in Chapter 3, this research is concluded with answers to the research questions, an assessment regarding the limitations of the research and a list of recommendations regarding usage of research and further research. The conclusion is structured around answering the research questions along with providing some short summarised general findings and closes with a verification on the achievement of the research objectives. The conclusion is followed by addressing the assumptions and weaknesses of this research. In closing, recommendations are presented to what could improve this research and could be investigated in future researches.

6.1. Conclusions

The research questions were developed to guide the focus of the research. To guide the conclusion of this research, answers will be provided to these questions. Leading in to the answers for each of the questions, a short summary of the methods and findings regarding the question are presented.

6.1.1. Research Main Question

To answer the main question, the stochastic behaviour between stations in terms of realised running times had to be verified. These results were obtained from the network timetable with both scheduled and realised times for station departures and arrivals. The verification of the stochasticity in realised running times bridged into the validation of the stochastic nature within deceleration performance indicators, known as deceleration loss times. The deceleration loss times were calculated from the track occupation times and realised arrival times at the stations, obtained from the train describer system TROTS and the network timetable respectively.

To estimate the applied deceleration behaviour from realised location tracking data, a deceleration reconstruction model (described in Section 4.4) was developed to extract the estimated behavioural coefficients best suited to the data and to test assumption that even the braking application would differ between drivers, as described in Section 4.4.3. With these coefficients extracted, details of the deceleration behaviour could be tested for their stochastic nature.

How does the stochastic behaviour, observed in the deceleration approach of a planned station stop, impact the infrastructure occupation in the network corridor between stations?

This research has empirically shown that there is indeed measurably significant stochastic behaviour, ranging from the realised running times and deceleration loss times to the estimated braking rates and the applied deceleration regimes to the impact the realised behaviour has had on track section occupation duration. It has shown to have a strong and linear correlation between the known spread of realised running times and the deceleration loss times, which in turn relates to the realised track section occupation duration.

Distributions for each of the performance indicators adhere very similarly to a distinct distribution shape, that being either the shape of a Gamma-distribution or the shape of a Exponentially modified Gaussian (EMG)-distribution with their respective shape coefficients of at least greater than 1, as seen in the example plot of Figure 6.1. The distributions were in this research only fitted to the EMG-distribution family as they appeared to fit slightly better at a more consistent rate than the Gamma-distribution family. The shape, scale and location coefficients were strongly dependant of the sample size and binning range of the datasets and are therefore left as a detail in their respective plots. All of the distributions, however, have at least a skewed Normal distributed appearance and will be between a Gaussian/Normal-distribution and an Exponential-distribution, hence the selection and fitting of the EMG-distribution.

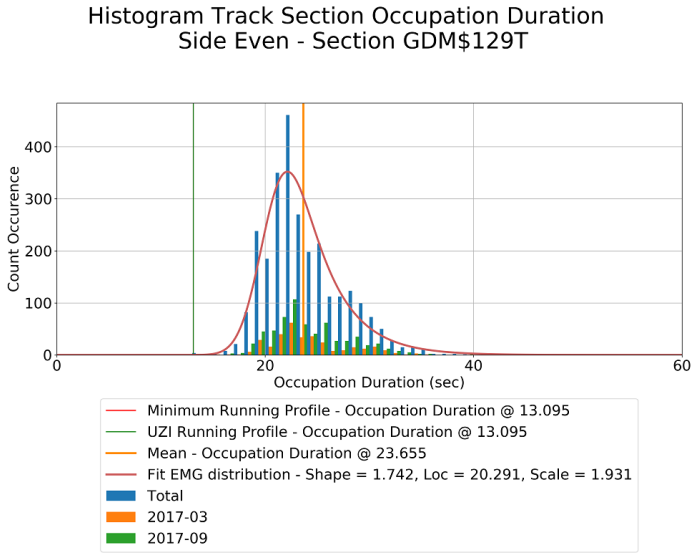


Figure 6.1: Distribution of track section occupation duration at track section GDM$129T

With regards to the impact of this stochastic deceleration behaviour on track section occupation, the occupation duration distributions as seen in Section 5.2.6 and in Figure 6.1 show both the histograms as well as the box plots of their realised distributions. Both distribution plotting types show that the majority of the realised train runs operated close to the expected track section occupation duration, but also show that some of the occupation durations lasted nearly double that of the expected duration in the particular section. In absolute terms the distribution spread differs at most 20 *s*, however in relation to the expected, this translates effectively to doubling the occupation duration and halving the potential infrastructure occupation. The exception of the track sections highlighted was the track section GDM$129T (Figure 6.1) just before station Geldermalsen, as it performed worse and clearly underestimated required occupation duration by about 50 to 100 %. This could have been caused by the realised running speeds being lower through the station switching area than the speed limit allows over this area of the track.

6.1.2. Research Sub Questions

The research sub-questions were made to break down the main question into simpler more specific questions, in support of answering the main question. Each of the sub-questions are given slight summarised but more specific findings regarding the used methods and results related to the sub-question before providing an answer.

What vehicle parameters are required to fully describe the behaviour in a deceleration approach of a planned stop?

To fully describe the deceleration behaviour of a realised train run, one would need to ascertain the applied deceleration regime and break the deceleration behaviour into sub-regimes defining different operational states (i.e. coasting, cruising, braking). These sub-regimes are subject to the differential equation, discussed in Section 4.4. As assumptions were made with regards to the non-uniform braking behaviour, this differential equation was expanded to become the equations specific to the braking variants, discussed in Section 4.4.3.

These differential equations require the vehicle's resistance coefficients, mass, velocity and applied brake rate. The velocity and details of the applied braking rate are obtained from realised location tracking data and estimated through the optimisation of the deceleration reconstruction model over said realised location tracking data respectively. The vehicle's resistance coefficients and mass are obtained from the table of nominal values, which was derived from the vehicle type's performance tests. The vehicle resistance coefficients were defined by the Davis Equation (example equation seen in Section 3.2) of which the constant vehicle resistance r_0 and quadratic resistance r_2 were given a margin of 10% to allow for some compensation of external influences, while the linear resistance coefficient r_1 has set to a constant as previous research (Bešinović et al., 2013a,b,c) stated this variable to be stable enough to be considered as one.

This leads, for β-vectors (i.e. a list or sequence of variable parameters describing the vehicle's behaviour within the DR model) required for the optimisation model, to be formed according to the tested deceleration regime and braking variant, which is shown for the braking variant 'Constant Coefficient' in Table 4.2 with the respective boundaries for the optimisation given in Table 5.1. Therefore the parameters required to fully describe the deceleration behaviour, besides the vehicle parameters r_0, r_1, r_2, braking rate, vehicle velocity and mass, are the ones describing the deceleration regime and subsequently its sub-regimes. The sub-regimes are described by braking variant, sub-regime length S_i, sub-regime duration Δt_i and velocity difference Δv_i.

How would the collection and processing of the required data for the behaviour analysis be defined?

The analysis methodology, described in Chapter 4, provides the process structure used to turn the required data into the desired information. This chapter also states the data required for the analysis. The data sources can be grouped as 'Location tracking Data', 'Infrastructure Data' and 'Nominal Vehicle Characteristics'.

The required location tracking data for this research consists firstly of the rolling stock's GPS tracking data which was pre-processed with the corresponding train numbers, timetable points and signalling data (i.e. passage time and signal state) into the enriched dataset, known by the Dutch IM ProRail and TOC NS as, MTPS data. Secondly the infrastructure-sided vehicle tracking was done through the Dutch IM's train describer data system TROTS, which logs when a train ID is registered to occupy or release a track section on the Dutch infrastructure network. These logs provide a less dense dataset with the usually longer distances between datapoints, but have the advantage of the location and time of a vehicle to be more accurately defined and are not dependant on a fickle wireless connection. Lastly, the network timetable was procured from the Dutch TOC NS and it provided both the scheduled and realised times for station departure, station arrival and

drive-through timetabling points.

The required infrastructure data was obtained from the Dutch IM's GIS system InfraAtlas, which stores the location of every switch, signal, timetabling point and track section weld in terms of distances to the starting point of each train path in the system. Along this dataset, a table of GPS coordinates for each train signal in the network and a table for translating the track section IDs from InfraAtlas to TROTS were provided. These two tables were built by the data analysis department at the Dutch IM ProRail.

The nominal vehicle characteristics grouped the nominal parameters describing the vehicle resistances, mass(i.e. both static and rotational mass) and operational characteristics (i.e. traction force and braking rate). These were obtained from TU Delft and were derived from independent testing done by Ricardo Rail. These values are used by scheduling tools as expected performance to determine the feasibility of a newly developed network timetable. For this research it was used to determine the expected performance and used as a reference to the realised train run data. It also functions as an extra starting point for the optimisation algorithm.

The development of the data fusion from different data sources, is done to improve on the quality of the data sets supplied to the DR model, both in robustness and data point detail level and to create a finer mesh-grid of data points. The pre-processing of the data is described in Section 4.3. The pre-processing of the location tracking data covered processes, such as data formatting, creation of derivable data, data alignment, data fusion and filtering. As this data driven analysis uses several data sources, data had to be formatted accordingly to conform to the chosen data structure as to aid simplify further processing steps and the analysis (i.e. date-time structures, decimals, file structure). The creation of derivable data was to correct or fill in missing data points, such as average velocity in the train describer data, relative distances between the rolling stock's GPS points and the average velocities between the GPS data points.

The data alignment is done firstly through the use of a common GPS coordinate anchor point and later refined by adjusting for differences in distances in relation to the respective time-stamps to compensate for sensor offset to the rolling stock combination's first axle. After the alignment, the different data sources are fused into a single larger unified, DR-compatible formatted dataset, combining the strengths of both location tracking data sources while compensating for their respective weaknesses, as discussed in Section 4.3.5. Lastly the filtering of the train runs, was done using the approach of defining 'velocity boxes' in the speed profile space. The filtering on most restrictive allowed signalling combination on a station approach was foregone as it was too complex to implement, too difficult to properly determine the signalling profile to filter on and least reliable of the two filtering techniques.

How would the data-driven reconstruction model be defined to allow for the fusion of different database sources and to analyse the realised data to reconstruct the deceleration behaviour of vehicles in detail?

Considering the purpose and approach of the model to be data-driven, it will have to process the location tracking data of realised train runs. This adds a layer of iteration to the DR model, as it is calculated for each individual train run. The approach chosen within the DR model, is to estimate each deceleration approach to a station stop for each deceleration regime profile and each braking variant, with each analysis scope restriction adding a layer of iteration. On top of all these layers, the optimisation algorithm tests unique combinations of β-vectors, testing with each evolutionary generation a group of potential solutions for several generations. This causes the programming of the DR model to have a highly nested for-loop, on top of a highly nested for-loop in the optimisation algorithm. The deceleration reconstruction model's process structure is elaborated on in Section 4.4.

The estimation of the realised train run's deceleration regime parameters, described by the β-vectors, is done by applying the optimisation algorithm to the minimisation problem de-

scribed in Section 4.4.6 and finding the β-vector best fitted to the realised data. As the realised data is derived from multiple data sources, the DR model is preceded with a substantial pre-processing stage to process and prepare the realised data for data fusion and the DR model. These processes are described in the previous research sub-question.

What are the distribution types and stochastic values of the realised deceleration behaviour of commuter heavy rail vehicles?

Observing the results in Section 5.3 for all performance indicators with histogram plots, they are all estimated to be fitted to either a Gamma-distribution or more likely to a Exponentially modified Gaussian-distribution (EMG) with their respective shape coefficient greater than 1. The shape, scale and location coefficients were strongly dependant of the sample size and binning range of the datasets and are therefore left as a detail in their respective plots. All of the distributions, however, have at least the appearance of a skewed Normal distribution and will range somewhere between a Gaussian/Normal-distribution and an Exponential-distribution.

Could any sub-regimes be determined in the realised deceleration behaviour?

Due to the noisiness of the data, the deceleration regime profiles were obtained through visually inspecting the realised data after the data alignment and fusion of both GPS and train describer data sources. This method was chosen as it helped with computational time by not having to apply the differential equation and optimisation algorithm to each of the datapoint pairings to estimate the possible sub-regime. Furthermore it helped with reducing code development time trying to deal with the noisy data and correctness of the sub-regime type estimation. The visually derived deceleration behaviour was clearly visible in the speed profiles of the realised train runs and the different behaviour patterns were noted as a combination or profile of sub-regimes (e.g. 'Br' or 'Co-Br-Cr-Br', see Table 4.2 for all 10 deceleration regime profiles selected in this research).

How do the parameter distributions of the realised deceleration behaviour compare to the nominal parameters of the expected deceleration behaviour?

The vehicle parameters r_0 and r_2 were allowed to vary slightly to consider slightly different external conditions. The results of the DR model (see Section 5.2.2) show that constant coefficient r_0 is deemed a consistent variable, with the majority of the tested runs to return a value near the nominal value or hitting the lower bounds. However, these results hitting the lower bound are assumed to be caused by the optimisation algorithm having a hard time trying to fit the tested deceleration profile and braking variant. The quadratic coefficient r_2 shows that it is a very sensitive variable, which is understandable considering the quadratic or cubic relation to the vehicle's velocity in the applied velocity differential equations. Other research would be required to investigate the relations of these parameters.

With regards to the braking rate distributions , it was defined in several perspectives to relieve any sense of bias by checking both the average braking rate in a braking regime and its maximum rate. This was further expanded by making a distinction between braking rate between stopping and slowing down to a lower velocity, when two or more braking regimes are estimated to be present. They all show a significant stochastic distribution in a similar shape as to the realised running time or deceleration loss time distributions and were fitted with the EMG-distribution. A lot of the estimated braking rates had a lower value than the nominal braking rate would have the TOC and IM expect. Making the distinction between average braking rate and maximum braking rate, provided a different view in which the realised and expected braking rates were closer to each other. However, even these distributions had a mean braking rate at these stations that was lower than what the operational expectations are for these vehicles.

What are the causalities for the stochastic variations found in a vehicle's deceleration behaviour?

The causalities tested in this research were the estimated applied deceleration behaviour and the effects of seasonality on operational performance. However, results discussed in Section 5.3.4, show that no direct correlation could be found. This could be due to the selection size of an entire month for the seasonality, allowing for undesirable season and weather conditions to obfuscate the results. However, the KNMI (Royal Netherlands Meteorological Institute) weather report for the months March and September show that these months were fairly consistent and for the significant majority had the right season and weather conditions.

With the regards to the estimated deceleration behaviour and partially the seasonality, the location tracking data might not have been consistent, refined or available enough to clearly define the deceleration behaviour. This will have to be further investigated when using different datasets that are either more extensive or from a different source.

6.2. Assessment Limitations Research

This research, while addressing the scientific gap, has its limitations in what or how it could address the scientific gap. Some limitations were addressed with founded assumptions, others with a simplification of the analysis methodology or data processing.

With regards the performance indicators, the indicator used in the estimation of infrastructure occupation, a simplification was made by defining the infrastructure occupation based on track occupation duration (i.e. duration between the occupation and the release of the track section), in order to validate the track section's performance as it is assumed to be scalable to the complete track section blocking times.

Other assumptions made within this research, were on the types of braking applications and the existence of the non-uniform braking rates within the braking regimes. The selected mathematical description of the non-uniform braking rates, that were tested within this research, were mostly derived from the scientific paper describing the braking behaviour in road vehicles on the freeway in unrestricted conditions (Maurya and Bokare, 2012). The second order polynomial time dependent braking rate and the 'Triple Step' braking variant were inspired by their variant cousins.

The reason for assuming a non-uniform braking rate application, was due to some visual observations in the curvature of the realised speed profiles not conforming to the otherwise constant and uniform application of the braking rate. However, the available realised data still was not refined enough to distinguish the nuances of the braking behaviour and to make any clear conclusion on the alternative braking application, especially with the smaller/shorter braking regimes. This would require further, more in-depth research in the application of the braking rate within the braking regimes of rail vehicles.

Some simplifications were made in this research with regards to the data pre-processing. The filtering of the realised train runs was simplified with the use of a v-box filter, instead of implementing a filter based on a signalling combination profile (i.e. a combination of signals settings in a chain of infrastructure signal) on the approach to a station stop. The filtering of location tracking data through the use of a Kalman filter was suggested. However, the used data was pre-processed by the IM incidental to no filtering applied as they stated the accuracy of the raw GPS data was of high quality and that minimal filtering was applied to data points 'jumping' for hundreds of meters, up to even two kilometres within seconds due to the loss and reconnection of the GPS sensors and mismatching timestamps. These same data points were filtered out within this research as they were sometimes left in the dataset, but were considered part of the pre-processing done by the IM when eliminating them from the analysis. The uncertainty of the GPS measurement only became visible in station stops, but only led to the estimated velocity of a vehicle never being considered 0 m/s. This was covered in the analysis by searching for the lowest estimated velocity

within the station area and to consider the found datapoint as the end point of the deceleration regime. Further refinement of the location tracking data near station stops is suggested in order to refine the estimation of the station stopping point of the realised train runs. This may require different data sources or estimation methods.

The limitation in the usage of the current location tracking data sources in the DR model, was the often big differences in average velocity between track sections, causing a significant shift in the averaged velocity of the fused dataset. The DR model handled well enough to smooth out the deceleration behaviour over both or multiple track sections, but would lead inherently to a larger error in the optimisation and fitting of the deceleration regime in the DR model. In the shorter deceleration regimes, such as at station Houten (Htn), it could lead to a misidentification of deceleration regime profiles, assuming two or more braking regimes to be present instead the actual single braking regime. A possible improvement is to introduce weights to the velocities of different data sources and use a weighted average velocity for the fused dataset of the realised train run. This was not considered during this research, in order to treat both the train describer data and the GPS data with equal importance.

With regards to the DR model and to its limitations, the DR model had to 'lock' both ends of the full deceleration distance, instead of freely searching for the deceleration regime's beginning and end in it's station approach. This was in order to account for the noisiness of the realised location tracking data without developing a complex, iterative search algorithm to find the deceleration distance outer ends. The current state of the Python code used for the DR model, has a high computation time (e.g. 1500 s on average for a single thread/process to test 1 realised train run by computing 40 optimisation generations per selection deceleration regime, braking variant and train station), despite the simplifications and limitations made to the model as described above. The DR model has also shown small stability problems with regards to handling certain β-vectors that break the computation cycle. Further computational hurdles include the optimisation convergence rate and the related solution accuracy. The problem with the convergence rate is due to the non-linear relation between vector variables and the minimisation problem, but also with the random nature of the Genetic algorithm in selecting unfeasible solutions/β-vectors and in the crossover and mutation of the offspring pool of vectors.

The Python coding of the DR model will need to be further refined and optimised to reduce the computation time while improving the deceleration regime identification, increasing the number of optimisation generations, improving optimisation convergence and increasing the number of stations tested in the corridor of the realised train run. The coding/model optimisation would make it viable to test larger sample sizes of train runs and to apply the model in a more generalised use case with regards to rolling stock types and infrastructure network.

6.3. Recommendations

The findings and conclusions of this research have led to several recommendations for future research and/or improvements to be made upon this research's analysis method. The recommendations following this research, with regards to continued research, are as follows:

- Investigation further refinement of data collection and processing for the data fusion
- Research to elaborate on specific braking behaviour (i.e. braking application and release) and possible non-uniformity.
- Investigate different data sources to be used for driver operational behaviour analysis.
- Further expansion on case studies in infrastructure and rolling stock to develop and test model generalisation.
- Research to investigate impact of station platform designs and signalling proximity to stopping points on final moments of the deceleration behaviour near the station.
- Research that elaborates on the causality of the applied deceleration behaviour.
- Research investigating the applied deceleration behaviour with regards to the relation between the policy surrounding the stopping signals (i.e. consequences of stopping past the 'red' signal due to underestimated required deceleration distance) and the deceleration guidance given in the approach of these signals.

In general, the future research recommendations can be categorised into two groups, one focused on the development of the DR model and the data processing or acquisition involved. The second group would be aimed at more accurately defining the realised deceleration behaviour and the causality behind the decisions leading to the applied deceleration behaviour.

The future research on the further development of the DR model, is expected to focus on improving the model's generalised implementation through testing different case studies, refinement of the currently used data sources (i.e. train describer data and GPS data) or implementation of different data sources within the data fusion of the DR model. These different data sets can come from sources like the time synced on-board accelerometers, speedometers or odometers, in-station tracking through detection loops within a track section or radar-based vehicle tracking, or improvements to the infrastructure data with regards to stopping points, specific to the rolling stock length, in relation to the rest of the infrastructure.

Future research regarding the deceleration behaviour into stations and braking application behaviour, is expected to be focused on defining a specific braking behaviour and elaborating on the possible non-uniformity of the braking application. This will have to coincide with the refinement of the data sources, as a more frequently and more specific data source would be required to improve on identifying the nuances in the braking application.

Another approach angle for research to further understanding the deceleration behaviour, is to elaborate on the causality (e.g. seasonality, delay state) of the applied deceleration behaviour. In particular, research is suggested with regards to investigating the impact of station platform design and signalling proximity to rolling stock length specific stopping points, and to investigate the relation between the policy surrounding the stopping signals and the deceleration guidance given in the approach of these signals and the impact it has on the applied deceleration behaviour. These last two research topics would investigate if there are any design or policy aspects that could influence the deceleration behaviour and how changes to these aspects would impact and possibly improve the deceleration behaviour.

With regards to direct implementation or consequences of this research, the following can be recommended:

- Expansion or adjustment of the current timetabling tools with the estimation method of the minimum running time expanded or account for a more realistic deceleration behaviour observed from this research.
- Drive interest in further understanding the deceleration behaviour to account for stochastic behaviour in the network or to develop methods to increase the consistency of the applied deceleration behaviour.
- Investigate and adjust signalling locations in the station approach in order to provide more guidance and achieve more consistency in deceleration behaviour on long distance, open-track station approaches.

This research has shown that the nominal values used to calculate the minimum running time and the method of determining the scheduled running times would benefit from some adjustments or expansion to cover the differences observed in the stochasticity with the running times and the deceleration loss times. A clear example of this is the estimated braking rates, showing a mean braking rate of 0.509 to 0.645 m/s^2 compared to the nominal counterpart of 0.8 m/s^2 that is used in timetabling tools and modelling programs. Empirically proving the stochastic nature in both the running times and the deceleration loss times, should provide encouragement to the development of alternative timetabling tools and methods in order to close the gap between expected and realised behaviour and resulting performance.

A goal of this research is to drive the interest of further understanding the realised deceleration behaviour and its stochastic nature, in order to account for the stochastic behaviour or to develop means to improve the consistency of the applied deceleration behaviour. This research has provided evidence to the existence of the stochastic nature of realised operations and provided an empirical link between the stochastic nature observed in realised running times and the deceleration loss times, which is a direct result of the realised deceleration behaviour, and the resulting impact of this stochastic nature on the infrastructure's track capacity through the track section occupation duration.

After investigating the differences in consistency of deceleration loss times between the stations Geldermalsen (Gdm), Houten (Htn), Zaltbommel (Zbm) and Culemborg (Cl) over the three measuring distances, it has shown that the longer distanced, open-track stations Zaltbommel and Culemborg have a large spread of loss times early on in the deceleration regime. Due to the significant differences in the start of the deceleration regimes, an investigation is recommended with regards to the signalling locations in their station approaches. It is assumed that this difference in behaviour is due to less detailed signalling over the entire deceleration distance as a consequence of larger track sections or block sections. Pending the results of the recommended investigation, a recommendation can be presented with regards the spacing of the track signalling, as it could improve the consistency of the deceleration behaviour.

Bibliography

Albrecht, T., Goverde, R.M.P., Weeda, V.A., and van Luipen, J. Reconstruction of train trajectories from track occupation data to determine the effects of a driver information system. *Computers in Railways X*, 2006.

Albrecht, T., Gassel, C., Binder, A., and van Luipen J. Dealing with operational constraints in energy efficient driving. *Proceedings of the 4th IET International Conference on Railway Traction Systems (RTS 2010)*, 2010a.

Albrecht, T., Gassel, C., Knijff, J., and van Luipen, J. Analysis of energy consumption and traffic flow by means of track occupation data. *Proceedings of the 4th IET International Conference on Railway Traction Systems (RTS 2010)*, 2010b.

Bešinović, N. *Integrated Capacity Assessment and Timetabling Models for Dense Railway Networks*. PhD thesis, Delft University of Technology, 2017.

Bešinović, N., Quaglietta, E., and Goverde, R.M.P. Calibrating dynamic train running time models against track occupation data using simulation-based optimization. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013) - The Hague, The Netherlands - 6 till 9 October 2013*, oct 2013a.

Bešinović, N., Quaglietta, E., and Goverde, R.M.P. A simulation-based optimization approach for the calibration of dynamic train speed profiles. *Journal of Rail Transport Planning & Management*, December 2013b.

Bešinović, N., Quaglietta, E., and Goverde, R.M.P. Calibrating and validating train dynamics characteristics against realisation data. *Proceedings of Models & Technologies for Intelligent Transportation Systems (MT-ITS 2013) - Technische Universität Dresden - Dresden, Germany - 2 till 4 December 2013*, December 2013c.

Daamen, W., Goverde, R.M.P., and Hansen, I.A. Non-discriminatory automatic registration of knock-on train delays. *Networks Spatial Econ. 9 (1)*, 2009.

Davis, W.J. *The tractive resistance of electric locomotives and cars*. General Electric, 1926.

De Fabris, S., Longo, G., and Medeossi, G. Automated analysis of train event recorder data to improve micro-simulation models. *WIT Transport Built Environment 103*, 2008.

Ghofrani, F., He, Q., Goverde, R.M.P., and Liu, X. Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C*, 2018.

Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Professional, Reading, MA, USA, 1989. ISBN 978-0201157673.

Goverde, R.M.P. *Punctuality of Railway Operations and Timetable Stability Analysis*. PhD thesis, Delft University of Technology, 2005.

Goverde, R.M.P. and Meng, L. Advanced monitoring and management information of railway operations. *Journal of Rail Transport Planning & Management 1 (2)*, 2011.

Goverde, R.M.P., Daamen, W., and Hansen, I.A. Automatic identification of route conflict occurrences and their consequences. *Computers in Railways XI*, 2008.

Hansen, I.A. and Pachl, J. *Railway Timetable and Traffic*. Eurailpress, Hamburg, Germany, 2008.

Hansen, I.A., Goverde, R.M.P., and Van Der Meer, D.J. Online train delay recognition and running time prediction. *IEE Conference on Intelligent Transportation Systems ITSC*, 2010.

Holland, J. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, USA, 1992. ISBN 978-0262581110.

Kecman, P. and Goverde, R.M.P. Process mining of train describer event data and automatic conflict identification. *Computers in Railways XIII*, September 2012a.

Kecman, P. and Goverde, R.M.P. Process mining approach for recovery of realized train paths and route conflict identification. *Proceedings of the 92nd Annual Meeting of Transportation Research Board*, October 2012b.

Kecman, P. and Goverde, R.M.P. Online data-driven adaptive prediction of train event times. *IEEE Transportation Intelligent Transportation Systems 16(1)*, 2015a.

Kecman, P. and Goverde, R.M.P. Predictive modelling of running and dwell times in railway traffic. *Public Transport 7 (3)*, 2015b.

Maurya, A.K. and Bokare, P.S. Study of deceleration behaviour of different vehicle types. *International Journal for Traffic and Transport Engineering*, 2012.

Medeossi, G., Longo, G., and de Fabris, S. A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management*, November 2011.

Mitchell, M. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA, 1996. ISBN 978-0585030944.

NS. Ns zet eerste stap richting hoogfrequent rijden. `https://nieuws.ns.nl/ns-zet-eerste-stap-richting-hoogfrequent-rijden/`, July 2016. Accessed: June 2018.

ProRail. Prorail en ns maken stap naar hoogfrequent rijden. `https://www.prorail.nl/nieuws/prorail-en-ns-maken-stap-naar-hoogfrequent-rijden`, July 2016. Accessed: June 2018.

ProRail. Programma hoogfrequent spoorvervoer. `https://www.prorail.nl/projecten/delft/programma-hoogfrequent-spoorvervoer`, June 2018. Accessed: June 2018.

Scheepmaker, G.M. Rijtijdspeling in treindienstregelingen: Energiezuinig rijden versus robuustheid. Master's thesis, Delft University of Technology, February 2013.

SpoorPro. Prorail: extra maatregelen bij tests hoogfrequent rijden. `https://www.spoorpro.nl/spoorbouw/2017/08/04/prorail-extra-maatregelen-aan-bij-tests-hoogfrequent-rijden/`, August 2018. Accessed: June 2018.

Tielman, W. An agent-based approach to simulating train driver behaviour. Master's thesis, Utrecht University, January 2015.

Trani, A.A. Cee 3604 rail transportation: Addendum, rail resistance equations. `http://128.173.204.63/courses/cee3604/cee3604_pub/rail_resistance.pdf`, 2018. Accessed: Feb 2018.

Vincenty. Vincenty's formulae. `https://en.wikipedia.org/wiki/Vincenty%27s_formulae#Inverse_problem`, 2018. Accessed: July 2018.

Appendix

A. Example Distribution Graphics Referred to in this Research

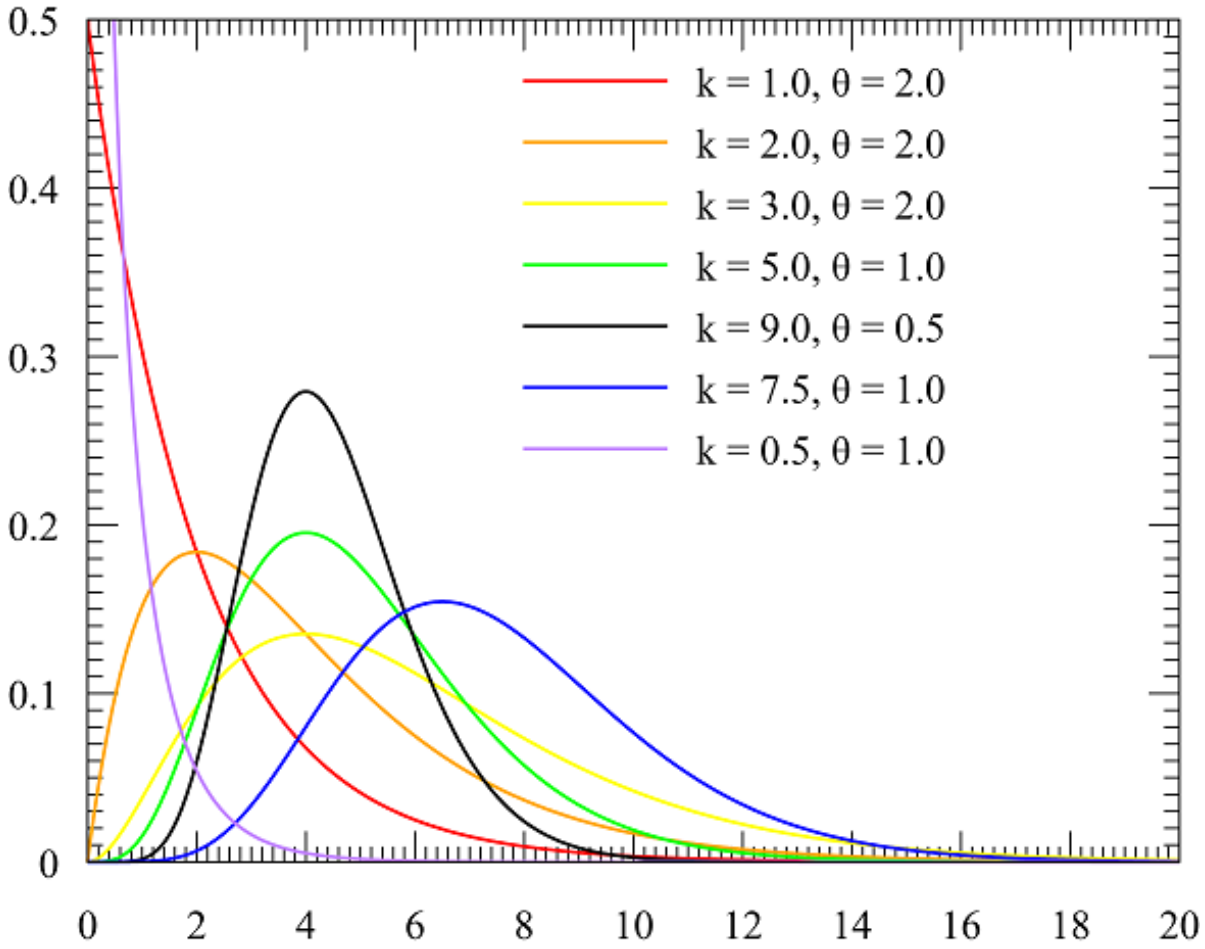


Figure A.1: Example plot showing different PDF plots of Gamma distributions with different shape coefficients (k) and different scale coefficients (Θ). Graph borrowed from wikipedia's page on gamma distribution (`https://en.wikipedia.org/wiki/Gamma_distribution`)

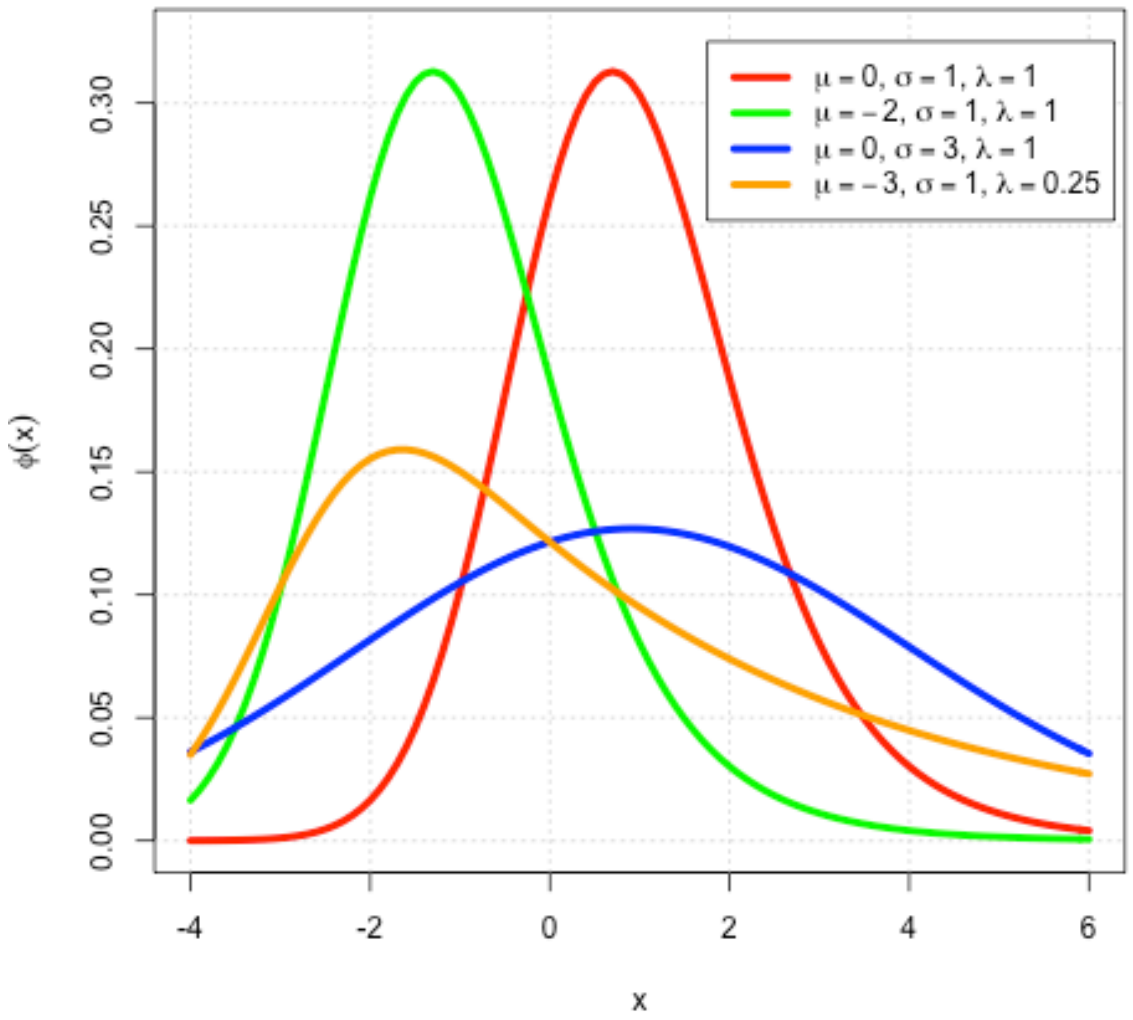


Figure A.2: Example plot showing different PDF plots of Exponentially Modified Gaussian (EMG) distributions with different mean coefficients (μ), standard deviations (σ) and exponential rates (λ). Graph borrowed from wikipedia's page on EMG distribution (`https://en.wikipedia.org/wiki/Exponentially_modified_Gaussian_distribution`). The function of SciPy and their shape coefficient relation ($K = 1/(\sigma\lambda)$), location (loc = μ) and scale (scale = σ) were used as seen in `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.exponnorm.html`

B. Summary Data Sources

Data Type	Data Usage	Source	Required parameters
Infrastructure	RSSLI Model	ProRail - PAB - InfraAtlas	section ID, length, GPS Location, Type, neighbours, static speed limit, curvature, gradient
Rolling Stock	RSSLI Model	TU Delft / Ricardo	Davis Resistance coefficients r_0, r_1 and r_2, braking, nominal vehicle mass
Train Describer	Location Tracking	ProRail - PAB - TROTS	section ID, train ID, occupation time, release time, current signal, following signal
Vehicle GPS	Location Tracking	NS - Data & Analytics	train ID, GPS location, GPS velocity
Network Timetable	Location Tracking	NS - Data & Analytics	Timetable day, train run ID, origin & destination details (i.e. timetable point, action type, planned execution time, realised execution time) rolling stock details(i.e. ID, type, position in couple)
Nominal Regime Characteristics	Nominal PI	NS - Prestatie & Innovatie	acceleration, braking, time allowance, applied driving regime, performance plans

Table B.1: Summary of the datasets used with their respective sources and required parameters.

C. Process Structure

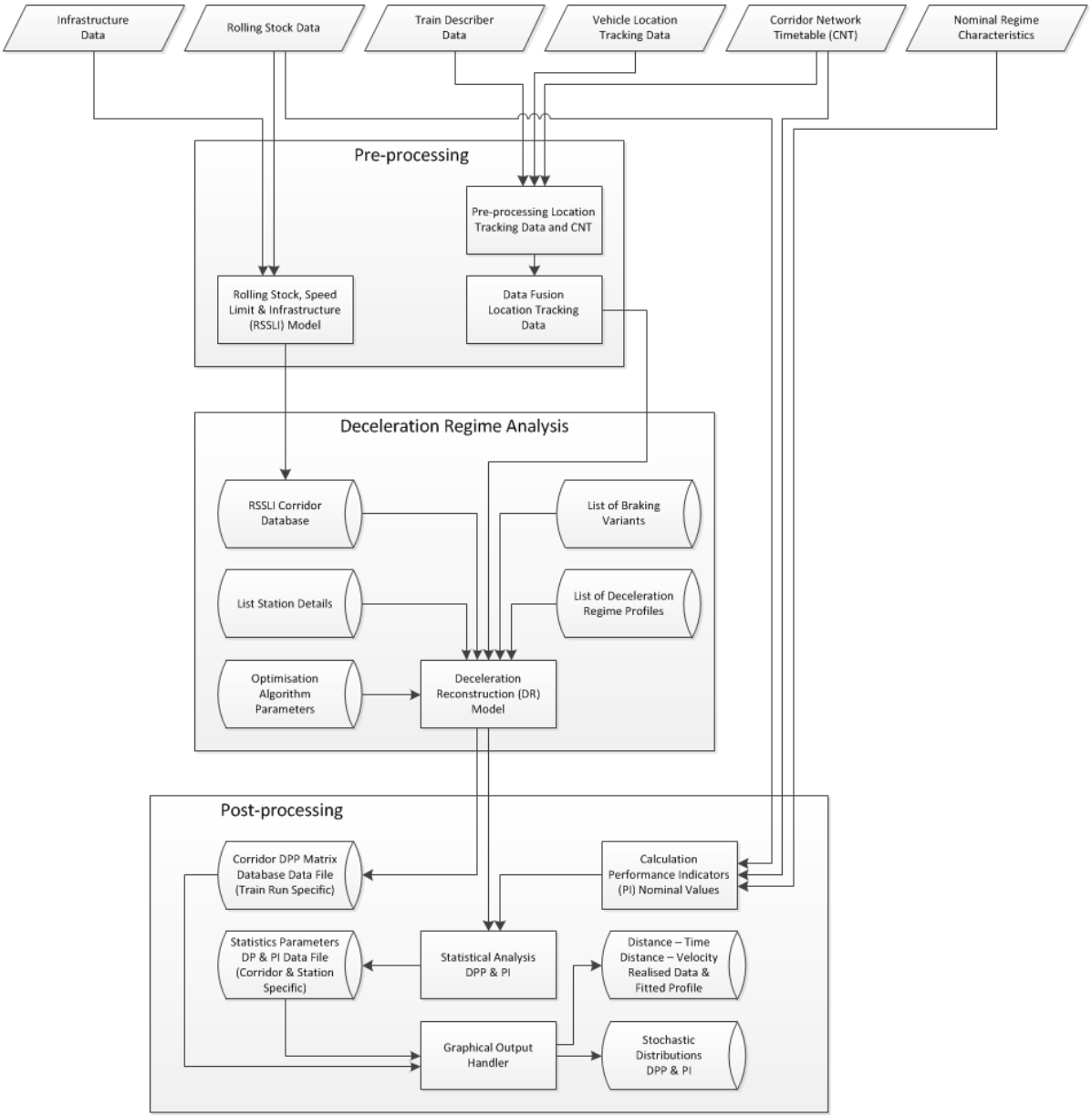


Figure C.1: Conceptual Framework Process Structure

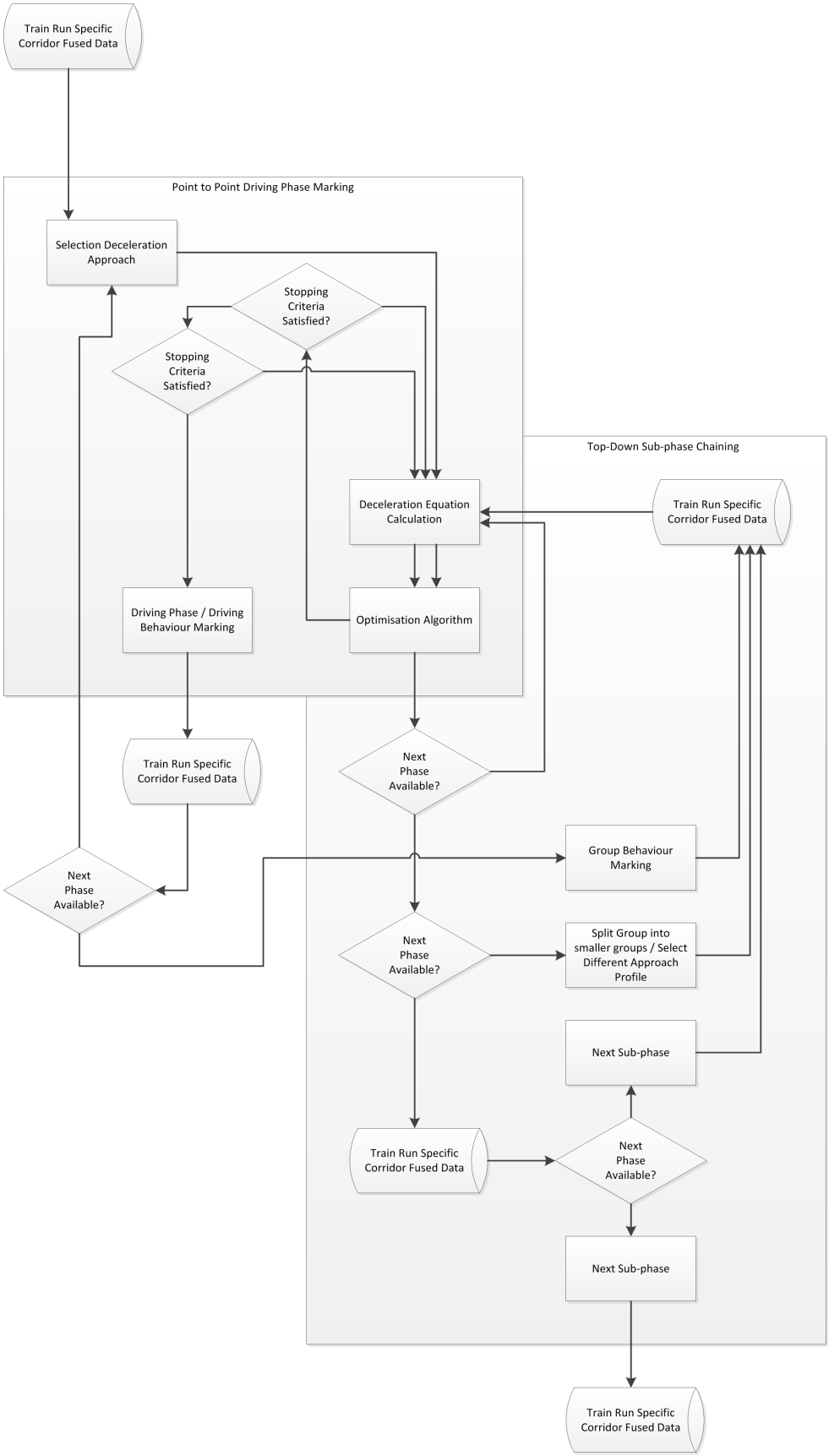


Figure C.2: Deceleration Reconstruction (DR) Model - Point-to-Point Differential Backtracking

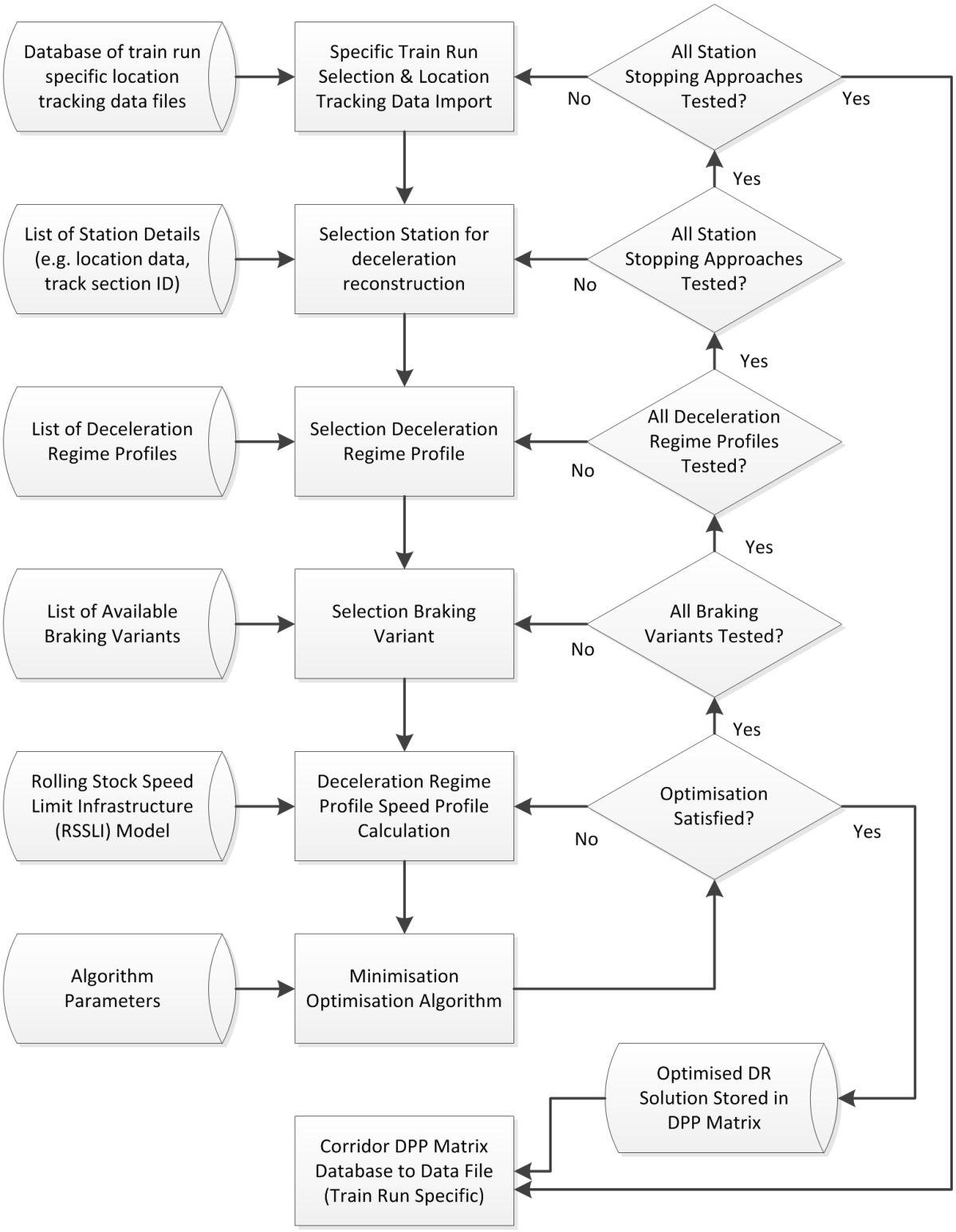


Figure C.3: Deceleration Reconstruction (DR) Model - Velocity Difference Backtracking

D. Braking Variants

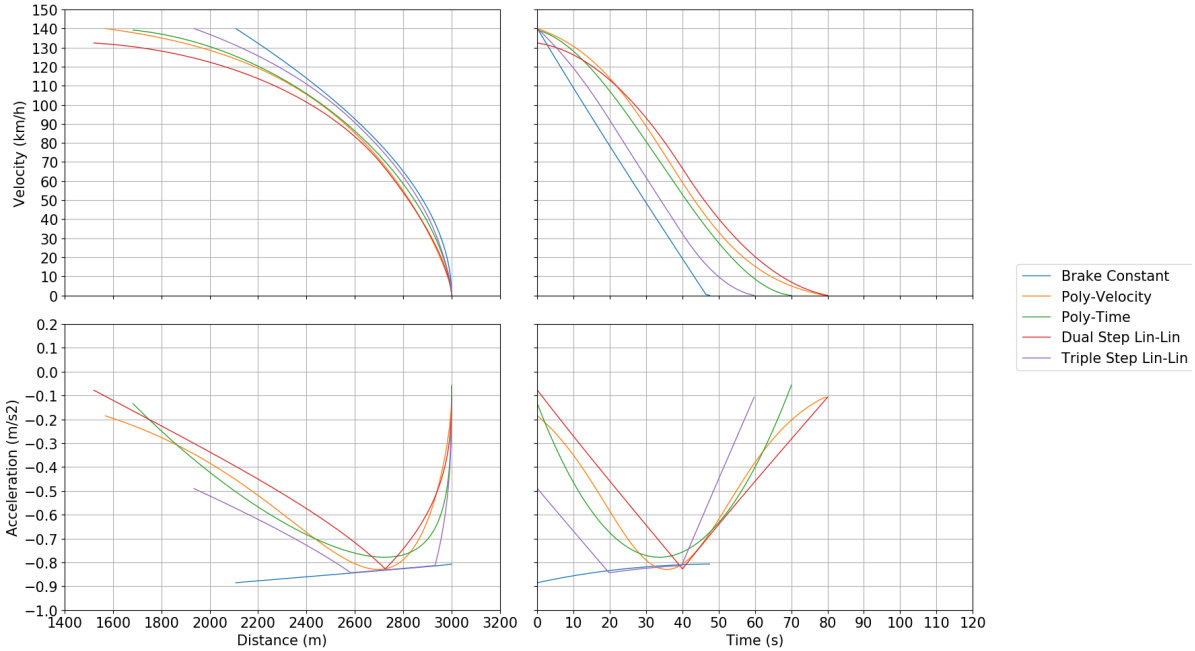


Figure D.1: All tested braking variants sampled to resemble constant braking coefficient deceleration in the distance domain and the effects when comparing the variants in the time domain side by side.

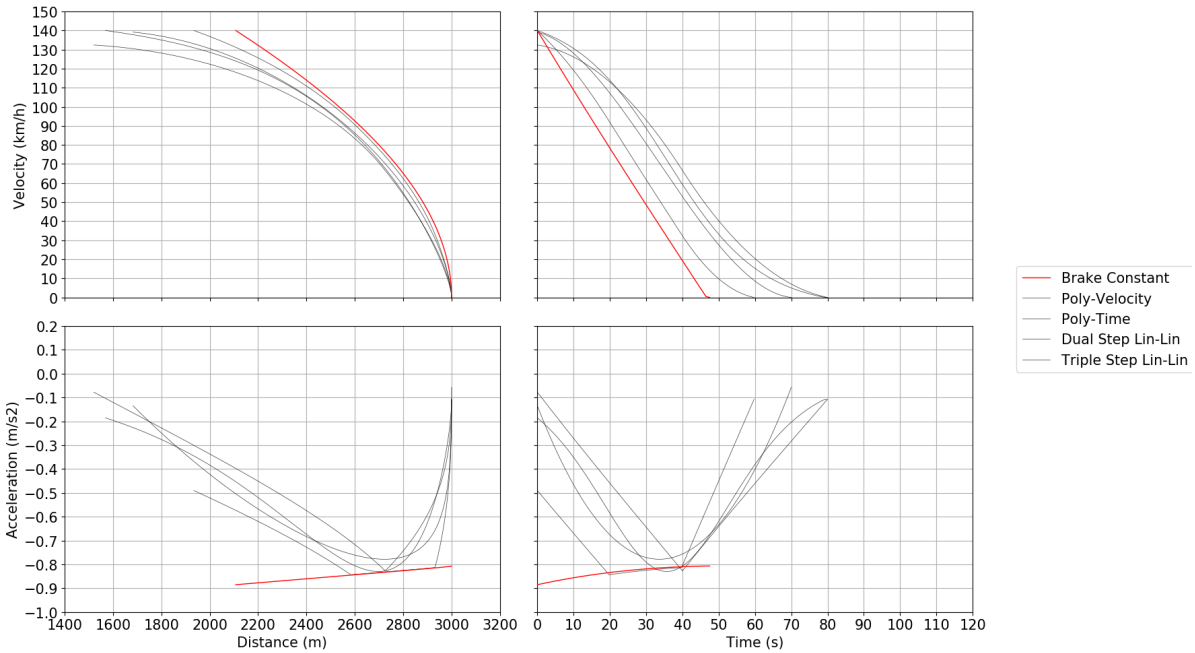


Figure D.2: Constant Braking Coefficient example shown in both velocity and acceleration in both distance and time domain, highlighted in red with other examples in black for comparison/contrast.

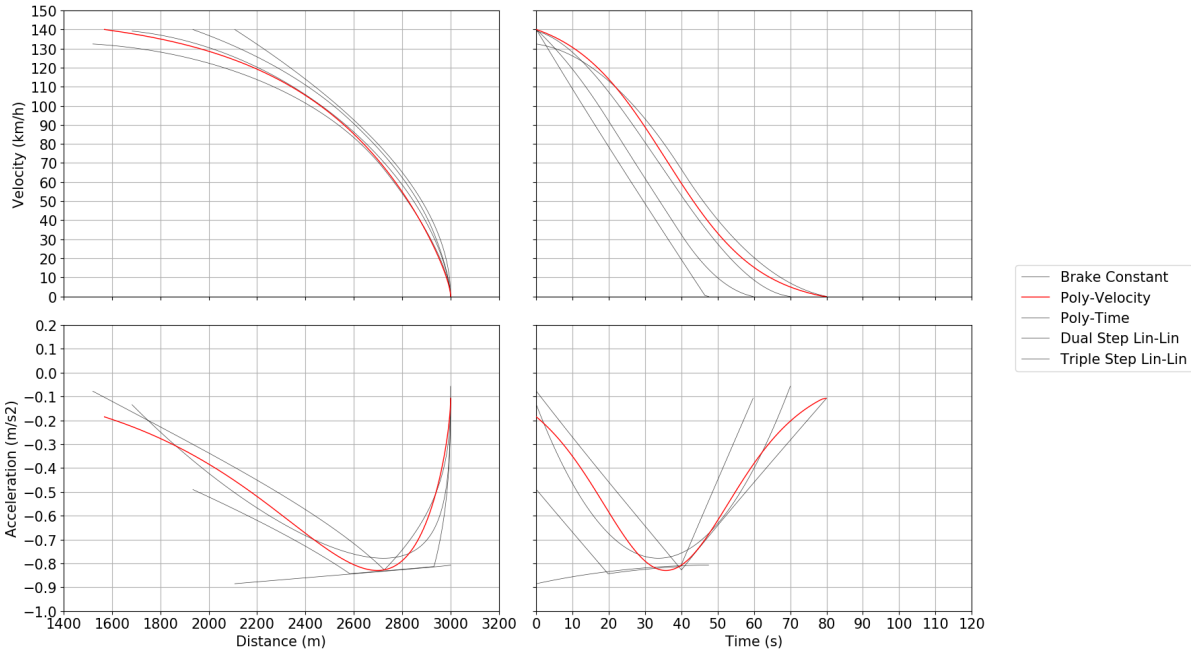


Figure D.3: Second order polynomial (velocity dependent) example shown in both velocity and acceleration in both distance and time domain, highlighted in red with other examples in black for comparison/contrast.

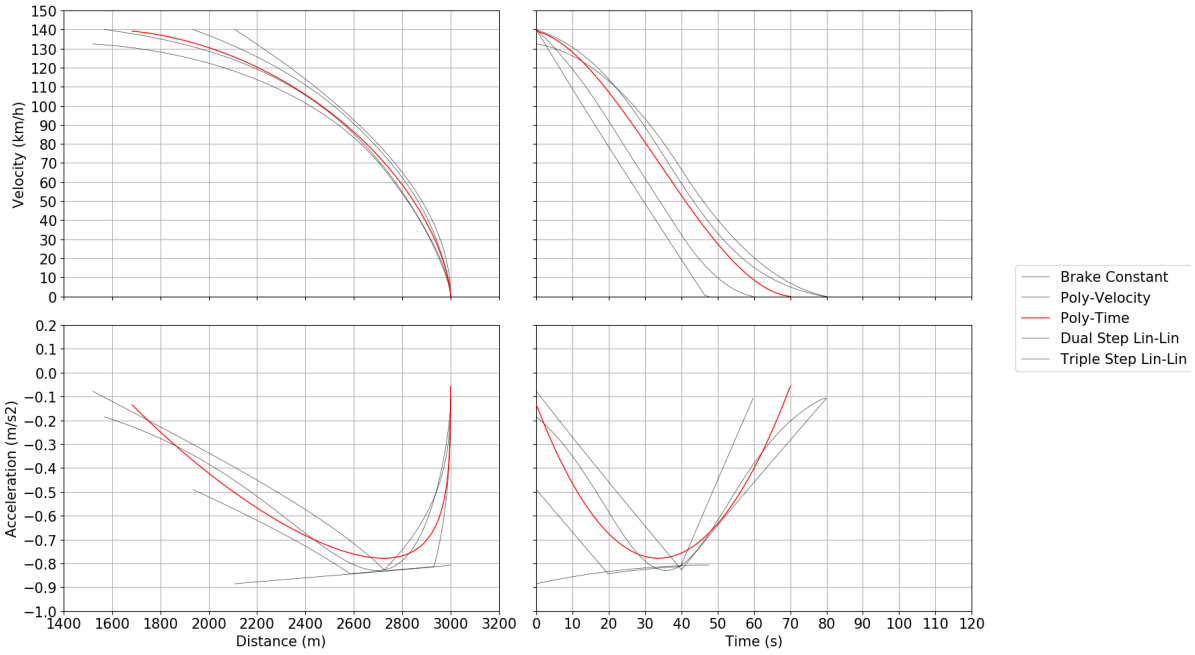


Figure D.4: Second order polynomial (time dependent) example shown in both velocity and acceleration in both distance and time domain, highlighted in red with other examples in black for comparison/contrast.

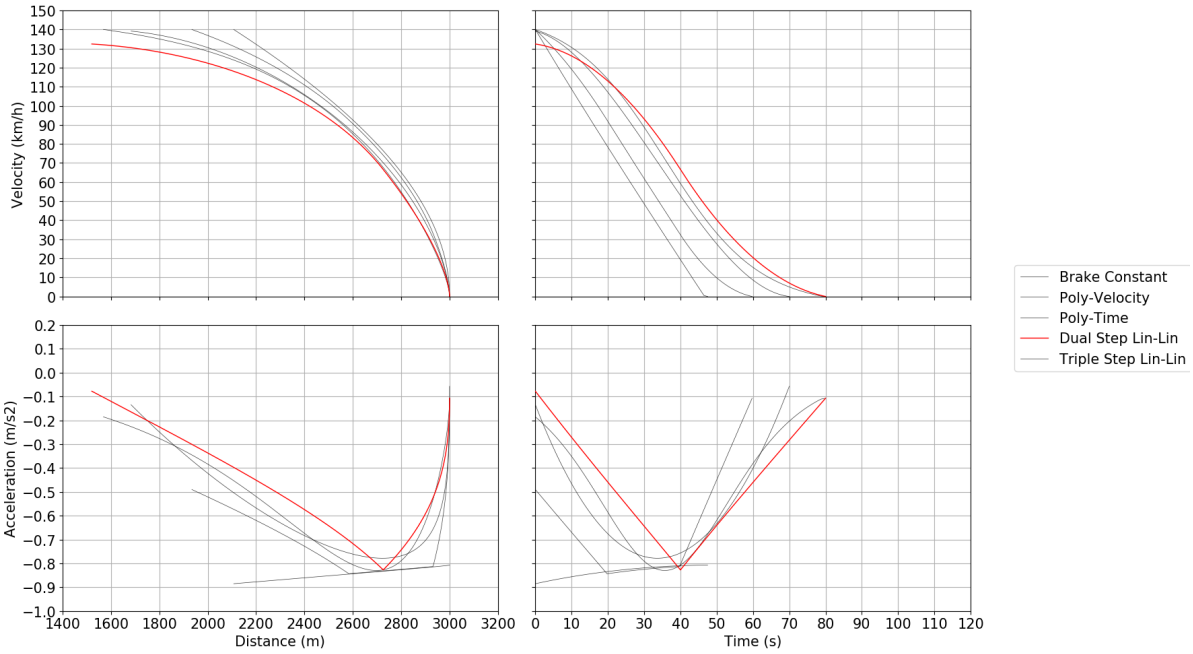


Figure D.5: Dual-Step ”Linear-Linear” example shown in both velocity and acceleration in both distance and time domain, highlighted in red with other examples in black for comparison/contrast.

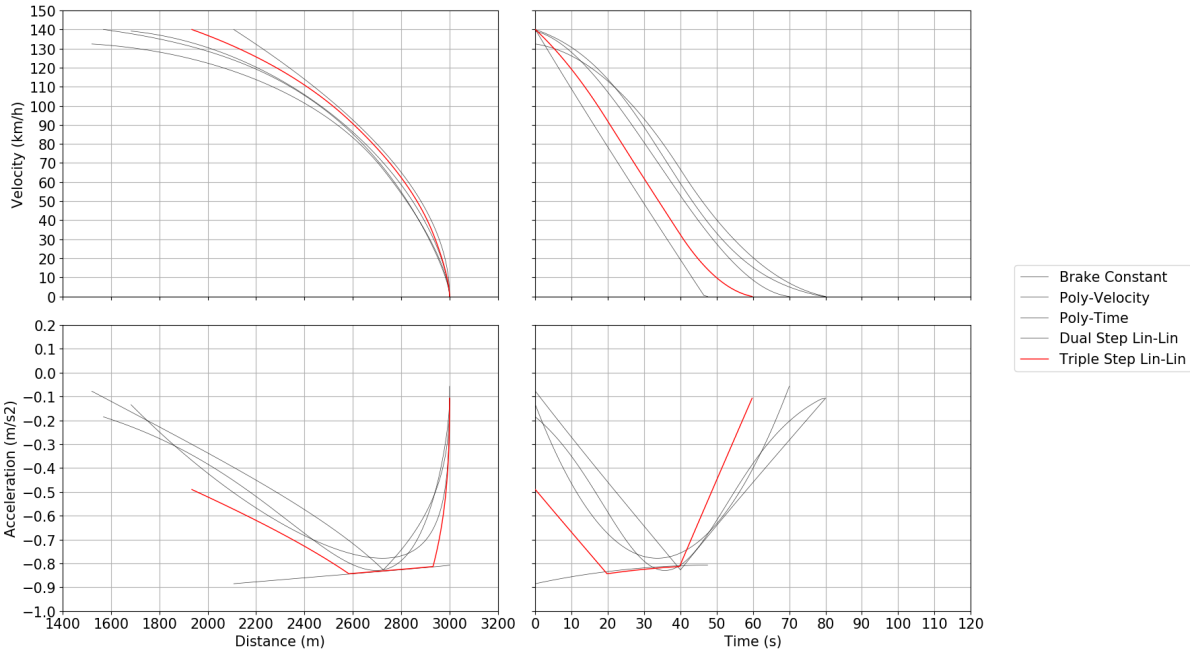


Figure D.6: Triple-Step ”Linear-Linear” example shown in both velocity and acceleration in both distance and time domain, highlighted in red with other examples in black for comparison/contrast.

E. DR Model Results - Model Accuracy

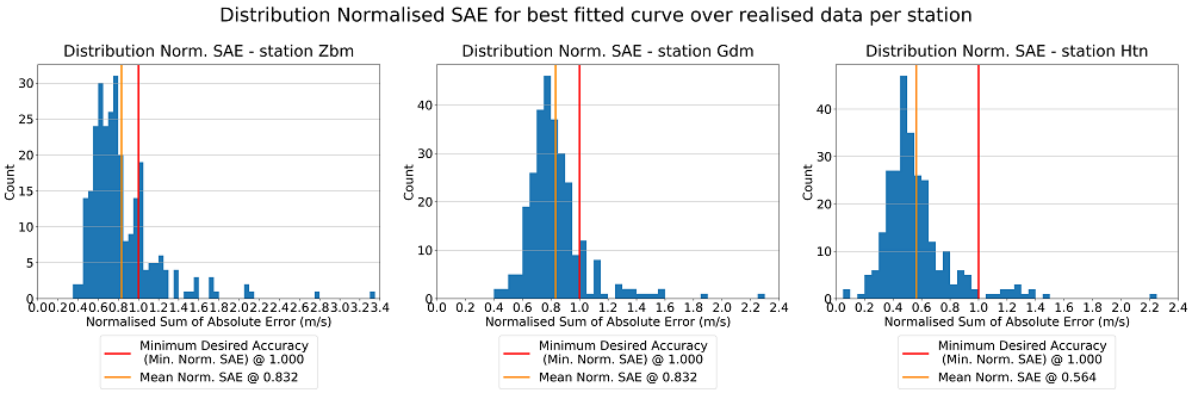
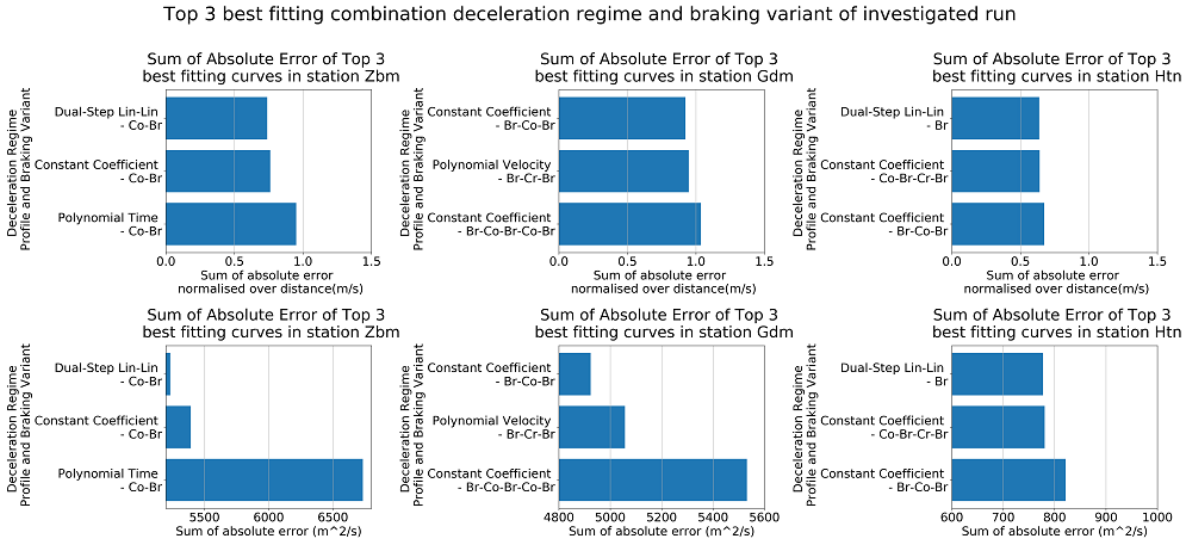
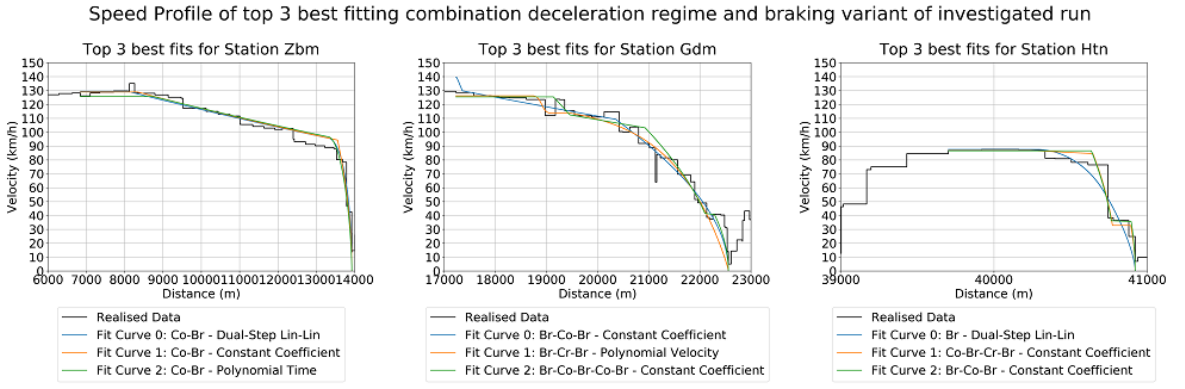


Figure E.1: Histogram Normalised Sum of Absolute Error - Best fitting deceleration regime curve on the realised running data, given per station.

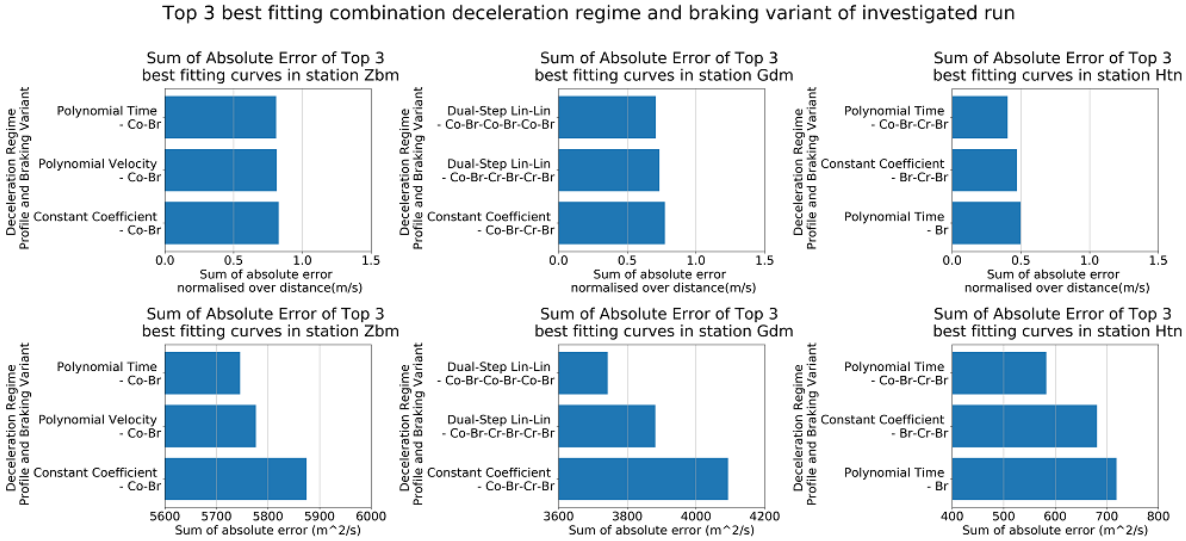


(a)

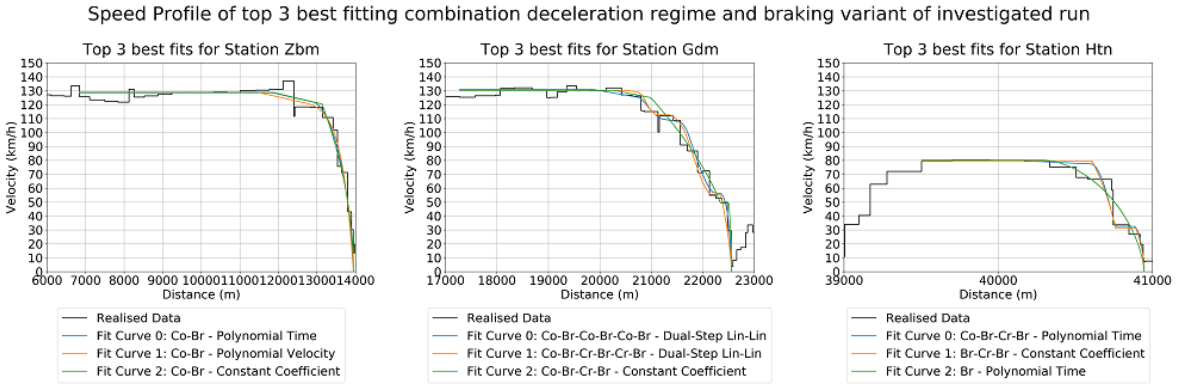


(b)

Figure E.2: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes

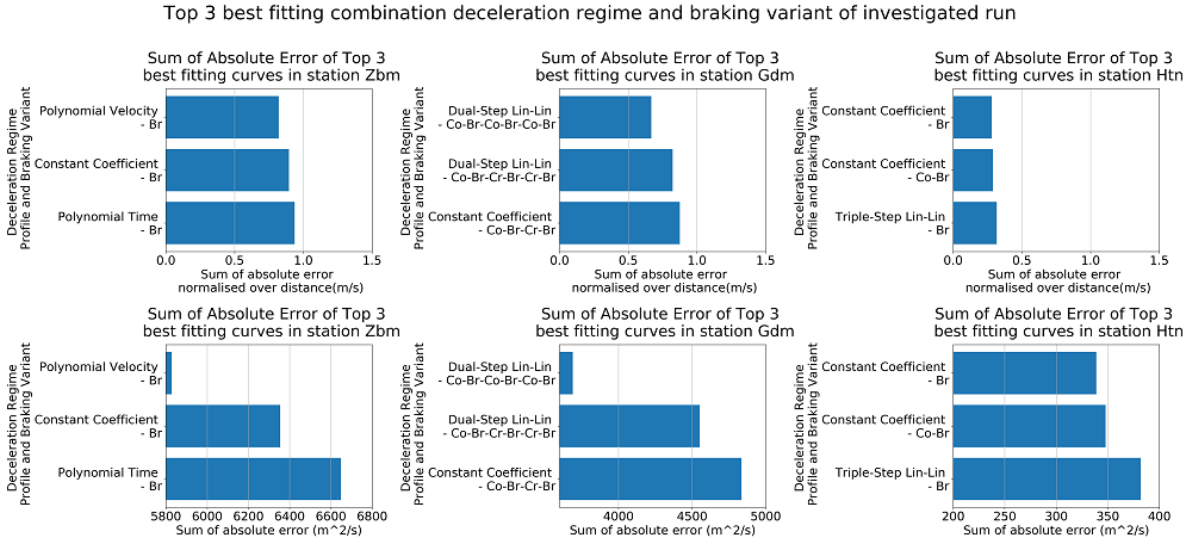


(a)

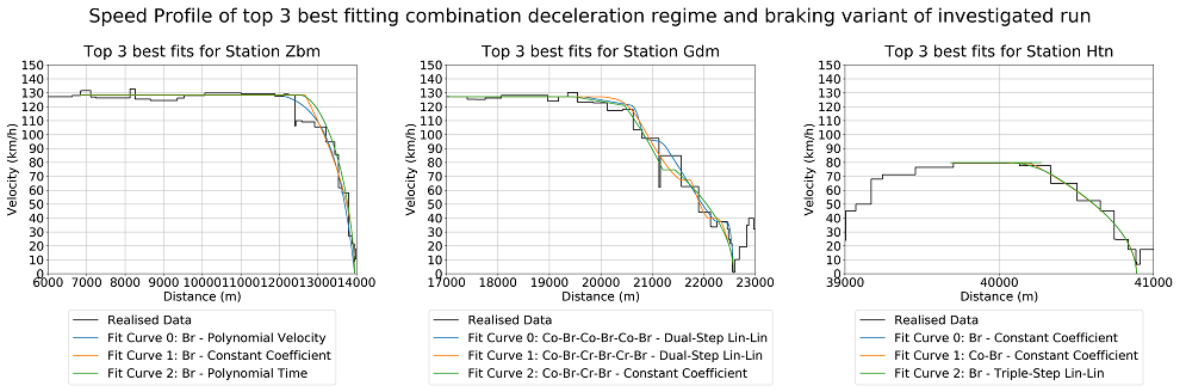


(b)

Figure E.3: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes

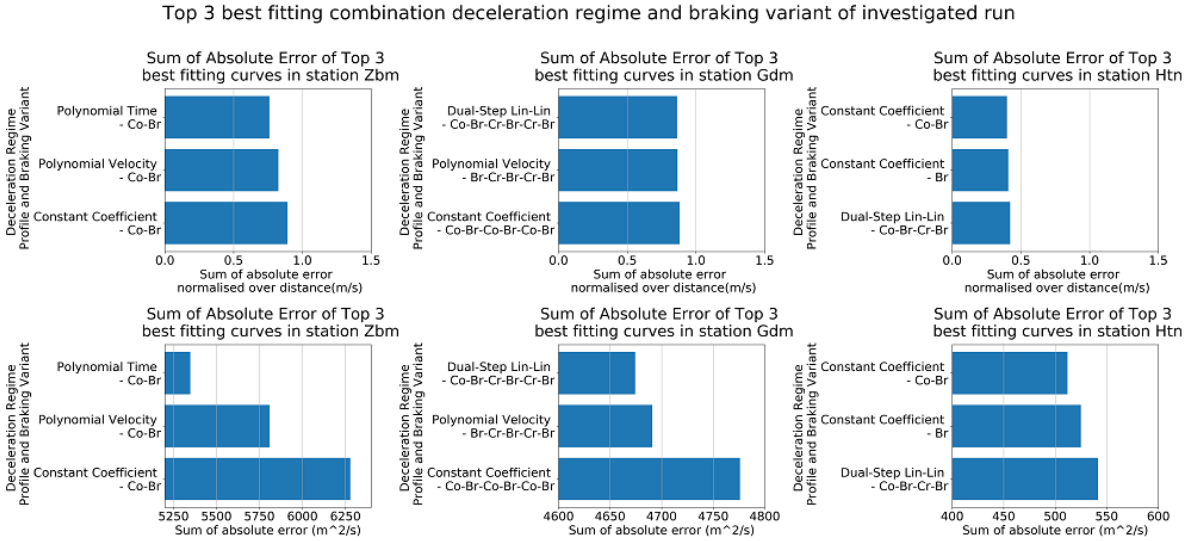


(a)

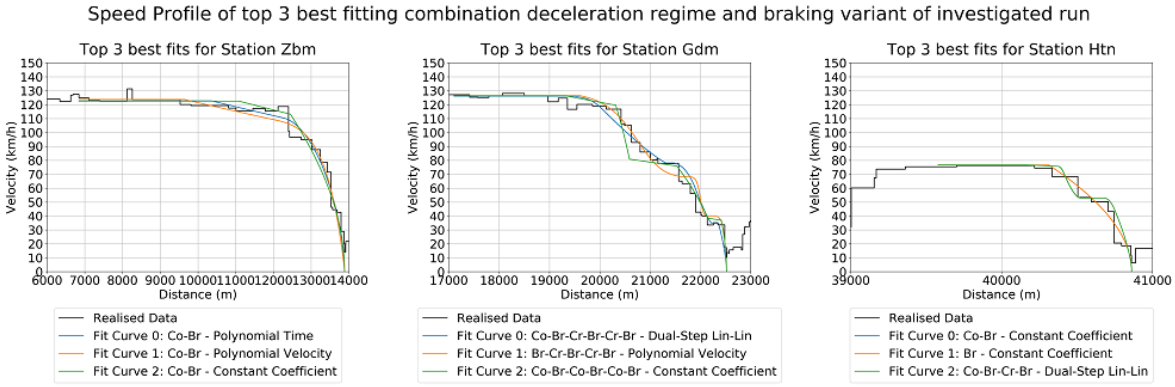


(b)

Figure E.4: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes

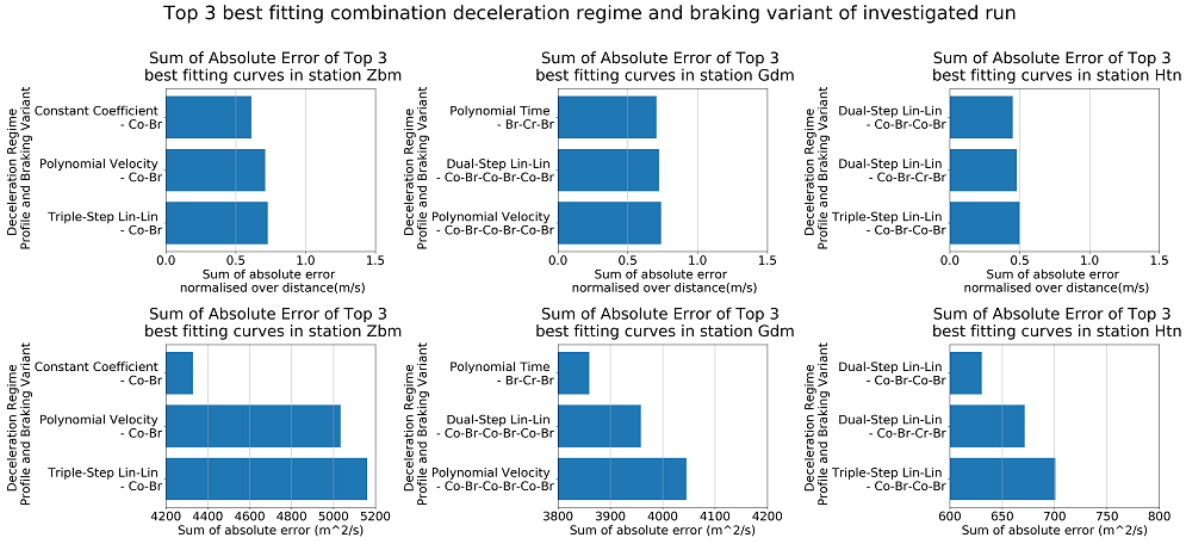


(a)

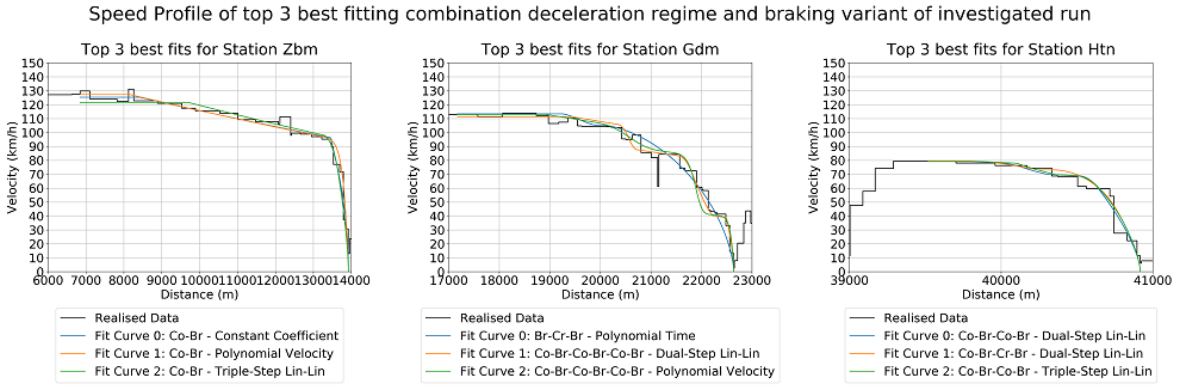


(b)

Figure E.5: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes



(a)



(b)

Figure E.6: Comparison Top 3 Best Fitting Deceleration Regimes for selected train run
(a): Comparison Sum of Absolute Error (SAE) and Normalised SAE
(b): Corresponding Speed profile for top 3 fitting deceleration regimes

F. DR Model Results - Overview Deceleration Regimes and Braking Variants

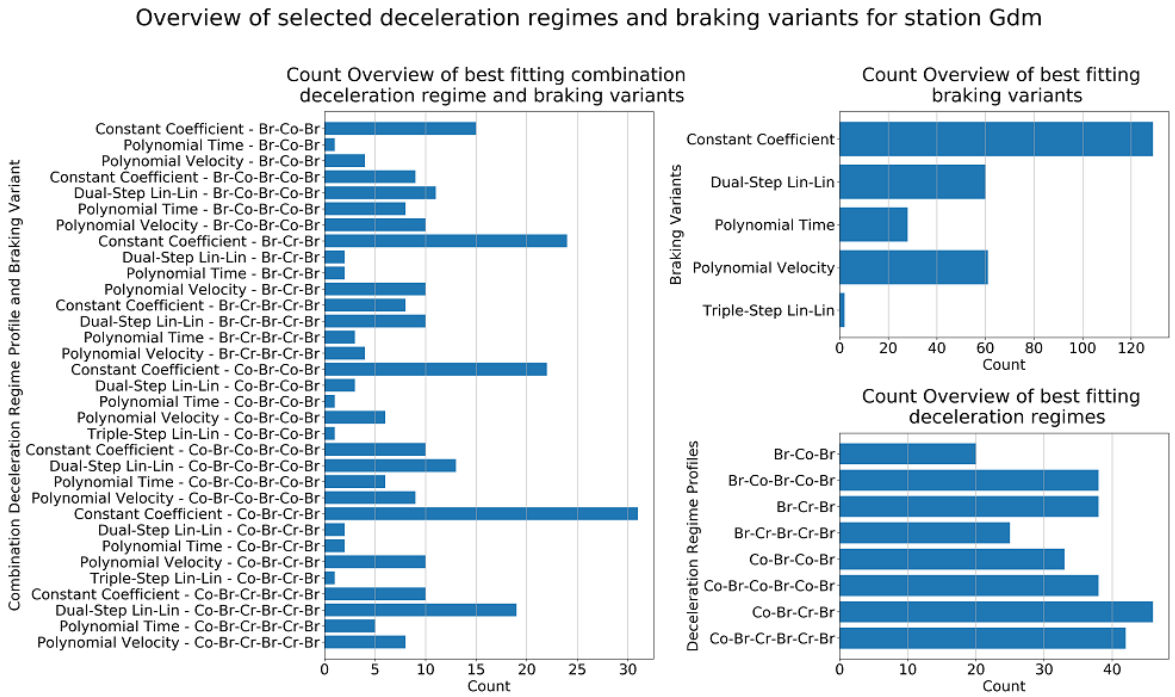


Figure F.1: Overview Estimated Deceleration Regimes and Braking Variants - Gdm

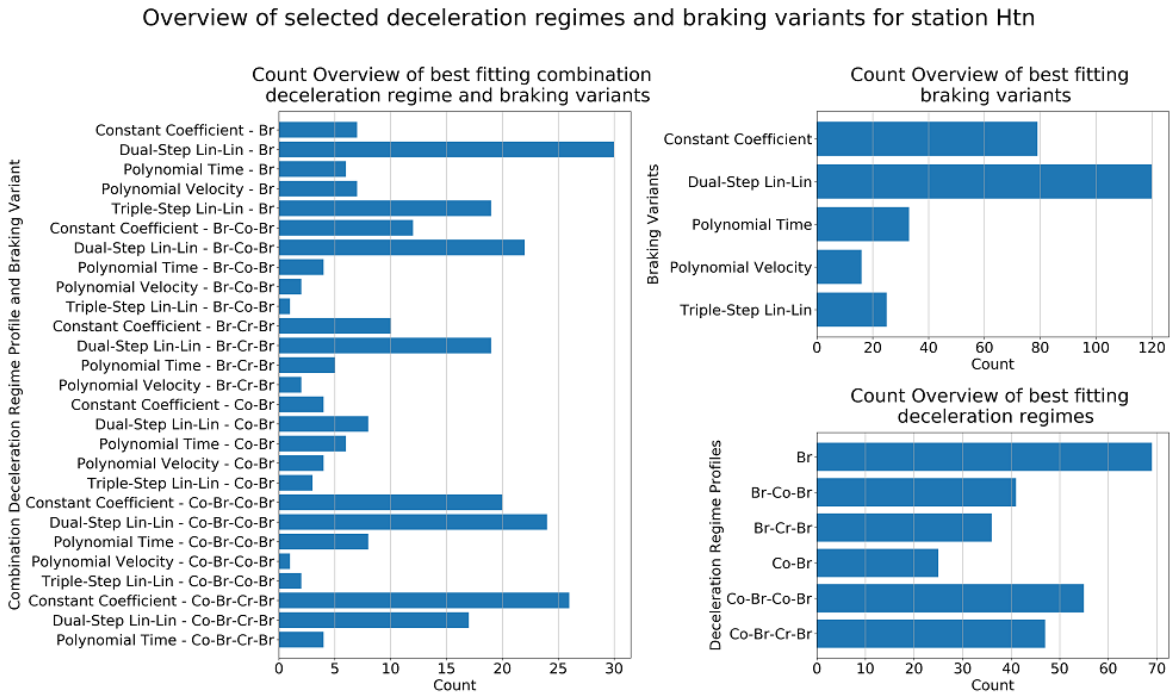


Figure F.2: Overview Estimated Deceleration Regimes and Braking Variants - Htn

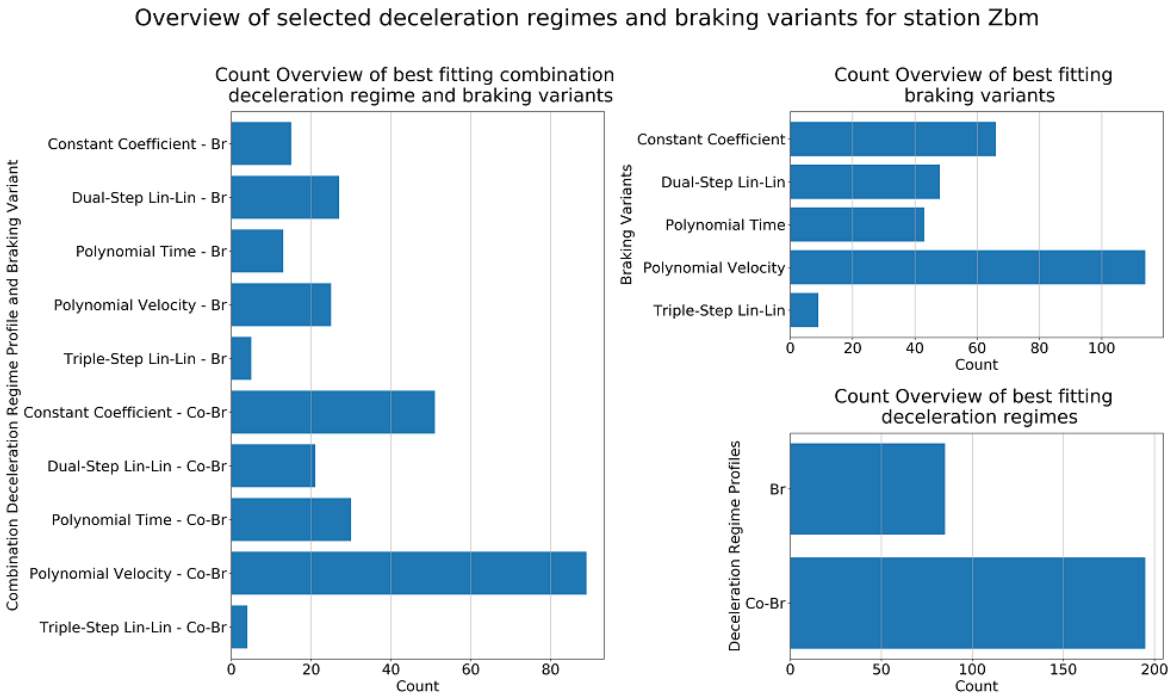


Figure F.3: Overview Estimated Deceleration Regimes and Braking Variants - Zbm

G. DR Model Results - Vehicle Coefficients

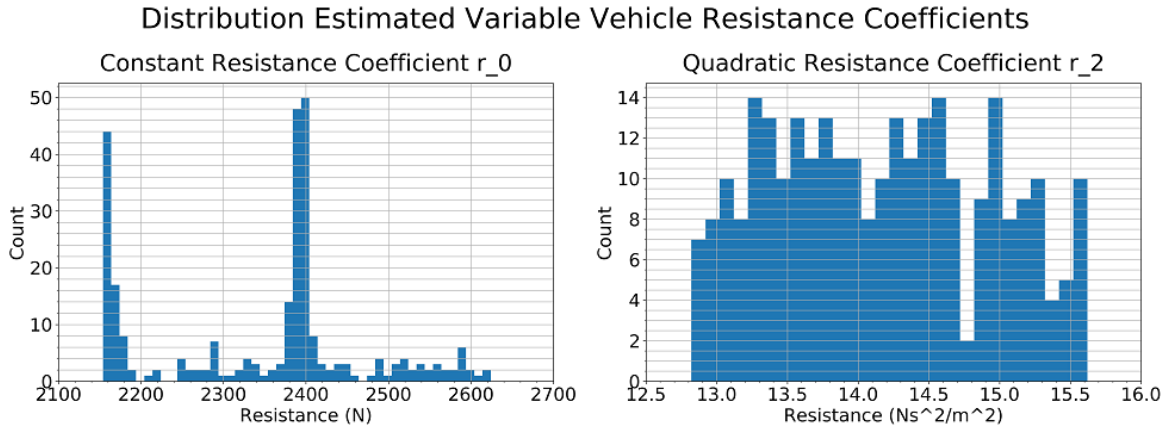


Figure G.1: Histogram Vehicle Resistance Coefficients - Station Gdm

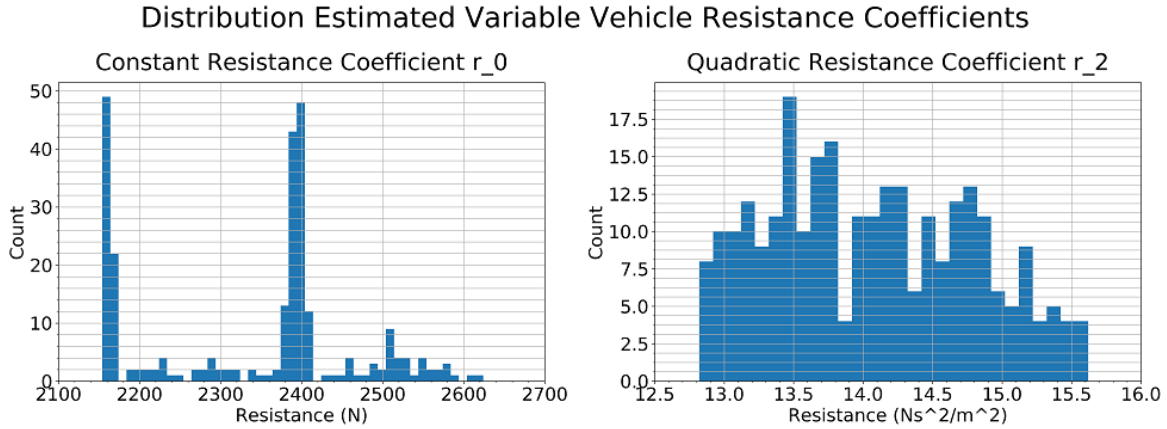


Figure G.2: Histogram Vehicle Resistance Coefficients - Station Htn

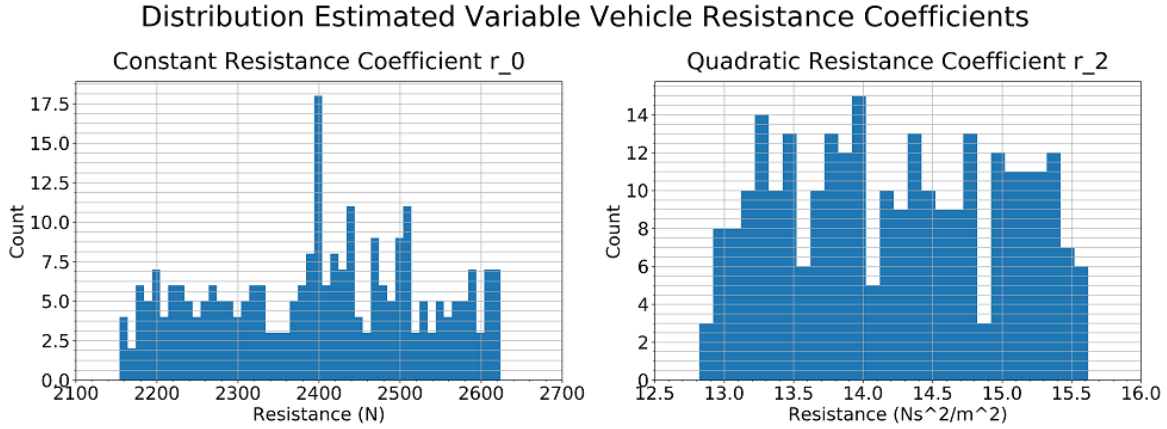


Figure G.3: Histogram Vehicle Resistance Coefficients - Station Zbm

H. DR Model Results - Braking Rates

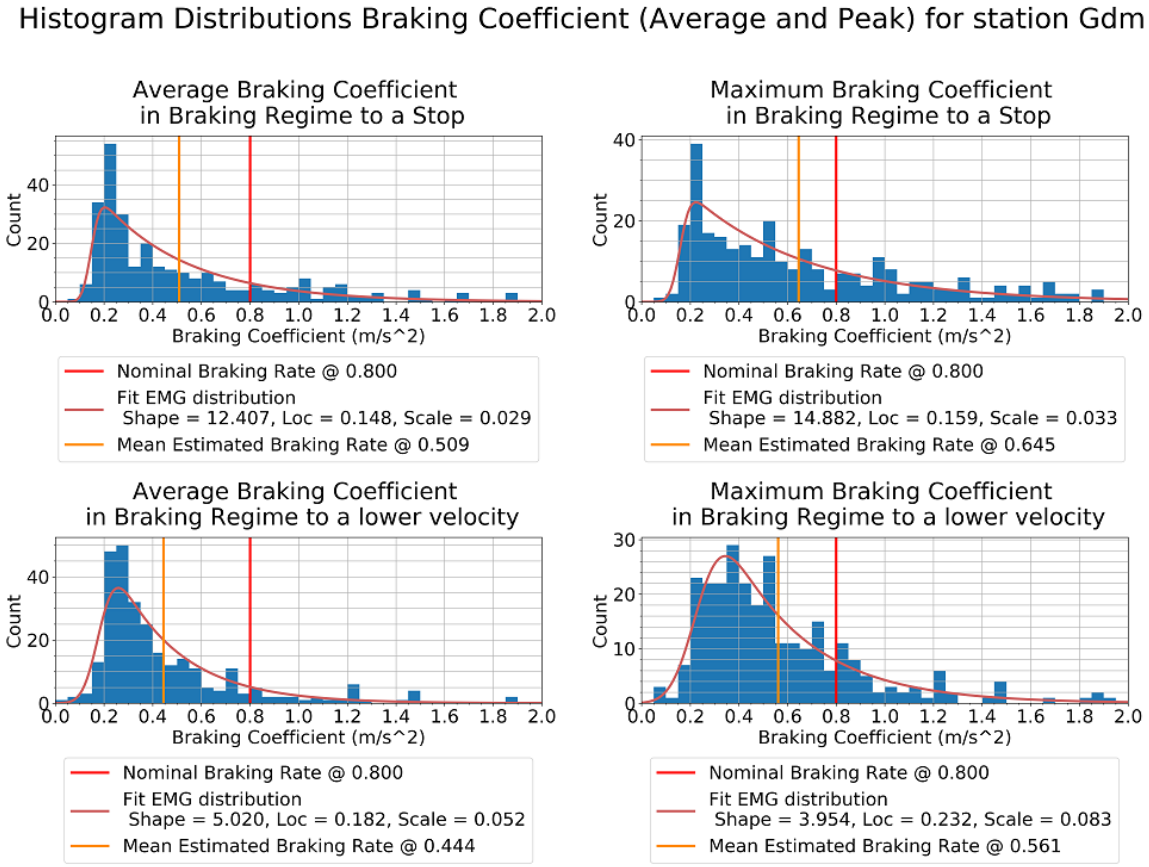


Figure H.1: Distributions Mean and Peak Braking Rate for stop and lower velocity - Gdm. Vertical red line is the value of the nominal braking rate at 0.8 m/s^2.

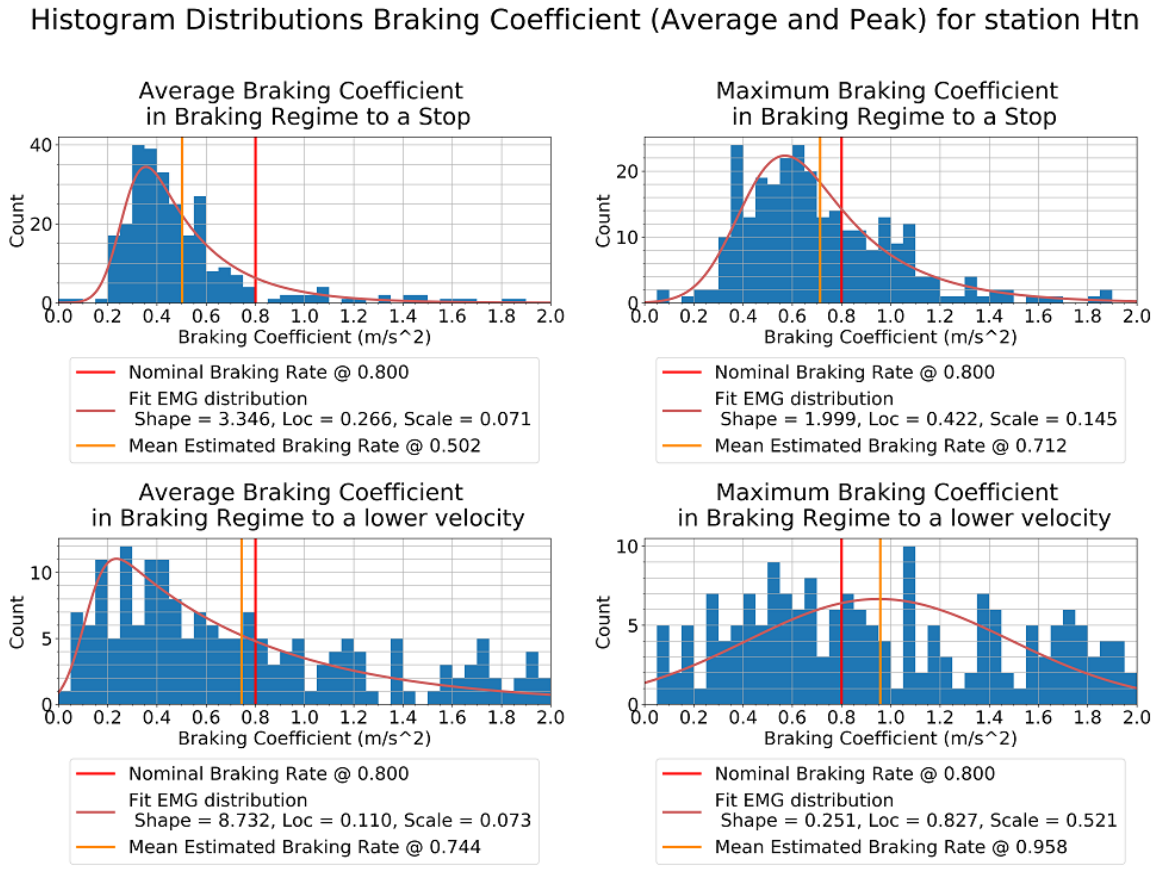


Figure H.2: Distributions Mean and Peak Braking Rate for stop and lower velocity - Htn. Vertical red line is the value of the nominal braking rate at 0.8 m/s^2.

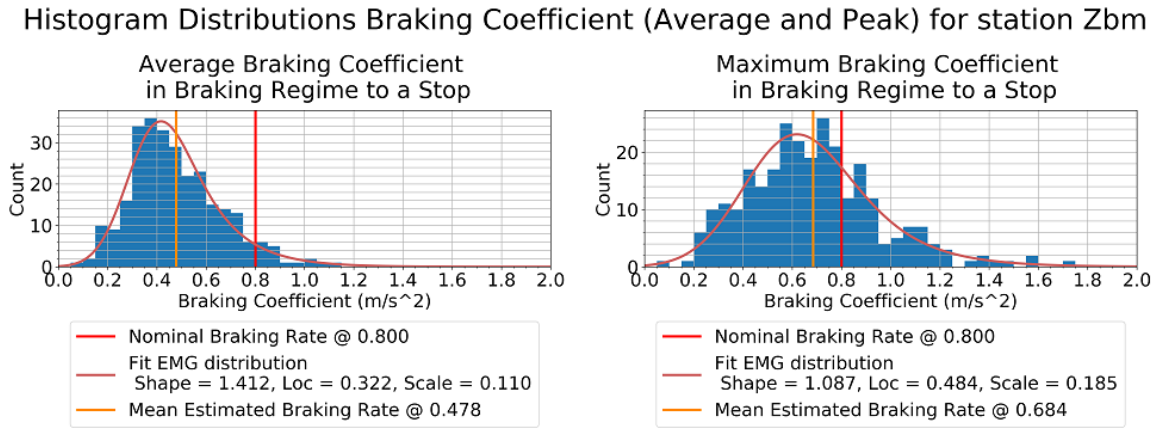
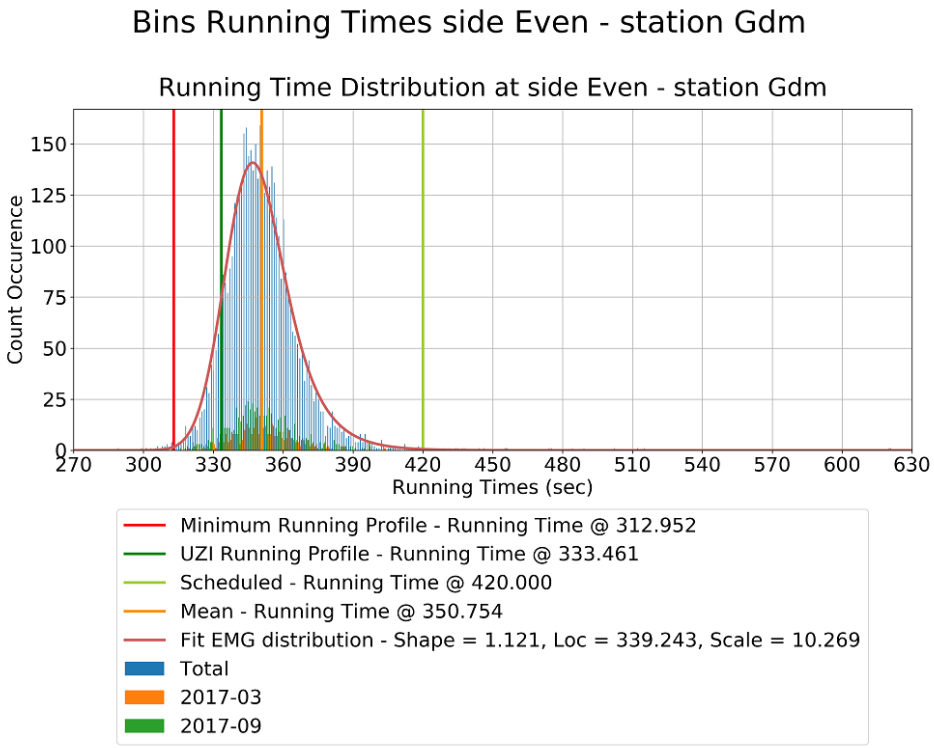
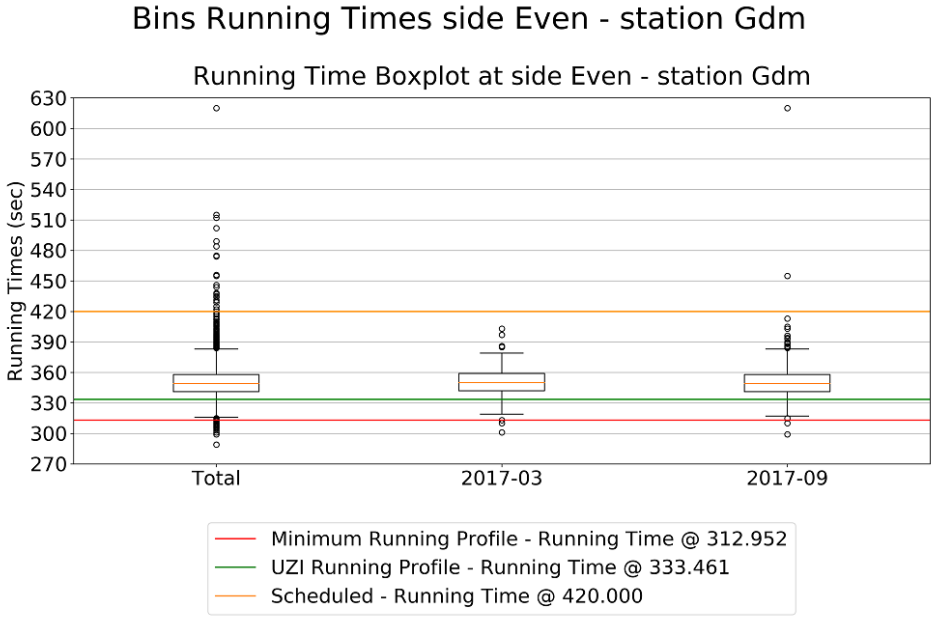


Figure H.3: Distributions Mean and Peak Braking Rate for stop and lower velocity - Zbm. Vertical red line is the value of the nominal braking rate at 0.8 m/s^2.

I. Analysis Results - Running Time Distribution

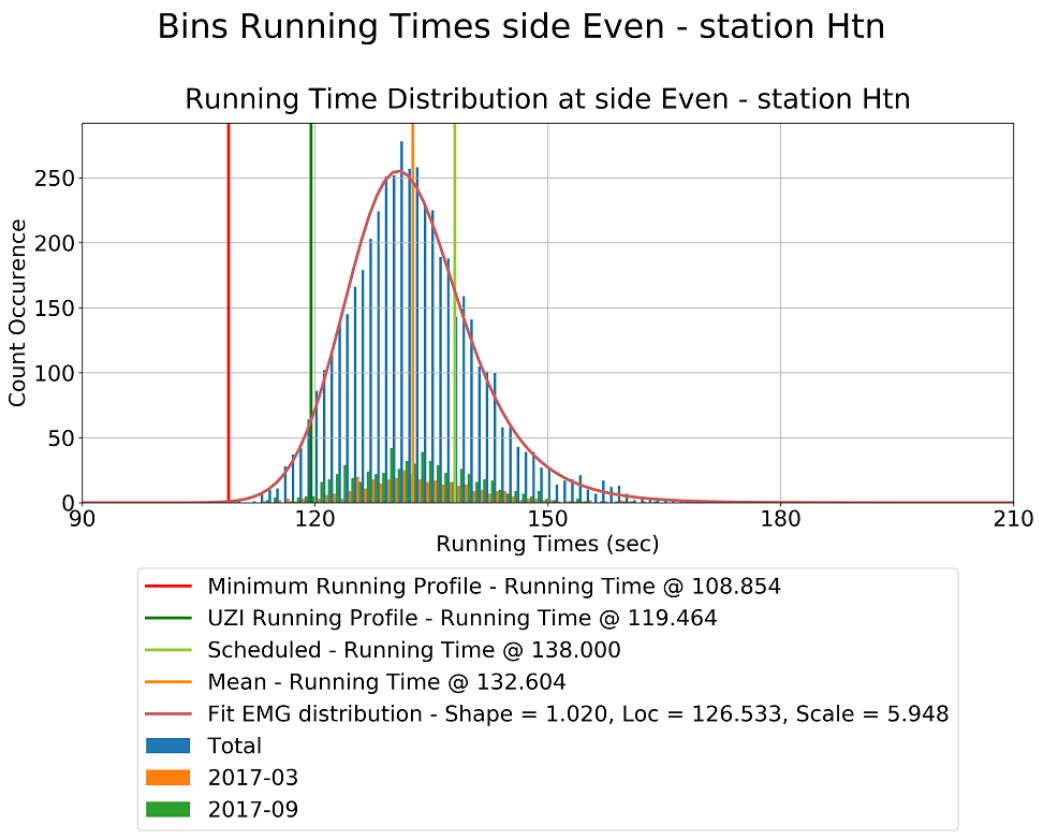


(a)

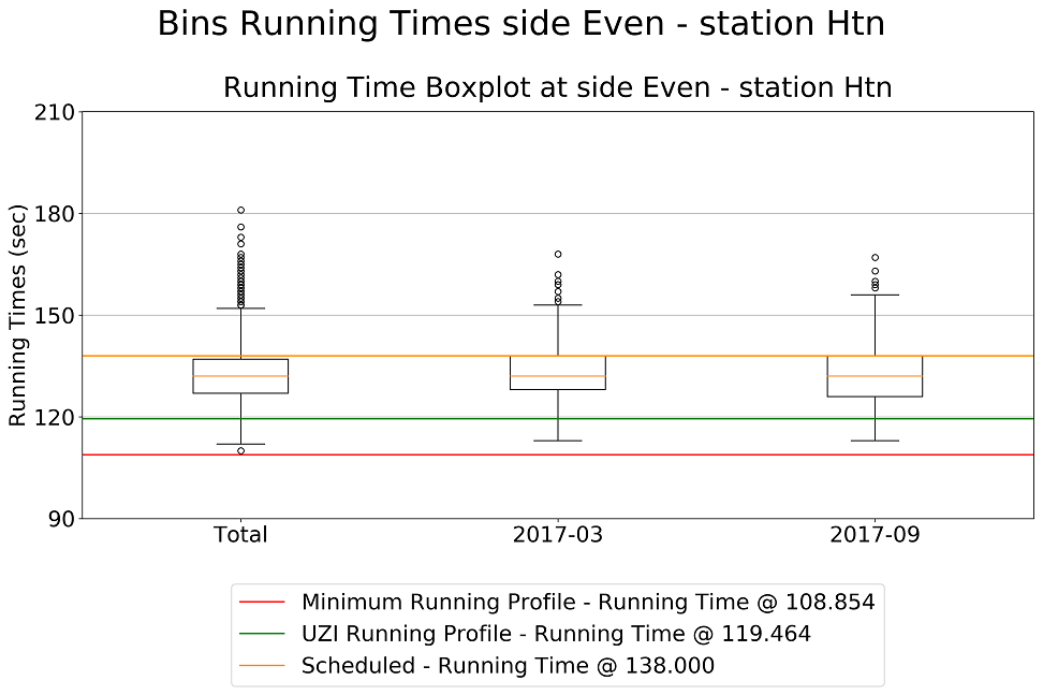


(b)

Figure I.1: (a): Distribution Running Times - Even Running Side Gdm
(b): Boxplot Running Times - Even Running Side Gdm

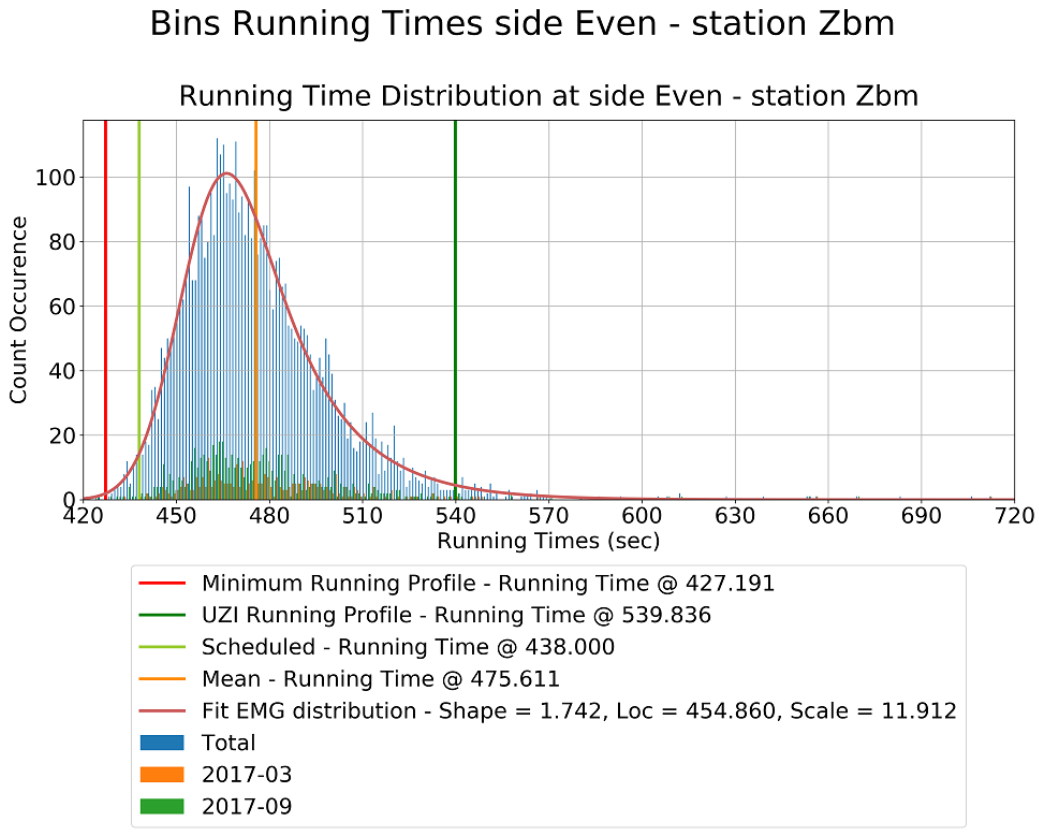


(a)

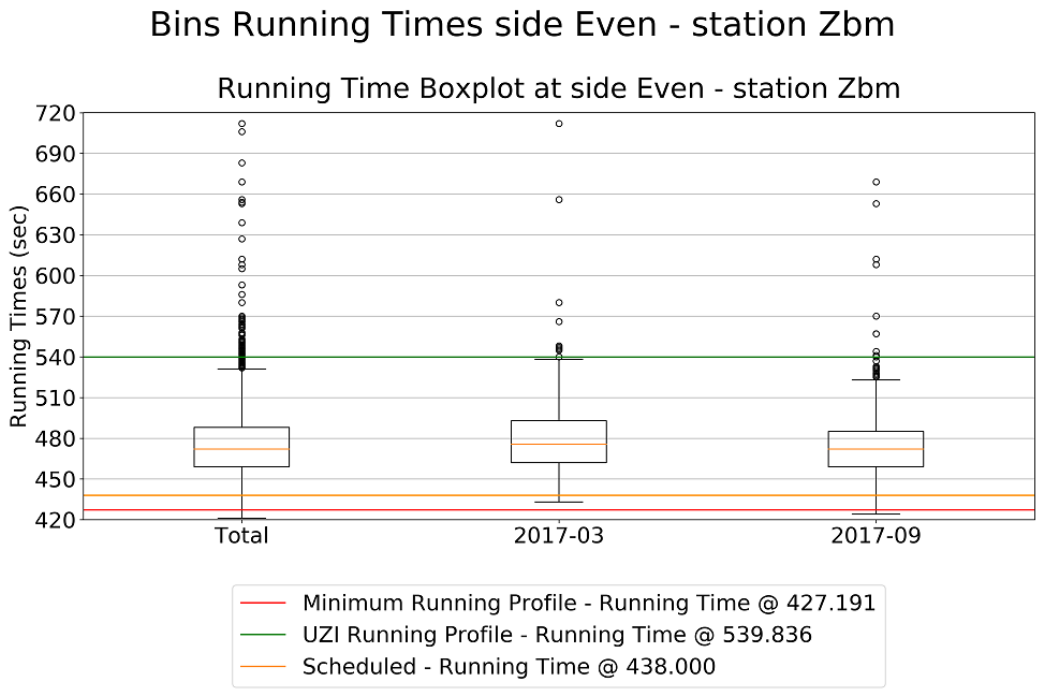


(b)

Figure I.2: (a): Distribution Running Times - Even Running Side Htn
(b): Boxplot Running Times - Even Running Side Htn



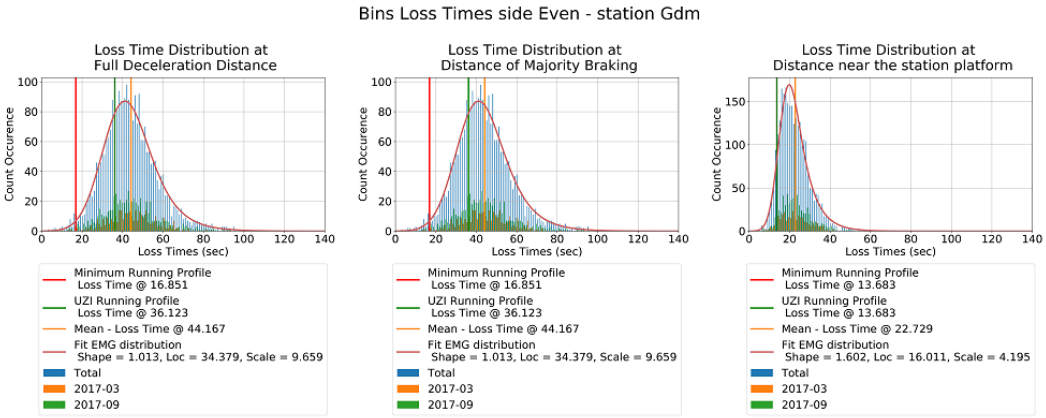
(a)



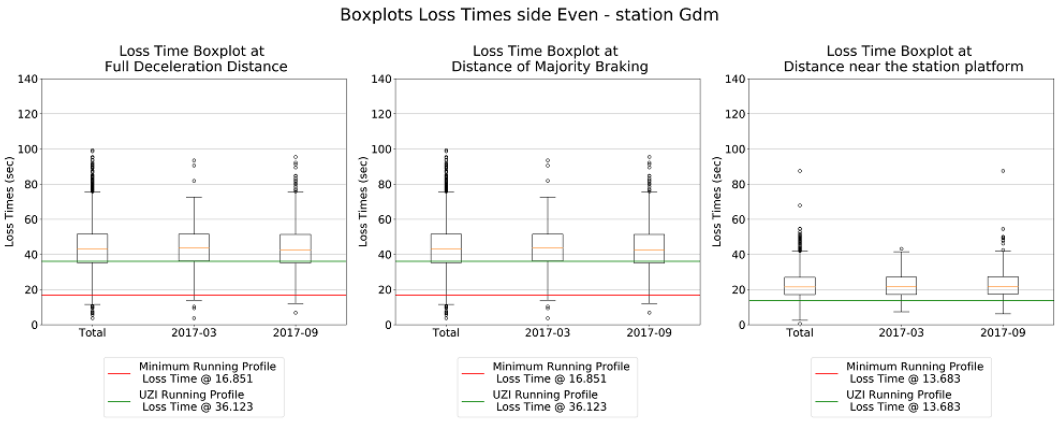
(b)

Figure I.3: (a): Distribution Running Times - Even Running Side Zbm
(b): Boxplot Running Times - Even Running Side Zbm

J. Analysis Results - Distribution Loss Times

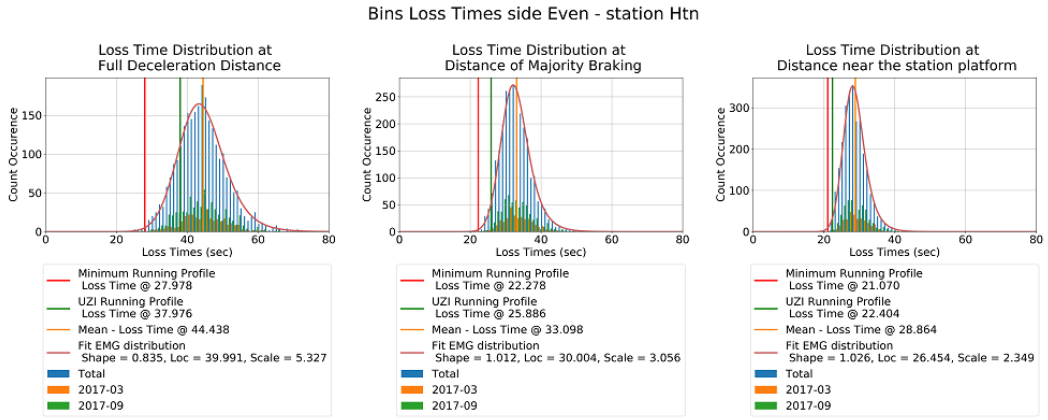


(a)

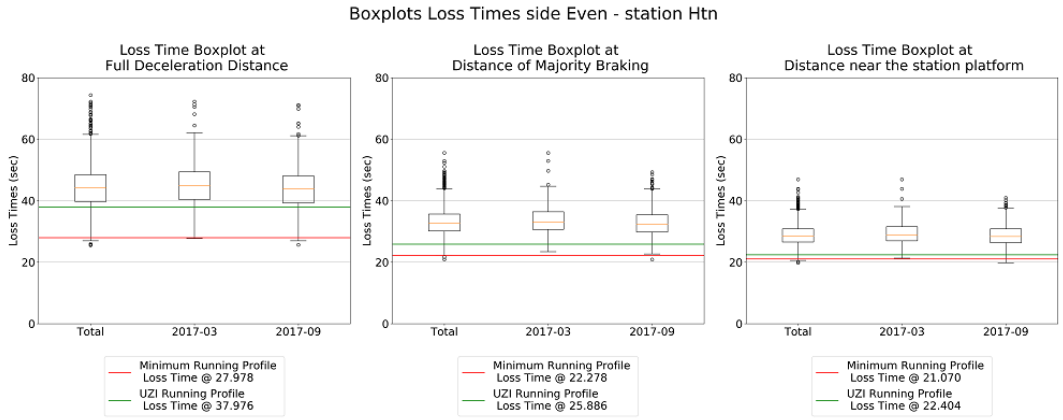


(b)

Figure J.1: (a): Distribution Deceleration Loss Times - Even Running Side Gdm
(b): Boxplot Deceleration Loss Times - Even Running Side Gdm

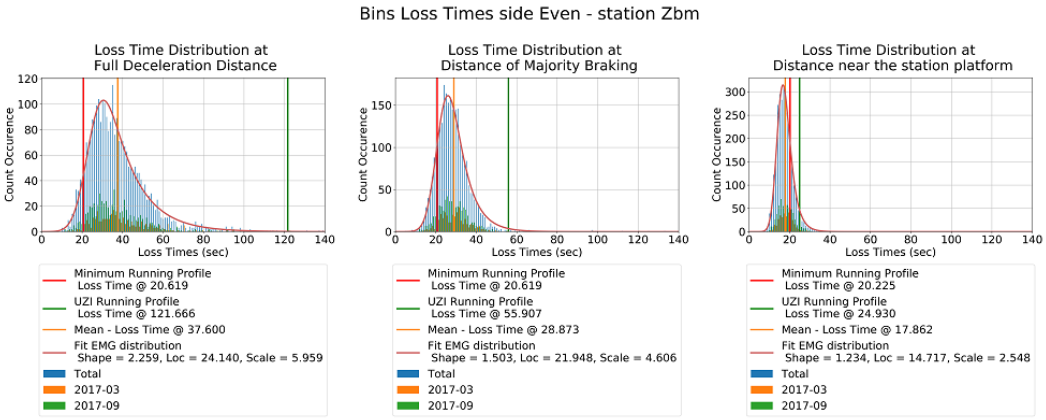


(a)

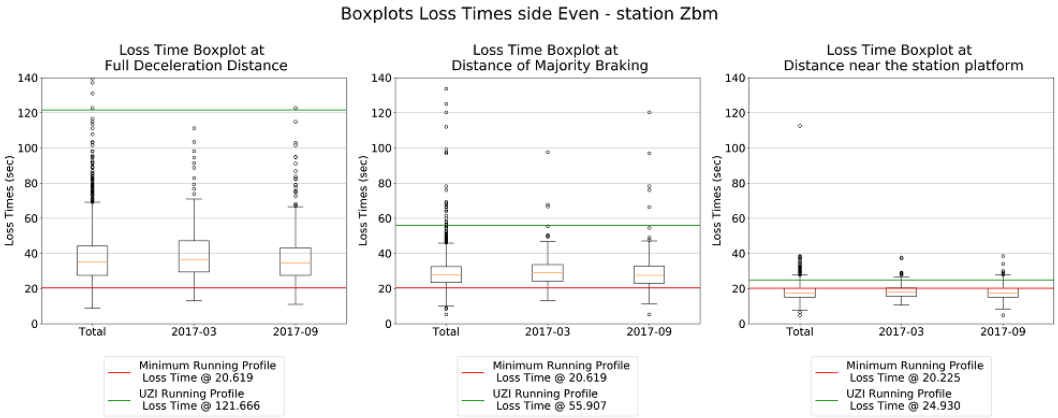


(b)

Figure J.2: (a): Distribution Deceleration Loss Times - Even Running Side Htn
(b): Boxplot Deceleration Loss Times - Even Running Side Htn



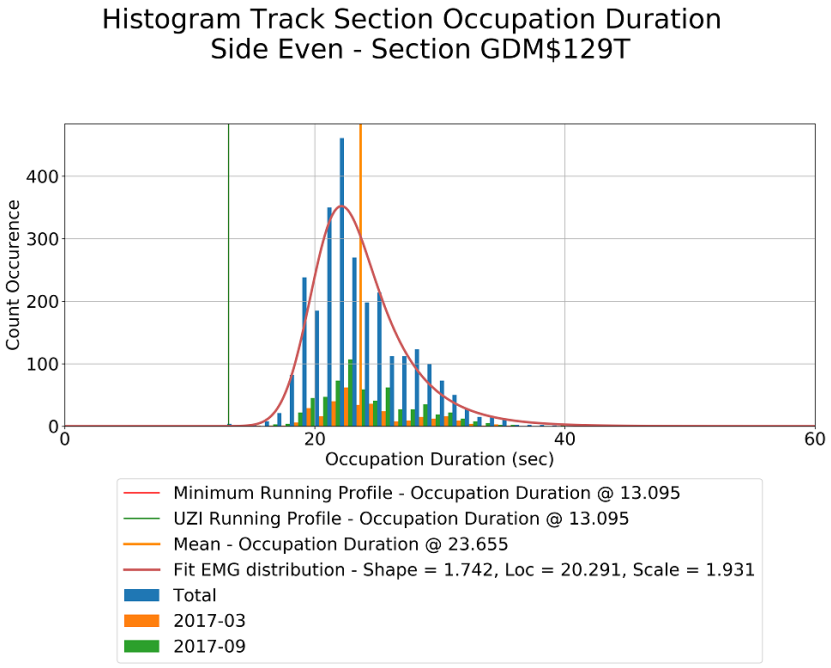
(a)



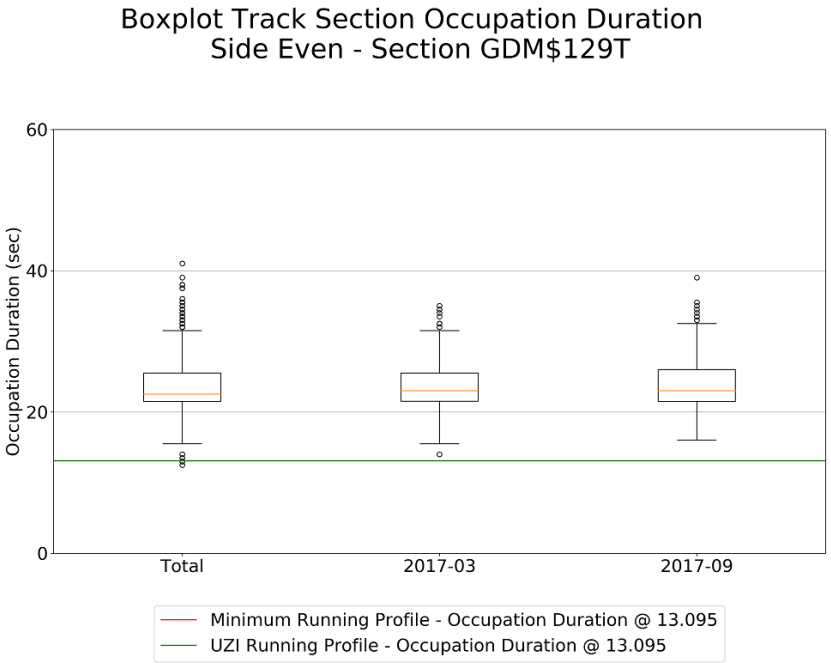
(b)

Figure J.3: (a): Distribution Deceleration Loss Times - Even Running Side Zbm
(b): Boxplot Deceleration Loss Times - Even Running Side Zbm

K. Analysis Results - Distribution Track Section Occupation Duration

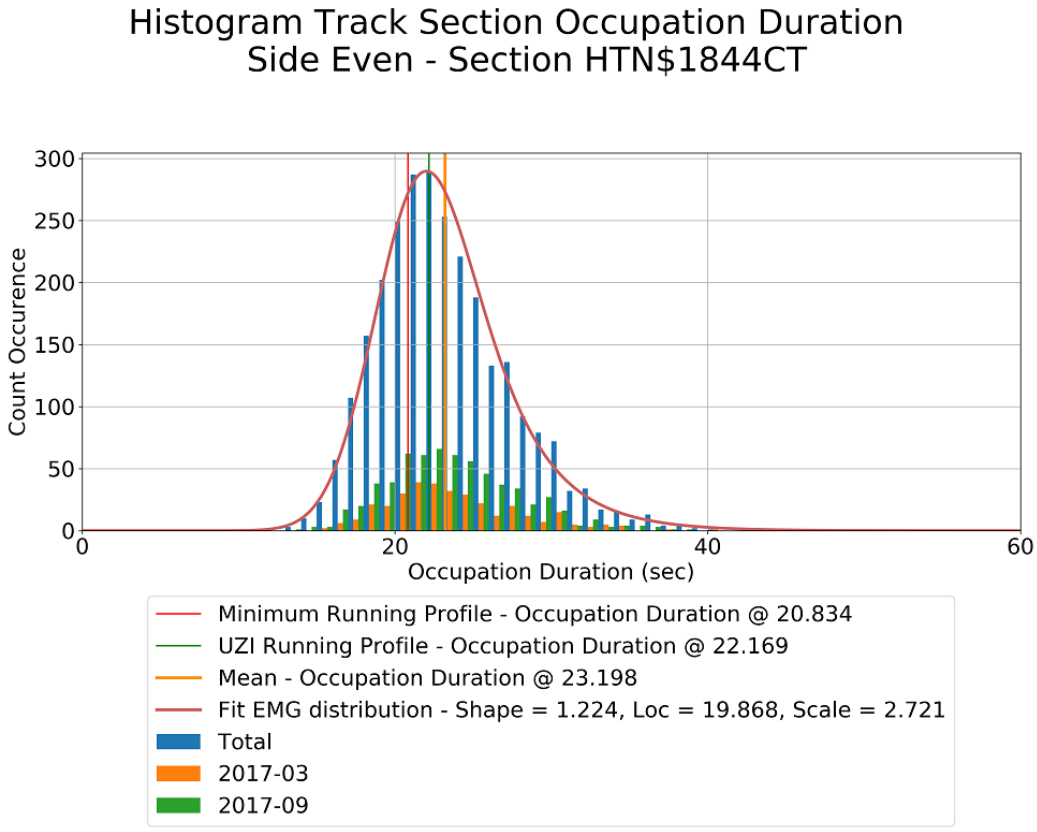


(a)

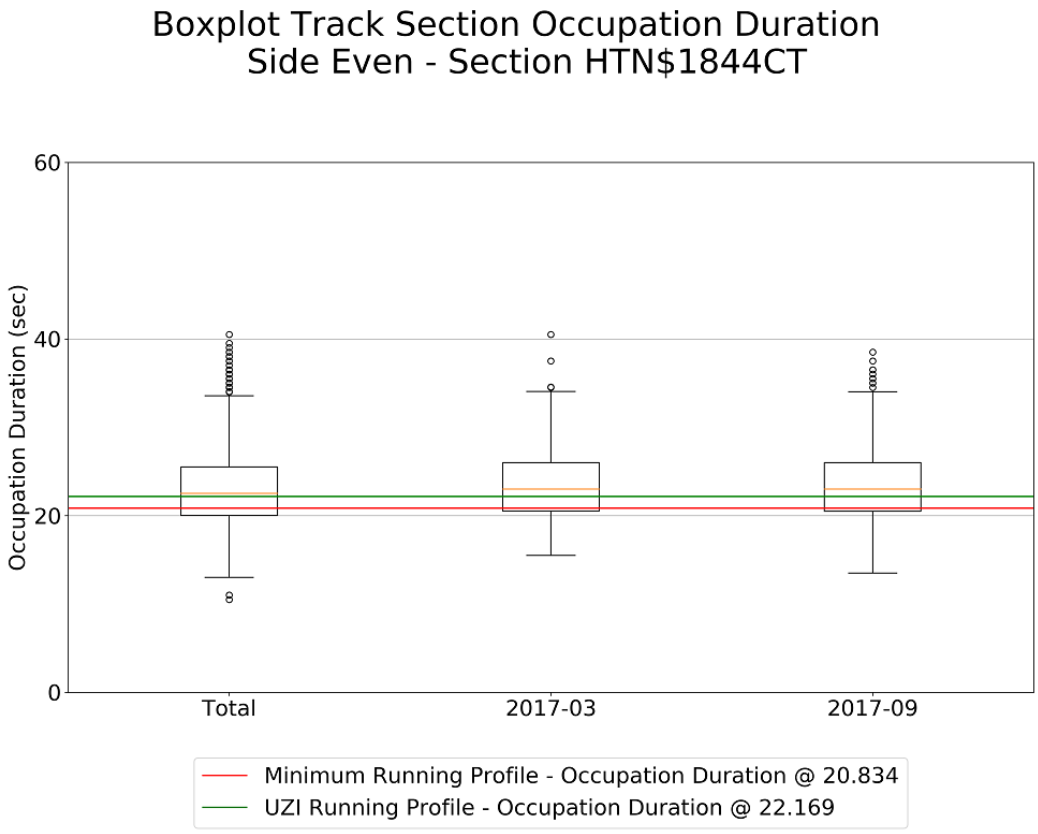


(b)

Figure K.1: (a): Distribution Track Section Occupation Duration - Even Running Side Gdm - Section GDM$129T
(b): Boxplot Track Section Occupation Duration - Even Running Side Gdm - Section GDM$129T

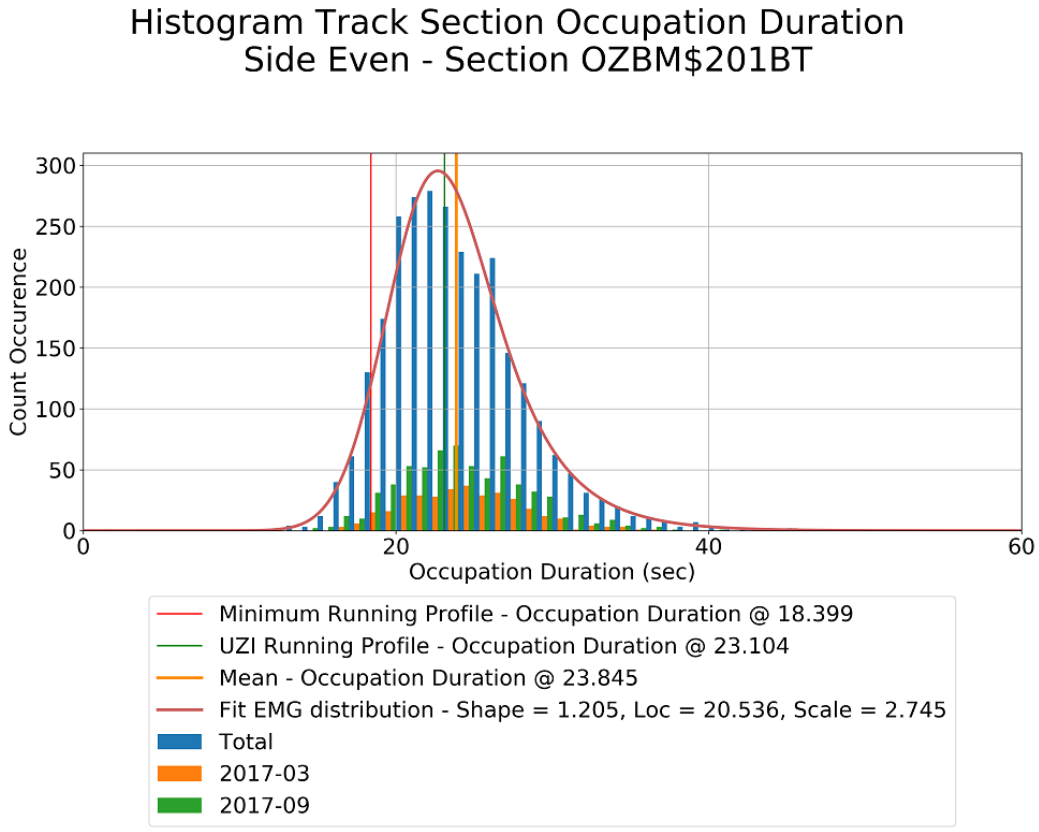


(a)

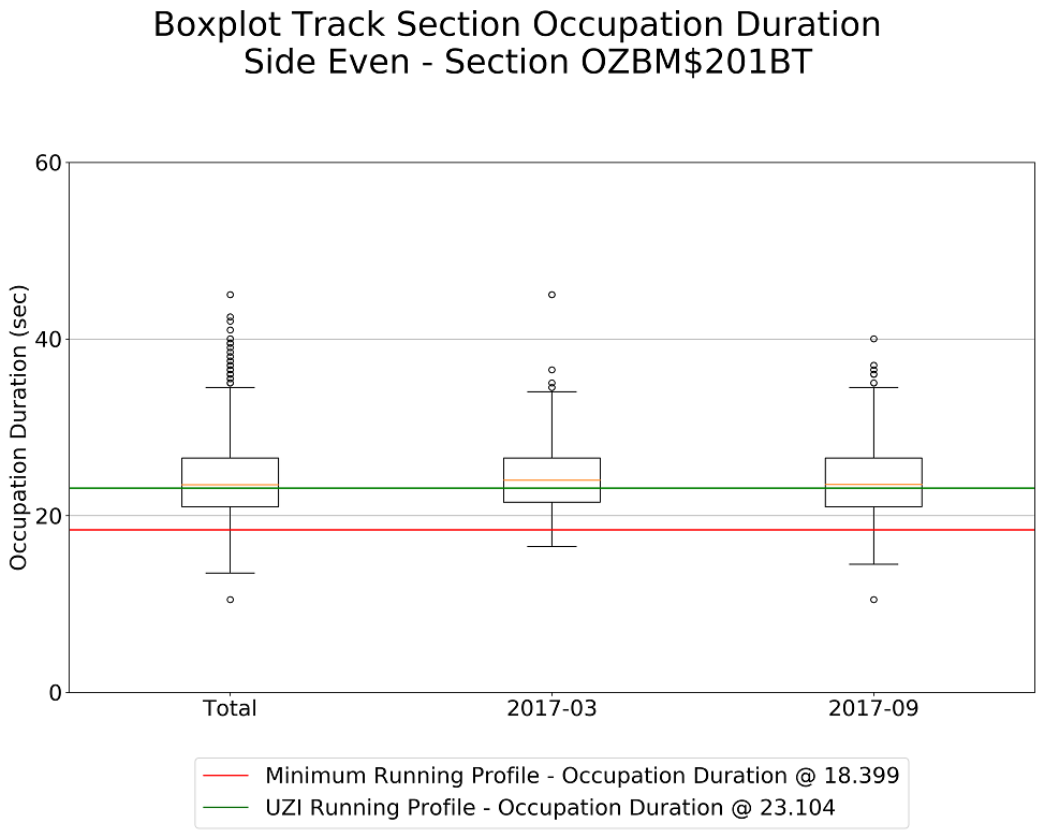


(b)

Figure K.2: (a): Distribution Track Section Occupation Duration - Even Running Side Htn - Section HTN$1844CT
(b): Boxplot Track Section Occupation Duration - Even Running Side Htn - Section HTN$1844CT



(a)



(b)

Figure K.3: (a): Distribution Track Section Occupation Duration - Even Running Side Zbm - Section OZBM$201BT
(b): Boxplot Track Section Occupation Duration - Even Running Side Zbm - Section OZBM$201BT

L. Analysis Results - Comparative Analysis

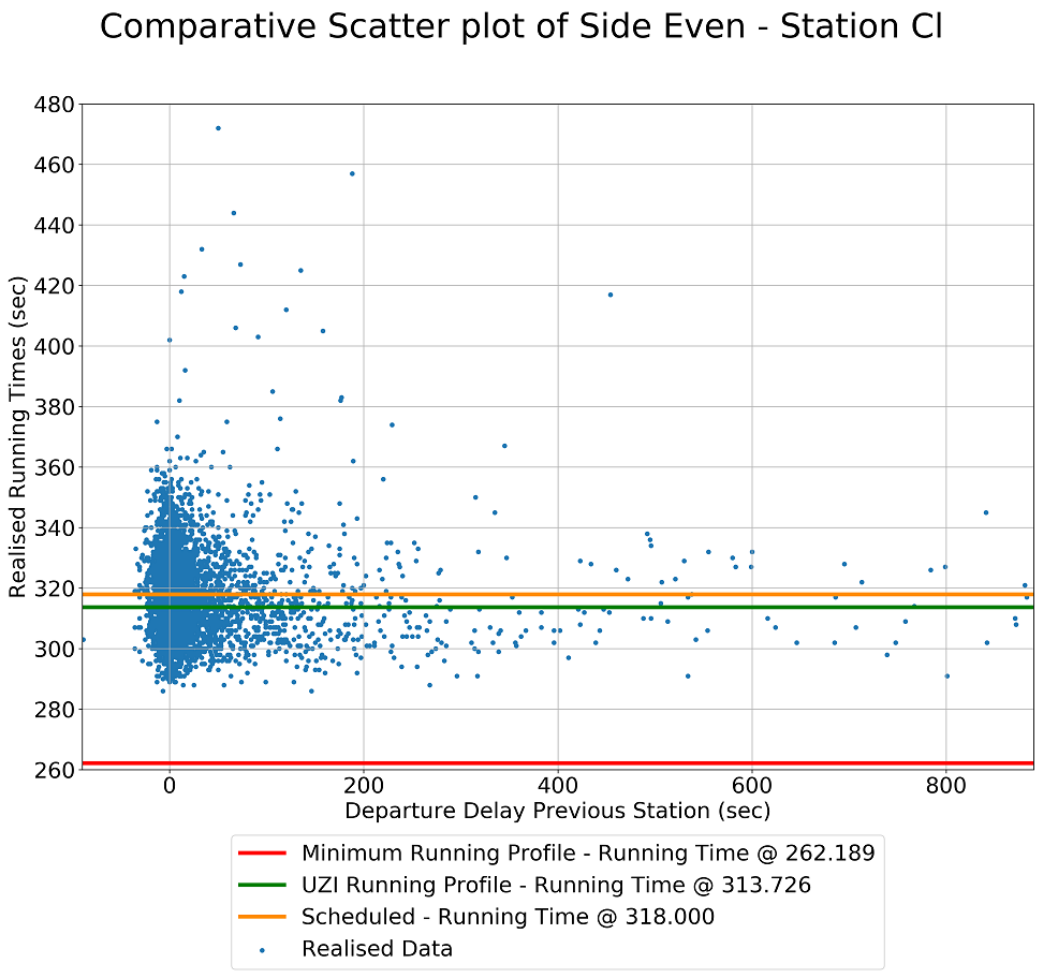


Figure L.1: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Cl

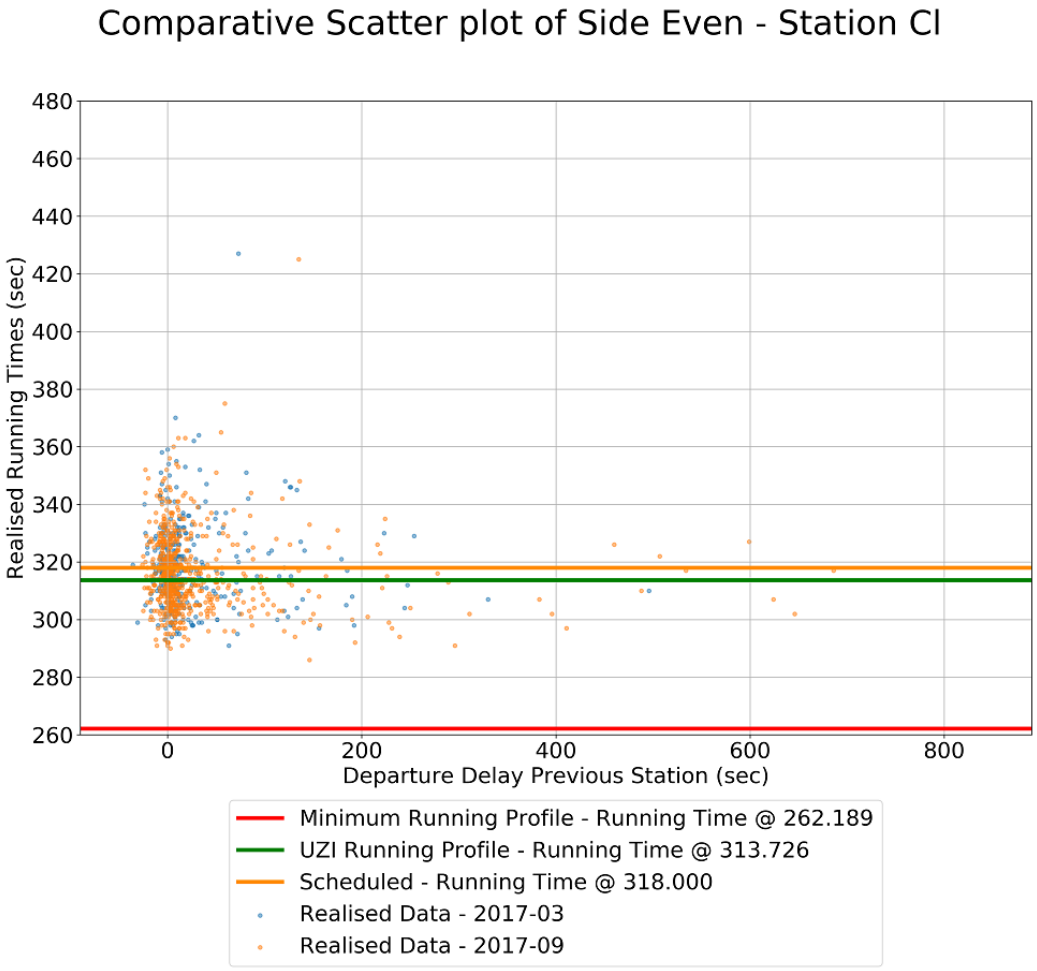


Figure L.2: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Seasonality Highlight - Cl

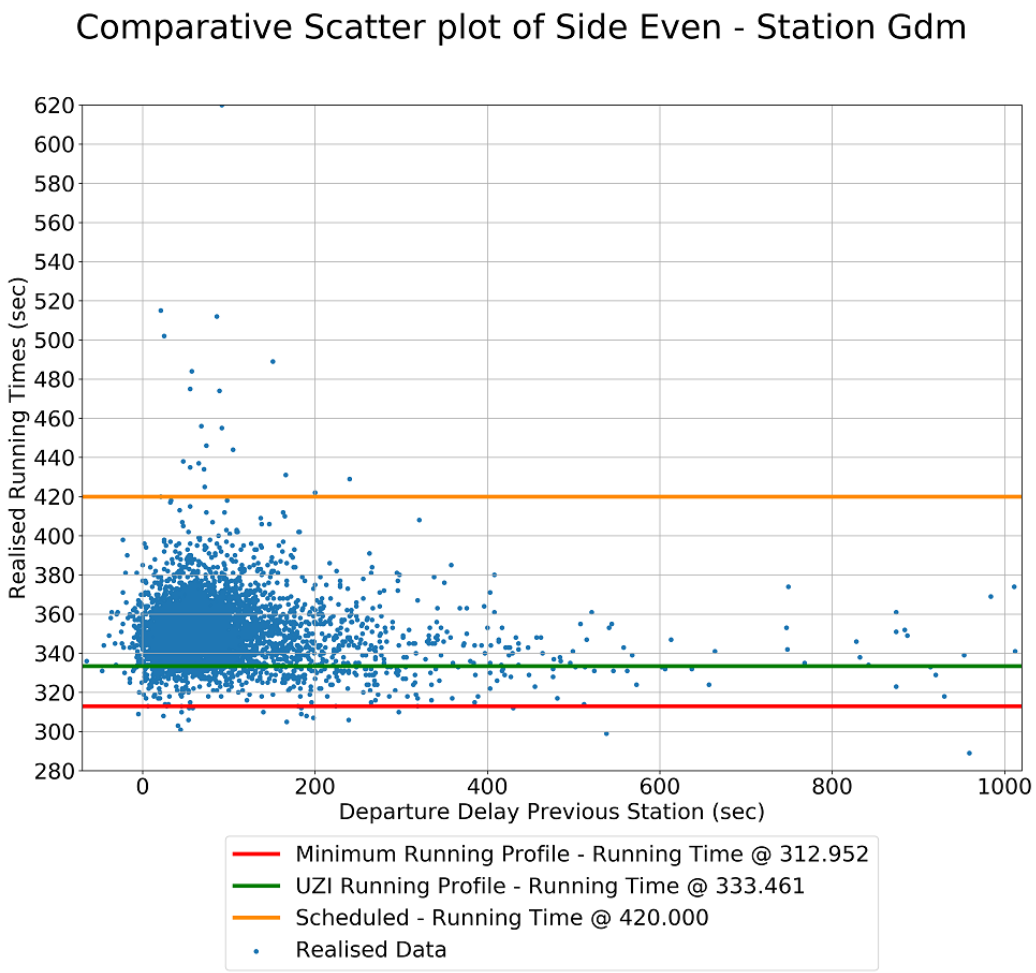


Figure L.3: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Gdm

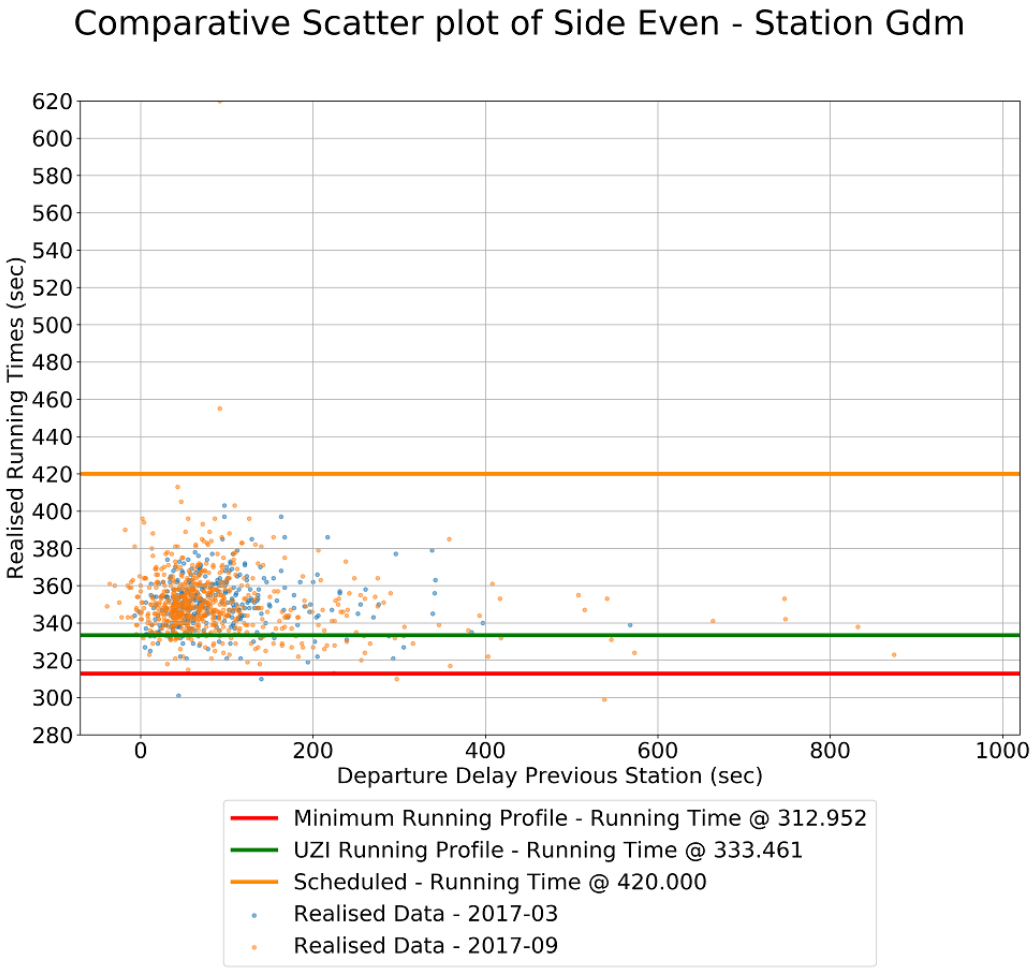


Figure L.4: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Seasonality Highlight - Gdm

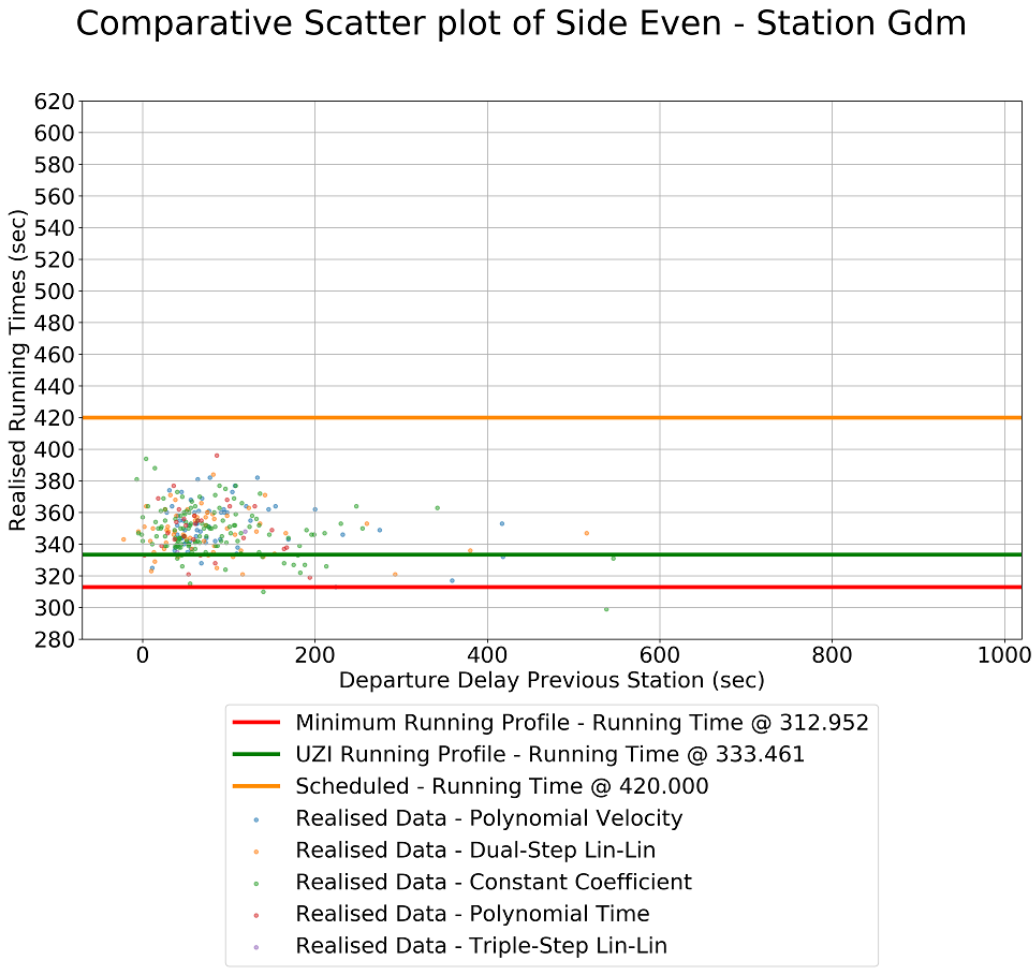


Figure L.5: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Brake Variant Highlight - Gdm

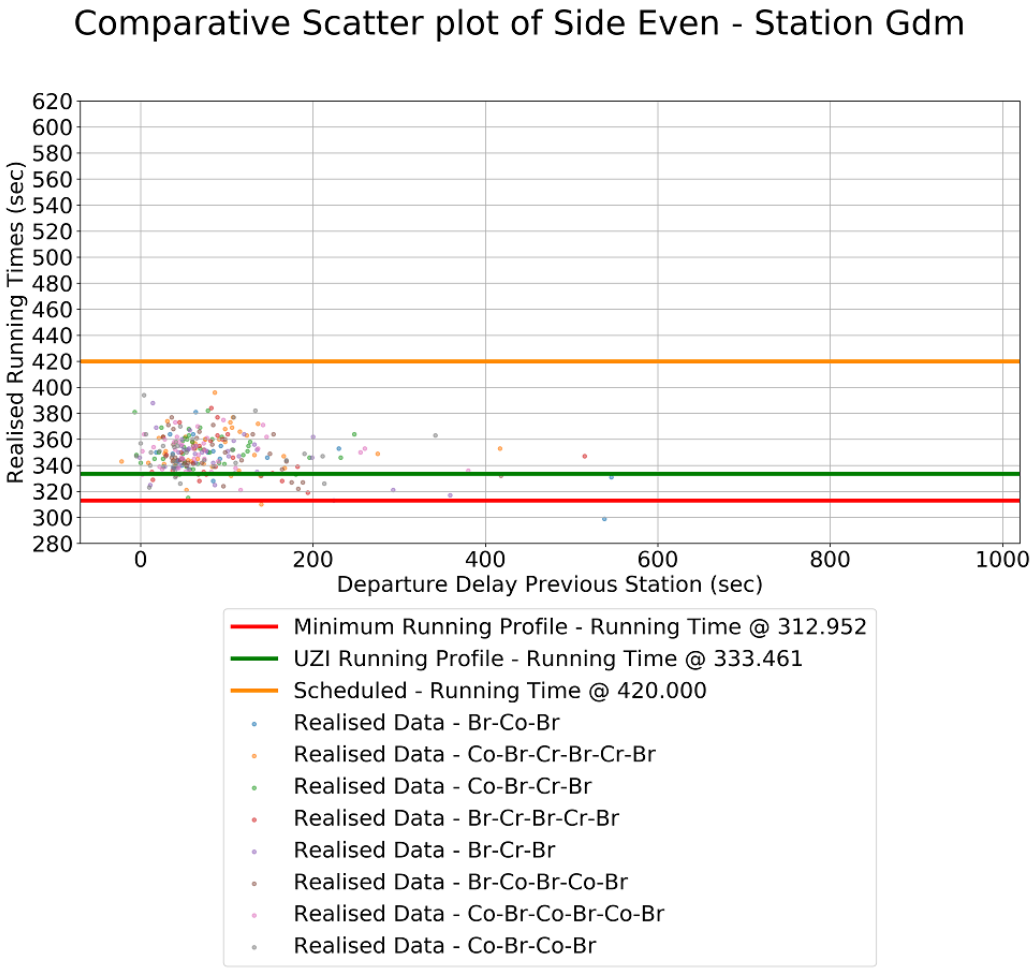


Figure L.6: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Deceleration Regime Highlight- Gdm

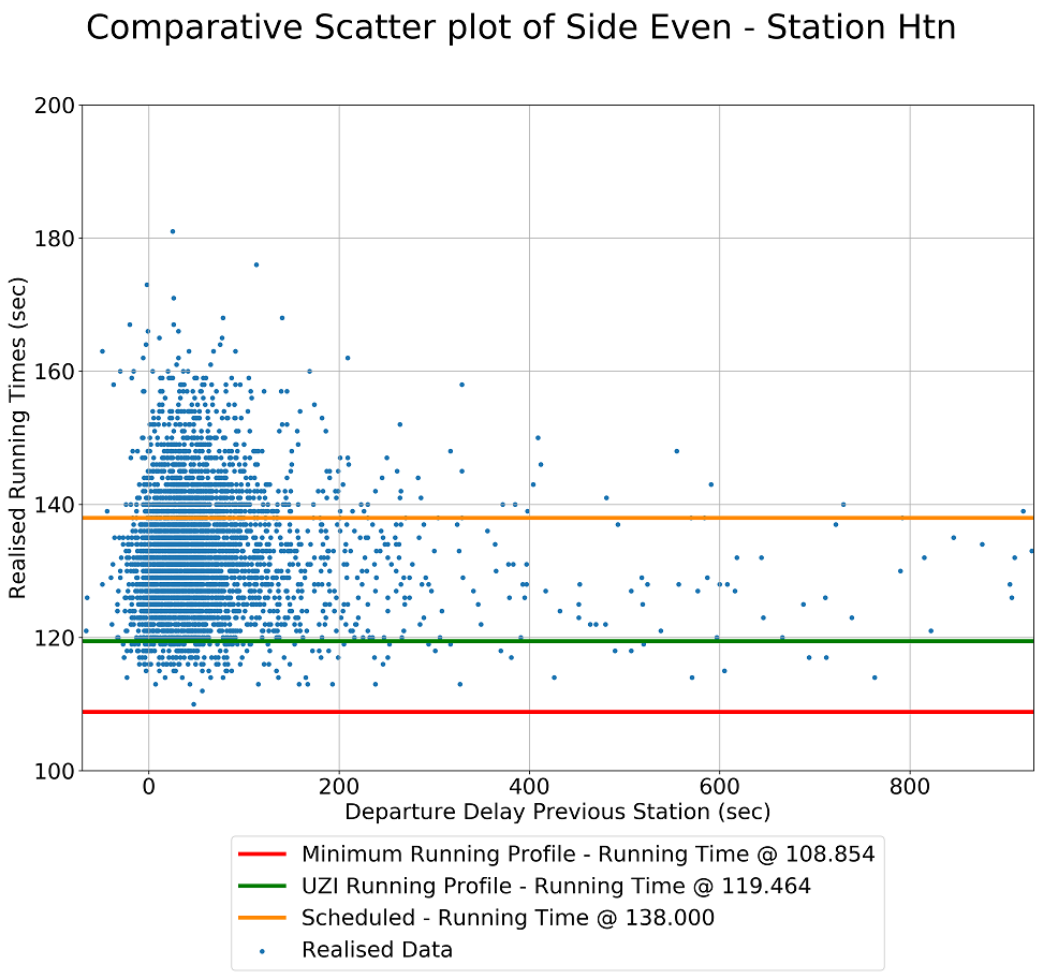


Figure L.7: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Htn

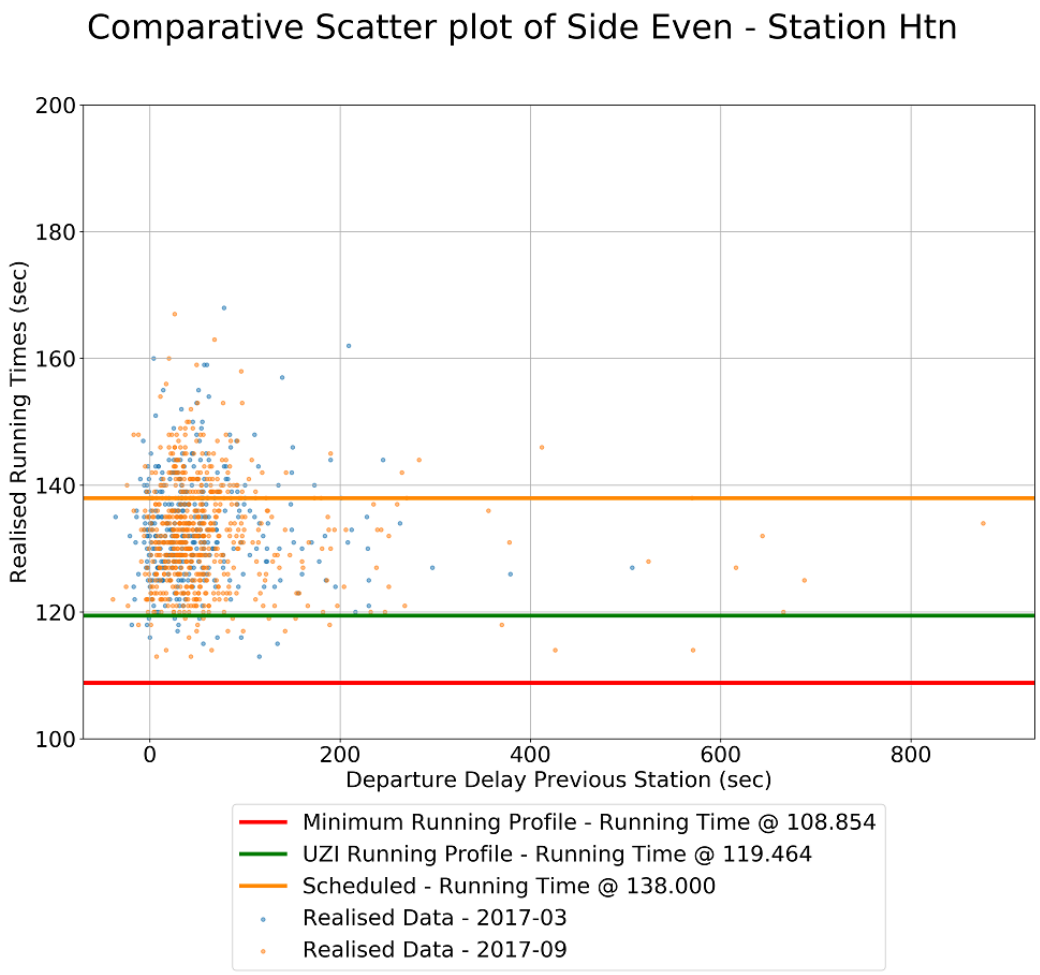


Figure L.8: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Seasonality Highlight - Htn

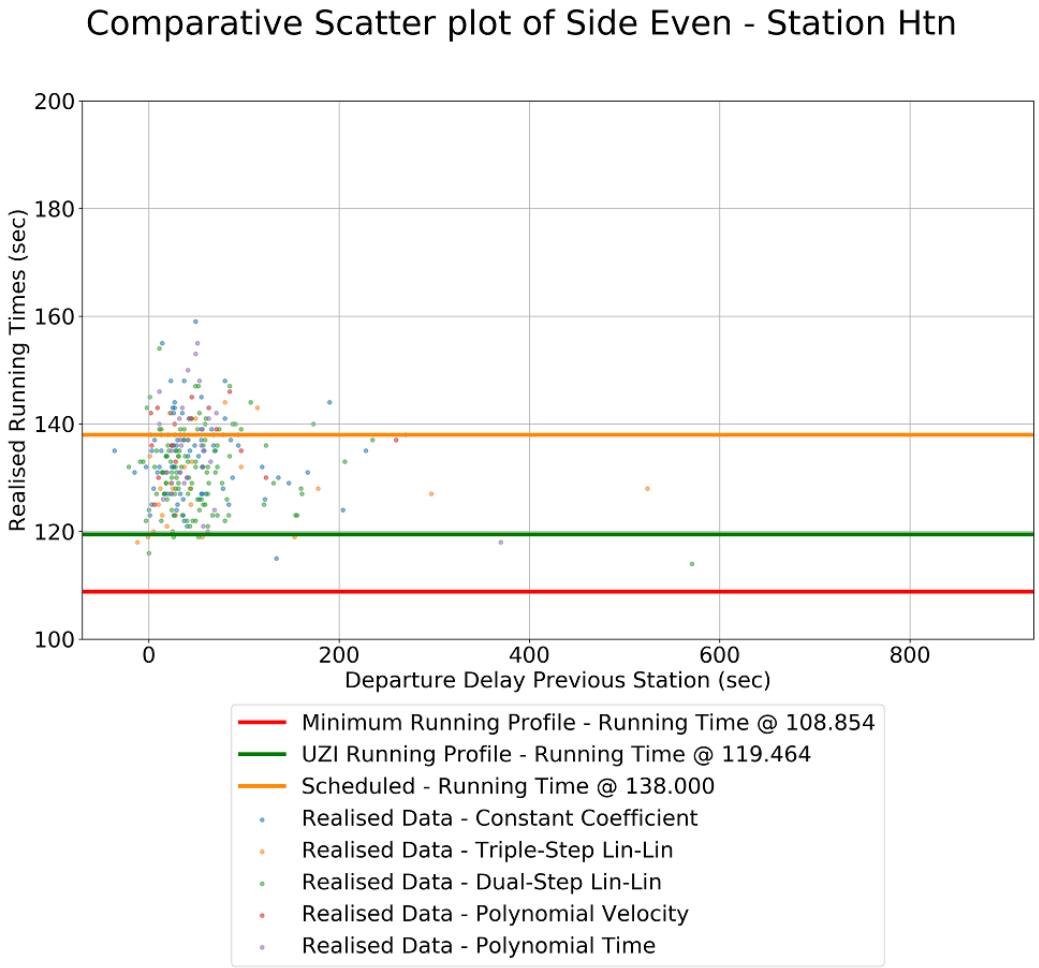


Figure L.9: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Brake Variant Highlight - Htn

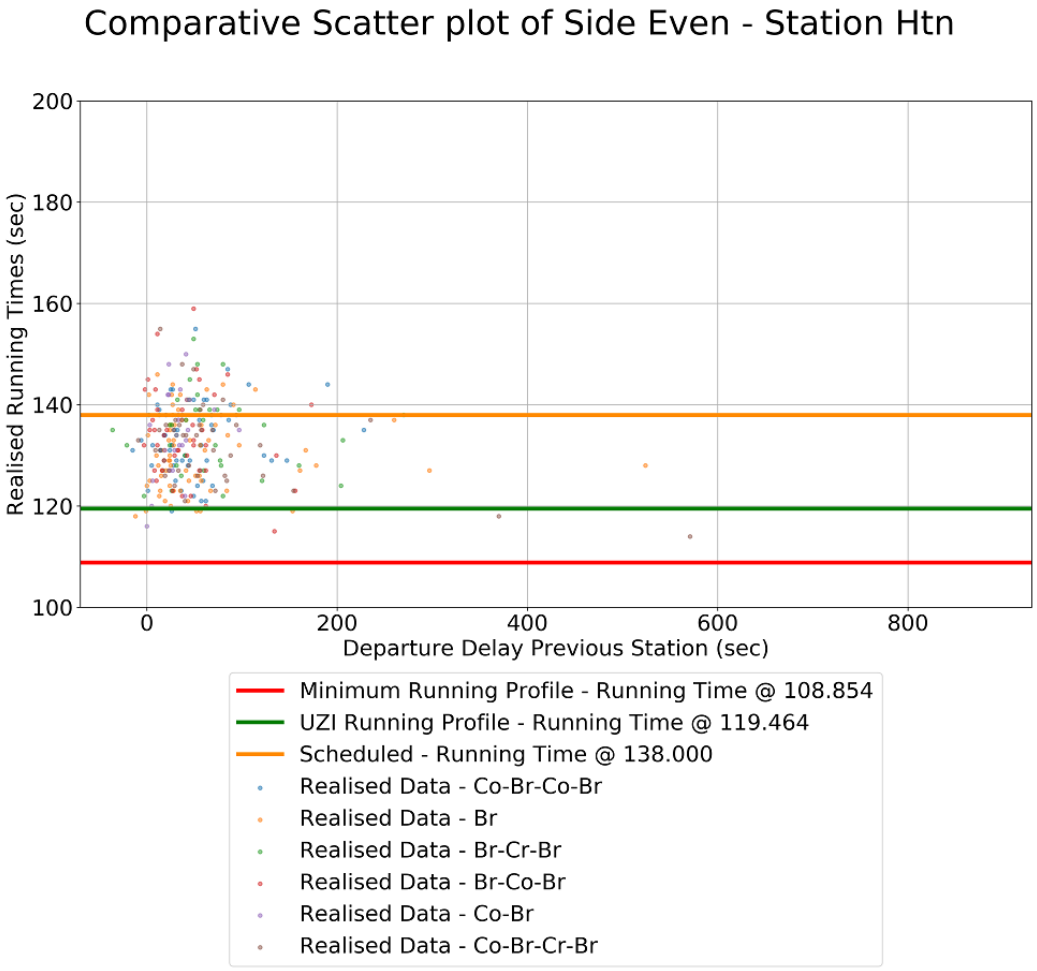


Figure L.10: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Deceleration Regime Highlight- Htn

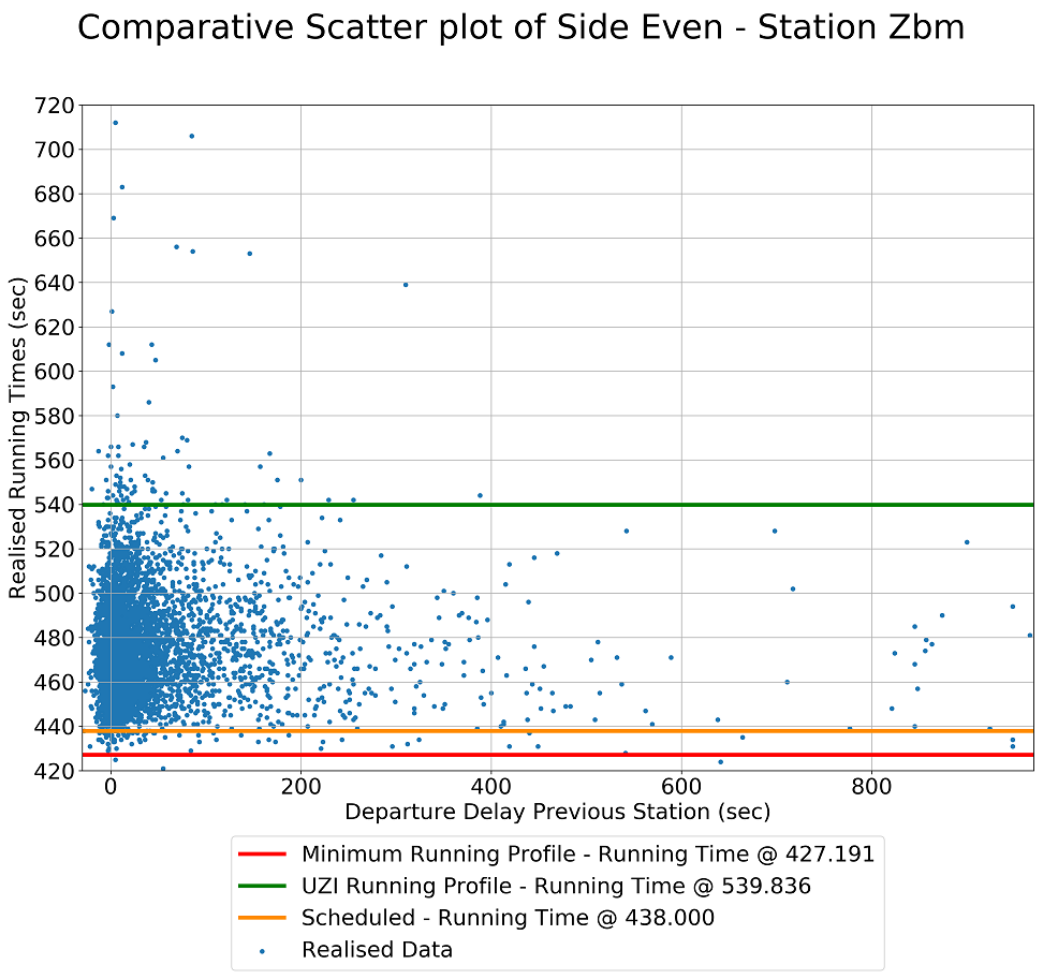


Figure L.11: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Zbm

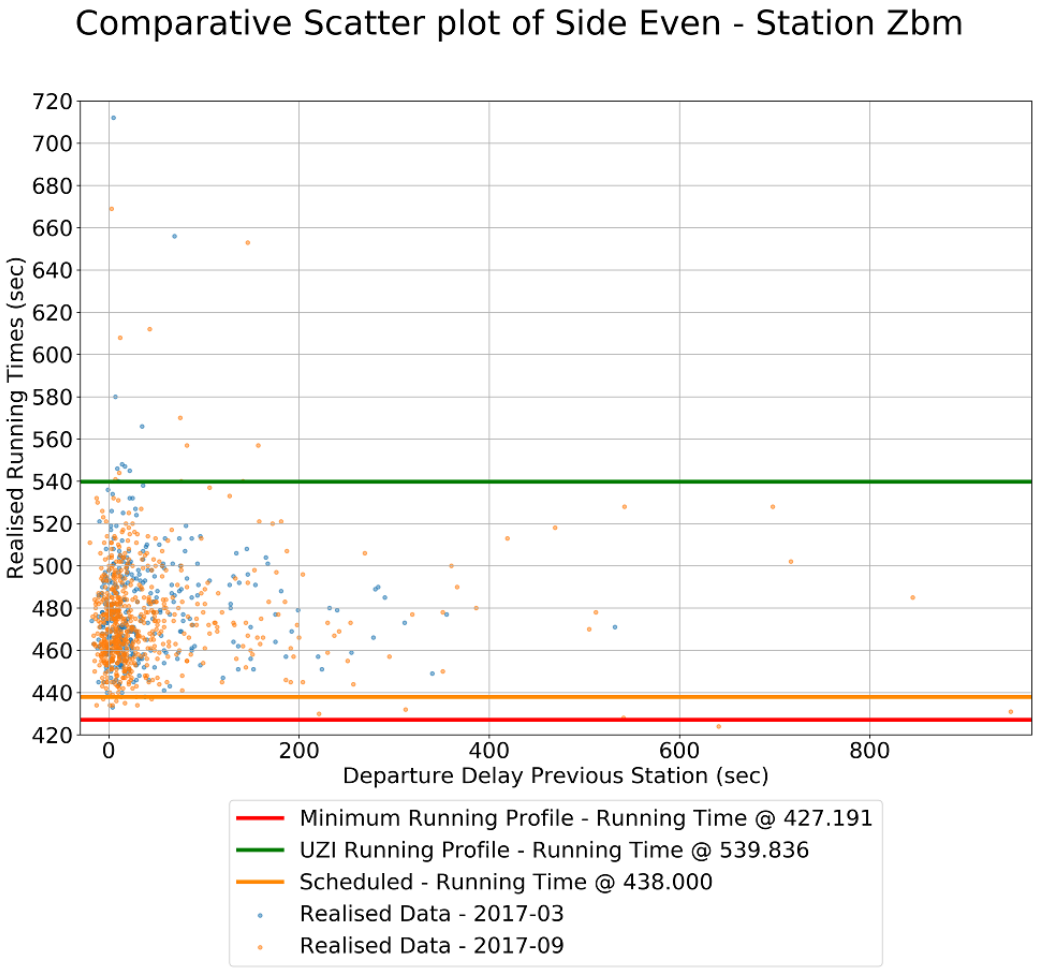


Figure L.12: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Seasonality Highlight - Zbm

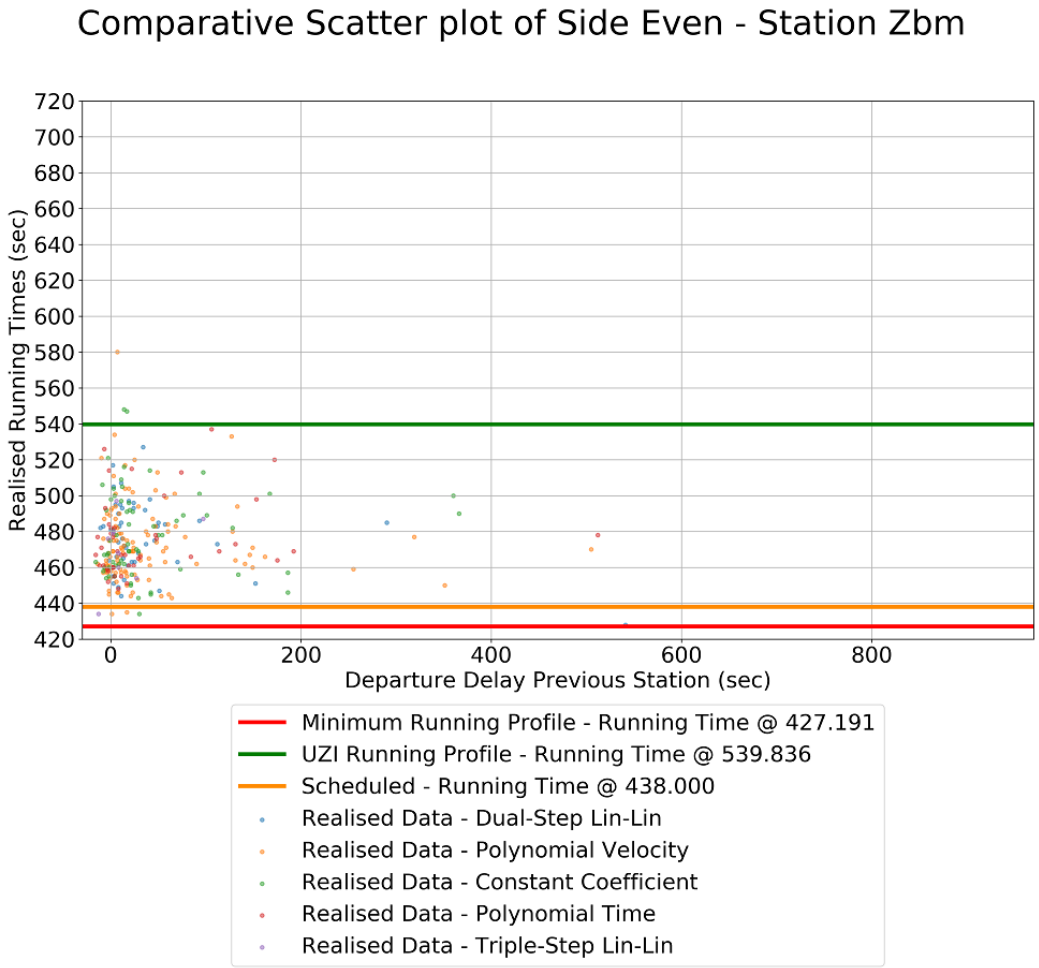


Figure L.13: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Brake Variant Highlight - Zbm

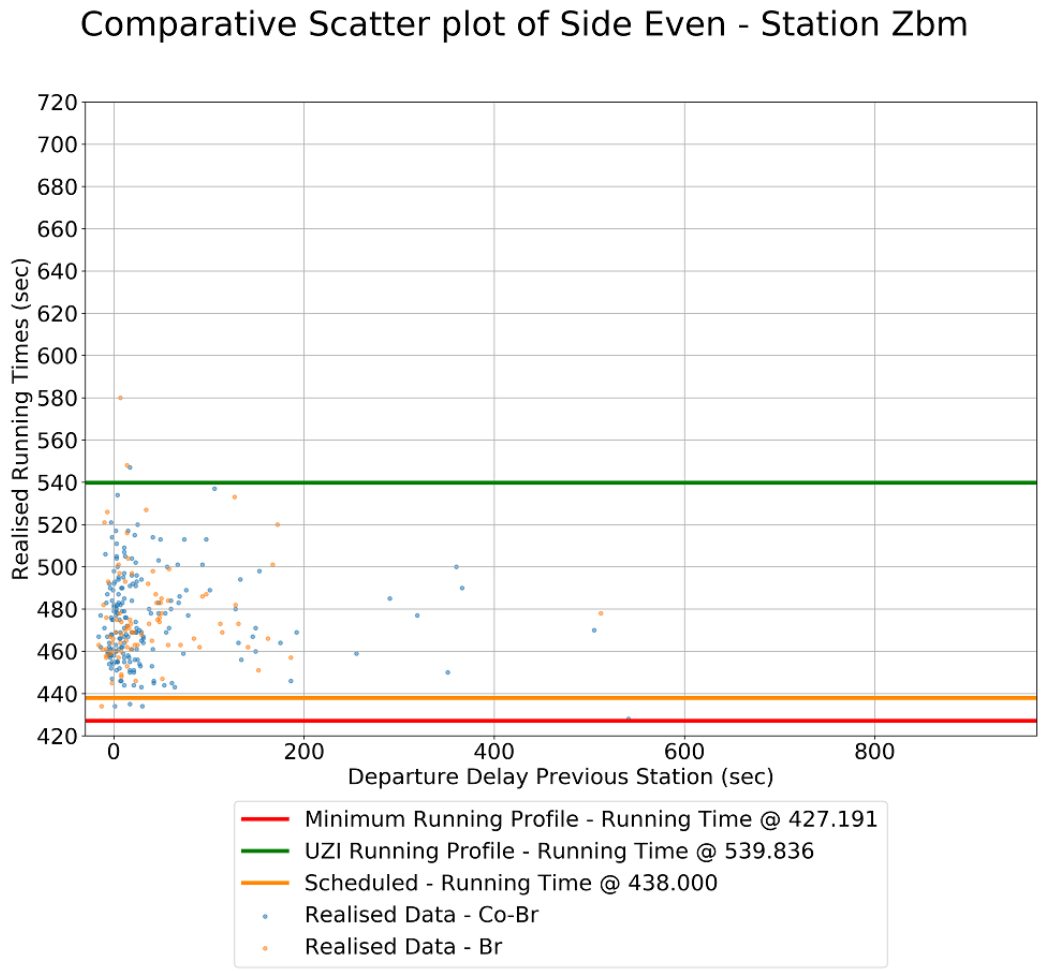


Figure L.14: Scatter Plot Comparative Analysis Departure Delay vs Realised Running Times - Deceleration Regime Highlight- Zbm

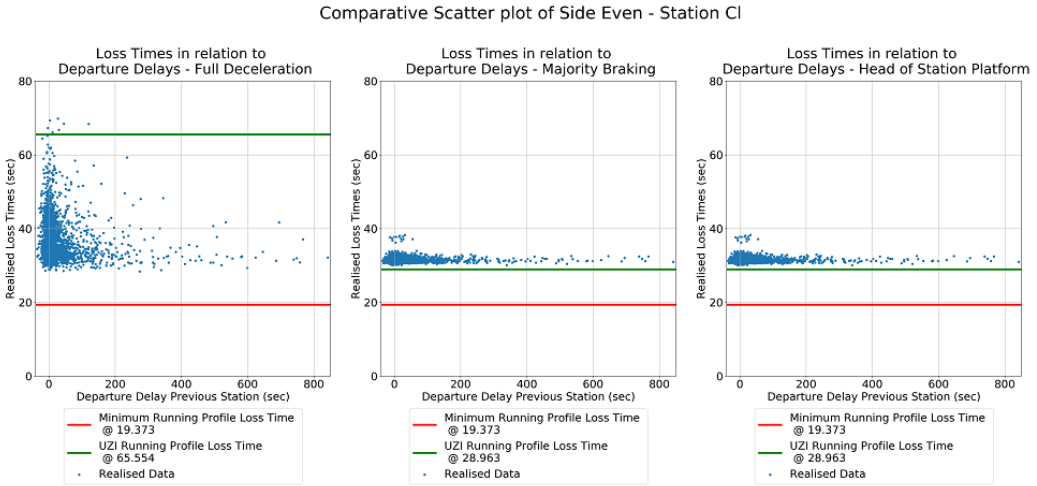


Figure L.15: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Cl

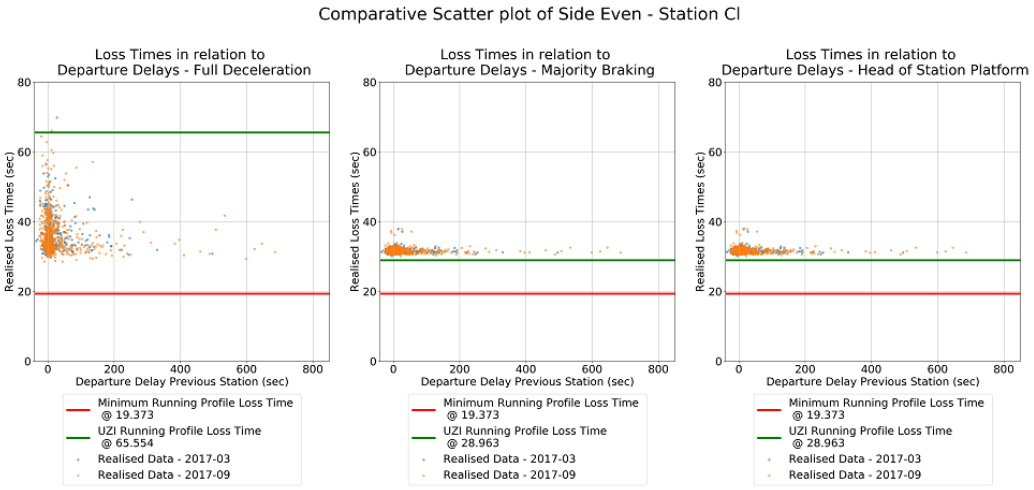


Figure L.16: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Seasonality Highlight - Cl

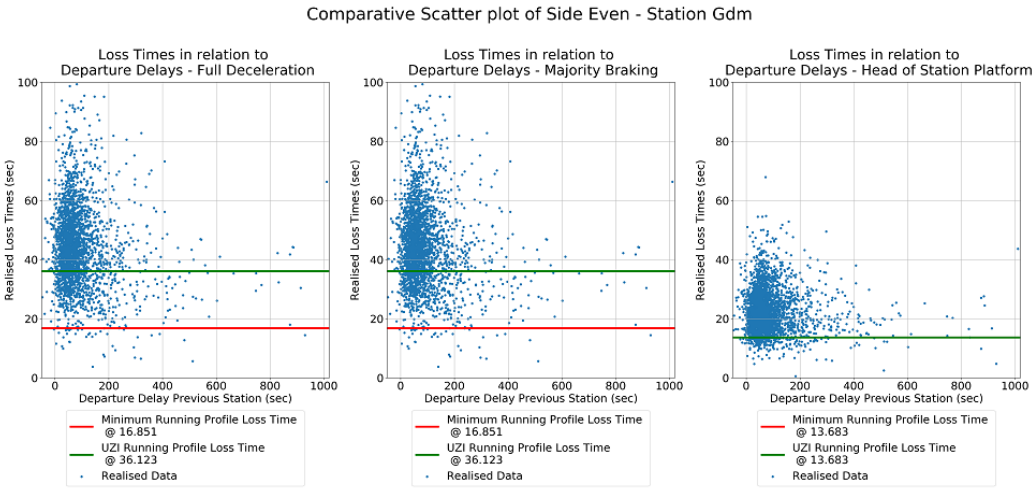


Figure L.17: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Gdm

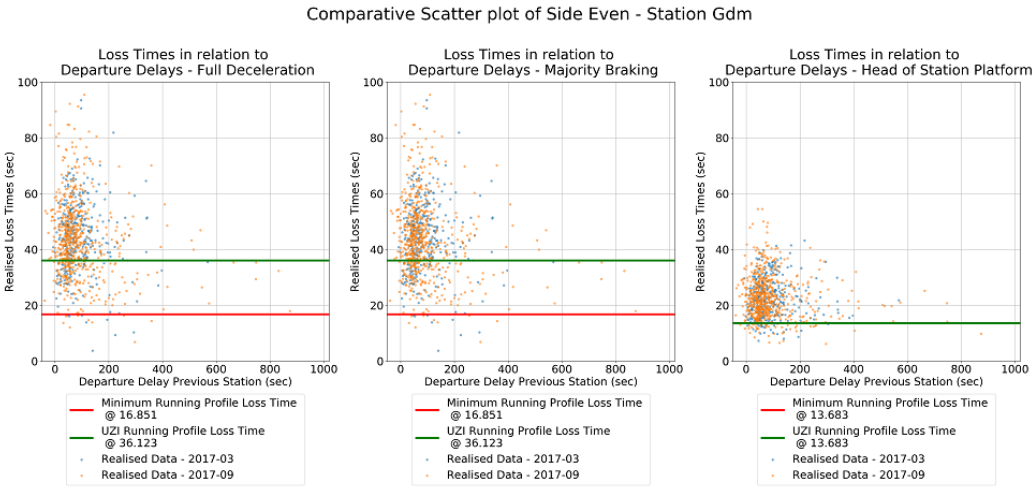


Figure L.18: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Seasonality Highlight - Gdm

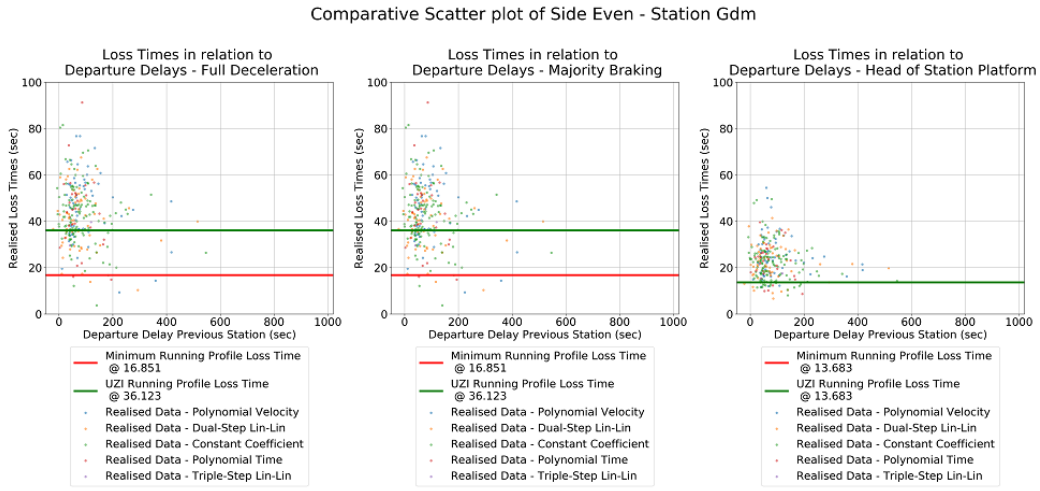


Figure L.19: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Brake Variant Highlight - Gdm

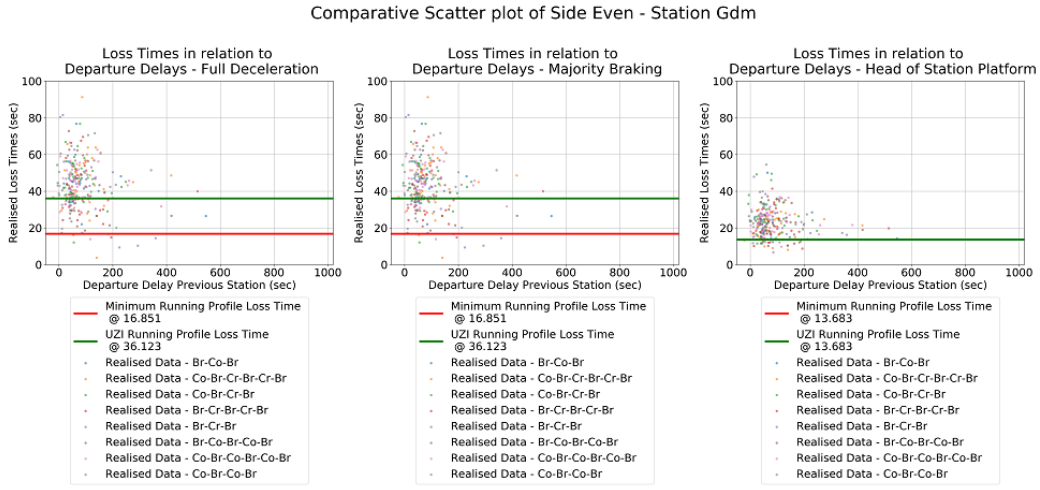


Figure L.20: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Deceleration Regime Highlight - Gdm

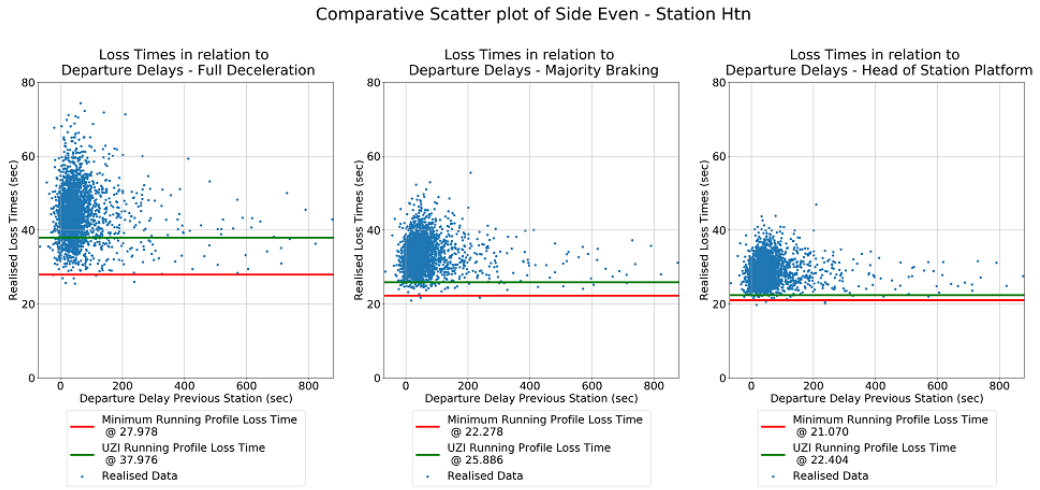


Figure L.21: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Htn

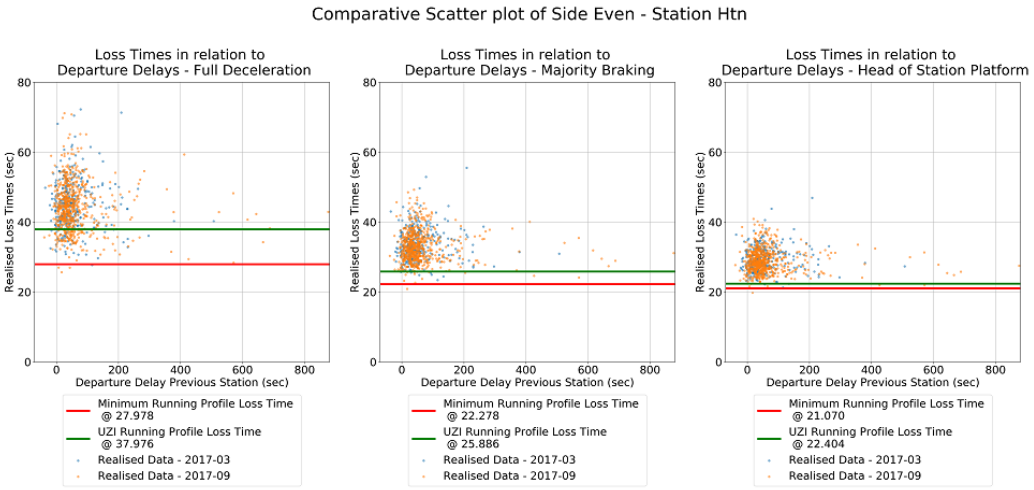


Figure L.22: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Seasonality Highlight - Htn

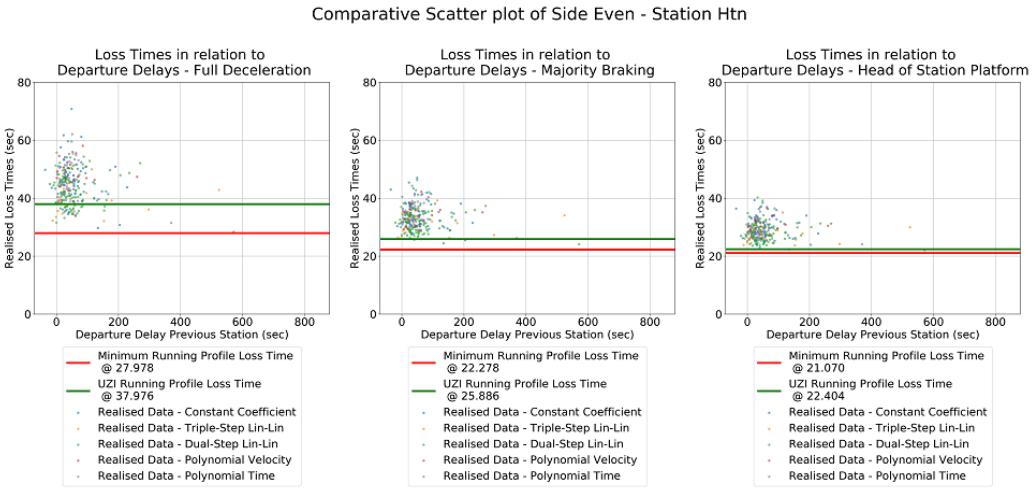


Figure L.23: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Brake Variant Highlight - Htn

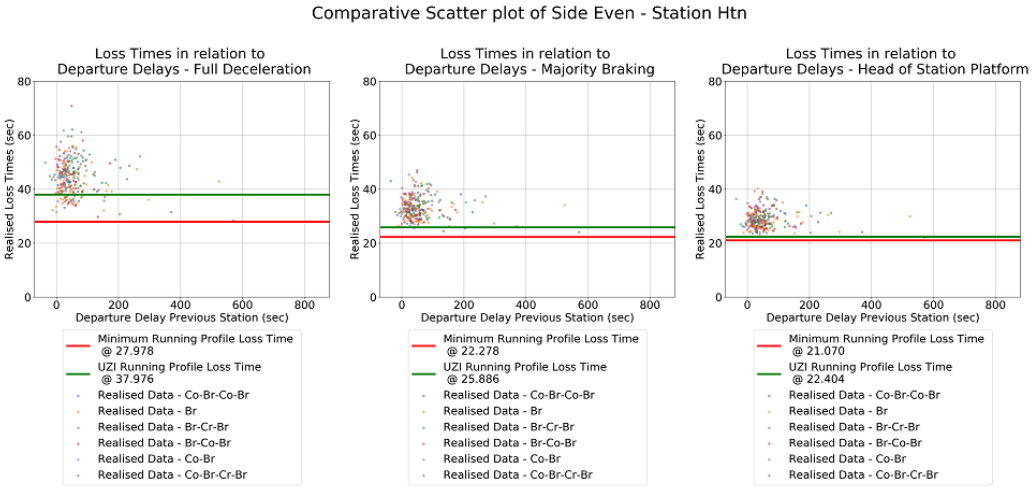


Figure L.24: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Deceleration Regime Highlight - Htn

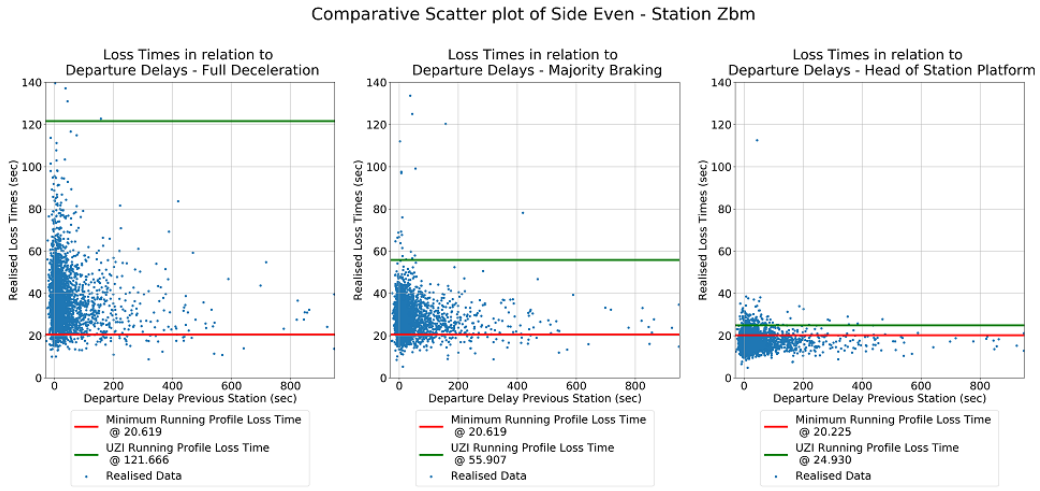


Figure L.25: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Zbm

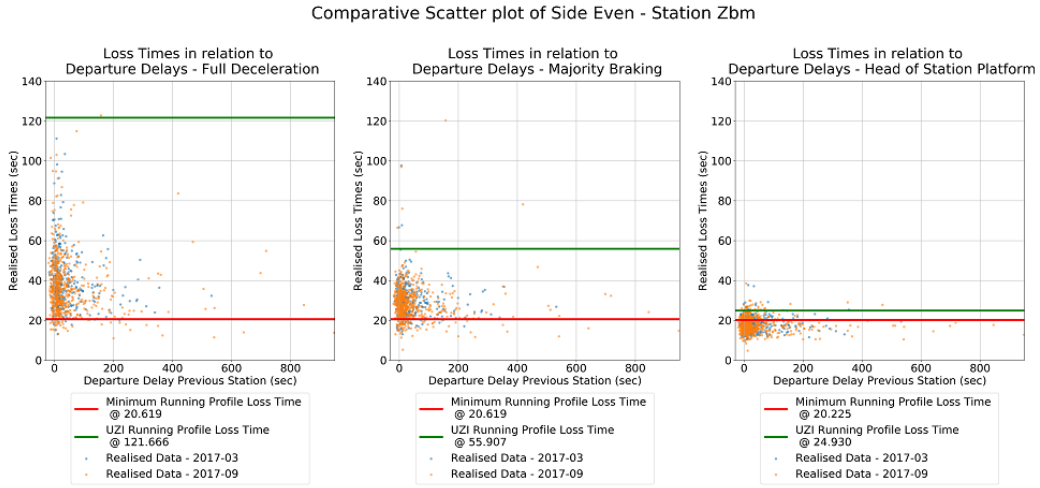


Figure L.26: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Seasonality Highlight - Zbm

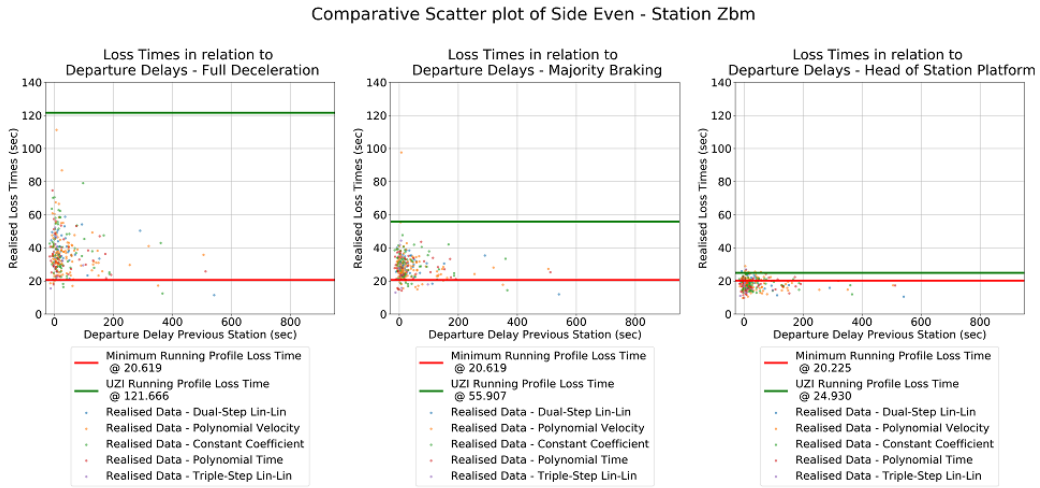


Figure L.27: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Brake Variant Highlight - Zbm

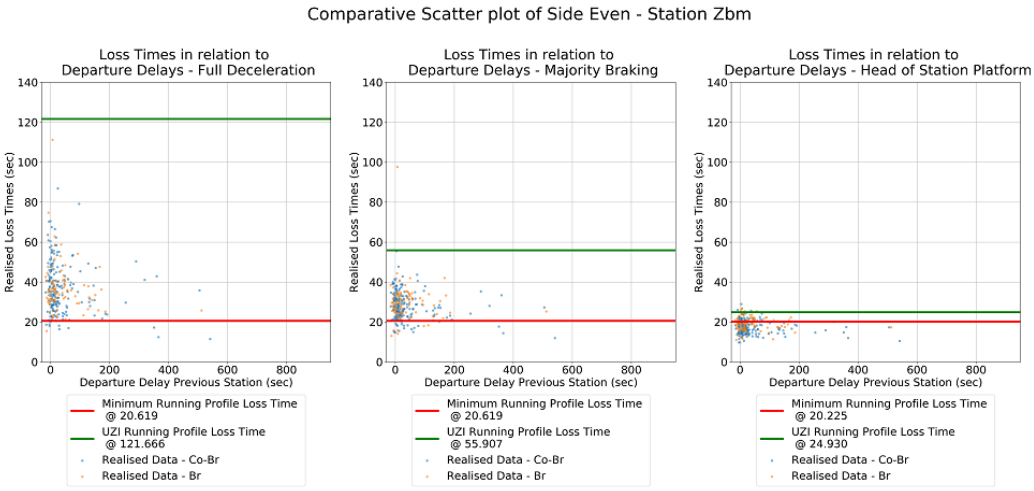


Figure L.28: Scatter Plot Comparative Analysis Departure Delay vs Deceleration Loss Times - Deceleration Regime Highlight - Zbm

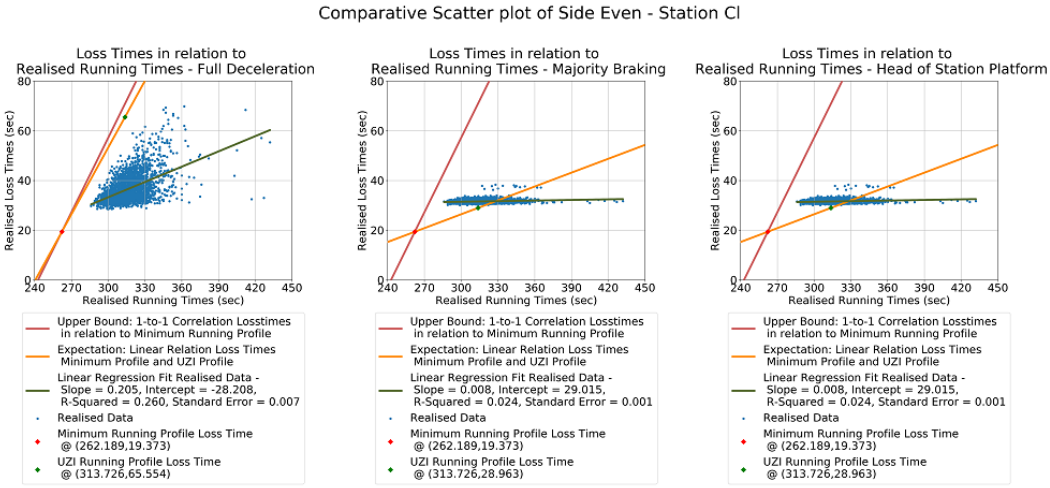


Figure L.29: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Cl

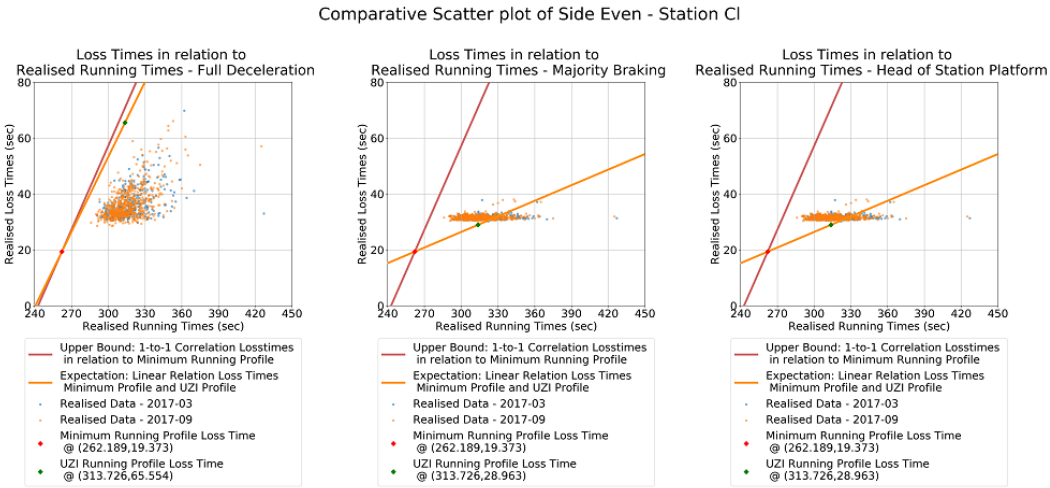


Figure L.30: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Seasonality Highlight - Cl

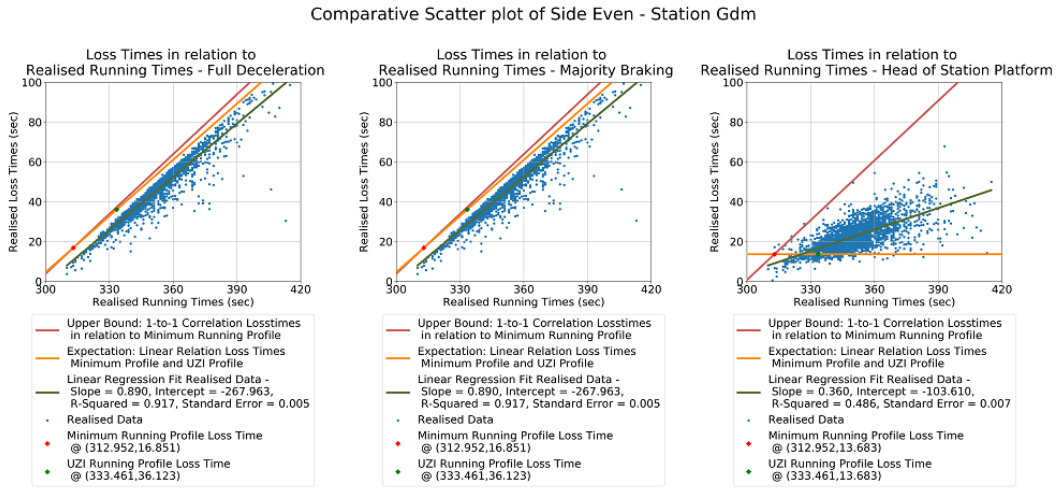


Figure L.31: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Gdm

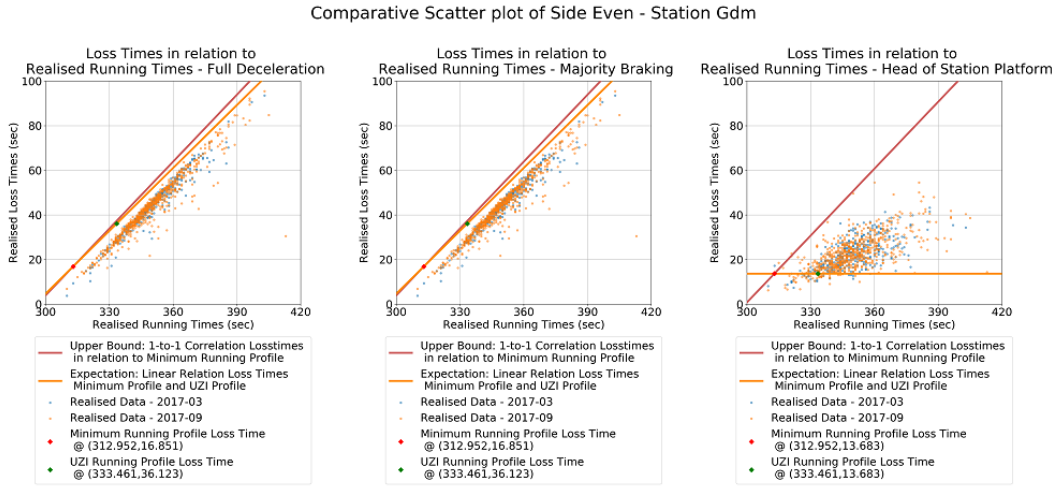


Figure L.32: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Seasonality Highlight - Gdm

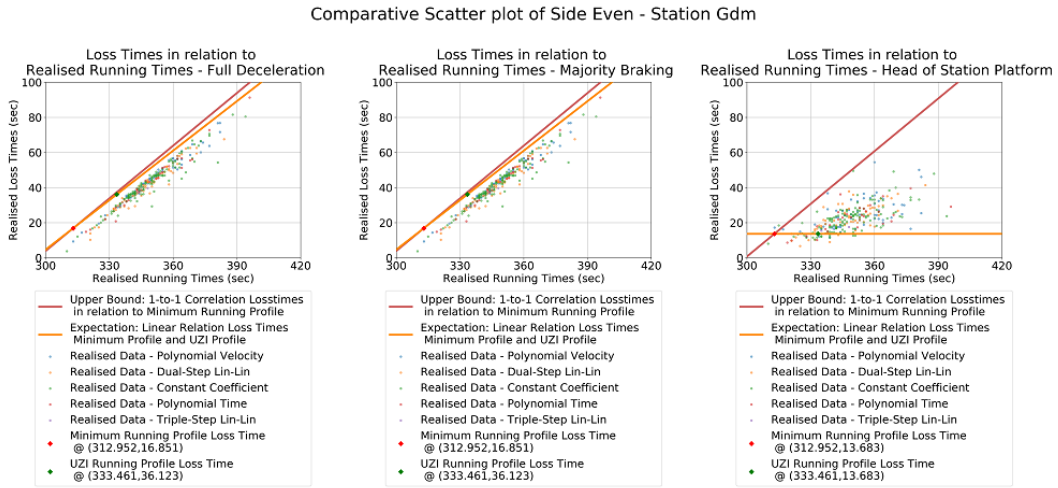


Figure L.33: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Brake Variant Highlight - Gdm

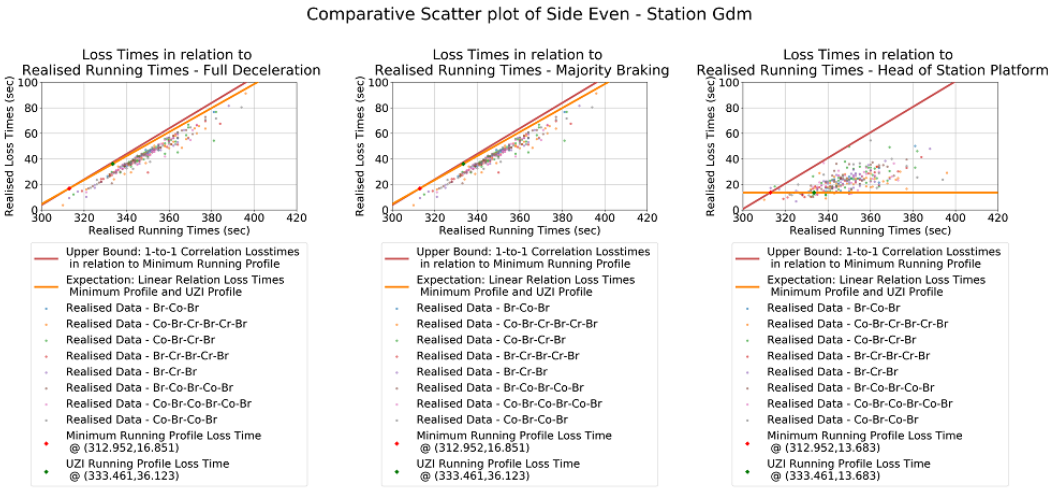


Figure L.34: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Deceleration Regime Highlight - Gdm

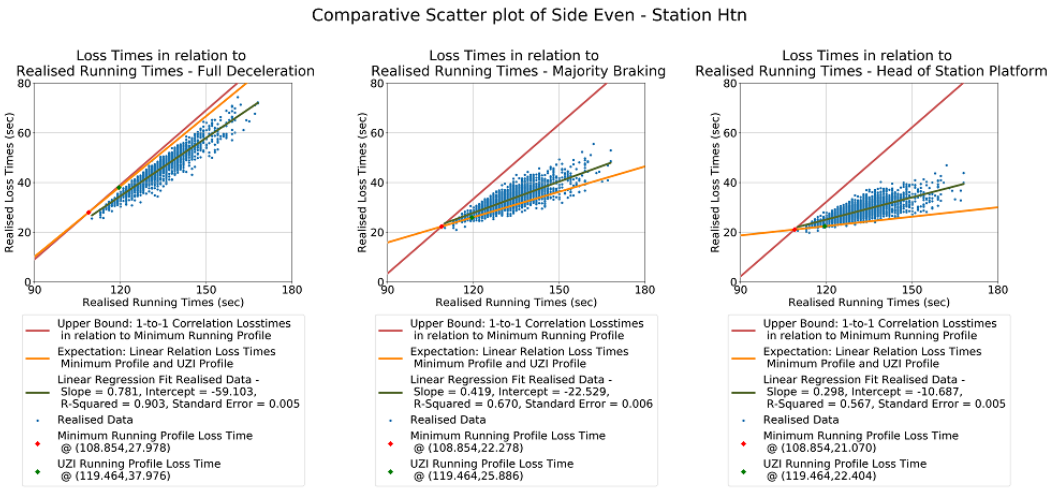


Figure L.35: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Htn

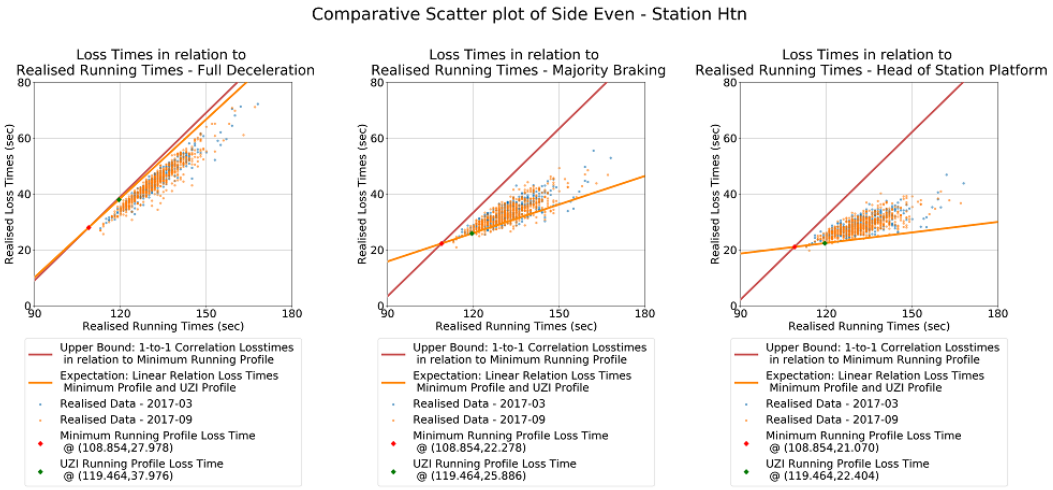


Figure L.36: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Seasonality Highlight - Htn

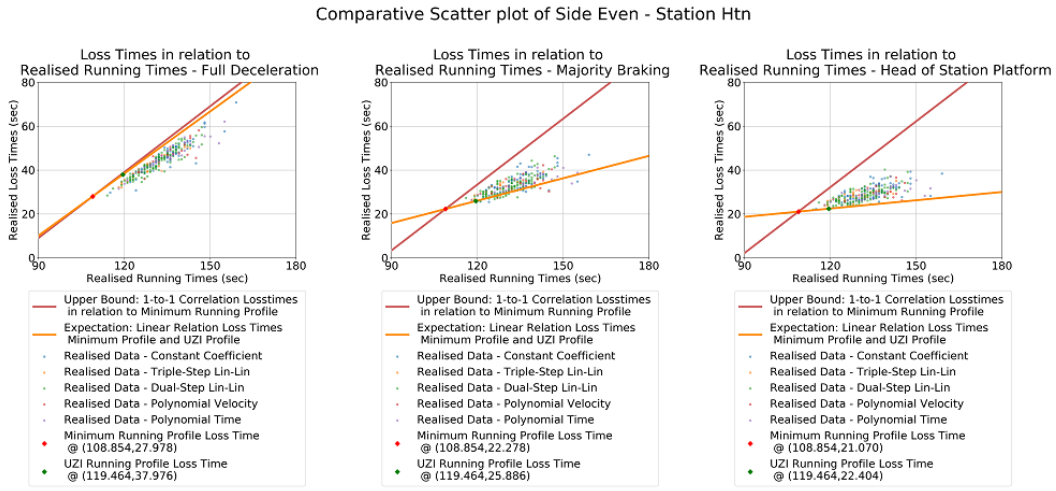


Figure L.37: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Brake Variant Highlight - Htn

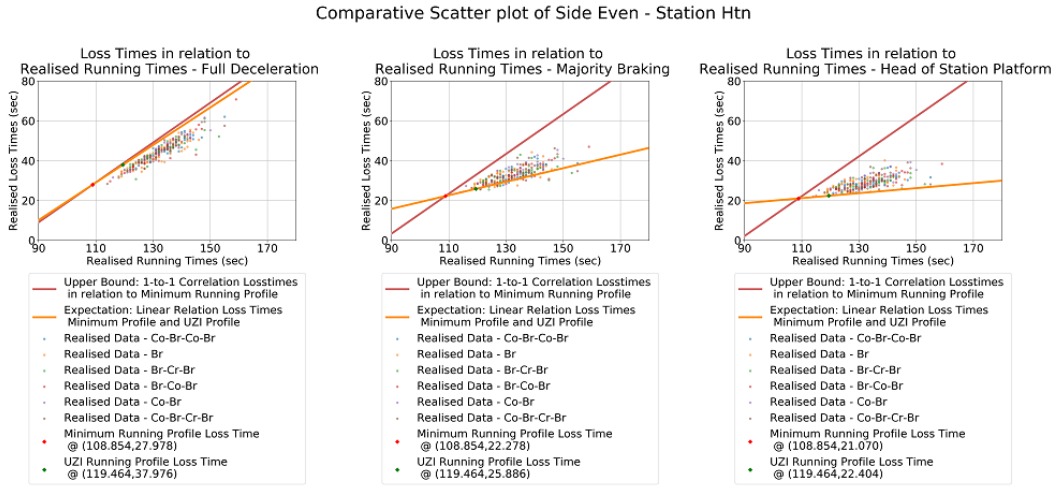


Figure L.38: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Deceleration Regime Highlight - Htn

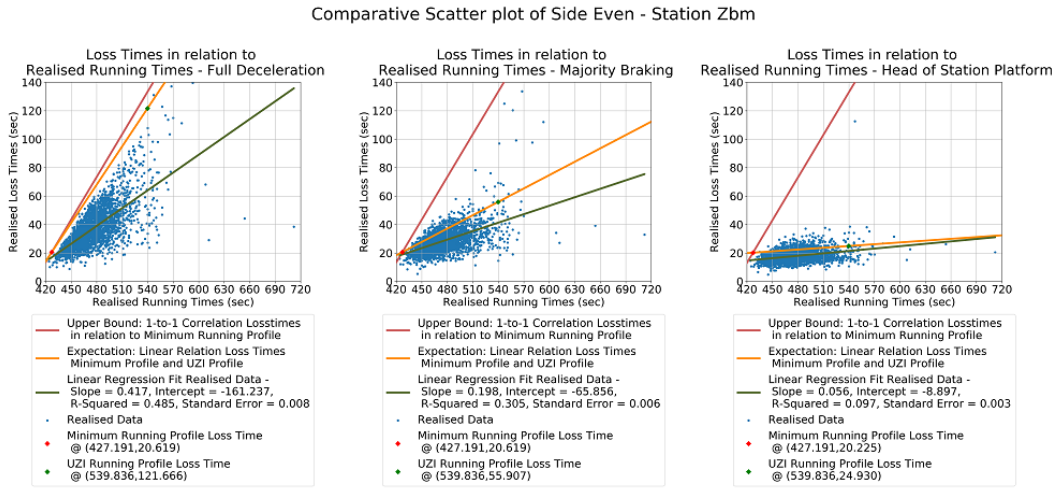


Figure L.39: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Zbm

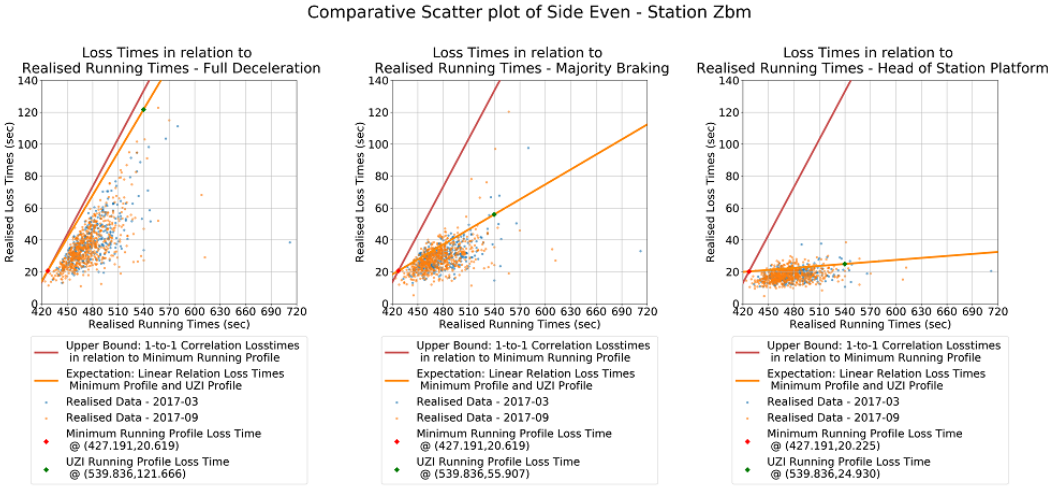


Figure L.40: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Seasonality Highlight - Zbm

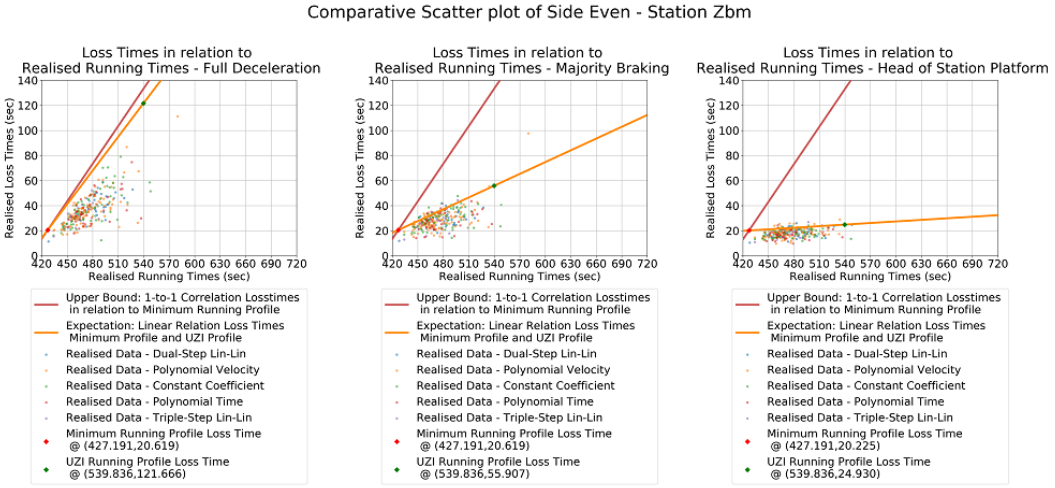


Figure L.41: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Brake Variant Highlight - Zbm

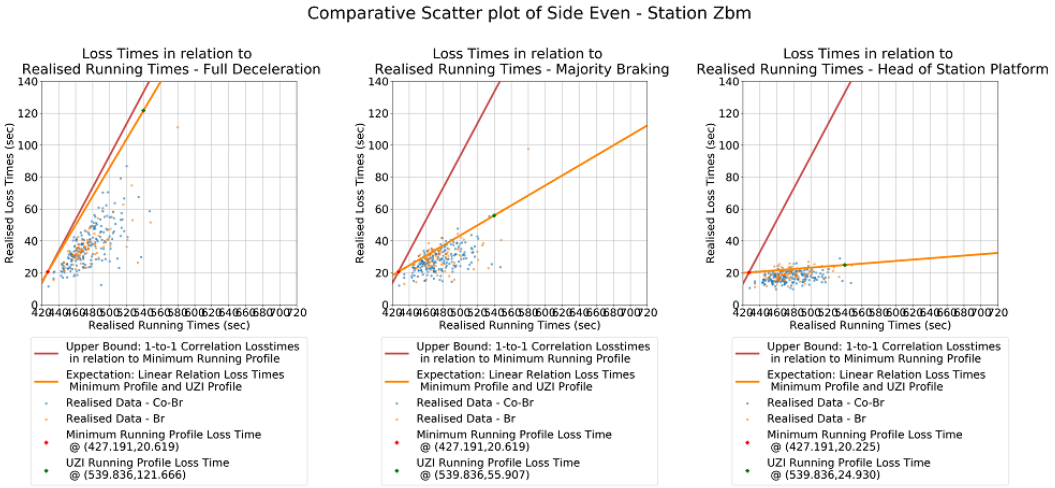


Figure L.42: Scatter Plot Comparative Analysis Realised Running Times vs Deceleration Loss Times - Deceleration Regime Highlight - Zbm