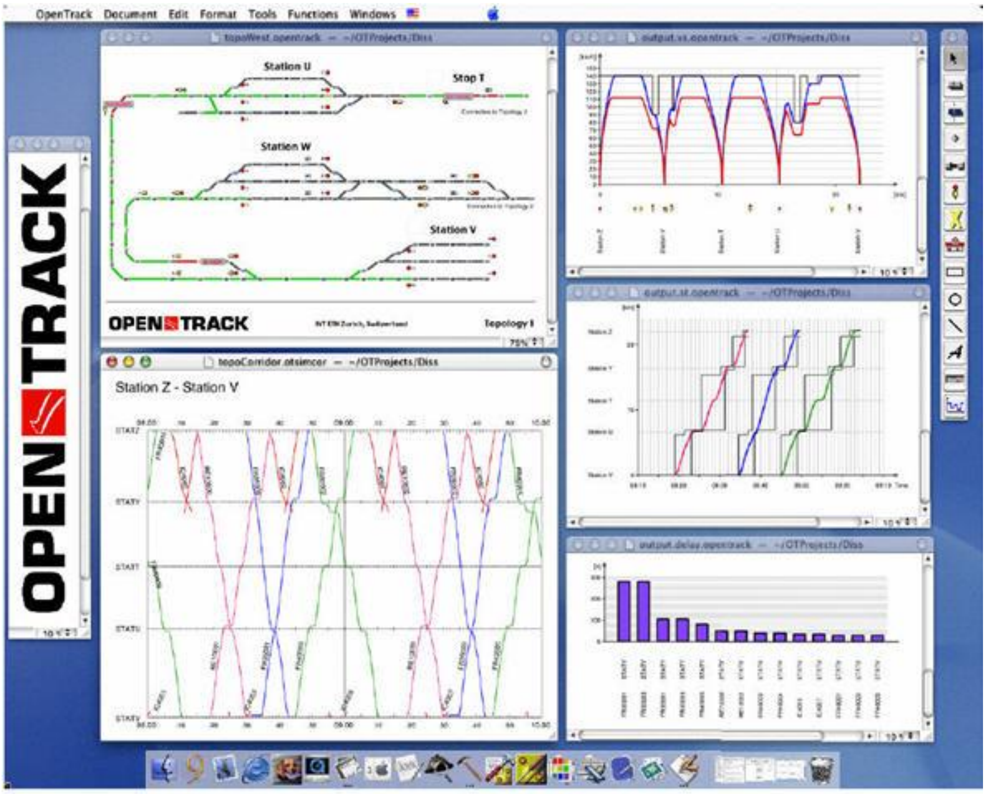
Movares
adviseurs & ingenieurs

Modelling Driving behavior in OpenTrack



Student: Haichao Nan
Student No: 4039610

Authorization page

	Name	**Department**
Student	Haichao Nan	TU Delft, Faculty of Civil Engineering, Department of Transport & Planning
Supervised by	Prof.dr.ing. I.A. Hansen	TU Delft, Faculty of Civil Engineering, Department of Transport & Planning
	Dr. E. Weits	Movares Nederland B.V.
	Dr. R.M.P. Goverde	TU Delft, Faculty of Civil Engineering, Department of Transport & Planning
	Ir. M. Wolbers	Movares Nederland B.V.
	Dr.ir. J.H. Baggen	TU Delft, Faculty of Technology, Policy and Management, Department of Transport and Logistics

Summary

Over the past decades, computer simulation is used more and more widely in every field of people's engineering industry for its cost efficiency, direct display of results and quick calculation speed.

OpenTrack is a typical railway microscopic simulation software tool implemented by the Swiss Federal Institute of Technology Zurich from mid 1990s. After 15 years of development, it has now become one of the most popular rail simulation tools and is used by a large number of railway operators, consulting companies and universities.

At the same time, as the world is experiencing global climate change and a growing population while the total amount of petrol and other natural fuel recourse is decreasing fast, the efficiency of energy consumption has become a critical requirement for present transportation system design and management.

Improving the energy efficiency of railways can be implemented in many aspects. In the domain of energy efficient driving, it had been proved many years ago by both theoretical and practical sides that in a train run, energy consumption can be reduced if drivers use energy effiecient driving strategies many years ago. However, in reality, the effect of driving behavior is difficult to evaluated because there are a lot of dynamics in a driver's driving: a driver is a human being and the process of how a driver determines his way of driving is not shown as easy as in the case of machines.

Therefore, it is important to investigate train drivers' driving behavior and check OpenTrack's modelling abilities for driving behavior.

This study aimed at following objectives:

1. From a theoretical point of view, study the background theory of OpenTrack modelling and energy efficient driving.
2. Analyze realization data and check which energy efficient driving behavior is applied in reality.
3. Analyze the capabilities of OpenTrack to model this behavior and formulate suggestions for OpenTrack's future improvement.

This study is divided into four parts as follows:
The first part mainly checks OpenTrack's simulation principles and other theories, including running time calculation and energy efficient driving strategies.

The second part refers to the driver's driving behavior investigation, which includes investigation method determination, data collection and processing procedures. That is, how to investigate the driver's driving behavior in reality, what data is selected and how accurate this is. Additionally, driving behavior theory and a summary of the author's study trip is presented to describe a driver's driving from both theoretical and practical aspects.

In this study, the Trento data derived from the TROTS (a Dutch train describer) log files is used. Due to its poor accuracy, a section groupig method is introduced to remove the deviations in the Trento data. Addtionally, this study also investigated the best approach (i.e., if the grouping method should be applied based on length or fixed section numbers) in applying the grouping method to the Trento data, and corresponding optimal coefficients are derived.

The third part focuses on the OpenTrack model construction based on given information about the real world. Because simulation is the virtual representation of reality, it is important to determine which factors have to be modelled and which will not.

In the last part, driving behavior in different driving regimes (acceleration, braking, cruising and coasting) and other aspects are investigated by using the Trento data recorded between 04-10-2010 and 08-10-2010 in the rail line of Geldermalsen -Den Bosch. Three train run groups are assembled based on different factors and studied seperately. Addtionally, OpenTrack's modelling abilities for driving behavior is reviewed.

Finally, within the scope restrictions of the data used in this study, this study mainly formulates the following conclusions:

1. In practice, the initial delay is not an important influencing factor for a driver to make his driving strategy, which means the relationship between a train's initial delay and its corresponding running time islacking;
2. Train drivers seldom do coasting during their driving times, even if they are not hindered by other trains;
3. It is proved that there is no evidence of a relationship between a train's acceleration/deceleration rate usage and cruising speed. In other words, it is not likely for a train driver to choose a higher cruising speed if he lost some time in the acceleration/deceleration phase.
4. The acceleration/deceleration rate used in one train series group generally shows an approximately symmetric distribution.
5. OpenTrack has a limited ability to model coasting trains in reality by using coasting boards, because the setting of coasting boards is not practical in reality.

Preface

This thesis is based on my master project of “Modelling driving behavior in OpenTrack”, and the latter is a graduation project to obtain a Master’s degree in Faculty of Civil Engineering, Delft University of Technology. Most of the works of this project are implemented during my internship at Movares, Utrecht.

Many people supported and assisted me during the process of this research. Firstly, I would like to use this opportunity to represent my thanks and gratitude to all my graduation committee memebers, for their suggestions and supervisions over the past eight months. This master project cannot be made without you! Thanks to Prof.dr.ing. I.A. Hansen for contacting the master project for me. Thanks to Dr. E. Weits for providing me the chance to do my master project in Movares and helped me a lot when I encountered diffciculties in my study. Thanks to Dr. R.M.P. Goverde for acting as my daliy supervisor from TU Delft and gave me many helps during my master project. Thanks to Ir. M. Wolbers for teaching me a lot everyday as my daliy supervisor in Movares. Thanks to Dr.ir. J.H. Baggen for being my supervior from the other department of TU Delft.

Thanks to Mr. Koning and other colleagues in Movares. They are all very nice to me and teach me a lot in a relaxed atmosphere. Their friendship and support is truly valuable. I will never forget my eight months’ internship in Movares and always be proud of once being a member of them.

Thanks to the people who also gave me brilliant ideas and suggestions in my master projects: Dr. Daniel Huerlimann from Switherland; Dr. Jan Hogenraad from NS; Mr. Schrader from NSB and MSc Hongfu Huang from Systransis.

Furthermore, I would like to thank all my schoolmates and friends especially Fangfang Zheng, Meng Wang. Thank you for your help and directions during my two years of study in the Netherlands. Thanks to Ling Ma for her supporting in my first year. Finally, I want to present my biggest thanks to my parents in China. Thank you for your understanding and I promise to you that I will never leave you again! I love you!

Haichao Nan
Utrecht, 2011

Table of contents

1 Introduction

1.1 Simulation software introduction

1.1.1. Background

Originally, simulation is defined as the reproduction of a real object or a process as a model, i.e. a railway laboratory in a scale of 1:90 about the reality. But in nowadays, it has a new meaning as representing the reality in a virtual representation. Computer simulation is used more and more widely in every field of people's engineering industry for its cost efficiency, direct display of results and quick calculation speed.

The application of simulation models in the railway domain has been implemented for many years, and after years of developing, modern railway simulation tools make it possible for software users to analyze a complicated network just in front of a computer. They can try many approaches to optimize the network's operational concept without the cost for expensive testing in reality.

1.1.2. Railway simulation software classification

In general, railway simulation models can be defined as three categories: macroscopic, mesoscopic and microscopic models. The differences between these three types of models are exemplified in Figure 1-1:

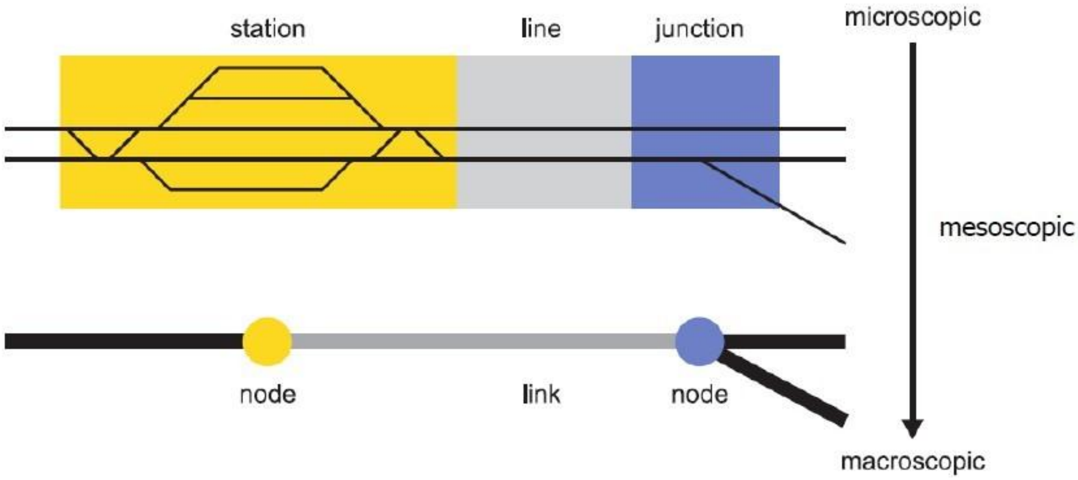


Figure 1-1 difference in the complex of node-link presentation of Macro, Meso and Microscopic models (Radtke 2008)

Macroscopic models are usually used for long-term planning, strategy study or special routing problems. In order to obtain a quicker response speed and more abstraction about the whole network, they require fewer details about link and nodes than microscopic models. However, in the contrary, due to the limited information about the track information such as gradients and and exact locations of speed changes, macroscopic models are not accurate enough for exact running time calculation and route conflict detection.

Microscopic models are more emphasized in detail and suitable for exact running time calculation, timetable construction and simulation, route conflict detection and dispatching problems. These models always contain detailed attributes of tracks, signalling systems and so on, thus they are suitable for determining reliable results for operational planning of networks, lines and junctions.

Mesoscopic models can be considered as a combination or “balance” between macroscopic and microscopic models. Other from macroscopic and microscopic models, mesoscopic models not only contain some details about nodes (stations), but also other information of links (tracks), such as track’s crossing, merging and splitting points. However, details about signal location are not included. The advantage of mesoscopic models is that they can detect and identify the conflict points of the network within a medium details simulation scale.

An example of infrastructure scale in mesoscopic model is given in Figure 1-2:

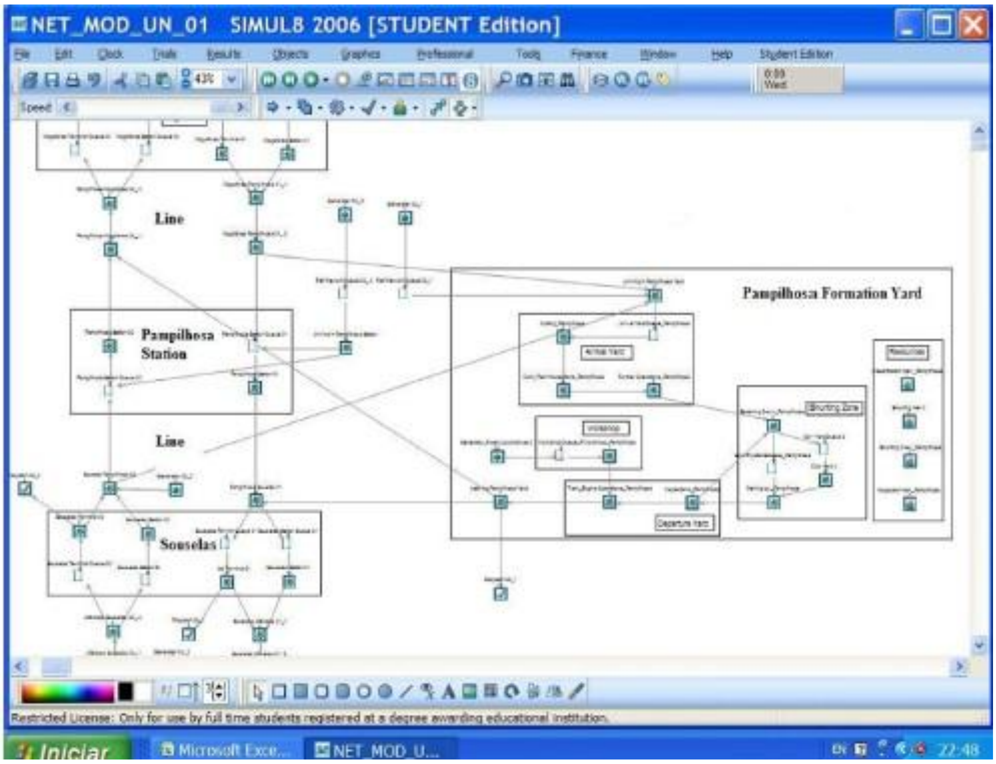


Figure 1-2 Example of infrastructure scale in Mesoscopic model (Marinov and Viegas 2011)

1.1.3. *OpenTrack software introduction*

OpenTrack is a typical railway microscopic simulation software tool implemented by the Swiss Federal Institute of Technology from mid 1990s. After 15 years of development, it has now become the most popular rail simulation tool and is used by a large number of railway operators, consulting companies and universities.

According to the OpenTrack’s release notes V1.6 (Hürlimann 2010), the main modules of OpenTrack is illustrated in Figure 1-3:

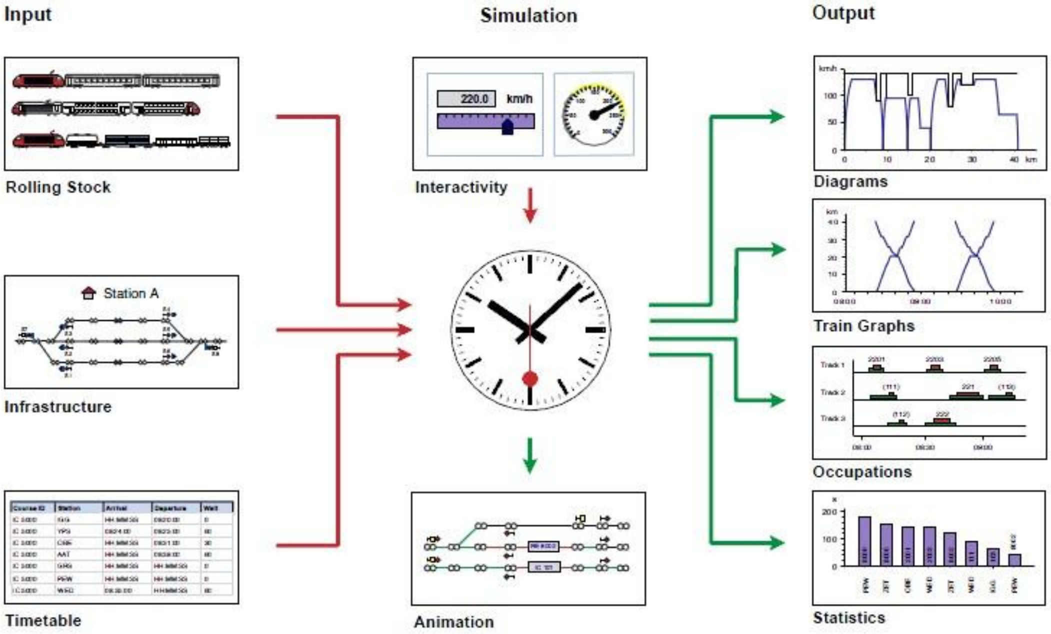


Figure 1-3 OpenTrack Process: Input-Simulation-Output (Hürlimann 2010)

Figure 1-3 shows a basic clue about the structure of OpenTrack. In OpenTrack, The input and simulation determines the output. Both of them are introduced in following chapters.

1.1.4. *Simulation model calibration*

Ane key condition for achieving the maximum effect of railway simulation is that the simulation result should reflect the real world; otherwise all the efforts of using the simulation tool would be useless.

Recently, it has been observed and reported that the running times calculated by OpenTrack may sometimes differ from reality, due to e.g. lack of knowledge of the formulae in OpenTrack and due to insufficient modelling capabilities in OpenTrack (Agricola 2009). In addition, as explained above, this kind of deviation from reality may significantly influence the efficiency and value of using OpenTrack.

1.2 Energy efficient driving introduction

1.2.1. *Background*

As the world is experiencing global climate change and a growing population while the total amount of petrol and other natural fuel recourse is decreasing fast, the efficiency of energy consumption has become a critical requirement for nowadays transportation system design and management.

Compared to road and air transportation systems, rail transportation is already considered as more energy efficient and thus more competitive in energy consumption. Therefore, further improvement of energy efficiency in rail transport will not only lower cost for railway undertaking companies, but also contribute to the whole rail industry's gloom in the future.

1.2.2. *Energy efficient approaches in railway industry*

The approaches of improving railway's energy efficiency can be classified into two categories. The first one is to improve railway transport's hardware, such as engines of rolling stock, locomotives, power system, etc; the other one is to improve the level of railway system's management. The latter category can be further divided into railway operation management on a big scale such as network's timetable design, and in a small scale of individual train driving. All of these classifications is shown in Figure 1-4:

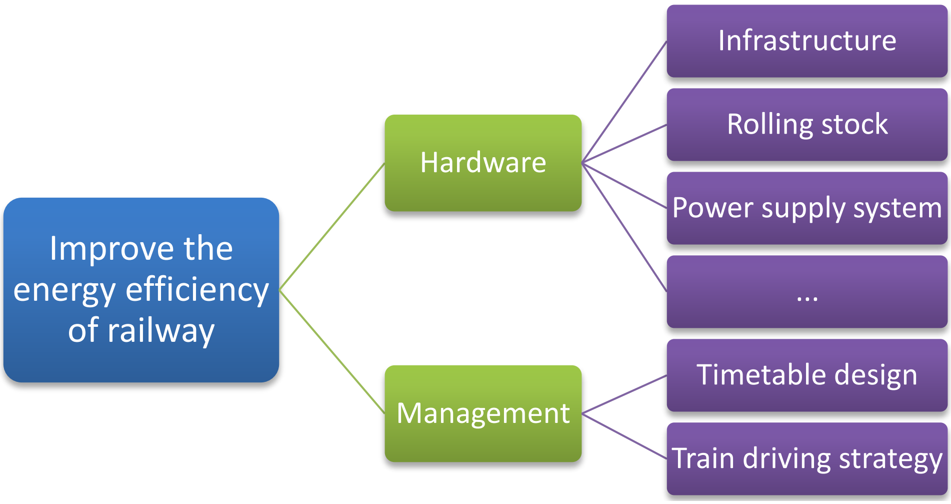


Figure 1-4 approaches classification in improving railway's energy efficiency

1.2.3. *Energy efficient driving*

Many years ago, it had been proved by both theoretical and practical sides that in a train run, energy consumption can be reduced if drivers use energy effieicnt driving strategies. These driving strategies will be further discussed in section 3.1.

Energy efficient driving can be implemented in practice in two ways. The first one is using some expensive equipment both in cab and control center to realize semi or totally ATO (Automatic Train Operation). In this case, a train may not even require a driver to control the train over OpenTrack sections from one station to another. With the high accuracy of train position and fast calculation speed of train's next strategy determination, a train could always drive in an optimal energy efficient way as the train is dynamically processing, and the driver only plays a role in the train's station dwell operation and safety. However, the drawback of this system is that the cost for introducing and maintaining this system is high.

The other implementation way is based on the condition that the trains are mainly controlled by drivers. In nowadays, it is the most common situation. A driver may obtain a training course to develop his skill in energy efficient driving, or a driving assistance system to obtain more real time information and driving advices.

1.2.4. *Problems in implementing energy efficient driving for drivers*

In the domain of implementing energy efficient driving strategies for drivers, the effect is difficult to be evaluated because there are a lot of dynamics in a driver's driving: a driver is a human being and the process of how a driver determines his way of driving is not shown as easy as machines.

Therefore, it is important to investigate train drivers' driving behavior and the relationship between driving behavior and train's energy consumption.

1.3 Study objectives

The objectives of this research are formulated as follows:

a) From a theoretical point of view, study the background theories of OpenTrack modelling and energy efficient driving.
b) Analyze realization data and check which energy efficient driving behavior is applied in reality.
c) Analyze the capabilities of OpenTrack to model this behavior and formulate suggestions for OpenTrack's future improvement.

1.4 Thesis layout

Based on this objective, this study can be divided into four parts as in Figure 1-5:

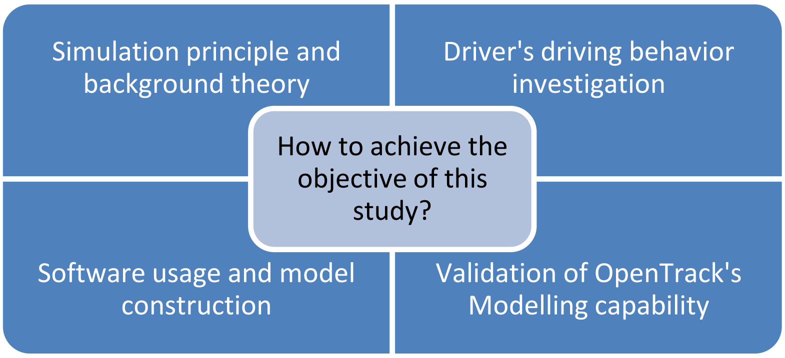


Figure 1-5 Research regime division

The first part mainly checks OpenTrack’s simulation principles and other theories, including running time calculation and energy efficient driving strategies. These two theories are introduced in chapter 2 and 3 separately;

The second part refers to the driver’s driving behavior investigation, which includes investigation method determination, data collection and processing procedures, that is, how to investigate the driver’s driving behavior in reality, what data is selected and why the selected data is needed for this study.

Driving behavior theory and a summary of the author’s study trip is presented in chapter 4, and driving behavior’s numerical investigation method is determined in chapter 5. Further explanations on how to process and filter selected realization data are given in chapter 6.

The third part focus on the OpenTrack model construction based on given information about the real world. As mentioned at the beginning of this chapter, simulation is the virtual representation of reality and therefore there is a selection on which factors have to be modelled and which will not. The process of this model construction is given in chapter 7;

In the last part, driving behavior in different aspects (acceleration, braking, coasting & relationship between a train’s initial delay and running time, etc.) are investigated in chapter 8 at first. After that, a comparison will be implemented between the constructed OpenTrack model and realization data, thus OpenTrack’s modelling ability for driving behavior is validated in chapter 9.

Conclusions of this study and further recommendations for OpenTrack’s future development are given in chapter 10.

2 Train dynamics and running time estimation

In order to investigate how the running time is calculated in OpenTrack software, it is essential to first study the theory of train motion dynamics and all the relevant factors that may influence a train's running time. In this chapter, firstly, all the efforts that train may take and their mathematical presentations will be discussed; after that, train's different motion behavior will be analyzed; finally, a running time deriving model si given in section 2.3. Some content of this chapter are based on the work of (Brunger and Dahlhaus 2008)

2.1 Train dynamics

2.1.1. *Tractive effort*

a) Induced tractive effort F_T [N]

Induced tractive effort F_T is defined as the effort that is intended to move the train, which is generated by the locomotive or the power equipment of the multiple units. Induced tractive effort is defined as the only internal effort from the train itself to control the train and it is related to the train's power P [W] and speed v [m/s] as shown in formula (2-1):

$$F_T = P / v \qquad (2\text{-}1)$$

b) Tractive effort limitations

In practical situations, not all amount of induced tractive effort can be transmitted to the train, mainly due to the following three conditions:

1) There will be a power efficiency factor η during the process of effort transmission within train's own mechanical structure. Usually this power efficiency factor η is estimated between 97% and 98%;

2) In order to prevent a train's engine from overheating, the maximum power is limited to a certain value;

3) Tractive effort should always be lower than an adhesion value, otherwise wheels will just do useless spinning.

Adhesion effort F_{adhes} [N] is determined by adhesion value μ and wheel load F_L [N], and it provides the maximum possible tractive effort without spinning wheels if there is no limitation for power or overheating reasons, it is calculated by formula (2-2):

$$F_{adhes} = F_L * \mu \qquad (2\text{-}2)$$

The adhesion value μ is influenced by many factors such as material characteristics of wheel and track, the weather and other reasons (falling leaf in autumn) and it is difficult to derive a reasonable and accurate value. However, according to (Brunger and Dahlhaus 2008), the adhesion value can be estimated by train's speed as shown in formula (2-3):

$$\mu(v) = 7.5/(3.6v + 44) + 0.161 \qquad (2\text{-}3)$$

One reminder here is the variable v in formula (2-3) only refers to the absolute numerical value of train speed (not considering train's running direction).

c) Effective traction effort $F_{Traction}$ and its presentations

From a) and b), the effective traction effort $F_{Traction}$ [N] can be presented as formula (2-4):

$$F_{Traction} = \min\{P * \eta / v, F_L * \mu\} \qquad (2\text{-}4)$$

However, for convenience, effective tractive effort could also be presented in hyperbolic or parabolic formulas as shown in formula (2-5):

$$\begin{aligned} F_{Traction} &= C_{0,k} + C_{1,k} * v + C_{2,k} * v^2, v_k \le v \le v_{k+1} \\ F_{Traction} &= C_{h,k} / v, v_k \le v \le v_{k+1} \end{aligned} \qquad (2\text{-}5)$$

Where $C_{0,k}$ [N], $C_{1,k}$ [Ns/m] and $C_{2,k}$ [Ns^2/m^2] are the coefficients for speed v with different power and v_k is the speed limit interval. Additionally, the relationship between effective tractive effort and train speed is exemplified by Figure 2-1:

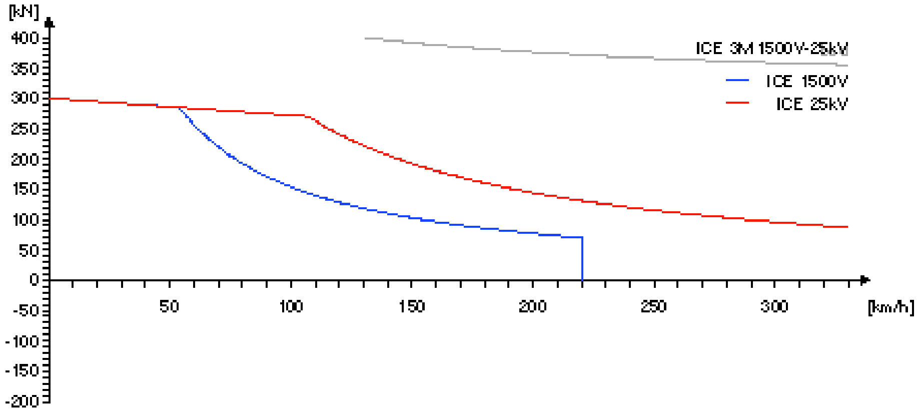


Figure 2-1 Relationship between tractive effort and speed

2.1.2. *Resistance effort*

Train resistance effort can be classified into two kinds: line resistance and train resistance. For line resistance, it refers to the resistance which is caused by gradient, curve or tunnel; for train resistance, it is a result of aerodynamic resistance and train's internal mechanism resistance. Train resistance can be further divided into traction unit resistance and wagon resistance (passenger wagons, freight wagons).

a) Line resistance F_{Rl} [N]
When a train is driving along the track, it may experience two kinds of resistances: gradient resistance and curve resistance.

Gradient resistance $F_{R\lg}$ [N] is given by formula (2-6) as follows:

$$F_{R\lg} = g * m * \sin\alpha \approx g * m * n \qquad (2\text{-}6)$$

Where n is the gradient factor defined as the rate of two points' vertical difference divided by their horizontal difference.

Curve resistance can be calculated by an empirical formula (Brunger & Dahlhaus, 2008), which is shown in formula (2-7)

$$F_{Rlc} = g * m * 700 / r \qquad (2\text{-}7)$$

Where r is the radius of the curve and 700 is an empirical coefficient. If r =700 m, curve resistance is the same as the gradient resistance for gradient n=1 in formula (2-6). Therefore, the influence of curve resistance is relatively small in overall resistance.

For tunnel resistance, no widely recognized formula is proposed to describe this resistance. Additionally, as this study is mainly focused on Dutch railway system and there are not so many tunnels in the Netherlands, no further discussion on this point is necessary.

b) Train resistance F_{TR} [N]
Train resistance may be caused by two reasons:
When train is moving, it will experience an air resistance;
Rolling resistance caused by train's internal mechanism interaction, such as wheel rims, axle-boxes, adhesion, etc.

Usually, train resistance is presented by parabolas as formula (2-8):

$$F_{TR} = r_0 + r_1 * v + r_2 * v^2 \qquad (2\text{-}8)$$

Where r_0 [N], r_1 [Ns/m] and r_2 [Ns^2/m^2] are the coefficient factors. In (Profillidis 2006), the term $r_0 + r_1 * v$ include various mechanical resistances, with r_0 only depends on rolling stock characteristics and represents the rolling resistances which are generated by friction between the wheel flange and rail on curves, $r_1 * v$ represents the various mechanical resistances which are proportional to speed such as rotation of axles and shafts, mechanical transmission, braking, etc. Term $r_2 * v^2$

represents the aerodynamic resistance with r_2 depend on the shape of the train and aerodynamic resistance generated along the surface.

In practical, there are several formulas can be used to calculate train resistance, and they are introduced as follows1

1) Train resistance in VPT formula
In the Netherlands, ProRail developed a system called VPT (Dutch name: Vervoer Per Trein) and formed a train resistance formula which is based on existing Dutch railway equipment (Agricola 2009). The VPT train resistance formula is given in formula (2-9):

$$F_{TR} = (A + N * B) * (V + \Delta V)^2 + M * (C + D * V) + N * E * (V + \Delta V) \tag{2-9}$$

Where		
	A	Rear-end aerodynamic coefficient[N·s^2/m^2];
	B	Length-dependent air resistance coefficient[N·s^2/m^2];
	C	Running resistance [N/kg];
	D	Speed dependent resistance [N·s/m·kg];
	E	Internal resistance [N·m/s];
	N	Number of wagons;
	M	Train mass in kilograms [kg];
	V	Train speed [m/s];
	ΔV	Wind speed in [m/s]

From formula (2-9), it is easy to recognize that the first part of the resistance represented by the quadratic velocity term and the second part describe the running resistance. The last term, the internal resistance, is often omitted by E=0. The speed-sensitive portion of the running resistance is also often omitted by D=0. For rear-end aerodynamic coefficient factor A, it is usually estimated under the condition of wind speed between 10 km/h and 15 km/h. Therefore, in reality, actual running time may even be smaller than calculated if the wind speed is smaller than 10 km/h. Formula (2-9) also proved that air resistance will play a major role in the overall resistances (Agricola 2009).

Comparing to parabolas format formula (2-8), the VPT the formula uses both speed of the train and wind together $(V + \Delta V)$ instead of just the speed of the train, which is a major characteristic of the VPT formula.

2) Train resistance in Dahlaus formula
In (Brunger and Dahlhaus 2008), a parabolas resistance formula specialized for traction unit (including multiple units) is given in formula (2-10):

$$F_{TR} = g * m_T * (a_0 + a_1 * v) + a_2 * v^2 + a_{2r} * v_r^2 \tag{2-10}$$

Where a_0, a_1 [s/m] and a_2 & a_{2r} [N·s^2/m^2] are the coefficient of the formula, m_T [kg] is the mass of the traction unit, and v_r is the relative speed between air and vehicle. In a situation of head wind equals to 15 km/h, it will have $v_r = v + 4.17 m/s$.

3) Train resistance in Sauthoff formula

The Sauthoff formula solves vehicle resistance calculation for passenger wagons, and its coefficients for formula (2-8) are determined by the mass of the vehicles, a factor relevant with the number of axles, number of vehicles and a value stands for the cross-sectional area of the vehicles weighted with their aerodynamic behavior. The Sauthoff formula is given in formula (2-11) (Brunger and Dahlhaus 2008):

$$F_{TR} = 1/1000 * m_w * g * (1.9 + c_b * 3.6 * v) + 0.0471 * (n_W + 2.7) * A_f * (3.6 * v_r)^2 \quad (2\text{-}11)$$

Where m_w mass of the vehicles [kg];

c_b Factor for number of axles;

n_W Number of vehicles

A_f Factor for the cross-sectional area of the vehicles weighted with their aerodynamic behavior;

v_r Relative speed between air and vehicle [m/s];

4) Train resistance in Strahl formula

For freight wagons, another formula called Strahl formula is used to estimate the vehicle resistance as formula (2-12):

$$F_{TR} = 1/1000 * m_w * g * (c_a + (0.007 + c_m) * (3.6 * v)^2 / 100) \quad (2\text{-}12)$$

Where m_w mass of the vehicles [kg];

c_a Factor for axle adhesion;

c_m Value for air resistance dependent on the kind of wagons;

5) Difference in using different Train resistance formulas

In (Agricola 2009), an experiment is implemented in OpenTrack to test different train resistance formulas' influences on some factors, i.e. running time, tractive effort. In this experiment, the ICM3 trainset is used as tested train in a rail section of 24 kilometres. However, as the train will reach the speed of 160 km/h after 10 kilometres, only train's dynamics in the first 10 kilometres are selected to plot the graph. Five different combinations of train resistance formula - Strahl and Dahlhaus, Strahl formula Filipovic, VPT, Strahl/Sauthoff formula for passenger wagons and Strahl for freight wagons are studied, the result is shown from Figure 2-2 to Figure 2-4 based on different axles as follows:

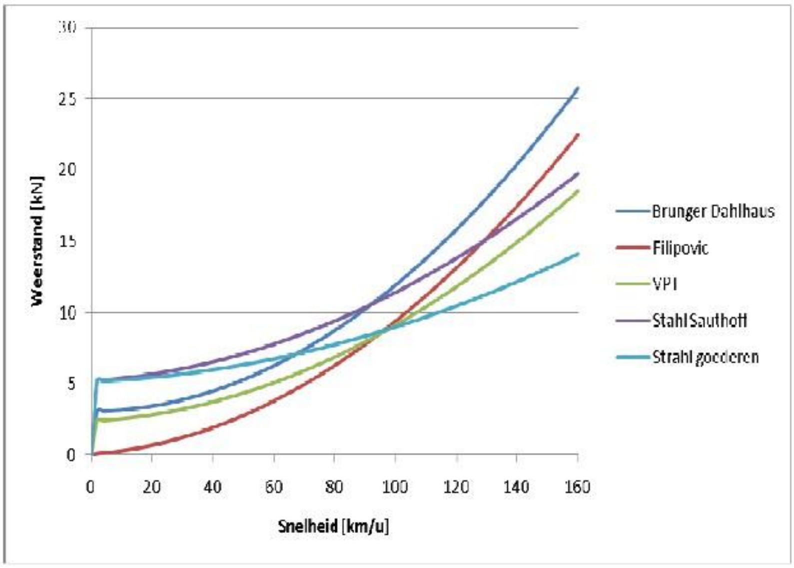


Figure 2-2 Relations between traction effort and speed using different formulas (Agricola, 2009)

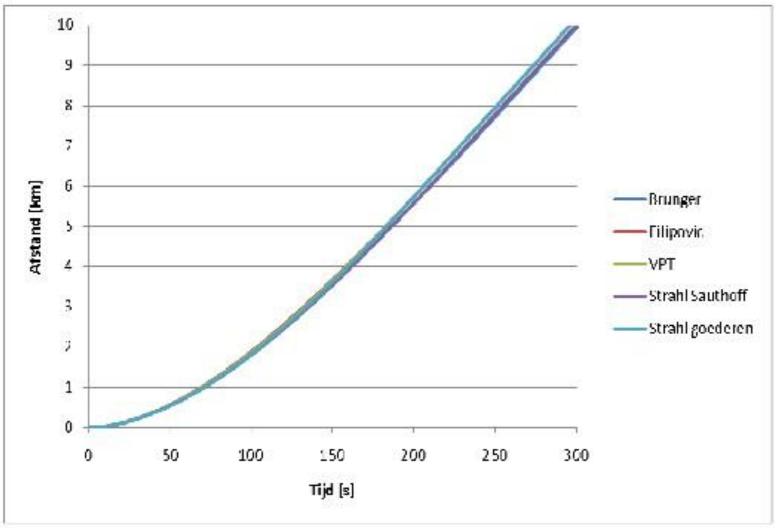


Figure 2-3 Relations between driving distance and time using different formulas (Agricola, 2009)

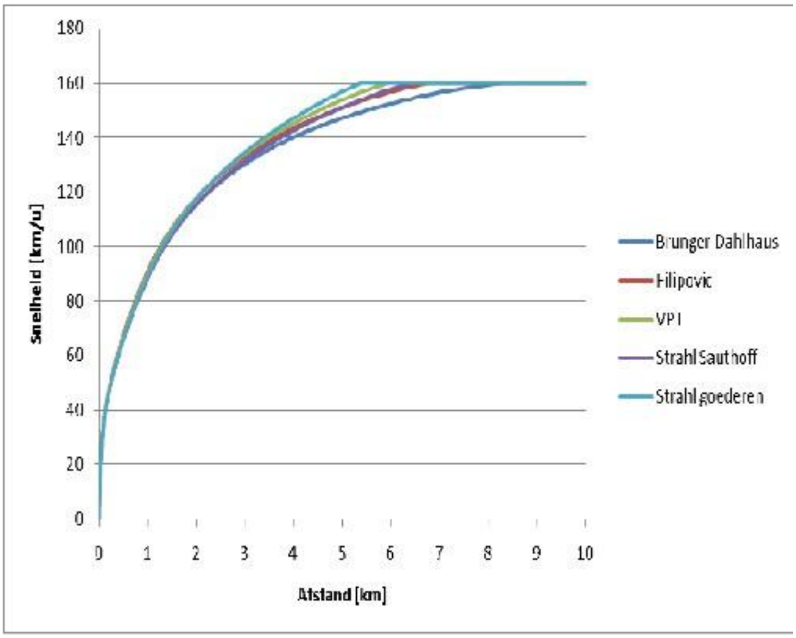


Figure 2-4 Relations between speed and distance using different formulas (Agricola, 2009)

The above three figures indicates that different usage of train resistance formula will show clear differences in tractive force, while minimal deviations in running time. This will lead to the result that by using different formulas, derived power usage will differ a lot while the running time will be almost the same.

c) Total resistance F_R [N]

Train's total resistance F_R is the sum of line resistance, train resistance and tunnel resistance, which is be presented in formula (2-13):

$$F_R = F_{Rlg} + F_{Rlc} + F_{TR} + F_{Tunnel} \qquad (2\text{-}13)$$

d) Rotating Masses

The rotating part of train will also consume some of the effort; therefore, a mass factor f_p is used to balance the difference of each part of the train. This value is calculated in formula (2-14):

$$f_p = (f_{pT} * m_T + f_{pW} * m_W)/(m_T + m_W) \qquad (2\text{-}14)$$

Where f_{pT} is a rotating factor for traction units;

f_{pW} is a rotating factor for passenger or freight wagons

e) Running time explanation

The final traction surplus is derived as in formula (2-15):

$$F_{Traction} - F_R = f_p * m * a = f_p * m * dv/dt \qquad (2\text{-}15)$$

Then train's running time is calculated as in formula (2-16) as following:

$$t = \int \frac{m * f_p}{F_{Traction}(v) - F_R(v)} dv \qquad (2\text{-}16)$$

However, formula (2-16) could not be directly used for running time in practical cases, because a train’s effective traction force and total resistance is depending on its speed and varies for every second.

Therefore, in OpenTrack, running time estimation is based on the Euler’s method. Euler’s method works by calculating the change in a variable from a given starting point and estimates each functional value using the preceding functional value, the preceding derivative of the function, and a fixed time step (Hürlimann 2010). This kind of stepwise computational method is widely accepted for train running time tools.

2.2 Train motion behavior

During a train run, only four kinds of possible basic motion behavior can be identified in practice: acceleration, cruising, coasting and braking, which are shown in Figure 2-5. All train’s motion behavior can be described by sequences of basic motion behavior.

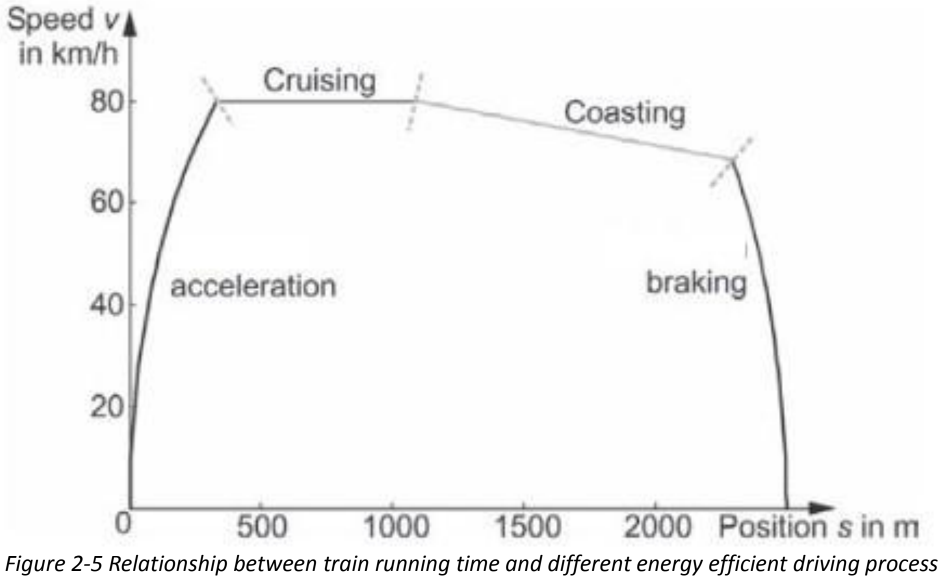


Figure 2-5 Relationship between train running time and different energy efficient driving process (Albrecht 2008)

For the acceleration phase, a train’s tractive effort should be higher than the resistance force. One reminder is that in a train’s acceleration process, its acceleration rate keeps changing instead of being stable, if the train is using constant engine power.

However, as the effective traction force is limited by several limitations, accelerate rate could not keep constant along train's acceleration phase. Therefore, this dynamics in train's speed and acceleration rate will bring more complex in calculating train's accelerating phase directly;

The cruising phase means that the train is running at a constant speed. In this case, $F_{Traction} = F_R$ and the running time of this phase t can be derived by simple Newton formula t=s/v, where s is the distance of cruising phase;

The coasting phase means that the train is running without any effort from the engine in a roll-out case where $F_{Traction}$ =0 [N]. In this case, the train will not consume any energy and its speed may increase or decrease depending on the gradient of the section and the value of train resistance;

The braking phase, nominally means that the train is reducing speed with traction force $F_{Traction}$ <0 [N]. However, from a practical usage point of view it can be classified as normal braking (comfortable braking) and emergency braking. Normal braking is used in normal situation, and its deceleration effort is smaller than that of the emergency braking in order to make passenger feel more comfortable, while emergency braking is used in unexpected situations and its objective is to make the train from moving to standstill as soon as possible.

2.3 A train's running time descriptive model

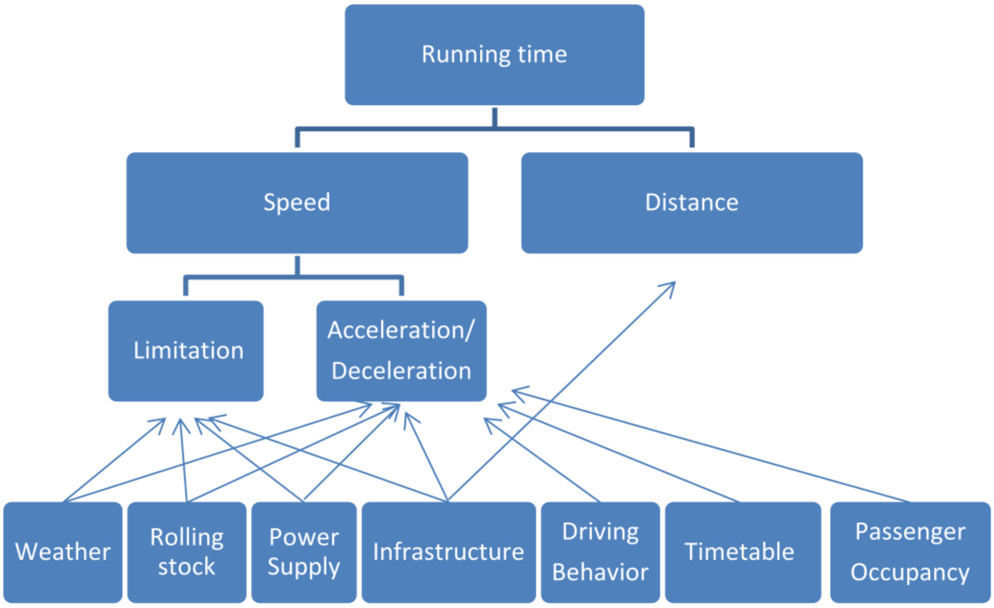


Figure 2-6 Descriptive model for a train's running time determination

Figure 2-6 illustrates the relationship between running time and actual factors. Basically, a train's running time between two points is only determined by two elements: the length of rail sections between these two points, and the train’s average speed. The length between two points is fixed, and is measured from infrastructure; while the speed may be always changing during a train running process.

The dynamics of speed can be presented by two elements as limitations and acceleration/deceleration rate usage. Limitations include maximum speed limitation and maximum acceleration/braking rates limitation, and this element is determined by infrastructure, rolling stock, power supply and weather.

A change in speed is realized by train’s acceleration or deceleration process. Within the limitations, a train’s acceleration/deceleration is determined by driving behavior, timetable, passenger occupancy (maybe slightly changes the weight of train set) and gradients/curves in infrastructure.

3 Energy efficient driving theory and development

3.1 Basic energy efficient driving theory

During a train run, if its energy regeneration in the braking phase is not considered, then the total amount of power consumption W [J] can be presented by formula (3-1) as follows:

$$W = \int F_T(t) * v(t) d(t) = \int P(t) * dt \qquad (3\text{-}1)$$

If the effect traction effort is always lower than adhesion effort F_{adhes} , then the relationship between induced engine power and effective traction effort is shown in formula (3-2):

$$F_{Traction}(t) = P(t) * \eta / v(t) \qquad (3\text{-}2)$$

Then the power consumption W is given in formula (3-3):

$$W = \int (F_{Traction}(t) / \eta) * v(t) d(t) \qquad (3\text{-}3)$$

By this step, the effective traction effort $F_{Traction}$ can be directly used to estimate train's power consumption in one formula.

If a train is running under following conditions:
a) Its traction force is between train's maximum positive traction force and braking force;
b) Its travelling speed is always lower than the speed limit of a section;
c) It starts its travelling from standstill and stops again at final destination;
e) In braking phase, the energy regeneration is not considered;

Then the Pontryagin's Maximum Principle can be used to solve the minimization problem of energy usage (Albrecht 2008) and the optimal train control strategies are four control strategies: maximum acceleration, cruising, coasting and maximum braking as shown in Figure 2-5 in chapter 2.

One reminder is the term maximum acceleration and maximum braking refers to maximum available and suitable acceleration and braking rate at that time instead of maximum capability of the train's engine. For example, for passenger trains, maximum braking strategy means maximum braking that the driver can use, and this braking process can still guarantee passenger's comfort. Therefore, this maximum braking is different from the maximum braking used in emergency braking.

3.2 Energy efficient driving in practical application

The result of optimal energy efficient driving strategies derived from the Pontryagin's Maximum Principle has been found for many years, and it is known by many drivers. Consider a simple case: if the speed limit for a section is constant and the train has enough time supplement, then it is easy for the train driver to do energy efficient driving as shown in Figure 2-5. However, if there is a low speed limitation area in between, more than one acceleration and braking driving regimes is needed. In this case, determining the switching point between different driving phases becomes difficult for drivers.

From a theoretical point of view, if there is no maximum time limitation, a train's running time can be very high if the driver decides to run at low speed. However, in reality, the train has to run according to the timetable in order to increase railway transportation's network capacity and service quality. Therefore, the timetable constrains a train's maximum running time. The relationship between train running time and different energy efficient driving process is illustrated in Figure 3-1:

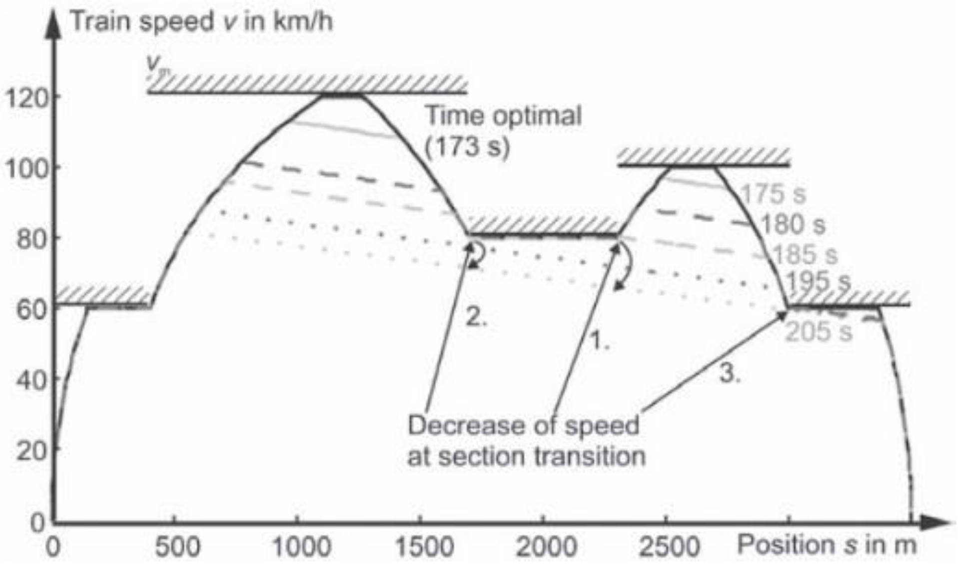


Figure 3-1 Relationship between train running time and different energy efficient driving process (Albrecht 2008)

Figure 3-1 shows if a train driver wants to save more energy, then he will spend more running time during his trip. A timetable includes two components: minimum running time and supplement. The potential of how much energy a train can save depends on how much time supplement is left for the rest of its trip. If a train is running in an undisturbed situation and does not have initial delay from its departure, then it will obtain the maximum ability to do energy efficient driving; if a train is experiencing a delay, then the first objective for the train is to recover this delay by using time supplement instead of doing energy efficient driving.

Additionally, Figure 3-1 indicates that if there are many speed limitation areas in a section, then more energy efficient driving regimes are required. In this case, driver has more than one option for implementing energy efficient driving strategies and it is difficult and confusing for a driver to determine an optimal comprehensive strategy process series. On the other side, a driver only wants to take minimum actions during his driving. A further description of this multi-sectional energy efficiency driving strategies determination problem given by (Albrecht, et al. 2010)

The objective of this problem can be defined as follows:
a) Find optimal energy efficient driving strategies for each sub section which is divided by different speed limitation;
b) Determine switching point for each different driving regime

The solution for this problem is based on subsection's length, infrastructure condition (speed limit), train motion characteristics, running time supplements, etc.

In the study of (Albrecht, et al. 2010), a constructive two-level algorithm based on a theory of optimal control and a gradient method is used to derive the energy-optimal regime sequence with the minimal number of regime changes. Finally, an advisory system called BEA is developed, which can be used to estimate energy consumption offline and perform driving simulation & derive onboard advices for real trains online.

In the study of (Chevrier, et al. 2011), a multi-objective evolutionary algorithm is used to search optimal solution for a train run. This algorithm contains three criteria: maximum reduction of travel duration, maximum reduction of delay and minimum energy consumption. Based on different weighting factors in the algorithm's objective function, various solutions can be generated and their differences for each criterion are exemplified in Table 3-1.

Table 3-1 Example of different solutions' performance for each criterion

Solution No.	Energy Consumption [KJ]	Deviation with S^1（%）	T [s]	Deviation with S^1（%）
S^1	507,618		484.13	
S^2	378,683	-24.50%	508.79	4.90%
S^3	246,245	-51.50%	555.63	14.70%

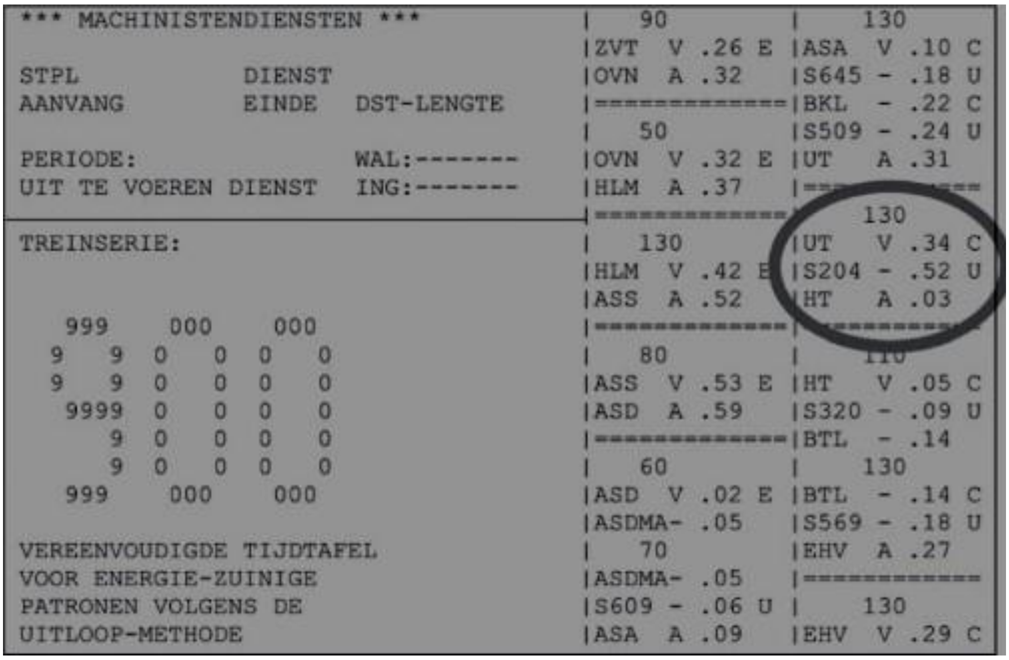
Table 3-1 contains three solutions: S^1, S^2 and S^3. S^1 has the higest weight value for minimum running time, S^3 has the highest value for minimum energy consumption and S^2 keeps a relative balance between above-mentioned criteria. Therefore, Table 3-1 illustrates that for a train run, energy consumption and running time keeps an inverse correlation.

3.3 Driver assistance system for energy efficient driving

Driver assistant systems for energy efficient driving have been introduced since at least 30 years ago and it can be classified into three categories as follows:

3.3.1. *Static systems*

Static systems provide a driver with offline information on the time and location to start coasting based on a given timetable. The advantage of this system is that it is easy for driver to follow, and the system itself is quite portable; the disadvantage is it could not provide an exact advice (i.e. exact time to start coasting) to a driver if his train is not running 100% in accordance with the timetable. One example of this kind of a system is illustrated in Figure 3-2:

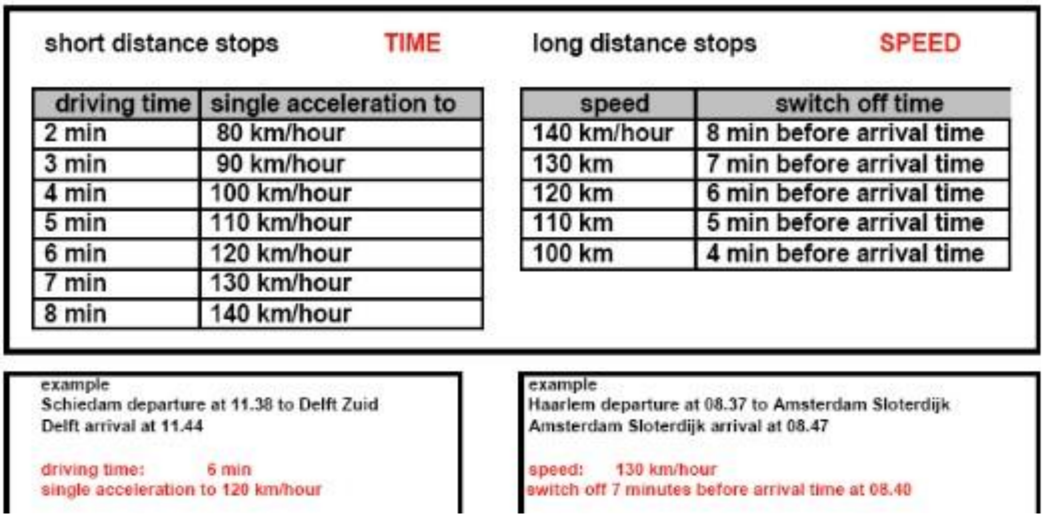


```
*** MACHINISTENDIENSTEN ***                 |    90          |     130
                                            |ZVT   V .26 E  |ASA   V .10 C
STPL            DIENST                      |OVN   A .32    |S645  - .18 U
AANVANG         EINDE    DST-LENGTE         |===============|BKL   - .22 C
                                            |    50          |S509  - .24 U
PERIODE:                 WAL:-------        |OVN   V .32 E  |UT    A .31
UIT TE VOEREN DIENST     ING:-------        |HLM   A .37    |=============
                                            |===============      130
TREINSERIE:                                 |    130         |UT    V .34 C
                                            |HLM   V .42 E  |S204  - .52 U
                                            |ASS   A .52    |HT    A .03
   999      000      000                    |===============|=============
 9    9   0    0   0    0                   |    80          |     110
 9    9   0    0   0    0                   |ASS   V .53 E  |HT    V .05 C
   9999   0    0   0    0                   |ASD   A .59    |S320  - .09 U
      9   0    0   0    0                   |===============|BTL   - .14
      9   0    0   0    0                   |    60          |     130
   999      000      000                    |ASD   V .02 E  |BTL   - .14 C
                                            |ASDMA- .05     |S569  - .18 U
VEREENVOUDIGDE TIJDTAFEL                    |    70          |EHV   A .27
VOOR ENERGIE-ZUINIGE                        |ASDMA- .05     |=============
PATRONEN VOLGENS DE                         |S609  - .06 U  |     130
UITLOOP-METHODE                             |ASA   A .09    |EHV   V .29 C
```

Figure 3-2 Dutch example of static coasting point indication system (Albrecht 2008)

The black ring area of Figure 3-2 indicates a case as following: a train departs from Utrecht at 34 minute past some hour value and passes the signal 204 at 52 minutes past the same hour (the hour values are not relevant in an hourly timetable). If this train starts coasting from the location of the signal 204, it can arrive at Den Bosch station at 3 minute past the next hour value. However, if a train arrives at signal 204 later than xx: 52, then static systems could not provide drivers with another accurate time to change its driving regimes.

In 2010, a static driver assistant system called UZI (Dutch: Universeel Zuinigrijden Idee) was applied in NS. The UZI system is developed by a train driver Freddy Velthuizen and is based on his driving experience of more than 25 years. The current version of UZI is the UZI Basic: a card that indicates a train driver when to start coasting based on scheduled running time between two stations. It assumes the running time supplement in the timetable is between 5%-7% of minimum running time. One advantage of UZI Basic is it can be applied for all kinds of train types in the Netherlands. An example of UZI Basic is shown in Figure 3-3:



short distance stops TIME

driving time	single acceleration to
2 min	80 km/hour
3 min	90 km/hour
4 min	100 km/hour
5 min	110 km/hour
6 min	120 km/hour
7 min	130 km/hour
8 min	140 km/hour

long distance stops SPEED

speed	switch off time
140 km/hour	8 min before arrival time
130 km	7 min before arrival time
120 km	6 min before arrival time
110 km	5 min before arrival time
100 km	4 min before arrival time

example
Schiedam departure at 11.38 to Delft Zuid
Delft arrival at 11.44

driving time: 6 min
single acceleration to 120 km/hour

example
Haarlem departure at 08.37 to Amsterdam Sloterdijk
Amsterdam Sloterdijk arrival at 08.47

speed: 130 km/hour
switch off 7 minutes before arrival time at 08.40

Figure 3-3 Example of UZI basic card

In the future, a second generation called UZI Pro will be developed, and this new UZI system will be in a software format installed in the rail pocket for train drivers. The UZI Pro can adjust its advices based on the train’s current delay, remaining time supplement and the speed limitation details for forwarding sections. However, it still cannot use train’s dynamics data, such as train acceleration rate/speed, etc.

3.3.2. *Individual train dynamic driver assistance systems*

A typical individual train dynamic driver assistance system should have the ability to collect the train’s status (location, speed, etc.) and calculate the best solution for energy efficient driving strategy dynamically, but it is still working based on a certain train and doesn’t consider this train’s interaction with the other trains. Therefore, it could not handle the situation if the timetable is changed due to route conflict.

In Australia, (Coleman, et al. 2010) developed an in cab driver advice system Freightmister to help drivers driving efficiently and stay on time. Their result shows that by using the Freightmister system, maximum 22% fuel saving can be achieved on a wide variety of railways including bulk grain trains, long-haul freight trains, high-speed intercity passenger trains and slow roll-on-roll-off trains.

Freightmister uses the Global Position System to collect train’s real time speed and location and comprises an onboard unit and a central server to prepare route data, such as gradient, curve, etc. The optimal driving sequence suggestions are based on remaining slack time and are updated when the train is running.

3.3.3. *Network-based train dynamic driver assistance systems*

For a network-based train dynamic driver assistance system, optimal energy efficient driving advice is based on all trains that are running in the network. In this situation, a train driver can always obtain optimal driving strategies even when there is a change in the timetable and keep maximum performance for the whole network.

In (Albrecht and Oettich 2002), an example is given as follows: a train A is running from station M to station N and train B is running from station P to N, passengers from train A will transfer to train B in station N. If train A experiences a delay of two minutes, then train B will obtain additional two minutes of slack time while still guarantee this connection. In this example, if train B could be equipped with a driver assistance system based on individual train, then train B would arrive at station N on scheduled time and these two minutes can only be used as dwelling time. However, if the driver assistant system for train B is based on the network, then train B can use these two minutes as time supplement for energy efficient driving and still keep the connection. From this example, the advantage of using the driver assistance system based on network is shown as follows: compared to the driver assistance systems that based on a individual train, using network-based systems can derive a better energy saving result and keep the same capacity level for a train network.

Another advantage of a network-based system is that it can detect a trains' coming route conflict in advanced and prevent unscheduled stops. This feature is shown in Figure 3-4:

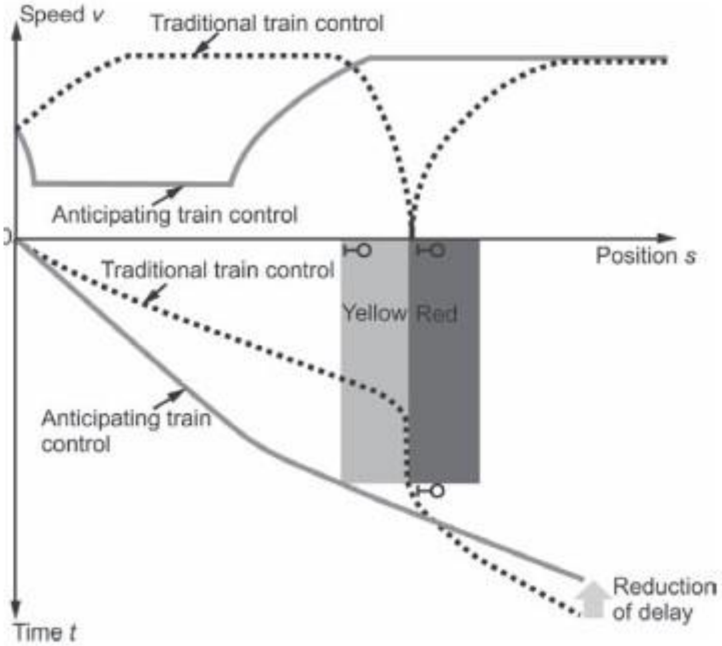


Figure 3-4 Train's optimal driving strategy in a disturbed situation (Albrecht 2008)

Figure 3-4 illustrates a situation as follows: when a train is running on an OpenTrack, a signal in front of the train turns to red due to route conflict, infrastructure defect, etc. In this situation, network-based driving assistance system can update its suggestions for drivers, which is to lower train's running speed and try to make the train pass this signal when the signal turns back to green again. In this way, a train can avoid driving sequences of keeping high speed – breaking- acceleration and save energy by just lowering down train's speed at the beginning.

4 Driving behavior introduction and study trip summary

4.1 Driving behavior model and its influence on running time

Assume a case that a train is running from one station to another station within the time given by the timetable, the rail sections between these two stations are not occupied, and the route has been set by operators. In this situation, a drivier's primary task is to drive the train and keeps the train's punctuality.

In a single train's operation level, the train control loop described by (Albrecht 2008) is given in Figure 4-1:

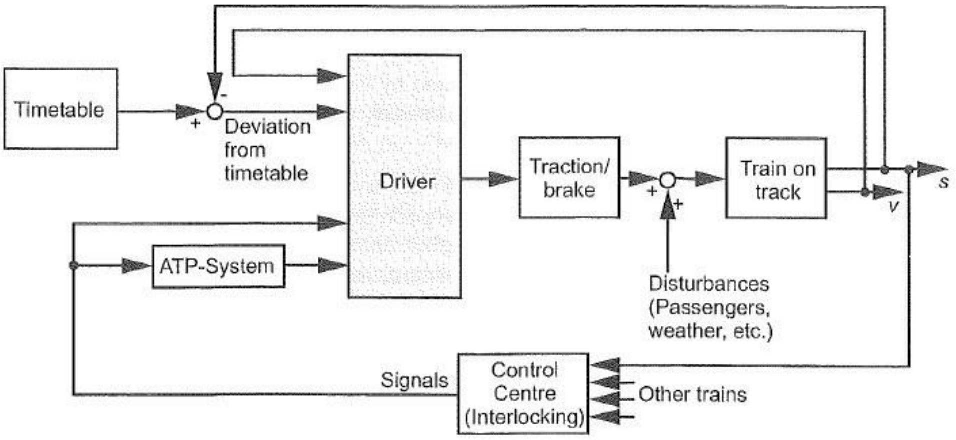


Figure 4-1 Understanding train control as a control loop)(Albrecht, Energy efficient train operation 2008)

A driver's influence on a train can be summarized as choosing the speed profile. By choosing different values of traction and braking forces, a train driver obtains the ability to change a train's running time. For example, if a train has a initial delay and the time supplement is not sufficient, then the drivier will run the train as quick as possible; if a train doe not have initial delay and the time supplement between two stations is sufficient, then drivers do not need to do "as fast as possible" driving. In the later situation, different driving styles may derive different running times and energy consumption results, that is also the reason why the driving behavior is studied by researchers.

4.2 Study trip with drivers

In order to gain some practical experiences on how a driver controls the train and reacts to the signal system, a study trip is implemented on 17-01-2010.

The whole study trip included one train run from Utrecht to Leiden; several train runs from Amsterdam to Rotterdam, Rotterdam to Amsterdam, Amsterdam to Leiden and the last one train run from Leiden back to Utrecht. Findings of this study trip are listed as follows:

4.2.1. Weather conditions

In the author's imagination before this trip, snow and rainwater are both important factors for train's resistance value. However, drivers proved that this perception is not true, at least they can hardly feel the difference between the tracks that under dry or wet condition. Instead, the falling leaves on the track will strongly influence train's track resistance and adhesion value, and this case has become a problem for driver's driving job in autumn.

4.2.2. Sight distance

There is no obvious difference between a driver's sight distance in the daytime and evening. For driver's sight distance value, the most important influencing factor is the road light that near the track. When a yellow signal is in a distance from a train driver and among a lot of road light (yellow), a driver may have difficulties in recognizing signals. However, as most of the train drivers are required to have a good "route knowledge" of the infrastructure, the driver's sight distance can be assumed as a constant value.

4.2.3. Cab signalling system

In driver's driving process, a situation sometimes happens that the speed limit indicated between the cab signal and trackside signal is different and this situation may confuse drivers. However, according to the driving regulations, trackside signal has a higher priority, so drivers all have a clear mind about which indications should be followed.

4.2.4. Passenger occupancy

In (Agricola 2009), he pointed out that the train drivers could feel the occupancy of passengers when they use traction forces and drivers will consider this passenger occupancy when they do coasting. However, during this study trip, all drivers only recognized the influence of the passenger occupancy to train resistance force (they may use more acceleration/braking force), but they did not think this factor is important enough to influence their decision to do coasting.

4.2.5. Acceleration and deceleration operation

In driver's driving control panel, deceleration force is selected among several steps. For each train type, the number of steps is different and the train drivers only need to select one of the steps. All drivers in this trip believed that because the braking rate for each step is obviously different from each other, there would be no difference in different driver's braking step selection result under the same circumstance (same train type, timetable, etc.), therefore, all driver's braking bahaviors should be the same.

For the acceleration phase, more variations in acceleration force value would be observed, because the acceleration rate is not only determined by driver's step choice but also by the computer adjustment in the locomotive. However, because the maximum exit speed for a station is always 40 km/h, the acceleration time needed for a train to accelerate from 0 km/h to 40 km/h between different drivers would also not differ so much under the same circumstance.

4.2.6. Buffer time between stations

Drivers do not know exactly how much buffer time is avaliable in the timetable. Therefore, the driver's perception about buffer time is only coming from their experiences.

For most of the stations in the Netherlands, planned dwell time is more or less equals to the dwell time that is needed for normal cases in reality. However, during the study trip, one exception is found in the train station of Schiphol airport. From the timetable, dwell time for this station is only one minute, which is obviously not enough for passengers to get on and get off the train. However, drivers have found that the section after Schiphol has more buffer time and this buffer time can be used to compensate train's delay in Schiphol train station.

The above case illustrates a phenomenon that in the Netherlands, buffer time may not be distributed equally along a route. Railway operation planners may assign more buffer time for some sections while less for the others. Then for the driver, he has to run fast in the sections with less buffer time and run slowly in the other sections in order to keep the accordance with timetable, and this driving style is not as energy efficient as in the situation that driver is driving in a route with buffer time equally distributed.

4.2.7. Uzi method applications

The Uzi basic card has been invented in 2009 and most drivers know the principles of this driving assistance system. Most of the drivers implied that the biggest limitation of the Uzi basic card is that it still lacks the possibility of providing quick and exact advices in a complex rail section.

According to the Uzi basic method, the suggested running speed is related with the planned running time, and this running time is calculated from the train's departure time in last station to this train's arrival time in next station. However, if there is a speed limitation area between stations, then it is easy to understand that train driver cannot totally follow suggestions from Uzi. Therefore, the application of the Uzi basic card is constrained by the complexity of railway infrastructure in reality. That is also the reason why the objective of Uzi Pro is to provide a more automatic and easy suggestion to drivers.

The findings above give an insight on the difference between theoretical study and practical operation for train driving. However, these findings could not be used in the numerical analysis of driving behavior directly. Therefore, a numerical driving behavior investigation is required for this study.

5 Driving behavior investigation method determination

In order to subtract train driver's driving behavior characteristics, the train speed profile is widely used as the study resources for its direct display and accurate reflection of the driver's operation. Constructing a train speed profile requires at least two elements: train's speed and its corresponding location. Empirically, there are two approaches by using different data resources: train describer and GPS system.

5.1 Deriving a train's speed and location based on train decriber

5.1.1. Train describer introduction and its application in the Netherlands

A train describer is a system that identifies a train at a particular position and keeps track of its progress from berth to berth (Goverde, 2000). The train describer is firstly used to support the investigation of incidents and support daily traffic management, but later its potential in tracking train's time and location is acknowledged by researchers.

In the Netherlands, the train describer is initially implemented as the TNV system. In this system, a unique train description (or a train number) is used to identify the train line service such as train's type, terminal stations, route, etc. The location of a train number is recorded by TNV-positions, which always includes more than one adjacent sections between two signals. The movement of a train is detected by track circuits. When a train enters the first associated track section, then a TNV-step of the train number to the next TNV –position is triggered, and through this way, the whole route and running time for a train can be derived. More details of the TNV system can be found in (Goverde, 2000).

Since last year, a new generation of train describer systems is implemented in Dutch railway, and the new system is called the TROTS system. Generally, the TROTS system holds the same principles of the TNV system, but TROTS log files contain more information, for example, each single section's occupancy time and clearance time are recorded and have already been coupled to train numbers (while TNV already provided this information, but not these information are not directly coupled to train numbers).

5.1.2. Literature survey in using train describer system

In (Goverde and Hansen 2000), the software TNV-Prepare is developed to analyze train's running time and other attributes. TNV-prepare traces events on a route of a train line, including TNV-steps, section entries, section clearances, signals, and point switches from the TNV log files, then indicates these events on tables.

(Goverde, et al. 2008) developed the software TNV-conflict, which can be considered as the successor of TNV-prepare. TNV-conflict can detect the train run which is hindered by other trains, therefore the unhindered train runs can be selected and used to do further studies on train's free running time distribution in reality.

In the study of (Albrecht, et al. 2006), the TNV log files are used to reconstruct train's speed profile to assess the effect of the Driver Information System (DIS). Although some deviations are found in the recording and processing procedure of train describer log files, this study proved that it is feasible to use train describer data to build train's speed profile.

5.1.3. *Empirical data processing methods for train describer log files*

a) Time deviations in the train describer system

In Dutch train describer systems, the precision of original recorded time is higher than one second (although these recorded times may be internally rounded off for practical reasons), but these time records only refers to the time recorded in train describer system, instead of the time when the actual event happened. Therefore, there are still time deviations between train's actual events and train describer data.

These deviations are usually assumed less than one second, but in some cases, it could still exceed one second. For example, when a train runs from one traffic control area to another one (in the Netherlands, there are 16 traffic control areas and each has a separate TNV-system (Goverde, 2000)), the deviation in the border sections of two traffic control areas may reach several seconds (according to the PAB office[1] of ProRail).

If the time deviation for each section stays the same in a specific train describer logfile, then the influence of this time deviation on a train's speed estimation will depend on the length of the section. Consider a train that runs through a section of 40 meters, the actual running time is 4 seconds and maximum time deviation for this area is 1 second. In this case, the actual train's average speed for this section is 10m/s, while in the train describer data, the minimum value becomes 40/(4+1)=8m/s, and thus creates a maximum relative deviation value of -20% comparing to train's actual value. But if the train still keeps this actual speed in a section of 400 meters and maximum time deviation for train describer system is still 1 second, then maximum relative deviation value becomes -2.4%. Nevertheless, this example clearly shows that the time deviations in train describer system can lead to a different speed estimation result even for the train runs with same train type and timetable, and this difference should not be neglected.

For this problem, Goverde (2000) developed a least square method. For a route of n sections, if train's entrance speed and exit speed for this route and the length of each section in this route is known, a minimum square function is constructed, based on train's practical limitations in maximum acceleration, deceleration, speed and time deviations (or some of these terms).

---

[1] The PAB office (Dutch name: Prestatie Analyze Bureau) belongs to one Prorail's department called Prorail Verkeersleiding, and its main task is to analyze the Dutch railway's operational performance by using the actual data, i.e. train describer log files.

There are three advantages in this method. Firstly, it is convenient to select proper terms in the object function; secondly, different weights can be assigned for each term if a user has a different data uncertainty for different terms; thirdly, this constrained nonlinear least square optimization can be solved by mathematical softwares, i.e. Matlab.

However, it should be noted that the difficulty of solving this method increases as the number of considered sections or terms increases. If a constraint for some terms is too tight, the constraint system may even become infeasible and no solution can be found. From this point, this problem becomes more like a computational problem.

In 2010, Albrecht, et al. (2010) proposed a method to speed up the procedure of solving the nonlinear least square function by dividing the whole route into several sub section sequences. These sub section sequences will be firstly considered separately by using least square method, and then the result for each sub section sequence can be assembled together. In the end, he concluded that setting twenty as the number of sections in a sub section sequence and five as the number of the overlap sections between two neighbour sub section sequences will give satisfying results.

b) Acceleration/ braking time estimations in the train describer system

In a train describer system, for either TNV positions in the TNV system or sections in the TROTS system, only train's entrance and clearance times will be recorded. Therefore, in the stations platforms, there will be no time records for a train to do braking until standstill or accelerate from standstill, and the estimation of train's acceleration/ braking time is essential for constructing train speed profiles.

Goverde (2000) introduced a method to solve this problem by comparing two distance factors. For braking phase, assume the distance between the halting point and the entrance point for the last section is S, train's initial speed and braking rate for the last section is v_0 and a_0, respectively after completing the least squares method. If the train keeps the braking rate as a_0 until it comes to standstill, then the required braking distance is given in formula (5-1):

$$l_0 = \frac{2 * v_0^2}{|a_0|} \tag{5-1}$$

If this train's maximum braking rate is $a_{\max}$, then the minimum distance $l_{\min}$ required for the train to brake from v_0 to 0m/s is given in formula (5-2):

$$l_{\min} = \frac{2 * v_0^2}{a_{\max}} \tag{5-2}$$

In Goverde's method, if l_0 < S, then the train will be assumed to do cruising at speed of v_0, then do braking at maximum braking rate $a_{\max}$; if $l_0 \geq S$, then train is assumed firstly do braking at braking rate a_0, then change to $a_{\max}$ until it comes to standstill.

For the acceleration phase, a similar approach holds. If the distance between train's halting point and the entrance point of next section is $S^{'}$, train's acceleration rate for the next section is $a_0^{'}$ and corresponding entrance speed is $v_0^{'}$. Thus the distance $l_0^{'}$ needed for this train to accelerate from 0 to $v_0^{'}$ is given by formula (5-3):

$$l_0^{'} = \frac{2 * v_0^{'2}}{a_0^{'}} \tag{5-3}$$

If train's maximum acceleration rate is $a_{\max}^{'}$, then train's minimum distance to accelerate from 0 to $v_0^{'}$ is given by formula (5-4):

$$l_{\min}^{'} = \frac{2 * v_0^{'2}}{a_{\max}^{'}} \tag{5-4}$$

Therefore, if $l_0^{'} \leq S^{'}$, then train is assumed firstly accelerate with $a_0^{'}$, then keep speed $v_0^{'}$ till it reaches the entrance point for the next section; if $l_0^{'} > S^{'}$, then train will firstly use the maximum acceleration rate $a_{\max}^{'}$, then switch to $a_0^{'}$ and enter the next section with speed $v_0^{'}$.

More details about this determination method can be found in (Goverde, 2000).

5.2 Deriving a train's speed and location by using GPS

The other method to track train's position is to use the GPS data (Global Position System). In study of (Medeossi, et al. 2011), GPS equipment are assigned on the train to record train's position, and these recorded data are used to reconstruct train's trajectories.

The most complicated part in dealing with GPS data is the data filtering procedure. In (Medeossi, et al. 2011), the Kalman filter is used to handle this issue. Their result proved the accuracy of GPS data is promising, but additional data, i.e. signal aspect data has to be added if the research goal is to study driving behavior, because if a train has an unscheduled stop in its journey, then this train run should not be sorted in the same group as the other free running trains.

In (Albrecht, et al. 2010) GPS is also used in verifying the speed profiles derived from the track occupation data. However, in this study, the GPS measurements are only taken in some fixed locations and the authors did not use these GPS measurements to build a speed profile totally based on the GPS measurements.

5.3 Driving behavior investigation data for this study

Due to the availably of data resources, the Trento data in version 1.3 is selected to construct train's trajectory and speed profile. Trento is a program that is developed by the PAB office and can be used to substract information from the TROTS log files/output them in csv files, which can be used in Excel directly. In this way, the Trento data can also be considered as a source based on train describer log files.

Further explanation on the Trento data selections and processing procedures will be given in chapter 6.

6 Trento and other data preparation and processing

In order to investigate driving behavior in a real situation, an investigation that is based on Trento data (Version 1.3) is implemented to study train's actual speed profiles. In this study, the line section between Geldermalsen (Gdm) and Den Bosch (Ht) is studied. Because firstly, this section is a busy line in the Netherlands, thus enough samples can be collected; secondly, investigaing this line section is feasible in the author's master project period; thirdly, it is more probable for a driver to do energy efficient driving in this line section than other busier line sections, i.e. the line section between Utrecht to Geldermalsen.

6.1 Trento data and OBE&OS drawings

From the Trento software, there are two data recording categories. The first category is based on section occupancy, i.e. section's occupancy time, exit time, etc; while the section category is based on the signal aspect, for example, signal's aspect changing status, etc. These two categories will be introduced in 6.1.1 and 6.1.2 separately.

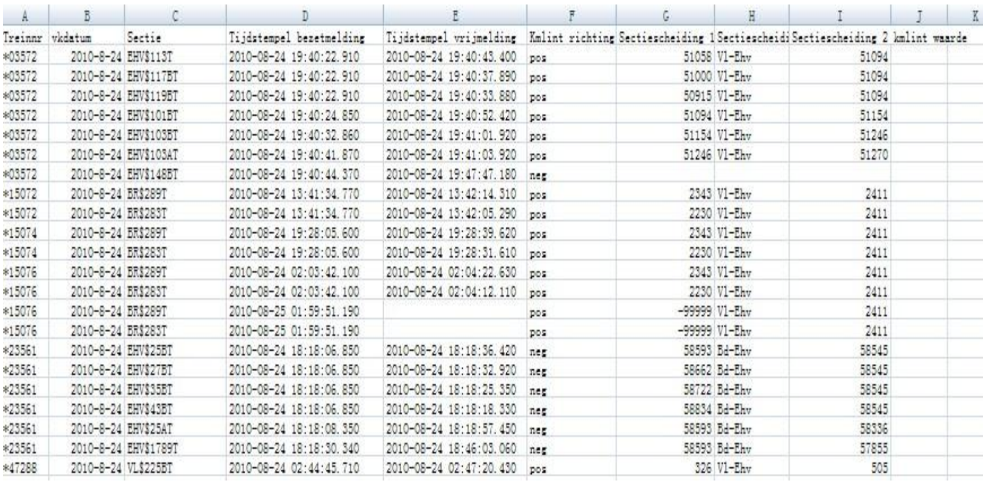
One reminder is that all the Trento data used in this study are baesd on Trento software version 1.3. In this chapter, the stuided Trento data are derived from the periods of 23-08-2010 to 27-08-2010 and 04-10-2010 till 08-10-2010 (but the Trento data for each day maybe not totally complete, which will be discuessed in following sections).

Additionally, the OBE drawings and the OS drawings are also essential for this study. Introductions on these two drawings will be given in section 6.1.3 and 6.1.4 respectively.

6.1.1. *Section occupancy data*

The track occupancy data is derived from the Trento software and presented in a csv sheet. In this data sheet, records are based on the unit of one section. Therefore, this data holds more than 80% percent of the total size of Trento data. (Because it is easy to understand that the number of signals is much less than the number of sections.)

Then an example of initial section occupancy data is shown in Figure 6-1:



A	B	C	D	E	F	G	H	I	J	K
Treinnr	vkdatum	Sectie	Tijdstempel bezetmelding	Tijdstempel vrijmelding	Kmlint richting	Sectiescheiding 1	Sectiescheidi	Sectiescheiding 2	kmlint waarde	
*03572	2010-8-24	EHV$113T	2010-08-24 19:40:22.910	2010-08-24 19:40:43.400	pos	51058	V1-Ehv	51094		
*03572	2010-8-24	EHV$117BT	2010-08-24 19:40:22.910	2010-08-24 19:40:37.890	pos	51000	V1-Ehv	51094		
*03572	2010-8-24	EHV$119BT	2010-08-24 19:40:22.910	2010-08-24 19:40:33.880	pos	50915	V1-Ehv	51094		
*03572	2010-8-24	EHV$101BT	2010-08-24 19:40:24.850	2010-08-24 19:40:52.420	pos	51094	V1-Ehv	51154		
*03572	2010-8-24	EHV$103BT	2010-08-24 19:40:32.860	2010-08-24 19:41:01.920	pos	51154	V1-Ehv	51246		
*03572	2010-8-24	EHV$103AT	2010-08-24 19:40:41.870	2010-08-24 19:41:03.920	pos	51246	V1-Ehv	51270		
*03572	2010-8-24	EHV$148BT	2010-08-24 19:40:44.370	2010-08-24 19:47:47.180	neg					
*15072	2010-8-24	BR$289T	2010-08-24 13:41:34.770	2010-08-24 13:42:14.310	pos	2343	V1-Ehv	2411		
*15072	2010-8-24	BR$283T	2010-08-24 13:41:34.770	2010-08-24 13:42:05.290	pos	2230	V1-Ehv	2411		
*15074	2010-8-24	BR$289T	2010-08-24 19:28:05.600	2010-08-24 19:28:39.620	pos	2343	V1-Ehv	2411		
*15074	2010-8-24	BR$283T	2010-08-24 19:28:05.600	2010-08-24 19:28:31.610	pos	2230	V1-Ehv	2411		
*15076	2010-8-24	BR$289T	2010-08-24 02:03:42.100	2010-08-24 02:04:22.630	pos	2343	V1-Ehv	2411		
*15076	2010-8-24	BR$283T	2010-08-24 02:03:42.100	2010-08-24 02:04:12.110	pos	2230	V1-Ehv	2411		
*15076	2010-8-24	BR$289T	2010-08-25 01:59:51.190		pos	-99999	V1-Ehv	2411		
*15076	2010-8-24	BR$283T	2010-08-25 01:59:51.190		pos	-99999	V1-Ehv	2411		
*23561	2010-8-24	EHV$25BT	2010-08-24 18:18:06.850	2010-08-24 18:18:36.420	neg	58593	Bd-Ehv	58545		
*23561	2010-8-24	EHV$27BT	2010-08-24 18:18:06.850	2010-08-24 18:18:32.920	neg	58662	Bd-Ehv	58545		
*23561	2010-8-24	EHV$35BT	2010-08-24 18:18:06.850	2010-08-24 18:18:25.350	neg	58722	Bd-Ehv	58545		
*23561	2010-8-24	EHV$43BT	2010-08-24 18:18:06.850	2010-08-24 18:18:18.330	neg	58834	Bd-Ehv	58545		
*23561	2010-8-24	EHV$25AT	2010-08-24 18:18:08.350	2010-08-24 18:18:57.450	neg	58593	Bd-Ehv	58336		
*23561	2010-8-24	EHV$1789T	2010-08-24 18:18:30.340	2010-08-24 18:46:03.060	neg	58593	Bd-Ehv	57855		
*47288	2010-8-24	VL$225BT	2010-08-24 02:44:45.710	2010-08-24 02:47:20.430	pos	326	V1-Ehv	505		

Figure 6-1 example of initial section occupancy data on 24-08-2010

The Initial section occupancy sheet has the following attributes:

a) Train number

In the Trento data, all the section's entrance and clearance times are grouped according to the train numbers.

b) Date

One reminder concerning the date attribute is that if a train starts its running schedule on a certain day but arrives at its destination on the next day, then its section occupation data for the next day are still stored with its first day's together, in order to keep the completeness of a whole train run.

c) Section name

The format of a track section name data is "station code" + code "$" + "section code"". The station codes and section codes are in accordance with the topological infrastructure databases.

d) Section entrance time

Section entrance time refers to the moment of the event that the first axle of a train enters the section. However, in Trento data, the section entrance time should be considered as the final recording event time instead of the acutual event time. Although the value of this time has a precision of 0.001 second in Trento data (not in TROTS data), there are still significant deviations between the final recording event time and the actual event time. Further discussions on this issue will be given in section 6.5.

e) Section exit time

Section exit time refers to the moment that the last axle of a train leaves the section, and this time value has the same characteristics as the section entrance time that inroduced in d).

f) Train running direction indication

The value for this attribute is binary: “pos” (positive) or “neg” (negative), which depends on the kilometration value for a train’s entrance /exit point of a section. If a train’s entrance point has a lower kilometration point than the exit point, then this train is considered as running in a positive direction.

g) Kilometration for train’s entrance point and exit point

In the Trento data, different kilometration standards are implemented for different railway lines, such as from Dordrecht to Eindhoven, Venlo to Eindhoven, etc. For a section that belongs to a big line, a unique kilometration number (in meters) is as given for its boundaries. If a train’s direction is indicated, then boundaries will become an entrance and an exit point for a train that runs through this section.

i) Line section name

In the Trento data, all sections’ kilometrations connected to a line section name, i.e. Venlo to Eindhoven, Dordrecht to Eindhoven, etc. However, this line section name indications does not really mean the section will be used for the route in the big line. For example, some sections in Utrecht station are indicated with the line from Utrecht to Zwolle, but these sections are in fact used for the trains that run from Utrecht to other cities.

j) Previous and next section name

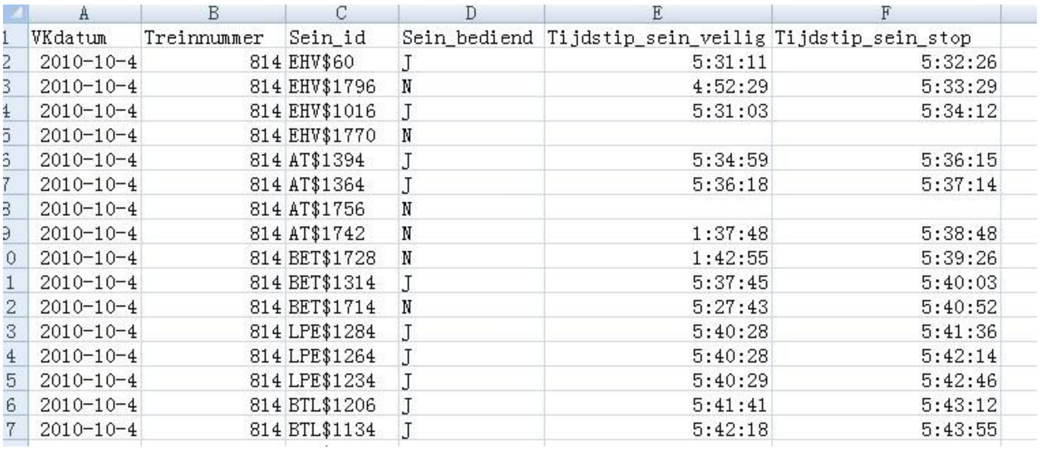
When a track occupation is recorded, its previous and next section are also given in the same row, thus the train’s route usage can be traced in the Trento data.

Although in the Trento data, the length of sections, train’s running times and average speeds are also given, their accuracy and reliability is unknown. Furthermore, these values can also be calculated by the existing data. Therefore, a decision has been made to delete the values that are generated by the Trento software.

For one day, there are more than 200,000 rows of section records for more than 16 big lines in the Netherlands.

6.1.2. Signal aspect data

The signal aspect data is another part of the data that derived from the Trento software, and it is presented in a csv sheet. In this data sheet, each record provides information of a signal event. An example of the signal aspect data is shown in Figure 6-2:



	A	B	C	D	E	F
1	VKdatum	Treinnummer	Sein_id	Sein_bediend	Tijdstip_sein_veilig	Tijdstip_sein_stop
2	2010-10-4	814	EHV$60	J	5:31:11	5:32:26
3	2010-10-4	814	EHV$1796	N	4:52:29	5:33:29
4	2010-10-4	814	EHV$1016	J	5:31:03	5:34:12
5	2010-10-4	814	EHV$1770	N		
6	2010-10-4	814	AT$1394	J	5:34:59	5:36:15
7	2010-10-4	814	AT$1364	J	5:36:18	5:37:14
8	2010-10-4	814	AT$1756	N		
9	2010-10-4	814	AT$1742	N	1:37:48	5:38:48
0	2010-10-4	814	BET$1728	N	1:42:55	5:39:26
1	2010-10-4	814	BET$1314	J	5:37:45	5:40:03
2	2010-10-4	814	BET$1714	N	5:27:43	5:40:52
3	2010-10-4	814	LPE$1284	J	5:40:28	5:41:36
4	2010-10-4	814	LPE$1264	J	5:40:28	5:42:14
5	2010-10-4	814	LPE$1234	J	5:40:29	5:42:46
6	2010-10-4	814	BTL$1206	J	5:41:41	5:43:12
7	2010-10-4	814	BTL$1134	J	5:42:18	5:43:55

Figure 6-2 example of signal aspect data on 04-10-2010

There are six attributes for each record (row). The date and train number attributes are similar as the track occupancy data, whereas there are three different attributes as follows:

a) Signal code

For each signal, a unique signal code (ID) is assigned to represent the signal. This code is always in a format of “station code” + code “$” +signal number”

b) Signal type

This attribute indicates where the signal is a manually controlled signal or an automatically controlled signal. Further differences between these two types will be discussed in next sections.

c) Time for the signal to show “proceed”

This attribute is originally called “Tijdstip sein veilig” in Dutch, but it can be understood as the time when a signal shows a “proceed aspect”. In the Netherlands, this proceed aspect can be regarded as the signal aspect which is better than red.

d) Time for the signal to show “stop”

This attributes indicates the time that a red aspect is shown in the signal. This time can be considered as a sign for the situation that the served train has entered the block. It is easy to understand that for safety reasons, when a block is occupied, its entrance signal should show a red aspect in order to make sure no other trains can enter this block.

Unfortunately, the signal aspect data for the period from 23-08-2010 to 27-08-2010 are incomplete due to the data avaliability reasons.

6.1.3. *The OBE drawing introduction*

The OBE drawing is a type of descriptive drawing that indicates the basic shapes for railway infrastructures. In the Netherlands, it is widely used as the standard reference for understanding the railway infrastructures because it can effectively show a larger area than the exact infrastructure drawings in a same page size. In the

OBE drawings, the names and general locations of all the sections and signals are given. The identifications of all the OBE drawings that are used in this study will be given in the appendix 2.

However, comparing to other detailed rail infrastructure drawings, i.e. the PVS (Permanente Vastlegging Spoorgeometrie) drawing, it has several unclear points as follows:

a) Exact locations of each element

In an OBE drawing, not all exact locations (kilometrations) of every element, i.e. section, signal, switch, station platform etc. are given. It can only present the geographical sequence of these elements and their basic layouts. As in the OBE drawings there is no standard scale ruler for all sections, therefore it is difficult to estimate an element's exact kilometration if this value is not initially given.

b) Length of a section or platform

For a section, even if the kilometrations of its boundaries are given as k_1 and k_2 ($k_2 > k_1$) separately, the actual length of this section l_0 may still be different from $k_2 - k_1$. There are two reasons: firstly, term $k_2 - k_1$ only means the distance of a direct link from a sections boundaries, therefore, if this section also contains a curve or a gradient, then this $k_2 - k_1$ is different from l_0; secondly, the directions of actual section rail and kilometration scale ruler could also be different.

c) Exact value for a gradient or curve

Although the OBE drawing also indicates gradients and curves in some sections, but these gradients and curves may be different from the real infrastructure. This phenomenon can be explained because the space in an OBE drawing is limited, then only the main (or average of many small gradients) gradients and big curves (for example, a gradient which is longer than 100 meters, or the radius of a curve is higher than 700 meters) can be shown in the OBE drawing.

These unclear points will obviously increase the difficulty in this study's speed profiles construction. The influence of these unclear points to this study will be shown in section 6.5.

6.1.4. *The OS drawing introduction*

The OS drawings is mainly used in presenting all the available signal aspects in a signal, and the aspect changing orders in a group of successive signals. The identifications of all the OS drawings that are used in this study will be given in the appendix 3.

An example of an OS drawing is given in Figure 6-3:

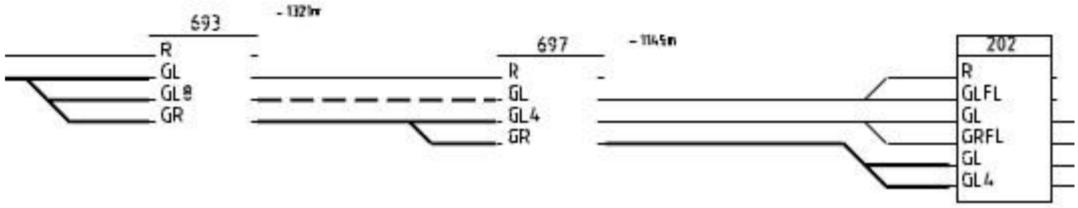


Figure 6-3 Example of the OS drawing for Utrecht-'s-Hertogenbosch, OS-blad 0322, version D

Generally, there are four elements in an OS drawing for one signal: its signal ID number, distance between this signal and its previous signal, all the aspects and the speed limit of the route from this signal and its neighbour signals. The last element is indicated by the type of the connection link between two aspects in neighbour signals (solid line, dashed line, etc.).

Figure 6-3 shows a signal changing sequence in signal 693, 697 and 202. An example is given as follows:

When a train A is running the block between signal 202 and its next signal (not shown in Figure 6-3), then train A's following train B will firstly see a red aspect in signal 202, a yellow aspect in signal 697 and a yellow with a speed limitation of 80 km/h in signal 693.

When the train A has left the block that starts at signal 202, it is feasible for train B to enter this block. Then the signal 202 can give a yellow aspect (or even higher) for train B. If a yellow aspect in signal 202 is given, then from Figure 6-3, the aspect in signal 697 can be derived from signal 202, so that is yellow aspect with a speed limitation of 40km/h (GL4). In this way, signal 693 will show a green aspect.

One reminder is that for a certain route, not all of the aspects in signal will be used. The number of available aspects in a signal depends on the number of routes that the signal may serve and other factors.

For each rail section, there are two OS drawings for different directions. The OS drawings use the same topological rail infrastructure database as the OBE drawings because singal ID and other elements' kilometration values are samely shown in the two drawings. Therefore, the distance that indicated in the OS drawing only represents the distance of a rail route between two signals, instead of the actual distance between these two signals in reality.

6.2 Trento data selection for study area

After all the data resources have been introduced, the next step is to select suitable train run samples for driving behavior investigation. So far, the only constraint is that the selected train runs must cover the railway section between Geldermalsen and Den Bosch. More constrains will be introduced in section 6.2.1.

6.2.1. *Data selection criteria*

As mentioned in section 2.3, the running time is determined by seven factors in reality: weather, rolling stock, power supply, infrastructure, driving behavior, timetable and passenger occupancy. If we wanrt to analyze the characteristics of different driving behavior, we have to guarantee that the other six factors should be the same for all train runs selected in one data group, otherwise different train runs are not comparable to each other. However, due to the data availability, we can not derive the weather, rolling stock usage, power supply condition and passenger occupancy for each train run, which is one of the limitations of this study.

Within the study area between Geldermalsen and Den Bosch, more than one train run groups may be constructed based on train running direction, timetable, etc. Within each train run group, at least three criteria have to be set as follows:

a) All trains in a group should belong to a same service

The objective of setting this criterion is to make sure that all trains in one group should have the same running time from timetable. If this condition does not hold, then different trains will have different buffer time, thus these train runs are not comparable to each other.

b) All trains in a group should use the same route

In a complex railway section from one station to another, for example, more than one track, a train may use different route to run through this section. For the trains that belongs to a same group, their route usage should be the same, otherwise the running distance for each train will be different from each other and make the train runs can not be compared with each other.

c) All trains runs should be free runs

If a train is hindered, driving behavior cannot be totally shown. Therefore, all hindered train runs should be excluded from the sample group.

Additionally, more criteria can be set to improve the similiarity of the trains that belong to one group, such as rolling stock type, train length, train composition, etc. However, for the data availability of this study, these criteria are not considered.

6.2.2. *Data selection based on section occupancy data sheets*

The railway network in Geldermalsen and Den Bosch area is given in Figure 6-4:

Figure 6-4 railway networks in Geldermalsen-Den Bosch area

Figure 6-4 shows that there are two merging/ split point and one crossing point in the line between Geldermalsen and Den Bosch, and most of the sections in this line are not shared by the other railway lines' route. Therefore, if a train is running between Geldermalsen and Den Bosch, it must go through the blue area indicated in Figure 6-4. Therefore, a quick procedure can be used to speed up selecting trains that run between Geldermalsen and Den Bosch area as given in Figure 6-5:

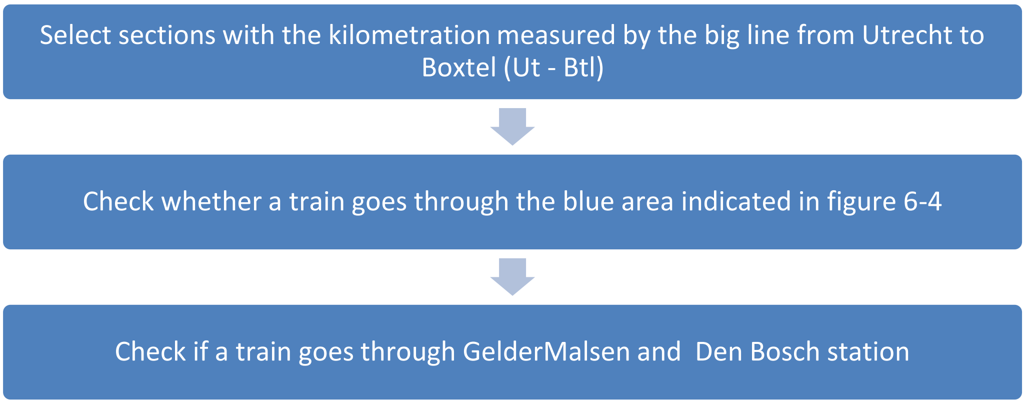


Figure 6-5 Procedures in selecting trains that run between Gdm and Ht

In order to check which train series will be selected in this study, an experiment on the section occupancy data on 23-08-2010 is exemplified, and the result is shown in Figure 6-6:

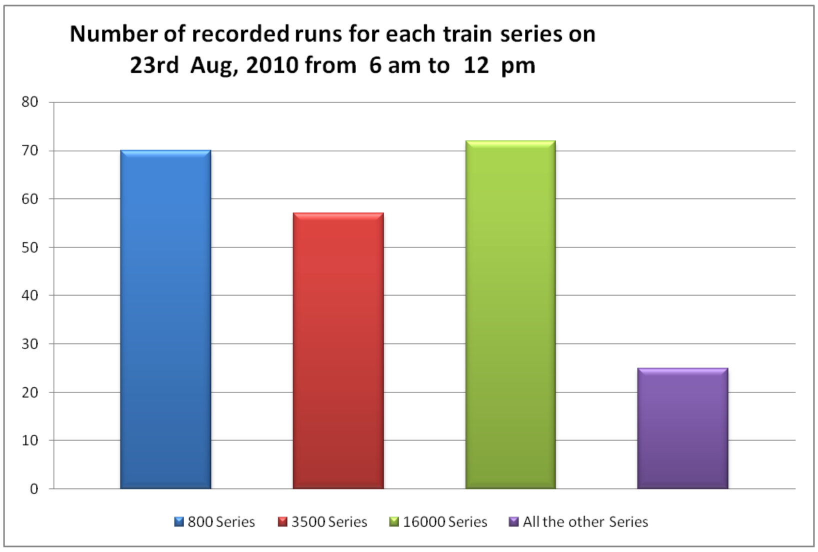


Figure 6-6 Number of recorded runs for each train series between Gdm and Ht on 23-08-2010

Figure 6-6 indicates that in the section between Utrecht and Den Bosch, there are mainly three train series, and the number of the trains in these three series is much higher the sum of trains which belong to the other train series (including freight trains). Therefore, train series 800, 3500 and 16000 are used as study groups. Series 800 and 3500 are intercity trains using various trainsets while series 16000 provide "stop train" service and use the sprinter trainsets. Considering a) the situation that the sprinter trainsets are not the major trainsets used in the Netherlands at present b) the intercity series trains have more chance to support energy efficient driving than the sprinter trains wich often have small margins, series 800 and 3500 are firstly selected.

For each train series, trains that run in both directions are originally included. Therefore, a train series has to be further divided into 2 groups by the train running direction. Therefore, there should be four groups for this study: 800 trains from Gdm to Ht, 800 trains from Ht to Gdm, 3500 trains from Gdm to Ht and 800 trains from Ht to Gdm. Within each group, train's running times are basically the same, while the only exception is found in 800 trains from Gdm to Ht. This exception will be considered in chapter 8.

The next step is to make sure that in each train group, every train uses the same route. This procedure can be solved by the PivotTable tool in Excel. The result of this step showed that most of these different route usage situations happen at the station areas of Geldermalsen and Den Bosch. There are two reasons for this result:

Firstly, most of the OpenTrack sections between Geldermalsen and Den Bosch are in double track level. In daily railway operations, each track will be specialized for only one traffic direction, while the other one will be kept as emergency track. Therefore, a train is not likely to use another route in double track level OpenTrack sections. Thus the route for a train is fixed. While in a station, there are more available tracks and the route planning for a train is more flexible.

Secondly, both Geldermalsen and Den Bosch stations are busy stations between the Utrecht – Den Bosch corridor. In these stations, a train is easier to get hindered. In a hindered case, the train may be arranged to run in another route.

After this step, we also observed that for 800 trains and 3500 trains that run from Den Bosch and Geldermalsen, these two groups have the same route and timetable running time. Therefore, these two groups are merged into one group together.

6.2.3. *Data selection based on signal aspect data*

a) Hindering case definition

The data selection process based on signal aspect data aims at excluding the train runs that are hindered by other trains.

Whether a train is hindered can be identified by the signal aspect that the train sees when it arrives at the location of signal. Furthermore, in order to avoid driver's anticipating driving, driver's sighting time (distance) should also be taken into consideration. An example is given as follows:

If there are n signal aspects for a signal as a_1 , a_2 ,… a_n where a_1 is the lowest order aspect and a_n is the highest order aspect. In the Netherlands, a block signal or a station signal should contain at least three aspects: red, yellow and a higher aspect (although for some advance signals, they do not have red. Thses signals are very rare thus will not be considered in this study).

Therefore, a_1 is red, a_2 is yellow and $n \geq 3$. In most of signals, a_n should be green (the exceptions will be discussed later in this section). If a train is about to enter a block and before the driver can see the entrance signal of this block (the distance between this train and the block is longer than the sighting distance), the signal has already shown a_n , then this train is considered as an unhindered train, otherwise this train is considered as a hindered train. Because we can imagine, if there is only one train running in the railway section, then the aspect it sees in every signal should always be a_n .

Therefore, signal aspect data can be used to detect hindered trains.

b) Difference between automatically controlled/ manually controlled signaling systems

In an automatically controlled signalling system, the signal aspects of a sequence of signals can be changed automatically according to train's track occupation status without additional route setting by the traffic controller".

For example, assuming a signalling system contains three successive signals as p_1 , p_2 and p_3 , p_1 is an automatically controlled signal, each signal has three aspects: green, yellow and red.

If a train enters the block between p_2 and p_3, then signal p_2 will turn red for safety. In this situation, its previous automatically control signal p_1 will turn yellow automatically.

Therefore, in a railway line equipped with automatic signalling system, if all signals' aspects are known when the first train is about to enter this line, then all signals' aspects for the rest of the trains during their run can be derived based on section occupation data and the OS drawings. Whether a train is hindered could thus be easily determined.

However, in a manually controlled signalling system, a signal aspect is controlled by the railway operator, and thus could not be accurately derived only from the section occupation data and the OS drawings.

Take the same example above except for the signal p_1 is a manually controlled signal: if a train enters the block between p_2 and p_3, then signal p_2 will turn red for safety and p_1 may still remians red if it does not receive the orders from the railway operator.

In this case, in order to derive the accurate aspects shown in manually controlled signals, it is also essential to check the signal aspect data, which contains the manually controlled signals' aspect switching event records.

c) Theory of detecting hindered trains by using the signal aspect data

Section 6.1.2 has already indicated that the signal aspect data sheet can provide a time t_1 when a signal p_1 shows a "proceed aspect" for a train. For p_1, this "proceed aspect" a_x should fulfil the condition that given by formula 6-1:

$$a_x \in \{a_2, ..., a_n\} \qquad (6\text{-}1)$$

Therefore, the lowest order aspect for a_x is a_2, and the highest order aspect for it is a_n.

Assume the next signal after p_1 is p_2, and at time t_2, signal p_2 also provides a "proceed aspect" for the same train. If this train still does not arrive at signal p_1, then at time t_2, a_x must has been switched to a higher order aspect a_{x+1} if $a_x \neq a_n$. Then the maximum time needed for signal p_1 to change its aspect from a_x to a_{x+1} is $(t_2 - t_1)$.

In this way, maximum possible time $T_{\max}$ (because the "proceed aspect" could directly be a_n) for signal p_1 to show a_n aspect can be given in formula 6-2:

$$T_{\max} = \sum_{i=2}^{n-2} (t_i - t_{i-1}) + t_1 \qquad (6\text{-}2)$$

Where t_i is the recorded “proceed time” that is given by the signal aspect data sheet.

In reality, one dilemma could happen to confuse the unhindered running: in some sections, a driver has a good sighting distance to see the next signal and do anticipating driving. For example, before a train enters the next block, the driver has already seen a yellow aspect shown in the next signal, he may slow down train speed immediately in order to make sure that the signal will turn to green again before he actually enters this signal. This kind of anticipating driving can not be distinguished from other “pure” free running records from the Trento data.

If the train’s actual entrance time for the first section between p_1 and p_2 is $T_{entrance}$, and in order to avoid anticipating driving, 9 seconds of driver’s sighting time is added, then if $T_{entrance} \geq T_{\max} + 9s$, the train can be considered as absolutely unhindered; if $T_{entrance} < T_{\max} + 9s$, then the train could be a hindered train. In order to make sure that all the selected train runs are unhindered, only the train runs that fulfil the first condition will be selected.

d) For the direction of Ht to Gdm, data of one signal is missing

For the direction of Ht to Gdm, data of one signal (signal number 748) in the border of Den Bosch and Hedel is missing

Because they are near Den Bosch station, it is not likely for a train to get hindered there, especially when the train has already depart from Den Bosch station. This conclusion can also be proved by speed profiles that are determined in the latter chapters.

e) For the direction of Gdm to Ht, two signals do not have green aspect for the route

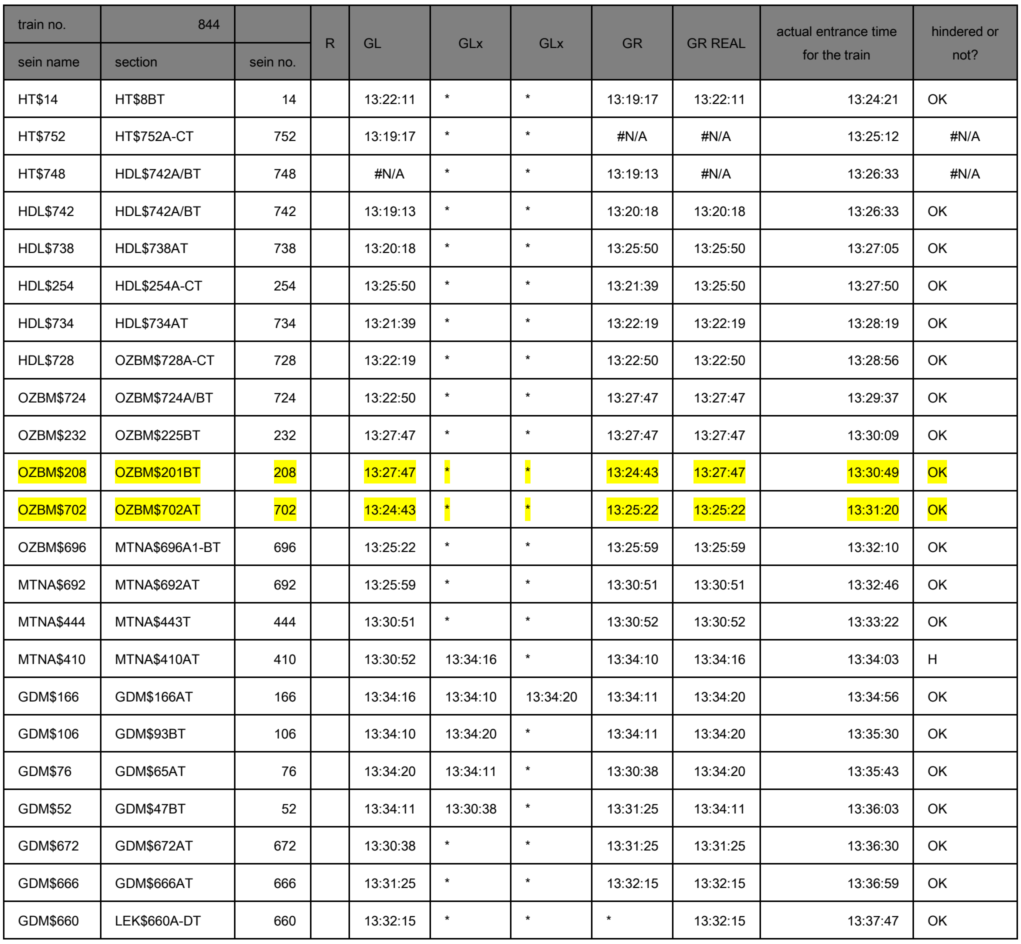
For the direction of Gdm to Ht, two signals before the last signal do not have the green aspect as the highest order aspect for the route for a train to stop in Den Bosch, and the highest signal aspects for these two signals are GL4 and GL, respectively.

Although from speed limitation point of view, the aspect GL4 does not have difference with GL, the trains which meet GL in this signal will still be considered as hindered trains. Because if a train sees GL instead of GL4 in this signal, this train is also expected to encounter a red aspect in the next signal, thus the train has to start braking and change its free driving style.

f) Hindered train runs detection in application

In order to use the above hindered train detection theory in practical, an automatic train detection Excel based on the VBA programming is constructed. The identifications of all the Excel sheets that are used in this study will be given in the appendix 4. An example of a hindered train detection sheet in Excel is given in Table 6-1:

Table 6-1 Example of automatic hindered train detection Excel sheet



train no.	844		R	GL	GLx	GLx	GR	GR REAL	actual entrance time for the train	hindered or not?
sein name	section	sein no.								
HT$14	HT$8BT	14		13:22:11	*	*	13:19:17	13:22:11	13:24:21	OK
HT$752	HT$752A-CT	752		13:19:17	*	*	#N/A	#N/A	13:25:12	#N/A
HT$748	HDL$742A/BT	748		#N/A	*	*	13:19:13	#N/A	13:26:33	#N/A
HDL$742	HDL$742A/BT	742		13:19:13	*	*	13:20:18	13:20:18	13:26:33	OK
HDL$738	HDL$738AT	738		13:20:18	*	*	13:25:50	13:25:50	13:27:05	OK
HDL$254	HDL$254A-CT	254		13:25:50	*	*	13:21:39	13:25:50	13:27:50	OK
HDL$734	HDL$734AT	734		13:21:39	*	*	13:22:19	13:22:19	13:28:19	OK
HDL$728	OZBM$728A-CT	728		13:22:19	*	*	13:22:50	13:22:50	13:28:56	OK
OZBM$724	OZBM$724A/BT	724		13:22:50	*	*	13:27:47	13:27:47	13:29:37	OK
OZBM$232	OZBM$225BT	232		13:27:47	*	*	13:27:47	13:27:47	13:30:09	OK
OZBM$208	OZBM$201BT	208		13:27:47	*	*	13:24:43	13:27:47	13:30:49	OK
OZBM$702	OZBM$702AT	702		13:24:43	*	*	13:25:22	13:25:22	13:31:20	OK
OZBM$696	MTNA$696A1-BT	696		13:25:22	*	*	13:25:59	13:25:59	13:32:10	OK
MTNA$692	MTNA$692AT	692		13:25:59	*	*	13:30:51	13:30:51	13:32:46	OK
MTNA$444	MTNA$443T	444		13:30:51	*	*	13:30:52	13:30:52	13:33:22	OK
MTNA$410	MTNA$410AT	410		13:30:52	13:34:16	*	13:34:10	13:34:16	13:34:03	H
GDM$166	GDM$166AT	166		13:34:16	13:34:10	13:34:20	13:34:11	13:34:20	13:34:56	OK
GDM$106	GDM$93BT	106		13:34:10	13:34:20	*	13:34:11	13:34:20	13:35:30	OK
GDM$76	GDM$65AT	76		13:34:20	13:34:11	*	13:30:38	13:34:20	13:35:43	OK
GDM$52	GDM$47BT	52		13:34:11	13:30:38	*	13:31:25	13:34:11	13:36:03	OK
GDM$672	GDM$672AT	672		13:30:38	*	*	13:31:25	13:31:25	13:36:30	OK
GDM$666	GDM$666AT	666		13:31:25	*	*	13:32:15	13:32:15	13:36:59	OK
GDM$660	LEK$660A-DT	660		13:32:15	*	*	*	13:32:15	13:37:47	OK

Firstly, all signals' name, number and their entrance section names is entered. For each signal, all possible signal aspects will be identified. If a signal aspect does not exist for a signal, then the cell of this signal under the aspect column (shown as R, GL, GLx, x represents the speed number indicated in the signal, GR) will be indicated with a "*".

Secondly, maximum time for a signal to show green aspect will be derived and inserted under the column of "GR".

One reminder is that for the situation in which a manually controlled signal locates next to an automatically controlled signal, it is possible that the derived green time for automatically controlled signal is even earlier than its "proceed aspect" time, which is obviously wrong from reality.

For example, in the yellow areas in table 6-1, signal OZBM$208 is an automatically controlled signal while signal OZBM$702 is a manually controlled signal. Derived green time for signal OZBM$208 is 13:24:43 while the "proceed aspect" time for this signal for this signal is 13:27:47.

This phenomennon is strange because in reality, if signal OZBM$702 has already shown the "proceed aspect" for a certain train, then its previous signal OZBM$208 should turn to green at the same time.

Therefore, a new column "GR REAL" is used to compare maximum time among all the signal indication times in a row, and this maximum value is considered as the real green time for this signal. The purpose of using the "GR REAL" is further make sure that the selected trains are not hindered by signals.

After the whole data selection procedure, finally three data groups are ensured: 800 series train from Geldermalsen to Den Bosch (85 train runs), 3500 series train from Geldermalsen to Den Bosch (65 train runs) and 800 & 3500 series train from Den Bosch to Geldermalsen. The percentage of excluded trains based on different dates and train series are given in Figure 6-7 and Figure 6-8, respectively:

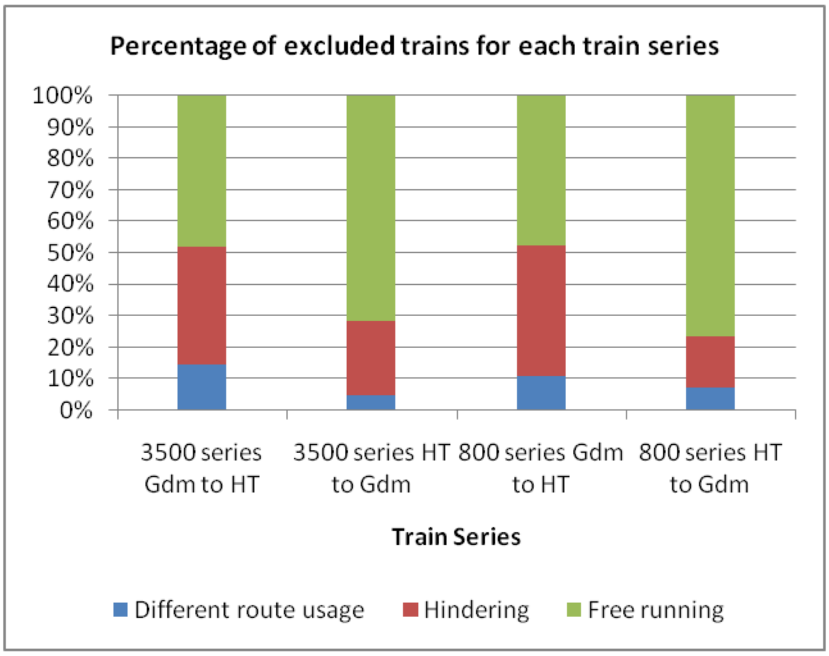


Figure 6-7 Percentage of excluded trains for different train series from 04-10-2010 to 08-10-2010

Figure 6-7 shows that for the direction from Geldermalsen to Den Bosch, more trains are excluded for both different route usage and hindering reasons (number of train runs for both directions are almost the same). Within one direction, there is no strong relationship between train series and number of excluded trains. One reason to this phenomenon could be that for the direction of Den Bosch to Geldermalsen, trains will depart from Den Bosch station. Therefore, they will have less possibility to get hindered in Den Bosch station.

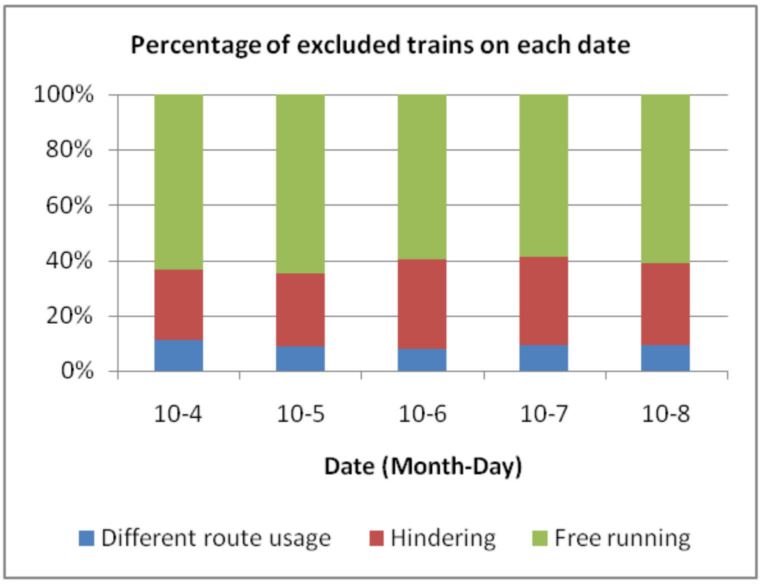


Figure 6-8 Percentage of excluded trains on each date from 04-10-2010 to 08-10-2010

Figure 6-8 illustrates that there is not obvious differences of percentages of excluded trains on different dates (each date (bar) in Figure 6-8 contains both 800 series and 3500 series trains for both directions).

One reminder is that for all train runs that are recorded during the period of 23-08-2010 till 27-08-2010, the procedure of detecting hindered train based on signal aspect data could not be conducted, because the signal aspect data for this time period are imcpmplete (data for all the signals in Den Bosch station area are missing) for the author.

6.3 Trento data errors correction

In original Trento's section occupation data, two kinds of obvious errors are found, namely wrong kilometration point indication and useless section records. These two kinds of errors will be introduced in section 6.3.1 and 6.3.2, respectively.

6.3.1. *Wrong kilometration point*

Examples of wrong kilometration point are exemplified in Table 6-2:

Table 6-2 Examples of wrong kilometration point in TROTS data on 23-08-2010

Treinnr	Sectie	Tijdstempel bezetmelding	Tijdstempel vrijmelding	Sectiescheiding 1 kmlint traject	Sectiescheiding 1 kmlint traject	Sectiescheiding 2 kmlint waarde
870	GDM$47BT	2010-08-23 20:15:57.950	2010-08-23 20:16:14.470	24750	Ut-Btl	24672
872	GDM$47BT	2010-08-23 20:46:51.360	2010-08-23 20:47:20.910	24750	Ut-Btl	24672
874	GDM$47BT	2010-08-23 21:06:09.510	2010-08-23 21:06:23.040	24750	Ut-Btl	24672
876	GDM$47BT	2010-08-23	2010-08-23	24750	Ut-Btl	24672

		21:36:43.390	21:36:52.410			
878	GDM$47BT	2010-08-23 22:05:38.590	2010-08-23 22:05:50.120	24750	Ut-Btl	24672
880	GDM$47BT	2010-08-23 22:35:17.400	2010-08-23 22:35:26.920	24750	Ut-Btl	24672
882	GDM$47BT	2010-08-23 23:07:10.950	2010-08-23 23:07:19.450	24750	Ut-Btl	24672
884	GDM$47BT	2010-08-23 23:39:50.570	2010-08-23 23:40:00.090	24750	Ut-Btl	24672
868	GDM$47BT	2010-08-23 19:36:43.600	2010-08-23 19:36:51.610	25037	Ut-Btl	24672

Table 6-2 indicates that the kilometration point indications for a same section are not always the same. In the last row for No. 868 train, the kilometration value for starting point for section GDM$47BT should be 24750 instead of 25037.

As the selected trains belong to a same data group, they should use the exact same route and thus same kilometration indication values for the boundaries of sections. Therefore, this kind of errors cannot be tolerated.

For this kind of error, two solutions can be used. The first one is to derive the section's kilometration values from the OBE drawing. However, if the OBE drawing does not contain the detailed kilometration indication, then the most frequently indicated kilometration value for the same section is used. For example, in table 6-2, the most frequently indicated kilometration vale for section GDM$47BT is 24750, then 24750 can be used as the correct value.

6.3.2. *Useless section records*

Some useless section occupancy records are exemplified in Table 6-3:

Table 6-3 Examples of useless section records on 23-08-2010

Treinnr	Sectie	Tijdstempel bezetmelding	Tijdstempel vrijmelding	Sectiescheiding 1 kmlint traject	Sectiescheiding 1 kmlint traject	Sectiescheiding 2 kmlint waarde
814	HT$752A-CT	2010-08-23 05:55:11.130	2010-08-23 05:55:47.200	46201	Ut-Btl	45375
814	HT$748B52DT	2010-08-23 05:55:42.690	2010-08-23 05:56:31.810	45375	Ut-Btl	43927
814	HDL$748B52DT	2010-08-23 05:55:42.760	2010-08-23 05:56:27.120	45375	Ut-Btl	43927
814	HDL$742A/BT	2010-08-23 05:56:27.120	2010-08-23 05:57:04.190	43927	Ut-Btl	42791

Table 6-3 shows that section HT$748B52DT and HDL$748B52DT have the same starting point and endind point but their entrance and exit times are different. For HDL$748B52DT, its clearance time equals to the train's entrance time for the next section, which is impossible in reality. Therefore, this section record is useless and should be deleted. In Trento data, this kind of useless record exists in every train runs.

6.4 Speed profile construction method

In a train describer system, if we consider each single section, only the occupancy and clearance times are recorded. Therefore, for a train that runs through this section, we can only derive its average speed. However, if a series of sections are considered together, more information could be extracted by applying the Newtonian mechanics theory (Goverde 2000). Constructing a speed profile requires both train speeds and locations. As the length of a section varies from less than 100 meters to more than 1000 meters, it is difficult to identify a train's deceleration dynamics only by the section length and train's running time if a section is too long.

In (Goverde 2000), trains are initially assumed to do constant acceleration/ braking rate running in a section, and this acceleration/ braking rate a is derived by formula 6-3:

$$a = \frac{v_{end} - v_{start}}{t} \qquad (6\text{-}3)$$

Where v_{end} is the speed of the train when it enters the next section and this value is estimated in the least square method ;
v_{start} is the speed of the train when it enters the section and this value is estimated in the least square method;;
t is the train' s running time in the section

The disadvantages of this method are:

Firstly, it assumes accurate section occupation times (in a second) and section lengths (or at least maximum deviation of the passing time data should be known);

Secondly, this method is time consuming in dealing with long line sections.

Therefore, this method is more applicable for the data that do not contain large inaccuracies. Otherwise, this method may fail.

For the Trento data used in this study, maximum deviation of the recorded time is not given. In some sections, this deviation can be higher than 1 second. One possible reason for this result is explained as follows:

In the TROTS system, two kinds of time data are recorded: the first one can be understood as "event time", which refers to the actual recording time for an event in TROTS system, while the other one can be understood as the "final recording time" after the TROTS system's final internal processing. For the latter one, the deviation is higher because it also contains the delays that are generated by the data processing procedure in the TROTS system. Trento software selects this “final recording time” instead of “event time” as resources, thus has unknown and higher deviations.

In this study, a crude speed profile construction method is used: for each section, a coordinator of a trains’ average speed and the section’s middle point is used to represent this train’s running status, and trains are assumed to do constant acceleration/deceleration rate running between these sections’ middle points. The whole process can be explained as follows:

Assume a study area has n sections as S_1 , S_2 ,…, S_n (platform sections for train to do initial acceleration or final braking are excluded), and starting point kilometrations and running times for these sections are l_1 , l_2 , …, l_n and t_1 , t_2 , …, t_n , respectively. For a section $i \in (1,n)$, train’s average speed $v_i = (l_{i+1} - l_i)/t_i$, and the kilometration value for section’s middle point $X_i = l_i + (l_{i+1} - l_i)/2$. Thus, a coordinator (X_i, v_i) is created.

Therefore, in a speed distance graph, the speed profile for S_1 , S_2 ,…, S_n can be drawn as the linear connections between coordinators (X_i, v_i) .

Additionally, for a train’s acceleration/ braking phase, this study uses the same method as (Goverde 2000), which is mentioned in section 5.1.

The constructed speed profiles in Excel by using linear connections are exemplified in Figure 6-9:

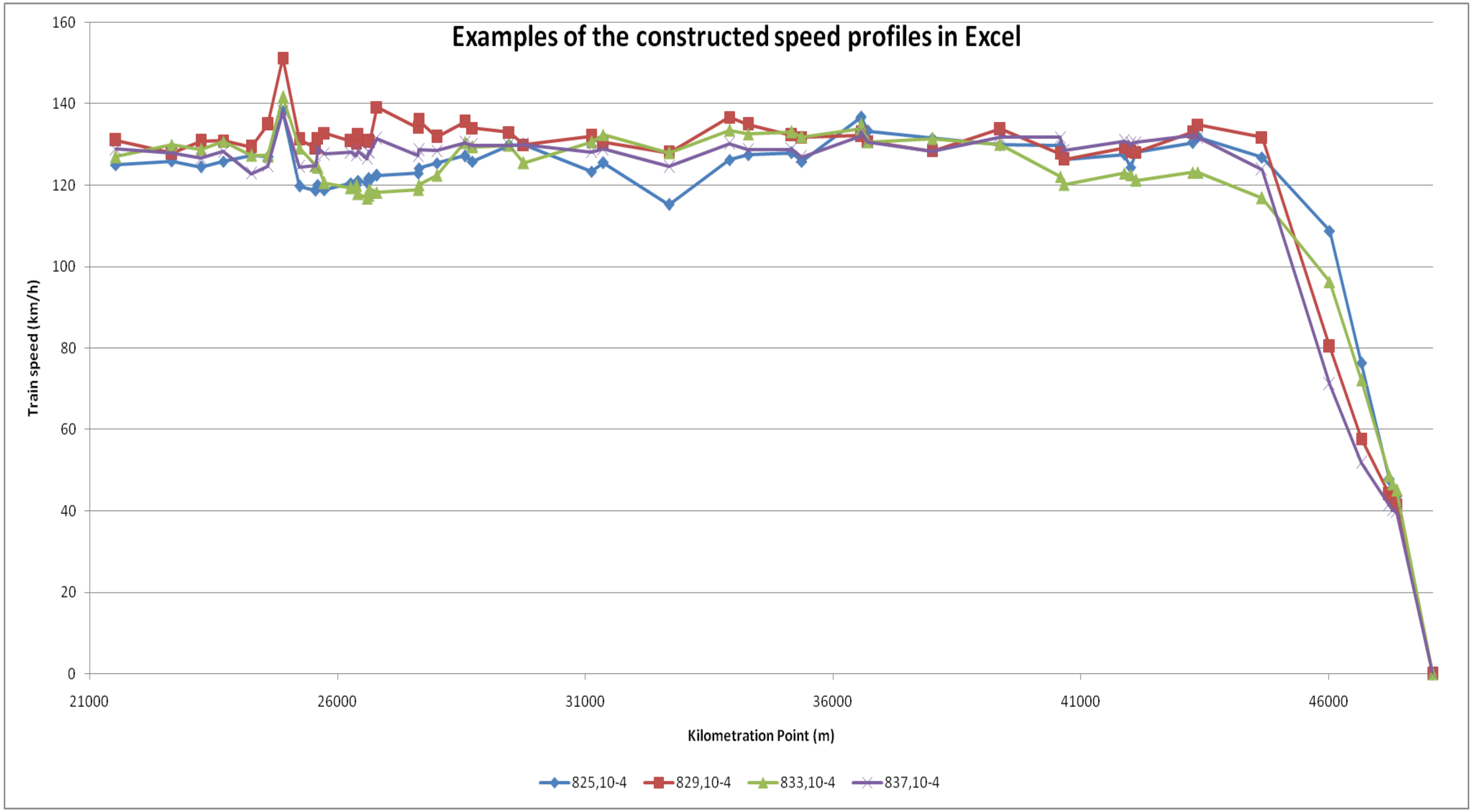


Figure 6-9 Examples of the constructed speed profiles in Excel by using linear connections

In Figure 6-9, different train runs are indicated by different colors and with the same format of “train number, month-date”. For each train run, the corresponding speed profile is the linear connections between coordinators (X_i, v_i). One reminder is that in order to improve the computational speed of Excel, train’s speed profiles in acceleration/braking phases are presented by linear connections in Figure 6-9 (but in the running time computation the method introduced in (Goverde, 2000) is used).

However, if there are many train ran sample are shown in a speed profile figure, it is difficult to recognize all coordinators.Therefore, coordinators are not indicated in the further speed profiles.

Furthermore, in order to make the speed profile more realistic (at least a speed profile should be polynomial instead of linear), the cubic Bezier curve is used in Excel to connect the coordinators for each train run.

Finally, the final speed profiles used in this study are exemplified in Figure 6-10:

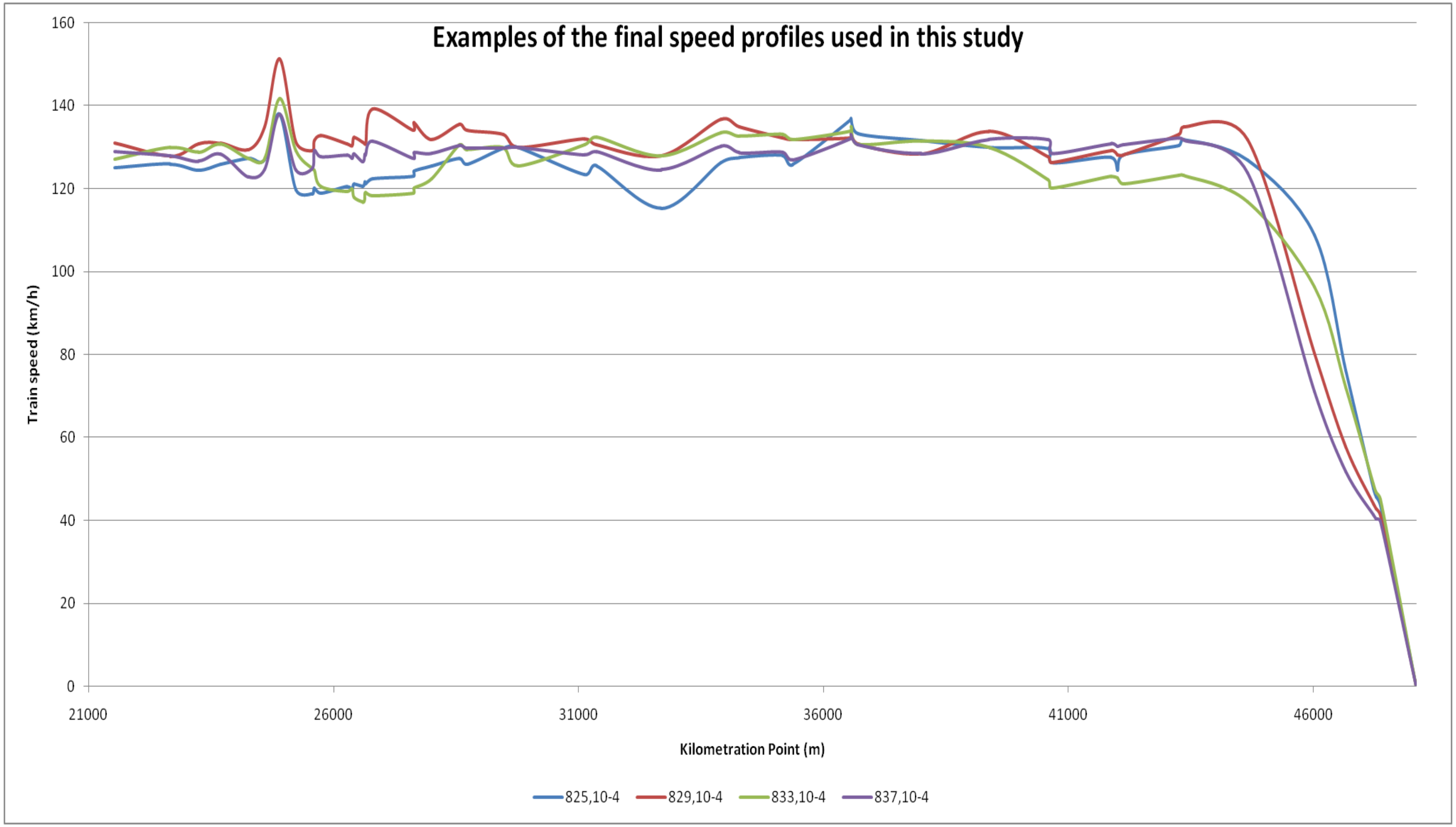


Figure 6-10 Examples of the final speed profiles used in this study

In fact, Figure 6-9 and Figure 6-10 use the same train run samples. We can see that the speed profiles shown in Figure 6-10 are clearer than in Figure 6-9. Therefore, the speed profile construction method used in Figure 6-10 will also be used in the further speed profile figures of this study.

Running time deviations arising from this crude speed construction method will be assessed in section 6.7

6.5 Assessment on the original Trento data

By the assumptions made in section 6.4, the speed profiles of original Trento data can be derived. Take the groups of 800 trains that run from Geldermalsen to Den Bosch as example, the speed profiles are shown in Figure 6-11:

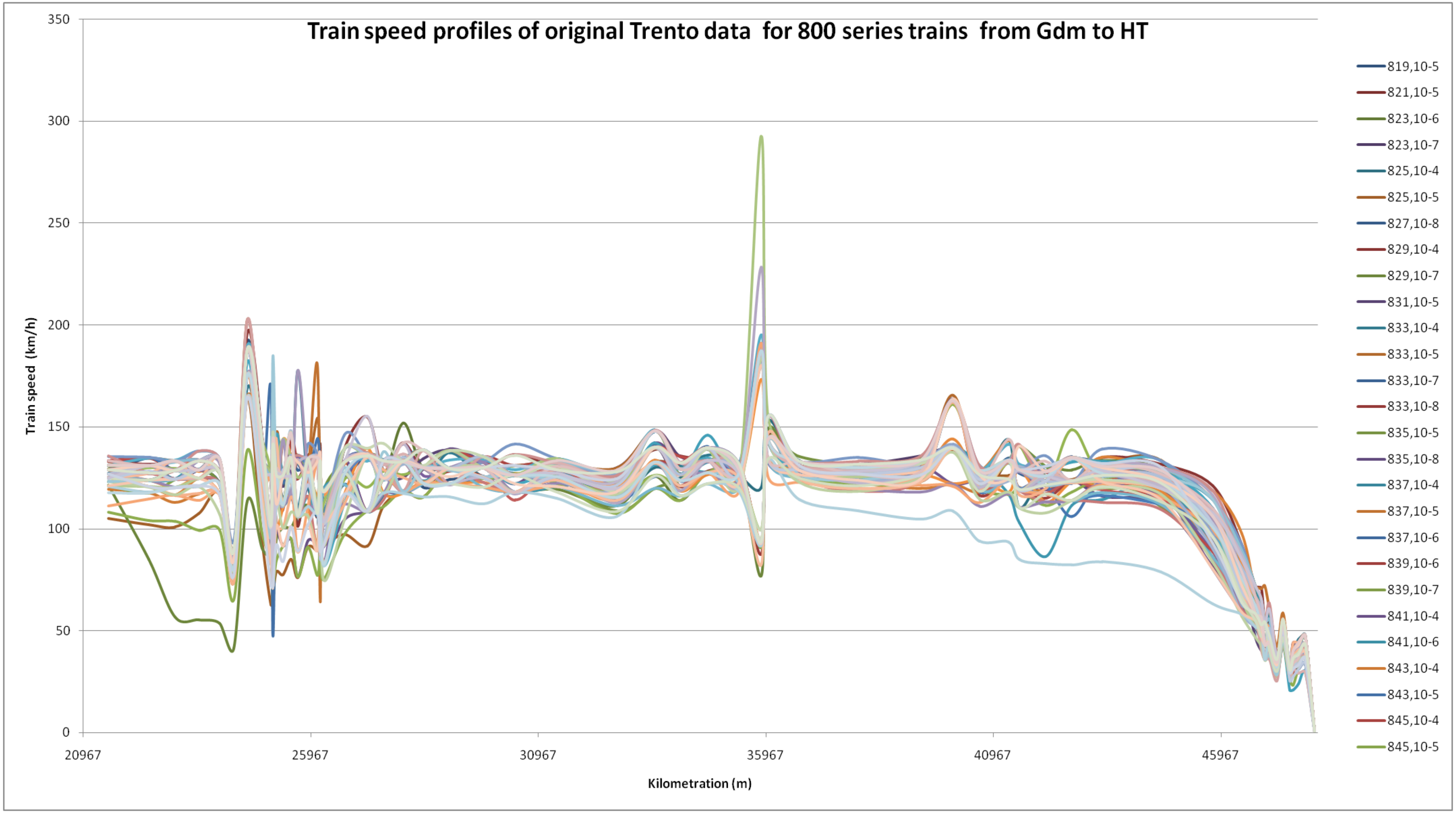


Figure 6-11 Train speed profiles of original Trento data for 800 series trains from Gdm to Ht, from 04-10-2010 till 08-10-2010

Figure 6-11 shows that the speed profiles that are determined by the original Trento data have a lot of fluctuations and errors. In the Netherlands, the common maximum speed limit for normal tracks is 130 km/h or 140km/h (not including the high speed lines). However, in Figure 6-11, many speed records exceeded this value.

In another experiment, 800 series trains that running from Den Bosch to Geldermalsen on 23-08-2010 are used to investigate the relationship between number of high speed record (higher than 140km/h) and section length is given in Figure 6-12:

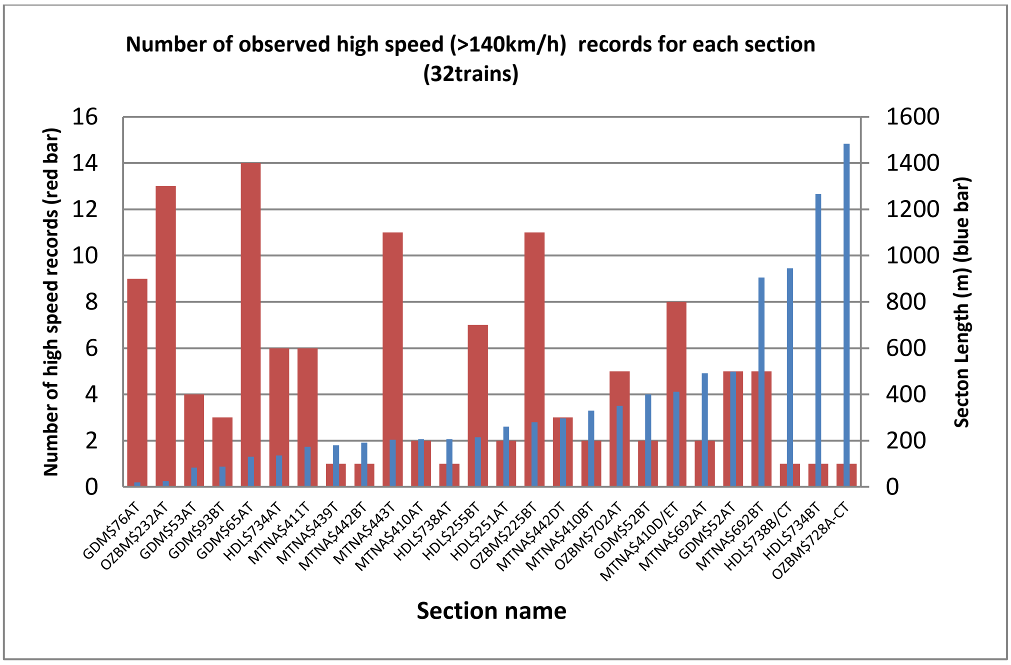


Figure 6-12 Relationship between number of high speed records and section length on 23-08-2010

Figure 6-12 shows a general trend that sections with shorter distance are more easily influenced by errors in Trento data than the sections with larger distance. Additionally, a large proportion of high speed records is located in the Geldermalsen station area, This phenomenon can be explained by the fact that the sections in the Geldermalsen area are shorter than the sections in the open line areas.

In Section 6.1, one conclusion has been made that there are still deviations in the time record of Trento data and section length estimation. Therefore, a Trento data processing is essential in improving the quality of speed profiles. The data processing method will be introduced in section 6.6.

6.6 Processing Trento data based on section grouping method

6.6.1. Grouping method introduction

As shown in section 6.5, the sections with larger distance will be less influenced by deviations than the sections with smaller distance. Therefore, the principle of grouping method is to group more than one section together to create a new virtual section, and use the new grouped section's average speed and central length point to create a new speed profile. In this way, high-speed record will be reduced and derived speed profiles would be more realistic. A sketch of this method is given in Figure 6-13:

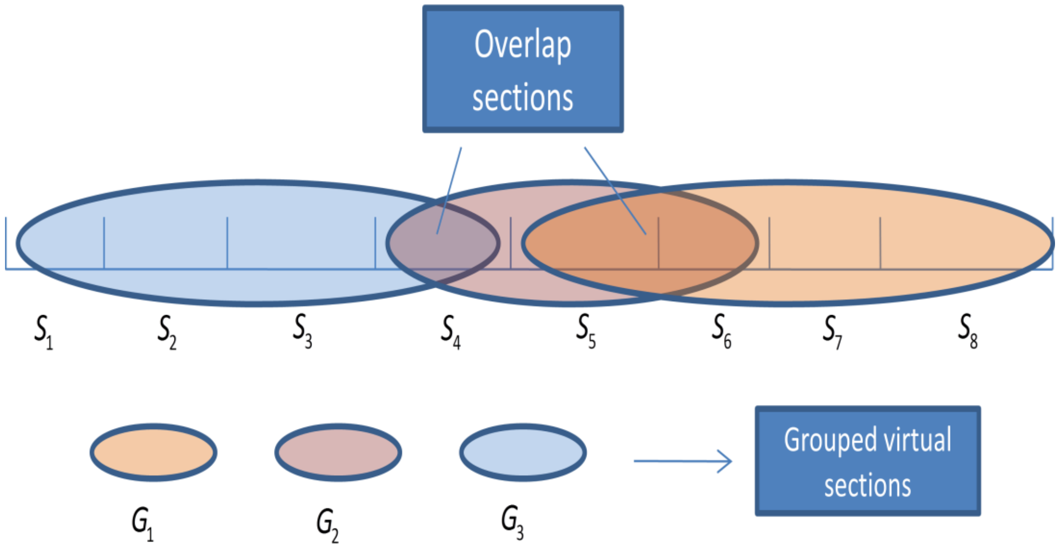


Figure 6-13 Example of section grouping method

If a route contains 8 sections (in this chapter, sections for train's final braking and initial acceleration will not be processed by this grouping method) which are shown as S_1 to S_8. If three grouped sections are created as G_1 to G_3. Then it is clear to see that G_1 contains four sections: S_1, S_2, S_3 and S_4; G_2 contains three sections: S_4, S_5, and S_6, etc. for each grouped section G_i, same speed profile construction method will be used as introduced in section 6.4 . The meaning of overlap sections is also exemplified by Figure 6-13: in this figure, Section S_4 is the overlap section for G_1 and G_2 and S_5 & S_6 are the overlap sections for G_2 and G_3.

One reminder for Figure 6-13 is that it only gives a general indication about the means of grouped virtual sections and overlap sections. It does not mean that in this study, the size of overlap sections will be varing between different grouped virtual sections.

For a grouping method, there are two control parameters: the size of a grouped section and the size of overlap sections between two grouped sections. For grouped section size, if this value is too small, then the created virtual section can still not absorb the deviations in Trento data; but if its value is too high, then some details of train's actual movement will be lost. This statement will be proved later in this section.

For parameter of the size of overlap sections between two grouped sections, it determines the number of grouped sections we could derive from a fixed study area. For example, if a route contains 8 sections, and the size of grouped section is 4. In this case, if size of overlap sections is 3, then 5 grouped sections can be created; if this parameter becomes 0, then only 2 grouped sections can be created. Therefore, the higher number of overlap sections is, the higher number of grouped sections can be derived.

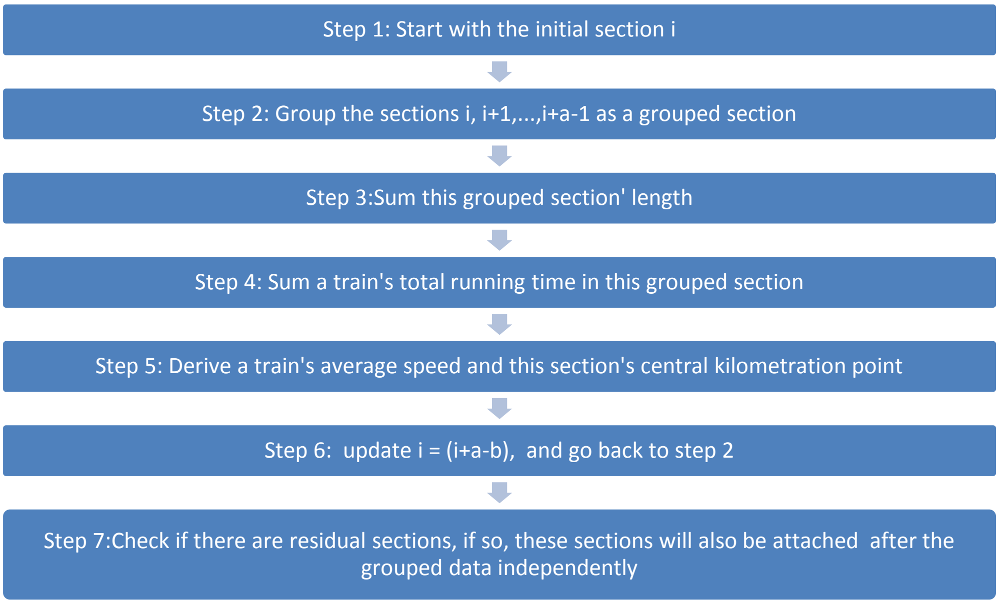
6.6.2. *Grouping sections based on number of sections*

In the grouping method based on number of sections, the values for the number of grouped sections and the number of overlap sections are fixed. Therefore, the length of grouped section may vary, i.e. if we compare a grouped section with short station sections and a grouped section with long open line sections.

If we define a section group contains a sections, and number of sections in a overlap between two adjacent section groups is b, for a studied route with n sections, number of section group m is given by formula (6-4):

$$m = \left\lfloor \frac{n-(a-b)}{a-b} \right\rfloor + 1 \qquad (6\text{-}4)$$

The automation of this grouping method can be realized by VBA programming in Excel. This automatic process is shown in Figure 6-14:



Step 1: Start with the initial section i

Step 2: Group the sections i, i+1,...,i+a-1 as a grouped section

Step 3:Sum this grouped section' length

Step 4: Sum a train's total running time in this grouped section

Step 5: Derive a train's average speed and this section's central kilometration point

Step 6: update i = (i+a-b), and go back to step 2

Step 7:Check if there are residual sections, if so, these sections will also be attached after the grouped data independently

Figure 6-14 Procedures in applying grouping method based on number of sections

One reminder is that it is essential to check of there are some sections which are not included in grouping method. For example, if n=8, a=4 and b=1, in this situation, the last section cannot be grouped. In order to keep complex of train run data, residual sections should also be attached after the grouped data independently (it means that the residual sections will not be included in the last section group).

Then the next question arises in how to select an optimal combination of grouping parameter values. Take the 818 train running from Den Bosch to Geldermalsen on 23-08-2010 as example, all possible parameter combination solutions and their corresponding speed profiles are shown in Figure 6-15:

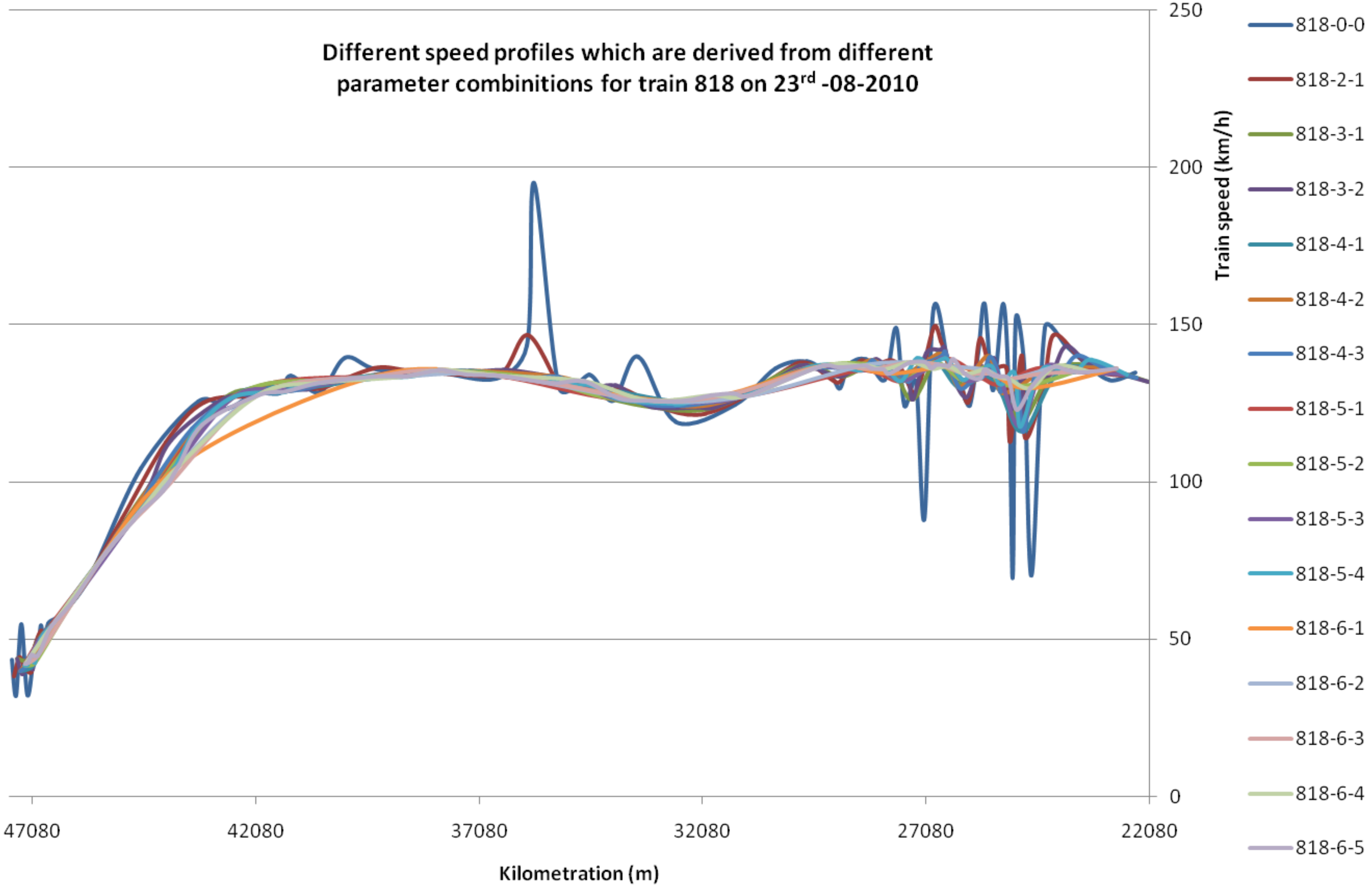


Figure 6-15 Example of different speed profiles that derived by different grouping parameter combinations for train 818 on 23 -08-2010

On the right side of Figure 6-15, a format of “train number + a +b” is used to indicate the meanings of various lines with different colours. For example, “818-4-1” means this speed profile is based on train number 818, number of original sections in a grouped section is 4 and number of sections in a overlap is 1. Moreover, the speed profile derived from the original Trento data is given by line “818-0-0” (according to the format definition, it should be “818-1-0”, and this line is the only exception from the format definition).

Figure 6-15 illustrates the following two findings:

Firstly, all the speed profiles that are derived from grouped sections are more realistic and have less fluctuations than the speed profile which is derived from original Trento data;

Secondly, if b is fixed, then higher a value will make the derived speed profile show less detail about train’s dynamics; but on the other side, it will also make the speed profile have less fluctuations. For example, if lines 818-6-1 and 818 2-1 in Figure 6-15 are compared, it can be observed that in train’s acceleration phase (kilometration area from 45000 to 39000) , line 818-2-1 shows more details in 818 train’s acceleration rate usage than line 818-6-1; but in the area of kilometration 28000 to 23000, it shows more fluctuations than line 818-6-1.

Therefore, to select an optimal parameter combination from all the possible combination sets is a very practical process, and different people may choose a different option.

Although from theoretical point of view, b value should be set as high as possible in order to derive more grouped sections, it is interesting to observe that the combination with lower b value may even create more realistic speed profiles than that with higher b value. An example based on the 818 train running from Den Bosch to Geldermalsen on 23-08-2010 is given in Figure 6-16:

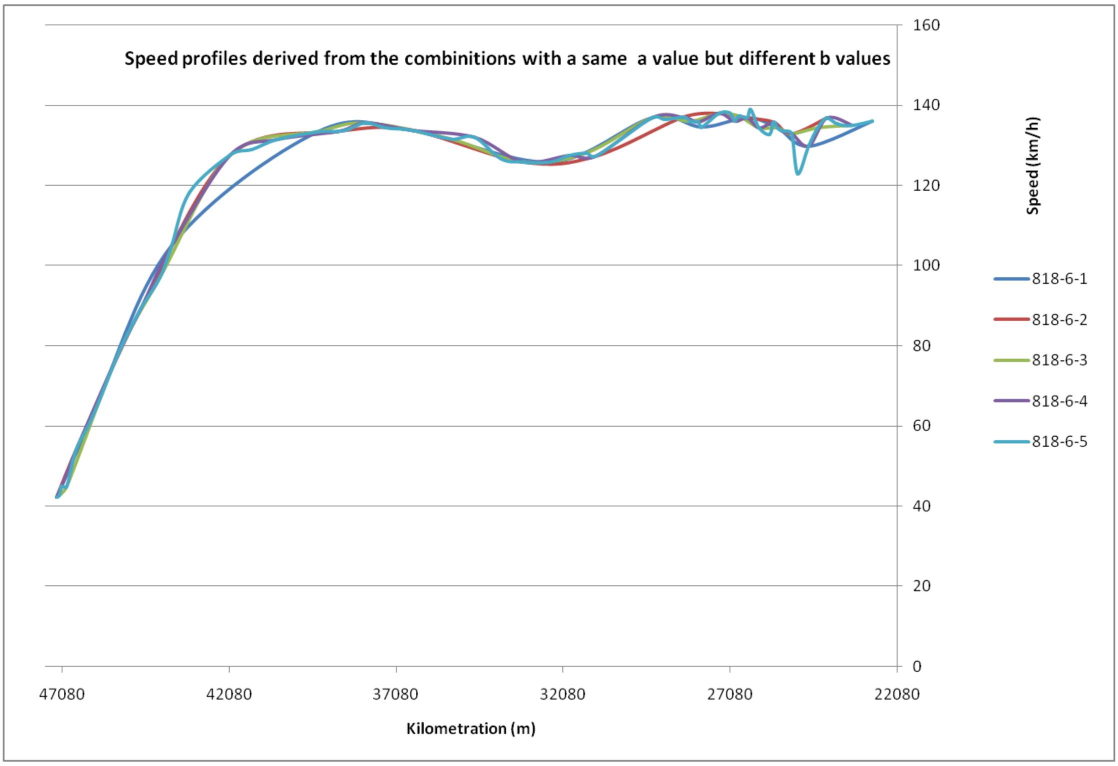


Figure 6-16 Speed profiles derived the parameter combinations with same a value but different b values for train 818 on 23-08-2010

Figure 6-16 indicates that the parameter combination of a=6, b=3 (line "818-6-3") shows less fluctuations in the area of kilometration points from 27080 to 22080 than the combination of a=6, b=5 (line "818-6-5"). This fact only shows that for a particular study area, some combinations with less b value can even derive a better speed profile than that with a higher b value, only because they are "lucky" enough to avoid some "bad recording sections", i.e. the border section between two TROTS system traffic areas. But it is recommended that for a new application line, grouping method parameters with maximum possible b value (always this value will be b=a-1) will be tested firstly.

Although the grouping method based on number of sections has proved its ability in improving the quality of speed profiles, the biggest defect is that this method could not control the length of grouped sections. This method will also create unbalance when it is used for both short station section groups and long open line sections. For example, if a open tack route contains 3 sections with equal distance of 1500m (it is possible in the Netherlands) and value a is set as 3, then the created grouped section will have a total distance of 4.5 km — in which a normal train can already finish a small trip with various driving strategies.

Therefore, an improved data grouping method based on the absolute measurement of length is introduced in the next section.

6.6.3. *Grouping sections based on section length*

The objective of the investigations in this method is to create an automatic VBA program that is based on the length of a grouped section and its overlap with previous grouped sections. However, as grouping method is still limited to include/exclude entire sections, it is difficult to define an exact and accurate value for both parameters. For example, if the length limit for a grouped section is set as 1000m and for overlap is set as 100m, and this parameter is applied to two sections with a length of 500m and 600m, respectively. Then it is impossible to split the second section into two small sub sections (as 500m and 100m) in order to guarantee the created group section has a length of 1000m.

In the same vein, the first section can not be divided into two sub sections with a length of 100m and 400 m, respectively to make sure the overlap must be 100 meters. Therefore, in this method, only **minimum lengths** of a grouped section and an overlap can be defined. These two variables will still be represented by a and b.

Assume we have the following variables:

P	number of sections in a route
i	section identification number
L(i)	length of section i
m	minimum length for a section group
n	minimum length for an overlap between two adjacent section groups
d	if a section grouping starts at section i, then d is the number of forwarding sections to the starting section of the next section group from i
e	if a section grouping starts at section i, then e is the number of forwarding sections to the end of this section group
D	if a section grouing starts at section i, then D(i) is the total length of the sections which have been grouped into a section group
j	if a section grouing starts at section i, then j is the number of the forwading sections being considered in a section grouping

The VBA program can be explained as follows:

```
For i=1 to P: Start grouping at section i, d=0, e=0, D=0, j=0

    For j=0 to P-i: D=L(i)+L(i+j)

        If D>= m And j=0 Then d=j, e=0, Exit for
        (If the first considered section i's length is already longer than m, then the second For loop has to be terminated
        If D>=m-n And D<a And j=0 Then d=1
        (If the first considered section i's length is already longer than (m-n), then the next grouping operation will start at section i+1)
        If D>=m-n And D<a Then d=j
        (If the D value is longer than (m-n) after section (i+j) has been considered, then the next grouping operation at section (i+j))
        If D>=m Then e=j, Exit For
```

```
        (If the D value is longer than m after section (i+j) has been considered, then
        e=j, and the second For loop is terminated)
        End If
        Next

i=i+d-1, d=0, e=0, D=0, j=0
Next
```

In order to speed up the process in determining optimal values for m and n, both m and n values will be in unit of 100 meters. Finally, two optimal parameter combinations and the speed profile which is derived from original Trento data for the trains that ran from Den Bosch to Geldermalsen, are given in Figure 6-17:

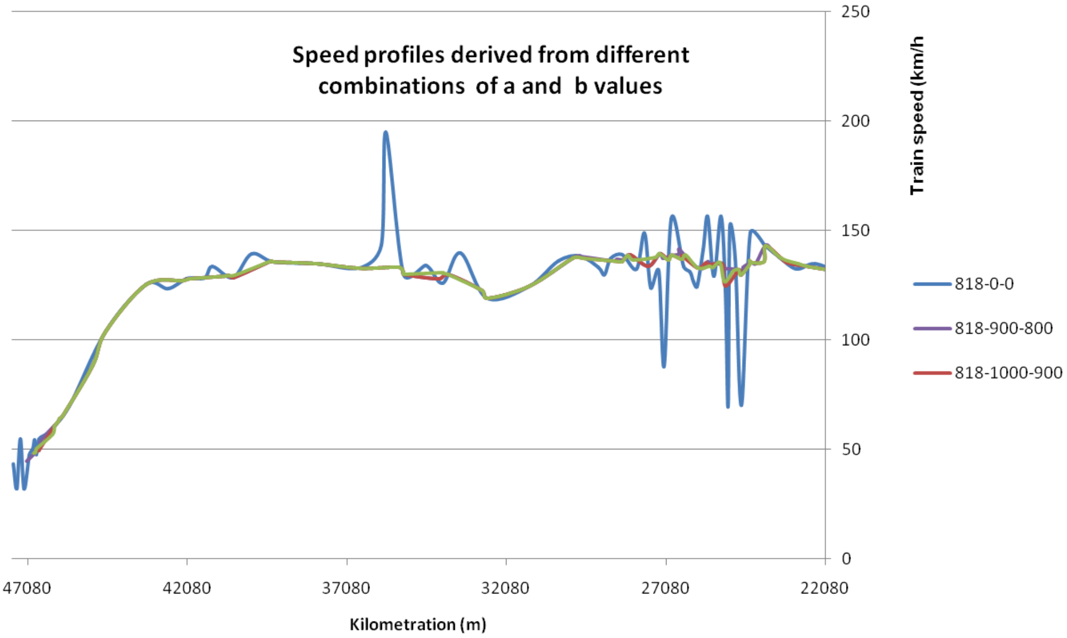


Figure 6-17 Speed profiles derived from different parameter combination factors for train 818 on 23-08-2010

In Figure 6-17 the legend in the right side of the figure indicates the “train number + a value +b value” combination that used for a speed profile. From this study, combinations of (a=1000, b=900) and (a=900, b=800) are both recommended as good combinations, **and factor combination of (a=900, b=800) will be used for further study for this thesis**.

6.7 Running time deviations from this crude speed construction method

In this section, a comparison is made to investigate the derived running times’ difference between the Trento data and the crude speed profile. After all the coordinators (X_i, v_i) have been calculated, the ruuning time between two adjacent kilometration points X_i and X_{i+1} and be determined by two methods.

The first method is based on the Trento's recording data by using formula 6-5 as follows:

$$T_{i\_i+1} = \frac{T_{i+1} + T_i}{2} - T_{overlap(i\_i+1)} \qquad (6\text{-}5)$$

Where T_i is the running time of X_i 's belonging section group i

$T_{overlap(i\_i+1)}$ is the running time of the section overlap between section group i and i+1

The second method is based on the derived crude speed profile. Although in this study, the crude speed profile will use the cubic Bezier curve as the connections between two coordinators, we still assume the connections are linear in here inorder to make the running time calculation easier (for train's initial accelration or final deceleration section, the running time are derived from the method mentioned in (Goverde 2000)). Therefore, the calculation formula is given in formula 6-6:

$$T_{i\_i+1} = 3.6 * \frac{X_{i+1} - X_i}{(V_{i+1} - V_i)/2} \qquad (6\text{-}6)$$

Where V_i is the section group i 's average speed [km/h]

Therefore, the derived running times' difference between the Trento data and the crude speed profile can be exemplified by several train runs from Geldermalsen to Den Bosch, which is given in Figure 6-18:

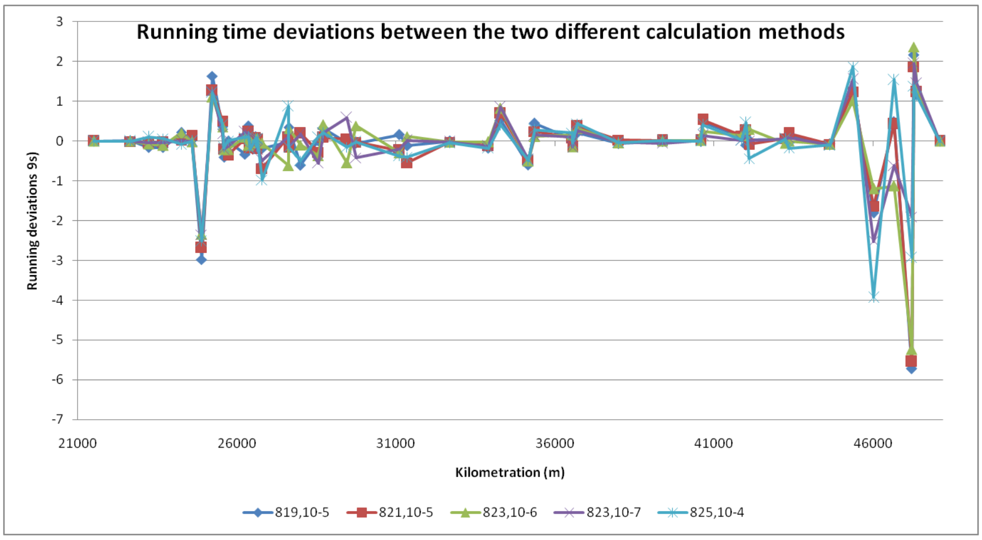


Figure 6-18 Running time deviations between the two different calculation methods

Figure 6-18 shows that the difference is small. For all the samples studied in this example, the running time deviations in all the section groups are lower than 3 seconds. The deviations in Geldermalsen area (location of the big fluctuation on the left side) and braking phase (location of the big fluctuation on the right side) is more severe than in open lines, but in general, the realiability of the crude speed profile is validated by the Trento data.

7 OpenTrack model construction

In this chapter, the process of constructing OpenTrack model that corresponds to this study's study area (from Geldermalsen to Den Bosch) will be introduced. In OpenTrack, this procedure is given in Figure 7-1:

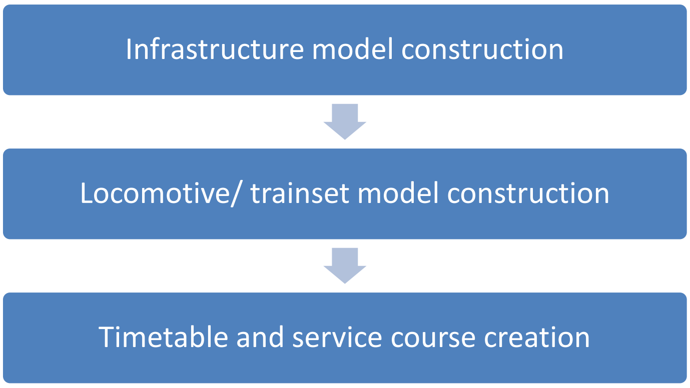


Figure 7-1 Procedures in OpenTrack model construction

These three procedures in Figure 7-1 will be explained in section 7.1, 7.2 and 7.3 separately.

7.1 Infrastructure model construction

All the elements for infrastructure model can be illustrated in Figure 7-2:

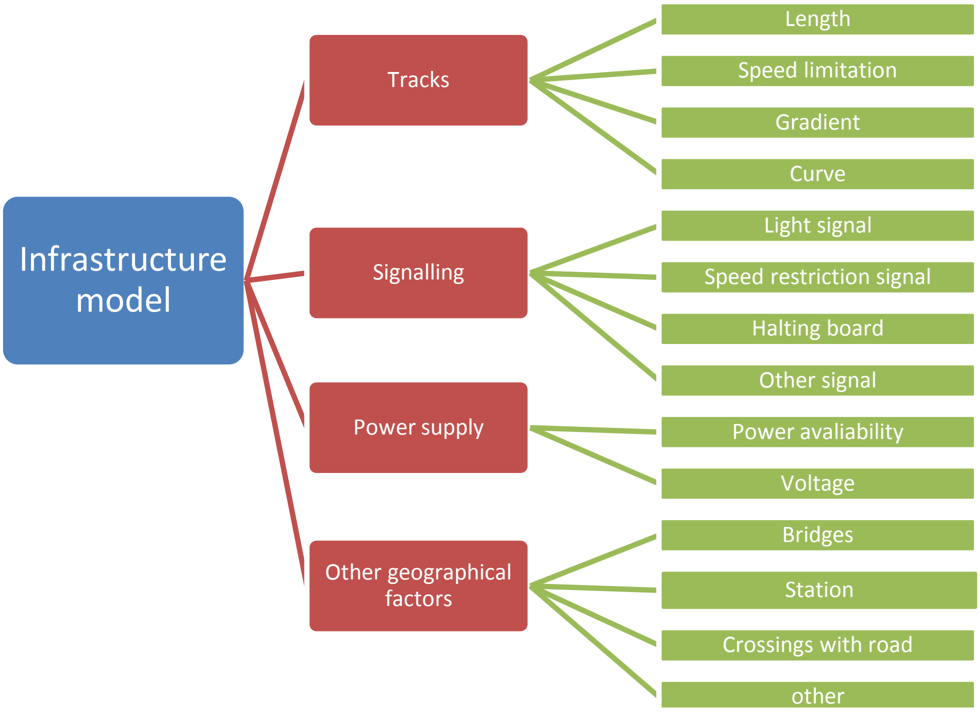


Figure 7-2 All the elements for infrastructure model

a) Rail

In reality, rails are divided into sections. In order to validate train's running time, it is essential to have each section's length and speed limit. However, for gradients and curves, their starting points and end points maybe different from sections. (in fact, for passenger trains, gradients and curves re more energy usage involved because all passenger trains are powerful enough to overtake all gradients in the Netherlands).

In OpenTrack, any part of rail can be presented as an edge and two nodes. In this study, all boundaries for blocks, gradients and curves will be a node in OpenTrack node in infrastructure model.

After the process of data selection, all selected train runs in one group should have the same route, and these train runs are not hindered by other trains. Therefore, it is not necessary to include all the switches and unused sections in infrastructure model.

b) Signal

Signals can be further divided as light signals (signals with colour aspects), speed restriction boards, halting boards and other signals. For this study, the light signals and speed restriction boards must be modelled in OpenTrack model because they are relevant with a train's running process. For each light signal, only the aspects for studied route will be selected.

The halting boards in station platforms should also be modelled because we have to know the total distance for a train run. In practice, this halting board can be selected as the location of a "Cijferbord". A "Cijferbord" is exemplified in Figure 7-3:

Figure 7-3 Example for a "Cijferbord"

The location of a halting board could also be estimated by the location of platforms in a station by researcher's own choice.

c) Power supply

In a railway network, for some sections, there can be no power supply for some reasons, i.e., transmission area between two electricity substations. Therefore, it is important to check the power supply availability for study areas and other factors that may influence a train's running time, i.e. supplied power voltage.

d) Other geographical factors

In OpenTrack model, it is also important to define a station area, which includes the entrance and exit signals, halting points. But this station area definition is not relevant for train's running time calculation.

Bridges and other geography factors are not necessarily shown in OpenTrack model, because these factors will not directly influence a train's running time. For example, if a train will pass a bridge and this bridge has a lower speed limitation than that of a normal track, then additional speed restriction signals will be added to inform the train in advance.

Finally, the constructed OpenTrack model is shown in Figure 7-4

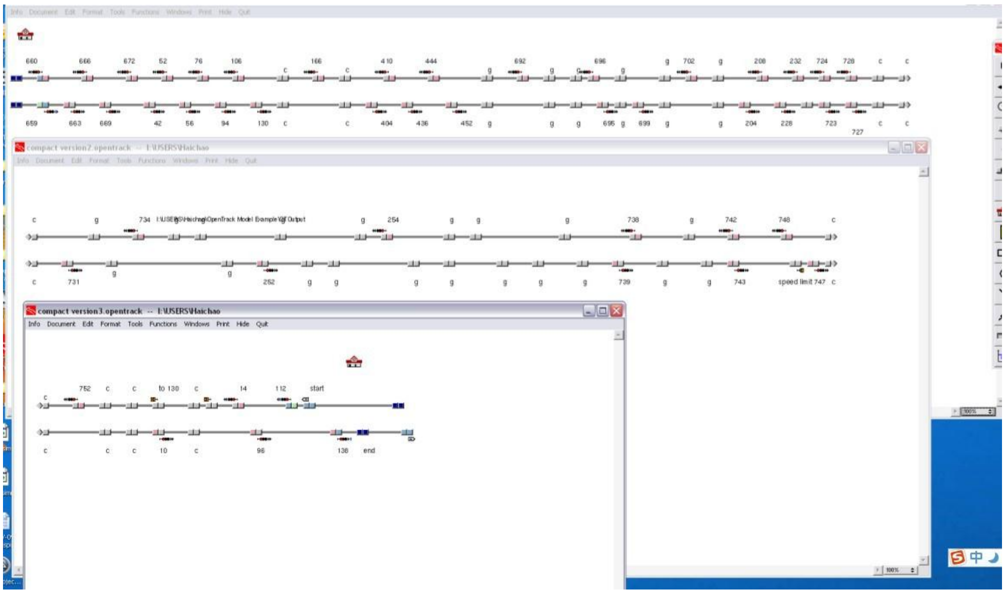
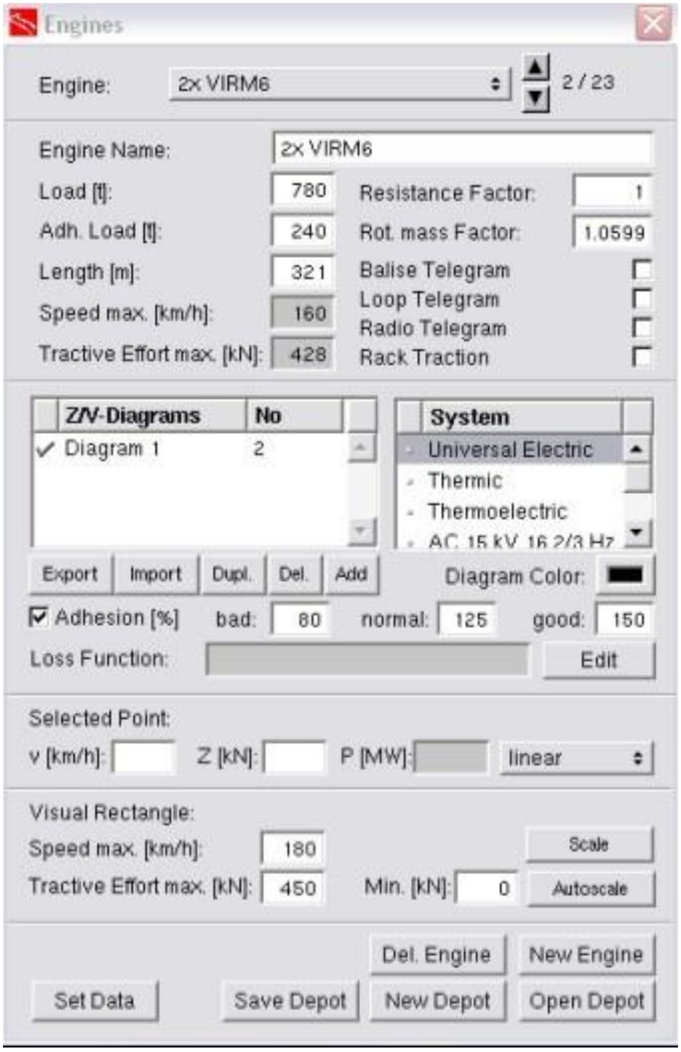


Figure 7-4 A sketch of the constructed OpenTrack model used in this study

7.2 Engine type and trainset determination

As has been explained in chapter 2, a train's running time is also affected by rolling stock type. In reality, two trains with a same train series may also have different rolling stock types and lengths. For each train group derived from Trento data, it is difficult to investigate each train's corresponding rolling stock type. Therefore, a virtual train set is assumed as the standard train type used for all train run samples.

In OpenTrack, the first step in this phase is select a train engine. For this study, a train engine "2*VIRM6" is created as shown in Figure 7-5 as follows:



Engine: 2x VIRM6 2 / 23

Engine Name: 2x VIRM6

Load [t]: 780

Adh. Load [t]: 240

Length [m]: 321

Speed max. [km/h]: 160

Tractive Effort max. [kN]: 428

Resistance Factor: 1

Rot. mass Factor: 1.0599

Balise Telegram ☐

Loop Telegram ☐

Radio Telegram ☐

Rack Traction ☐

	Z/V-Diagrams	No
✓	Diagram 1	2

System
- Universal Electric
- Thermic
- Thermoelectric
- AC 15 kV 16 2/3 Hz

Diagram Color: ■

☑ Adhesion [%] bad: 80 normal: 125 good: 150

Loss Function:

Selected Point:

v [km/h]: Z [kN]: P [MW]: linear

Visual Rectangle:

Speed max. [km/h]: 180

Tractive Effort max. [kN]: 450 Min. [kN]: 0

*Figure 7-5 Engine settings for the engine 2*VIRM6*

For this engine, the tractive effort-speed graph is given in Figure 7-6:

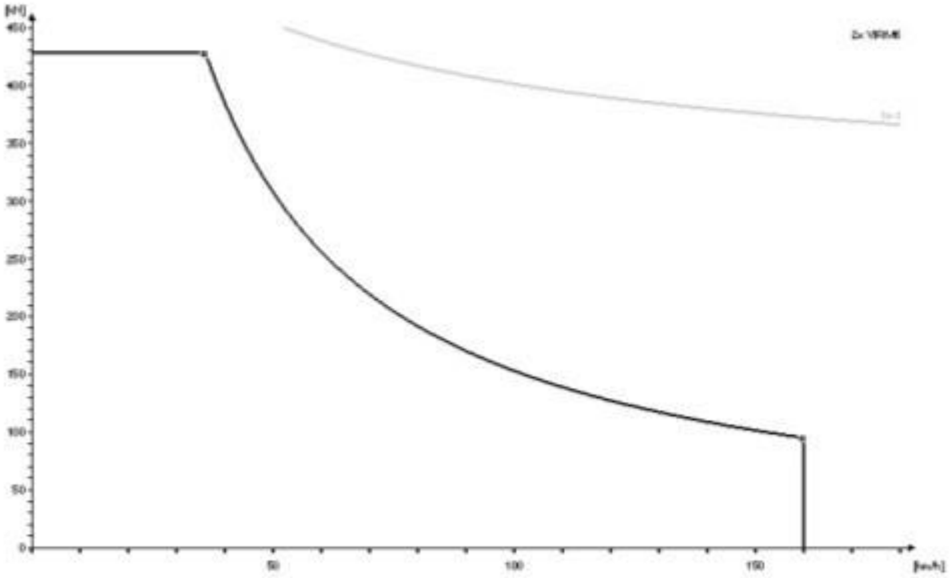


Figure 7-6 Tractive effort [KN]-speed [km/h] graph for VIRM 6

The second step is to define the trainset. The virtual train is composed by two VIRM 6 trainsets, and have a length of 321 meters.

For train resistance calculation, formula (7-1) can be used:

$$F_{TR} = N * M * 14.71 + (0.237 + N * 0.795) * (V + 10)^2 \tag{7-1}$$

Where M is the total mass of one train set (VIRM6)
N is the number of train sets
V is the train speed in [km/h]

Formula (7-1) is given by a literature that is published by the company NedTrain Consulting (NedTrain Consulting, 2001), and it can be considered as another representation of the general VPT formula (2-9). The relationship between all the coefficients used in formula (2-9) and (7-1) is given in Table 7-1

Table 7-1 The relationship between all the coefficients used in formula (2-9) and (7-1)

Coefficients	Formula (2-9)	Formula (7-1)
A	Rear-end aerodynamic coefficient [N·s²/m²]	0.237 [N·km²/h²]
N	Number of wagons	Number of train sets
B	Length-dependent air resistance coefficient[N·s²/m²]	0.795 [N·km²/h²]
V	Train speed [m/s]	Train speed [km/h]
△V	Wind speed [m/s]	10 [km/h]
M	Train mass for the whole train [kg]	Train mass for one train set [ton]
C	Running resistance [N/kg]	14.71 [N/ton]
D	Speed dependent resistance [N·s/m·kg]	0
E	Internal resistance [N·m/s]	0

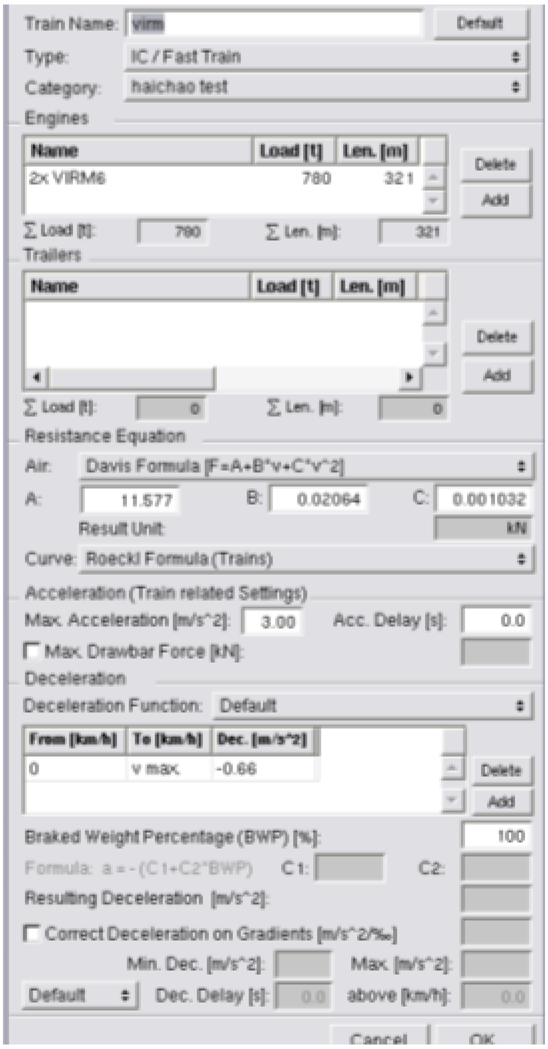
However, in OpenTrack, this VPT formula has to be converted into the Davis formula format, which is shown in formula 7-2:

$$F_{TR} = A + B * V + C * V^2 \qquad (7\text{-}2)$$

Where A, B and C are the coefficients .

For acceleration settings, max limitation value is set as 3m/s^2 . For deceleration function settings, this study firstly uses a constant factor of -0.66 m/s^2 . However, it has been proved in chapter 4 that in order to more accurately simulate the reality, a train should be set to use a periodical braking rate setting, because a train driver has to choose a braking step in practical situations.

Other train settings of the virtual train are given in Figure 7-7:



Train Name: virm
Type: IC / Fast Train
Category: haichao test

Engines

Name	Load [t]	Len. [m]
2x VIRM6	780	321

Σ Load [t]: 780
Σ Len. [m]: 321

Trailers

Name	Load [t]	Len. [m]

Σ Load [t]: 0
Σ Len. [m]: 0

Resistance Equation
Air: Davis Formula [F=A+B*v+C*v^2]
A: 11.577
B: 0.02064
C: 0.001032
Result Unit: kN
Curve: Roeckl Formula (Trains)

Acceleration (Train related Settings)
Max. Acceleration [m/s^2]: 3.00
Acc. Delay [s]: 0.0
☐ Max. Drawbar Force [kN]:

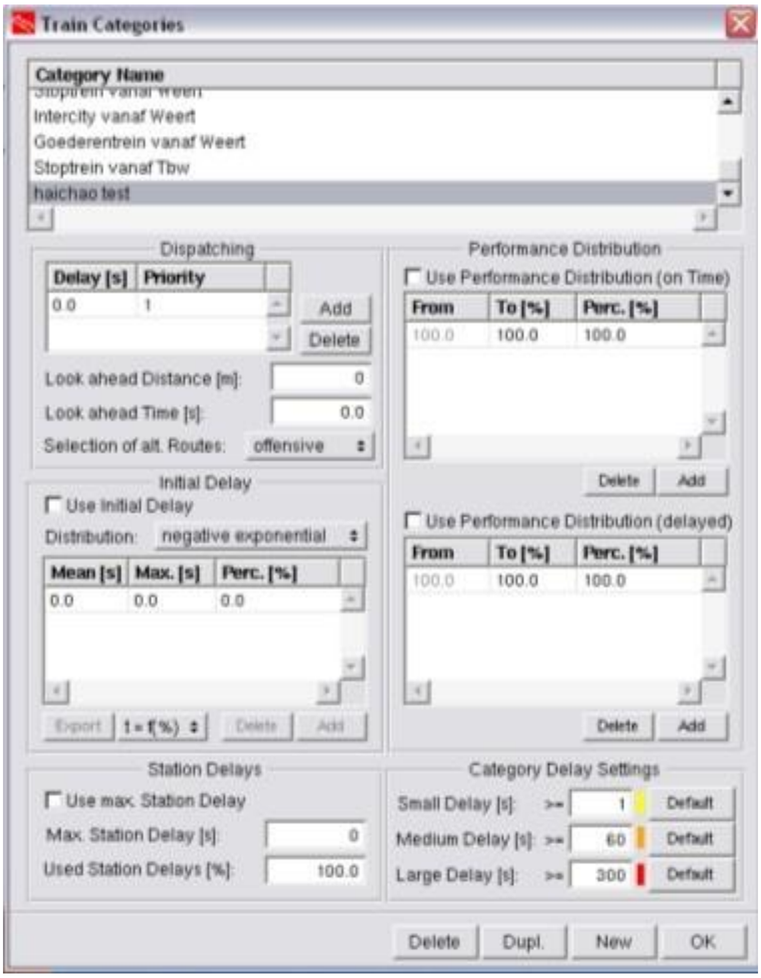
Deceleration
Deceleration Function: Default

From [km/h]	To [km/h]	Dec. [m/s^2]
0	v max.	-0.66

Braked Weight Percentage (BWP) [%]: 100
Formula: a = -(C1+C2*BWP) C1: C2:
Resulting Deceleration [m/s^2]:
☐ Correct Deceleration on Gradients [m/s^2/‰]
Min. Dec. [m/s^2]:
Max. [m/s^2]:
Default
Dec. Delay [s]: 0.0
above [km/h]: 0.0

Figure 7-7 Other parameter settings for the virtual train

The last step is to set train category factors. As 800 series train and 3500 series train are both intercity trains, virtual train is defined as IC/Fast trains in OpenTrack settings. As the main objective of this study is to validate certain types of driving behavior in OpenTrack, distributions for train category's initial delays, station delays, performance factors and delayed performance factors are not used. Additionally, as all selected train runs in Trento data group have been ensured to be unhindered runs, settings of dispatching factors are not relevant for this study. The whole train category setting is given by Figure 7-8:



Dispatching

Delay [s]	Priority
0.0	1

Look ahead Distance [m]: 0

Look ahead Time [s]: 0.0

Selection of alt. Routes: offensive

Performance Distribution

☐ Use Performance Distribution (on Time)

From	To [%]	Perc. [%]
100.0	100.0	100.0

☐ Use Performance Distribution (delayed)

From	To [%]	Perc. [%]
100.0	100.0	100.0

Initial Delay

☐ Use Initial Delay

Distribution: negative exponential

Mean [s]	Max. [s]	Perc. [%]
0.0	0.0	0.0

Station Delays

☐ Use max. Station Delay

Max. Station Delay [s]: 0

Used Station Delays [%]: 100.0

Category Delay Settings

Small Delay [s]: >= 1

Medium Delay [s]: >= 60

Large Delay [s]: >= 300

Figure 7-8 Train category settings

7.3 Timetable and service course creation

After infrastructure model and train set are determined, the next phase is usually to set the timetable for trains. But as the objective of using the OpenTrack model in this study is to derive the simulated train's running time while the timetable setting in OpenTrack model will not influence it, it is not nessesary to set the timatable for this study. Alternatively, a randomly estimated running time could be used.

After the running time has been set, the next step is to create a train course. In this case, the most important setting is the performance factor. In OpenTrack, train's final acceleration rate, braking rate and maximum cruising speed are calculated as their initial values times the performance factor. For example, if a train's maximum speed is initially set as 130km/h, and in a course, the performance factor is defined as 50%, then in the simulations, train will only do cruising at a speed of maximum 65km/h. In this study, different performance factors are used, which will be explained in chapter 8 and chapter 9.

After the process that introduced in section 7.1, 7.2 and 7.3, a simulation can be conducted in OpenTrack. Derived result can also be exported in various formats, and be compared with the Trento data.

8 Driving behavior analysis and OpenTrack modelling abilities

In this chapter, driving behavior analysis and a comparison between Trento data and OpenTrack result will be implemented for three data groups: 800 series trains running from Geldermalsen to Den Bosch, 3500 series trains running from Geldermalsen to Den Bosch and 800& 3500 trains running from Den Bosch to Geldermalsen.

One reminder is that in this chapeter, all the above three groups are derived from the time period from 04-10-2010 till 08-10-2010, and only consist of unhindered trains. Within each group, trains are using the same route.

8.1 Relationship between a train's running time and its initial delay

In this section, the relationship between a train's running time and its initial delay will be checked. As keeping the punctuality should be the highest priority in a driver's driving strategy decision making, if a train is experiencing an initial delay, train should drive as fast as possible.

In this study, a series of VPT timetables downloaded from the Prorail's website (www.prorail.nl) on 15-03-2010 are used to esimate a train's initial delay. A VPT timetable is examplified in Figure 8-1:

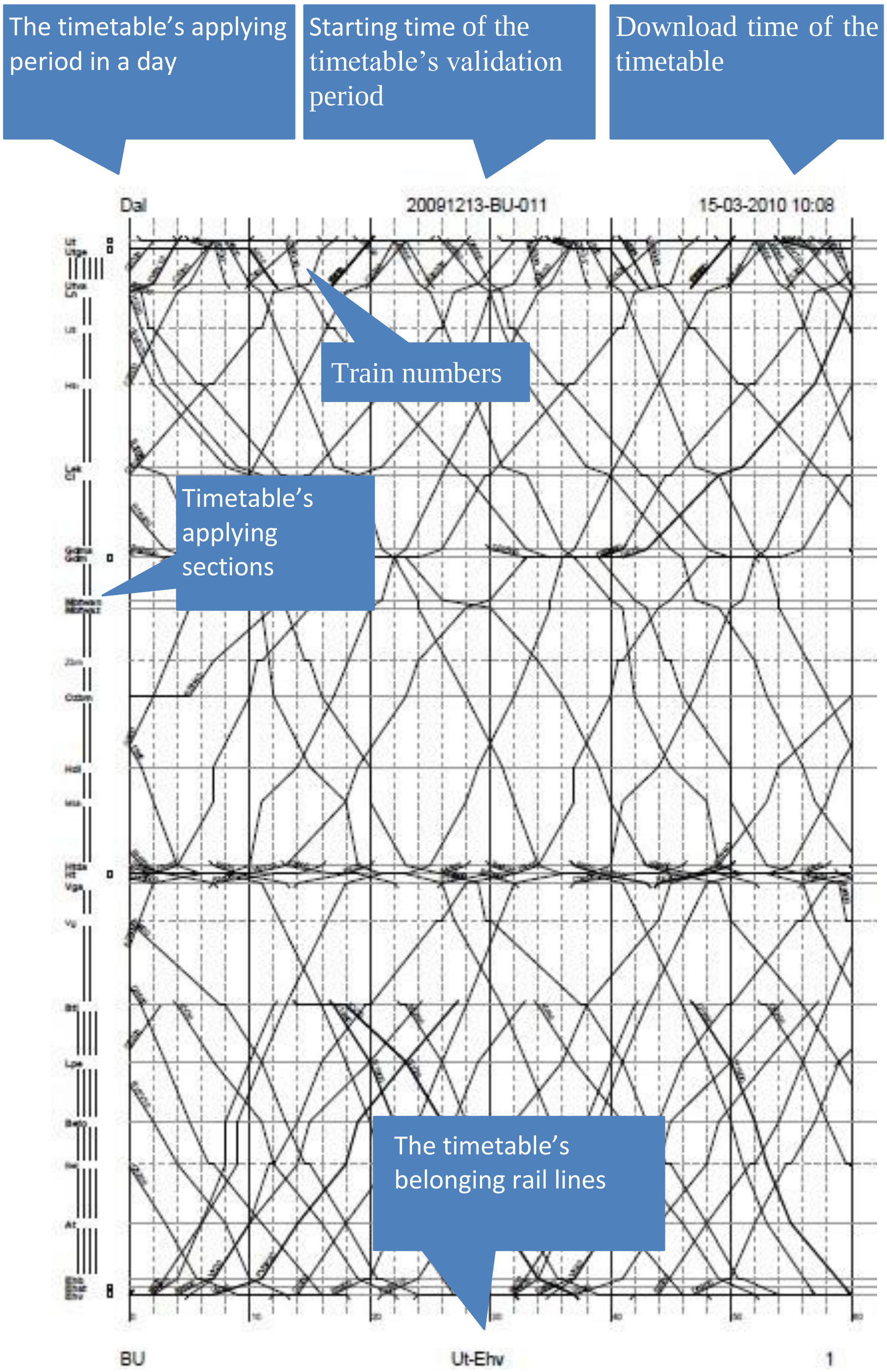


Figure 8-1 An example of timetables from the VPT system

A VPT timetable contains the following information:

a) Starting time of the timetable's validation period
b) Download time of the timetable from Prorail's website
c) The timetable's applying period in a day: there are different running schedules even in one day for different period, i.e. peak hours, off peak hours in the day time, off peak hours in the evening time.

d) Timetable’s applying rail sections: on the left side of a VPT timetable, the timetable’s applying rail sections are indicated and presented by the station names (or important junction names) and the number of tracks between two stations (in the indicated timetable’s belonging rail lines, i.e. Ut –Ehv).
e) The timetable’s belonging rail lines
f) Train numbers
g) Minute value in a hourly timetable

Figure 8-1 shows that a VPT timetable indicates the general schedules for trains, but it does not provide the exact time for a train to run from an exact point to another point. (But in the NS planning system, the exact running time from one station’s reference point to another station’s reference point is known). However, as the train’s punctual time will be applied for all train runs in one group, the deviations created in estimating the train’s punctual times from VPT timetables will not influce the investigation of the relationship between a train’s running time and its initial delay.

During the recorded time of the Trento data used in this study, there are three different running schedules: morning peak hour, evening peak hour and the rest of day. For the trains studied in this project, their running schedules do not show a difference in morning peak, evening peak and rest of day.

These schedules for all trains studied in this master project are given in Figure 8-2 (trains running from Den Bosch to Geldermalsen are not identified):

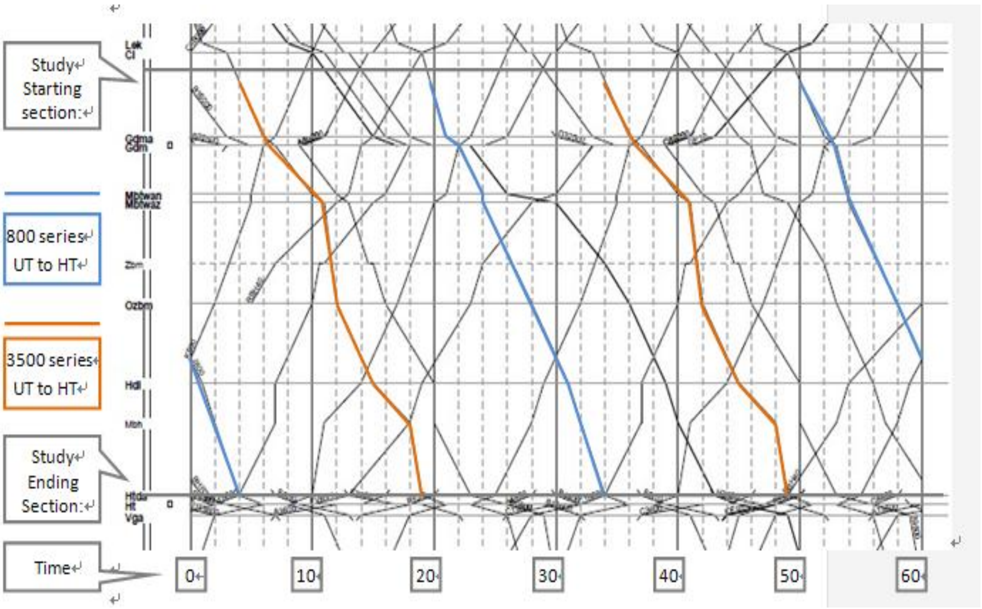


Figure 8-2 Running schedules for all trains

All the train series follow an hourly timetable and for each hour, two train runs will be operated. For example, 3500 series trains running from Geldermalsen to Den Bosch will enter the study area at 4:00 (minute:second) in first half hour and 34:00 in second half hour.

One interesting point for 800 series trains running from Geldermalsen to Den Bosch is that the two scheduled running times in each hour have a difference of 1 min (60 seconds) in the timetable. However, in reality, these two schedule's corresponding actual running time does not show this difference, which is shown in Table 8-1. Therefore, all 800 series trains can be grouped together.

Table 8-1 Running times for different departure times for 800 series train from Gdm to Ht

800 series	Average running time
Departure at first half hour	855 seconds
Departure at second half hour	845 seconds

For the other train groups, this scheduled running time difference between different departure times is not found. That means for the two trains in each series the scheduled running times are the same.

8.1.1. *800 series trains running from Geldermalsen to Den Bosch*

For this train series, if we assume punctual trains will enter the study area on 19:30 in first half hour and 49:30 in second half hour, then this group's initial delay distributions are shown in Figure 8-3:

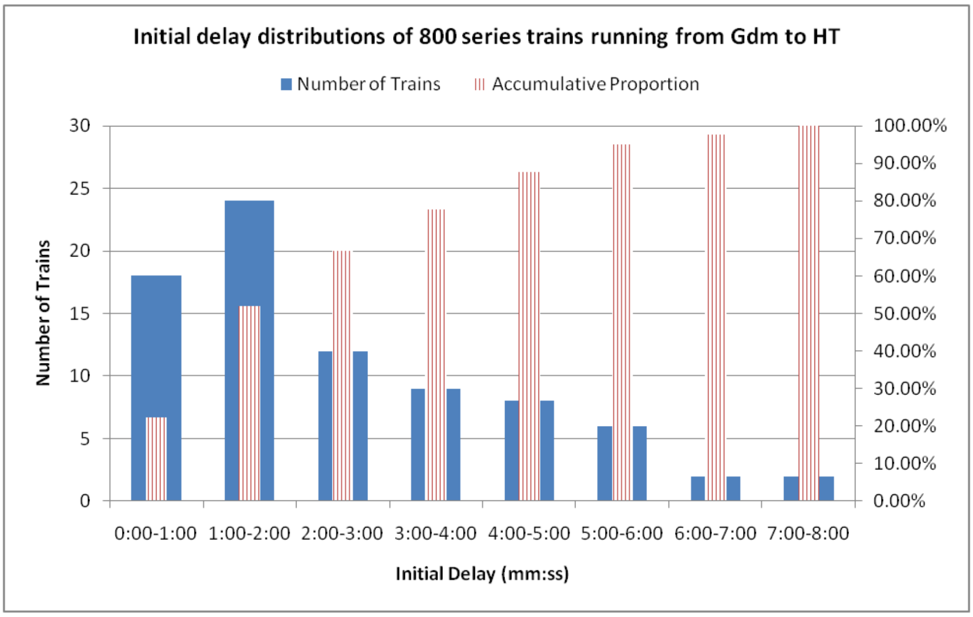


Figure 8-3 Initial delay distributions for 800 series from UT to Ht

Figure 8-3 illustrates that almost 70% of 800 series trains' initial delay values are lower than 3 minutes, and all trains' initial delays are smaller than 8 minutes. Furthermore, the number of trains keeps decreasing after the initial delay value increases after 2 minutes.

The relationship between train's initial delays and their corresponding running times is given in Figure 8-4:

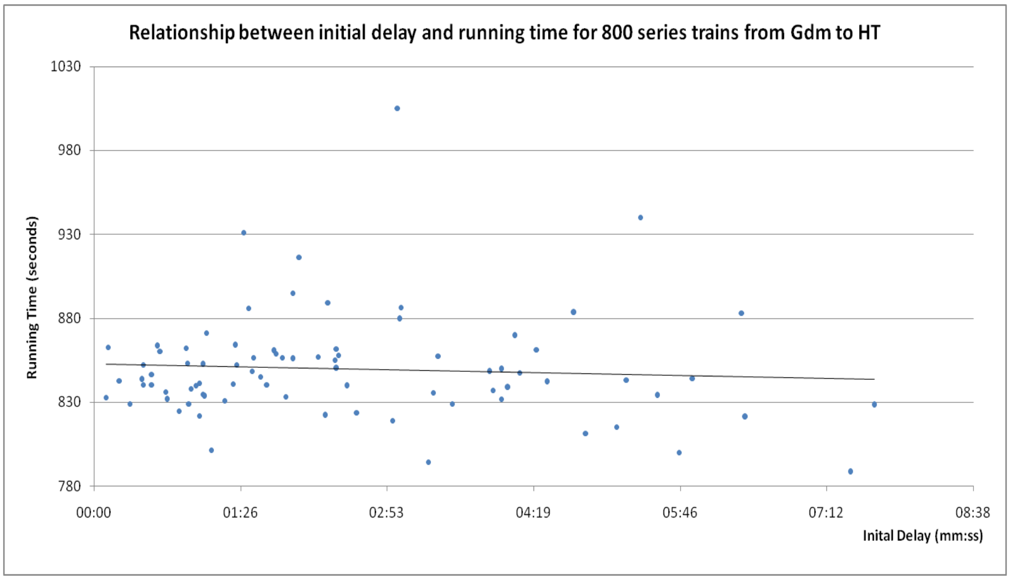


Figure 8-4 Relationships between running time and initial delay for 800 series trains running from Geldermalsen to Den Bosch

The summary of the linear regression analysis for this group is given in Table 8-2:

Table 8-2 Summary for regression analysis on 800 train series group from Gdm to Ht

Multiple R	0.070282
R Square	0.00494
Adjusted R Square	-0.00766
Standard Deviations	31.74261
Samples	81

Table 8-2 shows that R square value is only 0.00494, therefore the derived regression line can not accurately explain the relationship between train's initial delay and running times.

8.1.2. *3500 series trains running from Geldermalsen to Den Bosch*

By using the same method used in section 8.1.1, for 3500 series train running from Geldermalsen to Den Bosch, if we assume punctual trains will enter the study area on 4:00 in first half hour and 34:00 in second half hour, then 3500 series trains' initial delay distributions are shown in Figure 8-5:

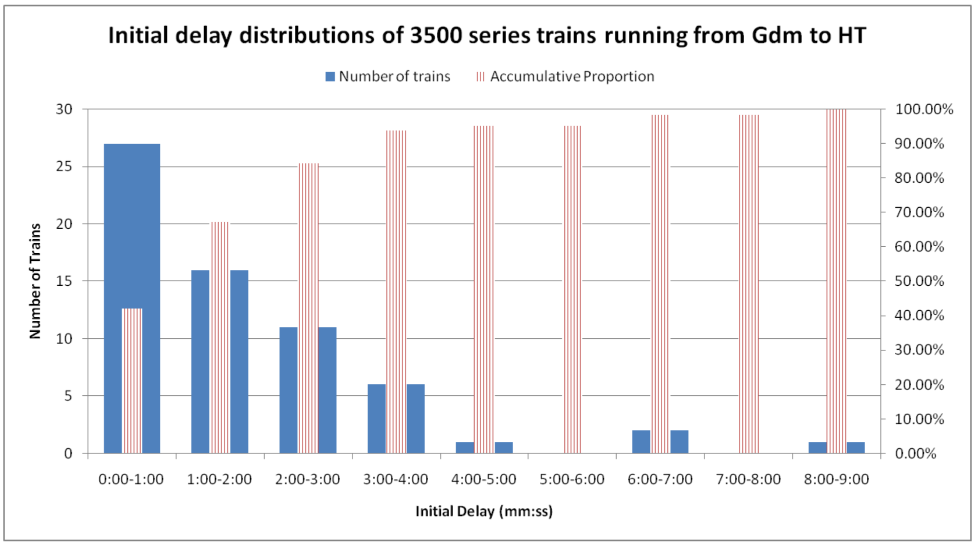


Figure 8-5 Initial delay distributions for 3500 series from UT to Ht

Comparing to Figure 8-3, Figure 8-5 shows that 3500 series train's initial delays is more dense than 800 series trains. More than 90% of trains' initial delay is lower than 4 minutes.

The relationship between train's initial delays and their corresponding running times is given in Figure 8-6:

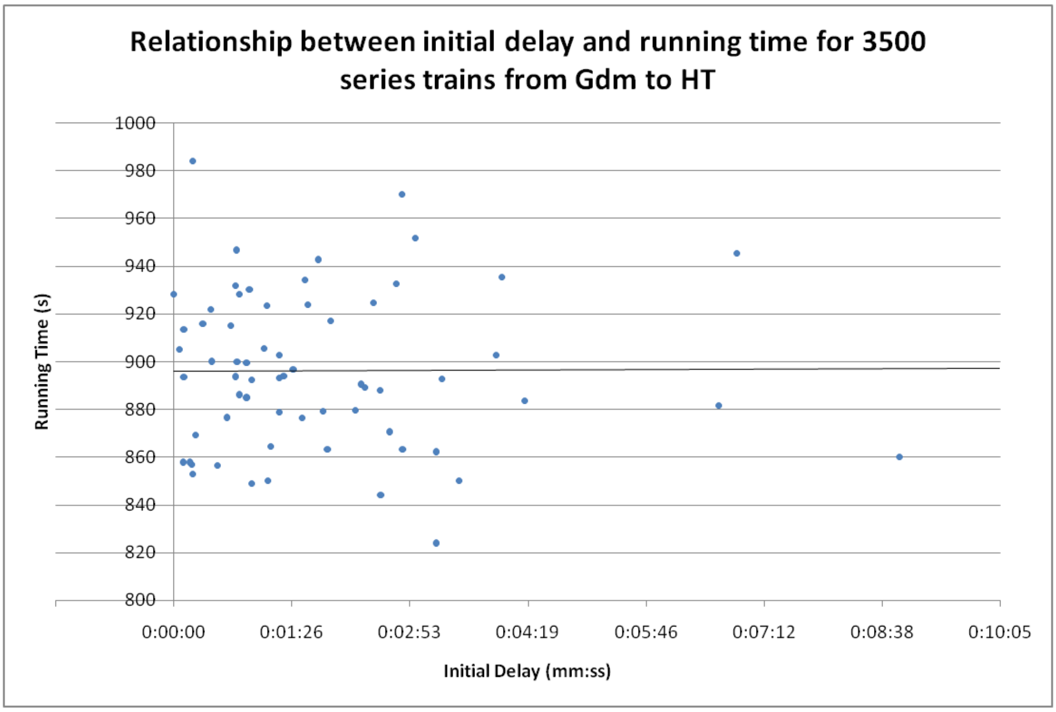


Figure 8-6 Relationship between running time and initial delay for 3500 series trains running from Geldermalsen to Den Bosch

The summary of the linear regression analysis for this group is given in Table 8-3:

Table 8-3 Summary for regression analysis on 3500 train series group from Gdm to Ht

Multiple R	0.008826905
R Square	0.000077914
Adjusted R Square	-0.015793865
Standard Deviations	33.12895621
Samples	65

Table 8-3 shows that R square value is low, therefore the derived regression line can not accurately explain the relationship between train's initial delay and running times.

8.1.3. *800& 3500 series trains running from Den Bosch to Geldermalsen*

If we assume for 800 series, punctual trains will enter the study area on 22:30 at first half hour and 52:30 in second half hour; for 500 series, punctual trains will the study area on 07:30 in first half hour and 37:30 at second half hour, then this train group's initial delay distributions are shown in Figure 8-7:

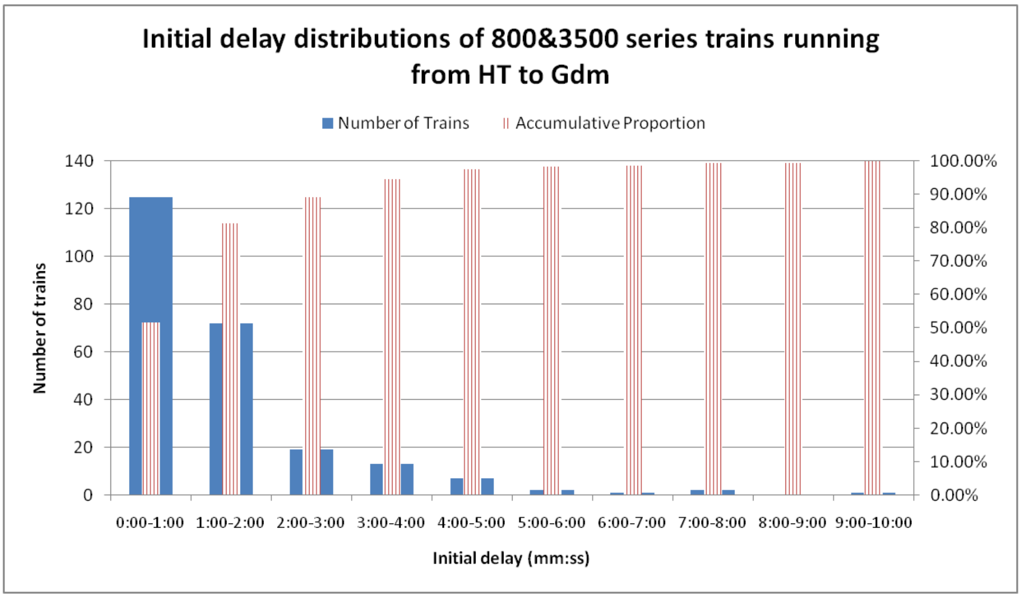


Figure 8-7 Initial delay distributions for 3500&800 series from Den Bosch to Geldermalsen

Figure 8-7 illustrates that for 800&3500 series trains running from Den Bosch to Geldermalsen, all trains' initial delays are lower than 10 minutes, and more than 80% of them are lower than 2 minutes.

The relationship between this group trains' initial delays and their corresponding running times is given in Figure 8-8:

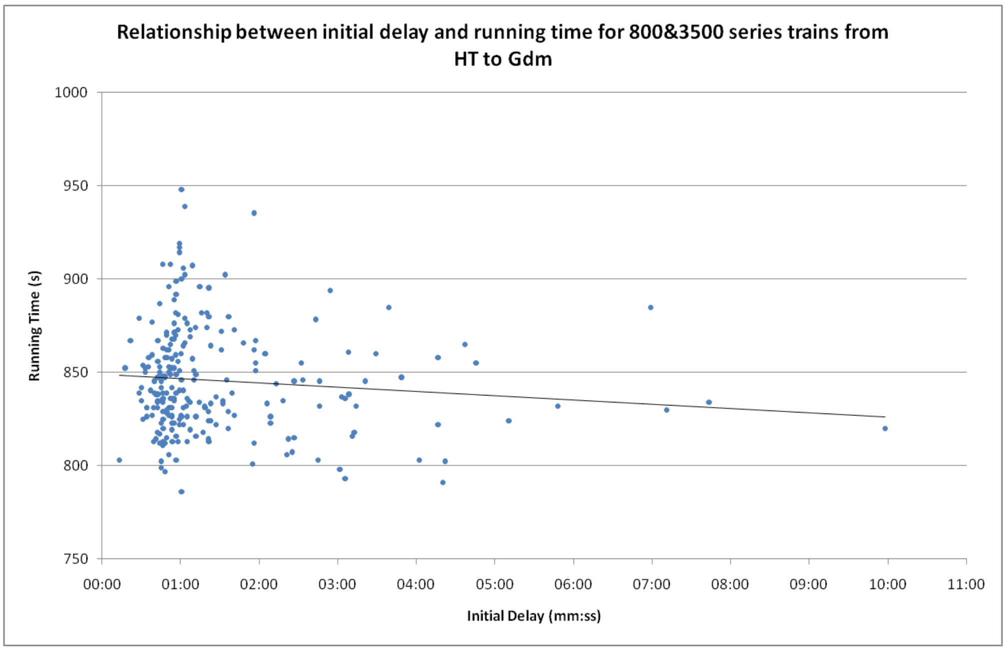


Figure 8-8 Relationship between running time and initial delay for 800&3500 series trains running from Geldermalsen to Den Bosch

In order to prove the relationship between a train's running time and initial delay in three data groups, a linear regression analysis is conducted, and 800&3500 series trains running from Den Bosch to Geldermalsen are used as test samples. The final regression line is also shown in Figure 8-8. The summary of this linear regression analysis is given in Table 8-4:

Table 8-4 Summary for regression analysis on 800&3500 train series group from Ht to Gdm

Multiple R	0.20520361
R Square	0.042108522
Adjusted R Square	0.038117307
Standard Deviations	14.41066035
Samples	242

Table 8-4 shows that R square value is only 0.042, therefore the derived regression line cannot accurately explain the relationship between train's initial delay and running times. Therefore, in this study the initial delay is not an important influencing factor for a driver to make his driving strategies. For each train group, the running time difference between the fastest train and slowest train is higher than 150 second, therefore it is not reasonable to explain this result by indicating that the slack time between Geldermalsen to Den Bosch is not sufficient from the timetable.

8.2 Speed profile segmentation

In order to do further study on a train's coasting behavior and acceleration/ braking rate usage, it is important to subdivide a train's speed profile into several segments.

For the running direction from Geldermalsen to Den Bosch, three segments are identified. This segmentation method can be illustrated by 800 series trains' speed profiles, which is given in Figure 8-9:

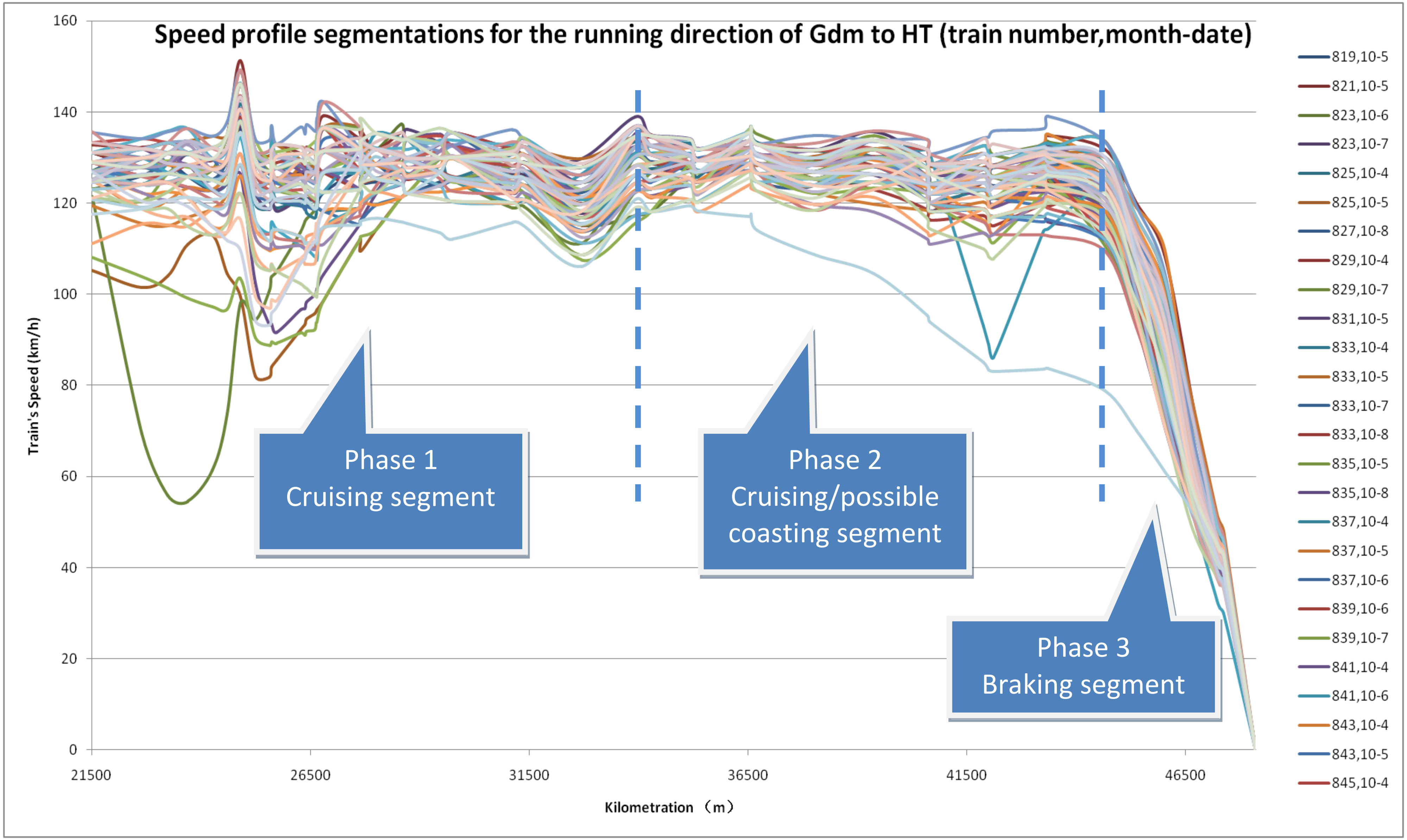


Figure 8-9 Speed profile segmentations for the running direction of Gdm to Ht

In Figure 8-9, a phenomenon is observed that all speed profiles drop down at kilometration point 32690, and climb up at 33911, after 33911, they become stable and show a general trend of going down. One reminder is that there are no signals or speed restriction boards in kilometration point 32690, and this pheomenon is due to the characteristics of the Trento data (deviations are not equally distributed along the studied rail sections).

Therefore, kilometration point 33911 can be selected as the border point for cruising segment and cruising/Possible coasting segment (location of the blue dashed line on the left side). For the railway section before 33911, the train's speed profiles are strongly influenced by the deviations of the Trento data and have many fluctuations, which are not realistic. Common sense says that, if a train driver would like to do coasting in a free running case, he would only do it before the final braking phase, instead of firstly do coasting then do cruising when there are no intermediate speed restrictions. Therefore, this railway section can be identified as the area of train's cruising segment.

One reminder is in this segment is that there are still some speed profiles indicating the corresponding trains had sudden braking operations for unknown reasons, i.e. train 829 on 7th October (the green line). Because they are not hindered by signals and also use the same route as other trains, these profiles are not excluded by this study.

In OBE drawings, a required speed limitation board with a speed limitation of 60km/h is located at kilometration point 45375 (location of the blue dashed line on the right side of Figure 8-9), and this speed limitation board also indicates to train drivers that they are about to enter Den Bosch station area. Therefore, kilometration point 45375 can be selected as the border point for cruising/coasting segment and braking segment. Figure 8-9 also shows that all trains have begun to do braking after kilometration point 45375.

For the running direction from Den Bosch to Geldermalsen, two segments can be identified. This segmentation method can be illustrated by 800&3500 series trains' speed profiles, which is given in Figure 8-10:

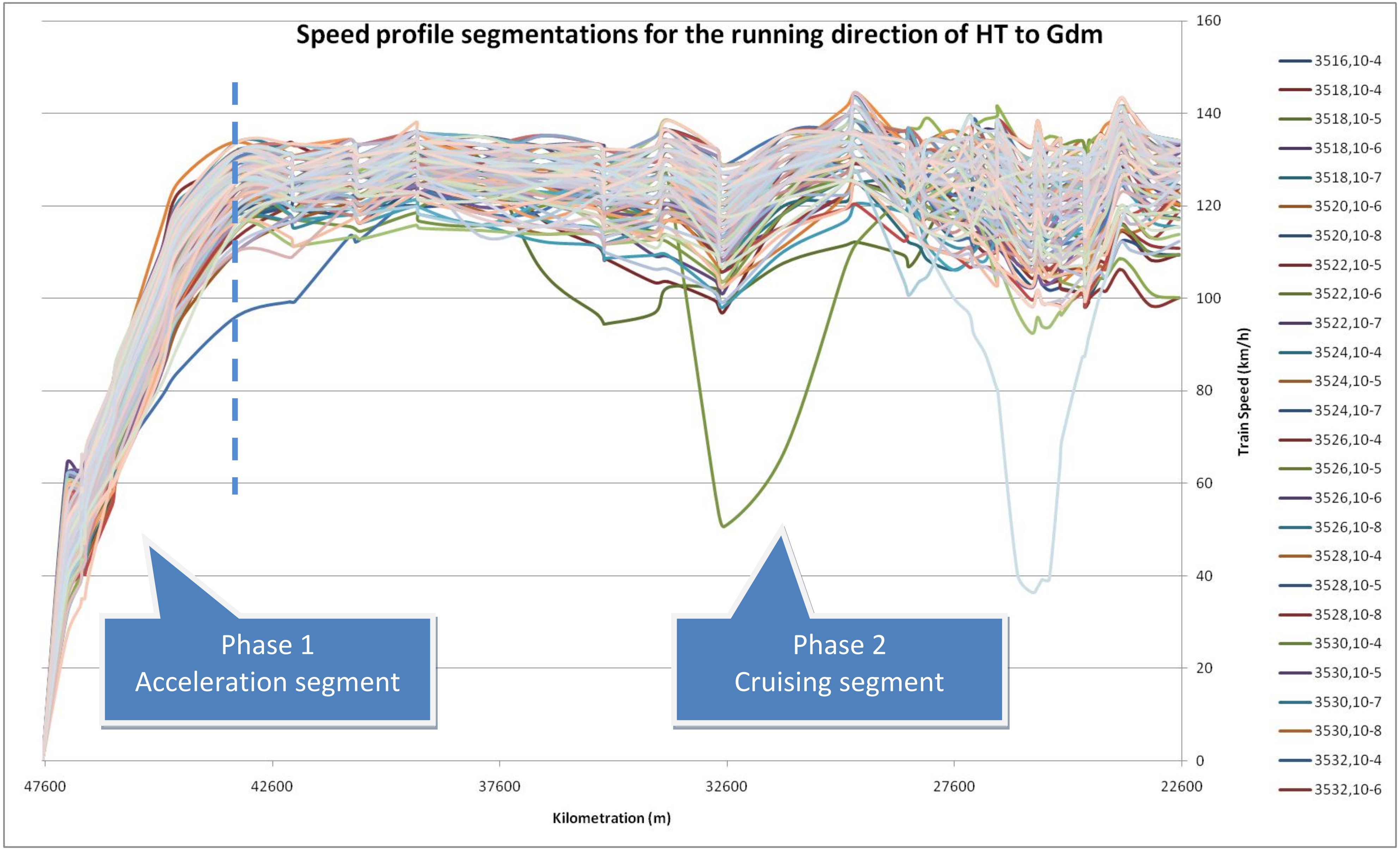


Figure 8-10 Speed profile segmentations for the running direction of Ht to Gdm

Figure 8-10 illustrates that after kilometration point of 43359, most of train’s speed become stable and do not increase as obvious as in the area of before kilometration point 43359. Therefore, kilometration point 43359 can be considered as the border point for the acceleration segment and cruising segment. For both 800 and 3500 series selected trains, they will not stop in Geldermalsen station but direct run to Utrecht station, therefore no it is not necessary to identify the cruising/possible coasting segment.

For the acceleration phase, the locations of speed limitation indication boards are given in Figure 8-11:

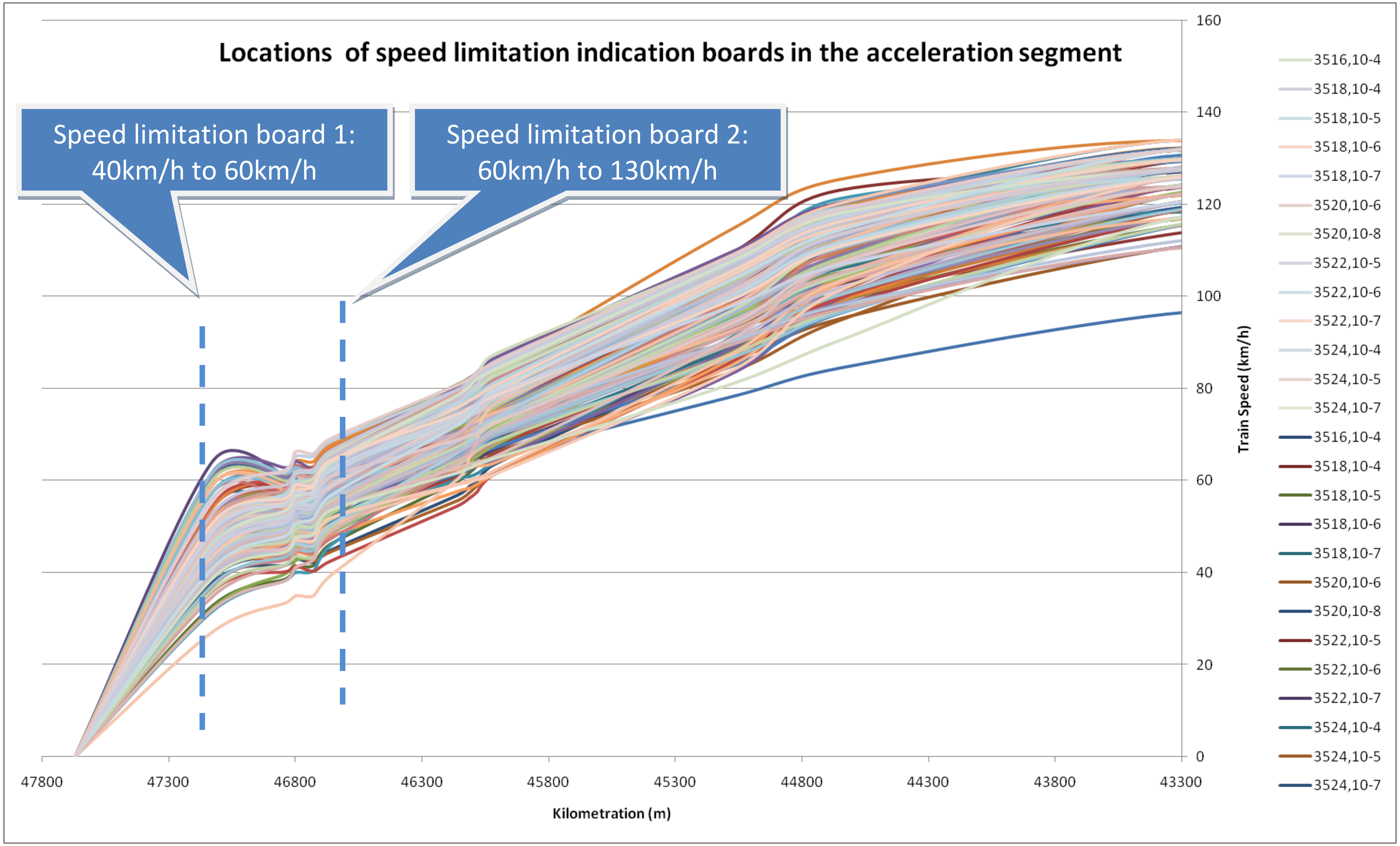


Figure 8-11 Locations of speed limitation boards in the acceleration segment

Figure 8-11 shows that the distance between two speed limitation boards only equals to (47130-46665)=465m, and the distance between the second speed limitation indication board and the end of the acceleration segment is longer than 3000 meters. The latter one seems to be much more than a train's required distance to accelerate from 60/km to 130km/h. Further investigations on selected train's acceleration rate usage are given in section 8.4.

8.3 Coasting behavior detection

For the running direction of Geldermalsen to Den Bosch, a cruising/possible coasting segment has been identified. In this segment, a coasting behavior detection method is introduced to detect if there are drivers doing coasting during their driving times. The process of this method is given in Figure 8-12

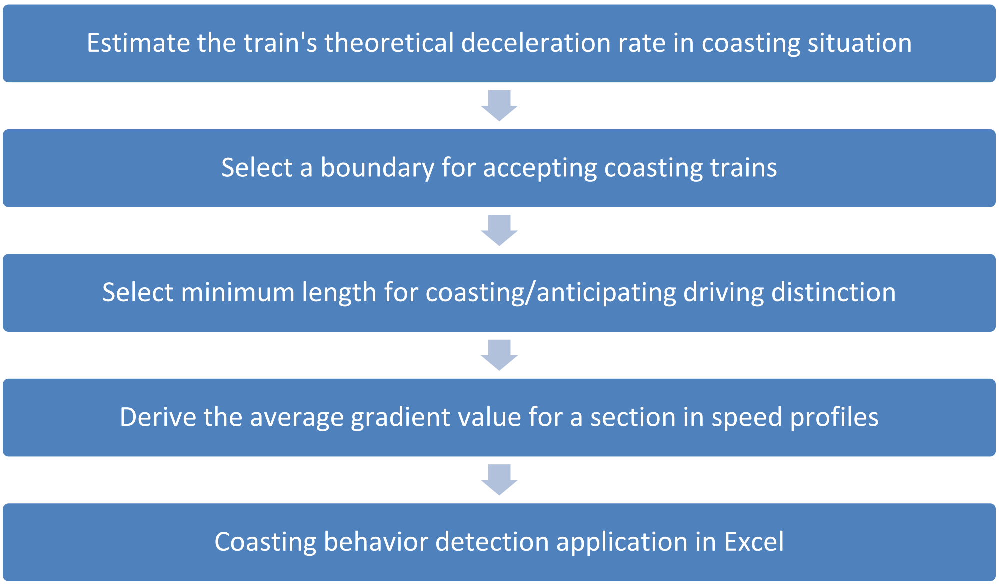


Figure 8-12 Process of coasting behavior detection method

8.3.1. *Train's theoretical deceleration rate in coasting phase*

As has been introduced in chapter 2, if a train is doing coasting, it will not experience the traction force from the engine. Therefore, it will only have two forces — train resistance and line resistance (mainly gradient resistance).

For estimating the train resistance value, formula (7-1) is used. We could see that in the coasting phase, the train resistance is relevant with the train speed in formula (7-1). In this study, if we use the train speed as 130km/h and take this speed value to formula (7-1), then we can derive the correspond train's deceleration rate for a coasting train running on a flat rail (gradient=0) is -0.04m/s^2 .

The line resistance can be calculated by formula (2-6). Therefore, a train's theoretical deceleration rate (the train's speed may increase or decrease, depends on the gradient value) in coasting situation $a_{coasting}$ is given by formula (8-1):

$$a_{coasting} = F_{TR} / m + F_{Rlg} / m = a_{trainresistance} + g * n \qquad (8\text{-}1)$$

Where $a_{trainresistance}$ is assumed as -0.04m^2 /s in this study
n is the gradient value

One reminder is that for calculating F_{TR} in this study, 130km/h is used as the standard speed value.

8.3.2. Boundary for accepting coasting trains

However, as the train resistance value is also influenced by speed, $a_{trainresistance}$ may differ from -0.04m^2 /s; Additionally, there are also deviations between different train types. Therefore, an interval for identifying that if a train is doing coasting is created, and is given in Figure 8-13:

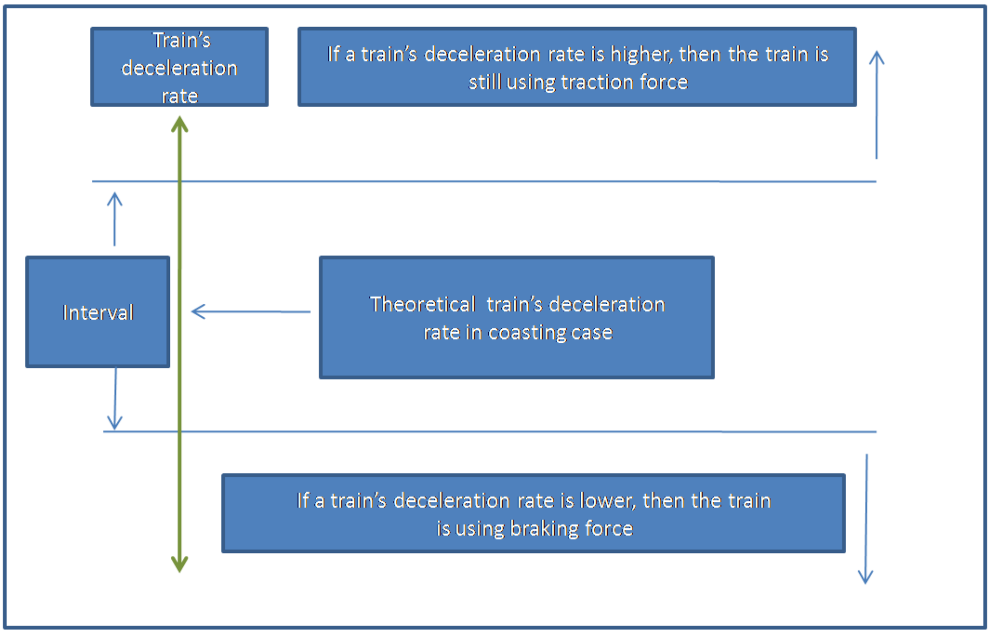


Figure 8-13 The principle used in detecting coasting trains

Figure 8-13 illustrates the principle that used in detecting coasting trains, and the usage of the boundary. In Figure 8-13, a green line is used to present the reduction rate. In a given section, there are three circumstances:

a) If a train's deceleration rate is higher than the upper bound, then this train is identified as it is still using the traction force;
b) If a train's deceleration rate is lower than the lower bound value, then the train will be identified as using braking forces;
c) If a train's deceleration rate is between the upper and the lower bound, then this train will be considered as a coasting train.

The procedure of choosing the boundary is a matter of careful choice. If this boundary is flexible, then more trains will be considered as coasting trains comparing to the situation in which the boundary is tight. In this study, the lower boundary is set to 0.5 times of the train's theoretical deceleration rate, while the upper boundary is set to 2 times of the train's theoretical deceleration rate ([0.5* $a_{coasting}$, 2* $a_{coasting}$]).

8.3.3. *Minimum length for coasting/anticipating driving distinction*

In reality, a driver may turn off power for a short distance in order to avoid sudden braking. For example, if a driver's sight distance is long enough to see the second signal ahead, and this signal is not showing green, then the driver may turn off power to slow down train's speed, and expect the signal ahead will turn green before the train enters that block. This kind of anticipating driving should not be considered as coasting eventhough the two driving behavior have the same effect in energy consumption, because the anticipating driving aims at avoiding sudden braking instead of saving energy. Additionally, this kind of anticipating driving will not turn off engine for a long distance, which is different from real coasting.

Therefore, the minimum length for coasting/anticipating driving is related to the block length. For the direction of Geldermalsen to Den Bosch, the average length for one block is 1165m, and in the other direction this value is 1189m. In this study, for both directions, this minimum value is set as 2 times of an average block length.That is, for direction of Geldermalsen to Den Bosch is set to 2330m, while for the other direction it is set to 2378m.

8.3.4. *Average gradient determination and method application in Excel*

Because the locations of gradient are not relevant with track section divisions, the average gradient value for the section have to be recalculated.

The application of this method in Excel can be exemplified by Table 8-5:

Table 8-5 Examples of coasting behavior detection method in Excel

Train	Starting point	Ending point	Gradient	Actual braking rate	Theoretical braking rate	Coasting detection	Running phase
881,10-5	33340	34482	0	0.104445583	-0.04	0	2
881,10-5	33690	34890	-3.90385	-0.06842098	-0.000961478	0	2
881,10-5	34482	35844	0	0.012390401	-0.04	0	2
881,10-5	34890	35844	0	-0.05427445	-0.04	1	2
881,10-5	35844	37260	0	-0.00906524	-0.04	0	2
881,10-5	35870	37260	0	0.329458526	-0.04	0	2
881,10-5	36150	37260	0	-0.25797515	-0.04	0	2
881,10-5	37260	38743	0	-0.03562719	-0.04	1	2

881,10-5	38743	40009	0	-0.02447728	-0.04	1	2
881,10-5	40009	41165	1.272832	-0.05837101	-0.052728324	1	2
881,10-5	40145	41165	0	-0.18604742	-0.04	0	2
881,10-5	41165	42585	3.988525	-0.05001773	-0.079885246	1	2
881,10-5	41425	42585	3.424615	-0.09397718	-0.074246154	1	2
881,10-5	41640	42585	-3.46944	-0.01950553	-0.005305556	0	2
881,10-5	42585	43927	-3.83806	0.003116662	-0.001619423	0	2
881,10-5	42791	43927	0	0.014364418	-0.04	0	2

One reminder in Table 8-5 is that the actual Excel sheet have more columns.

As is shown in Table 8-5 for a train run, each section's occupancy data will be recorded in one row. For each section, its average gradient, theoretical deceleration rate are calculated automatically. If a train is doing coasting, a value "1" will be assigned under the column of "braking detection", otherwise a "0" will given. After that, the VBA program will check if this train's total continuous coasting length is longer than the minimum length for coasting behavior identification.

Finally, for all trains running from Geldermalsen to Den Bosch, only train 881 on 5th October is identified as coasting trains, and its coasting area is indicated by blue color in Table 8-5 (rail sections between 37260 and 41165). Further discussion on this train's driving behavior will be given in section 8.5.

8.4 Train's acceleration/ braking rate usage

8.4.1. Train's braking rate usage in the direction of Geldermalsen to Den Bosch

For the running direction of Geldermalsen to Den Bosch, trains will first do cruising and then do braking within our study area; therefore, only trains' braking rate usage can be investigated in this direction.

An indicator of "average braking rate" is introduced to represented a train's braking rate usage in the braking segment. This average braking rate $b_{average}$ is given in formula (8-2):

$$b_{average} = L_{brakingsegment} * 2 / t^2_{brakingsegment} \qquad (8\text{-}2)$$

Where $L_{brakingsegment}$ is the length of the braking segment area [m];

$t_{brakingsegment}$ is train's running time in the braking segment

Formula (8-2) indicates that in this method, trains are assumed to do constant braking in the braking segment.

For 800 series trains, their average braking rates are shown in Figure 8-14:

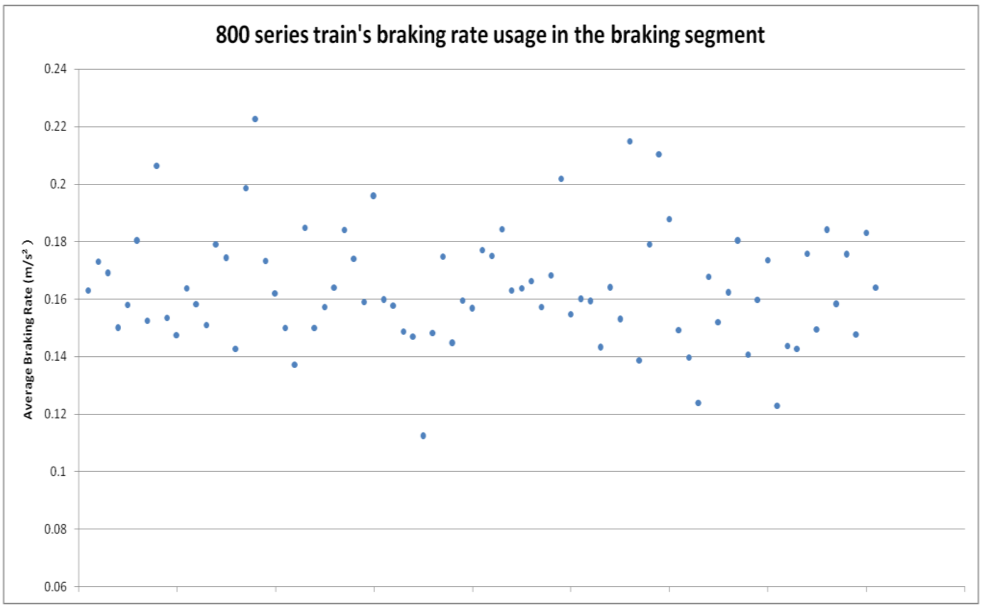


Figure 8-14 800 series trains' average braking rates in the braking segment

One reminder in Figure 8-14 is that the x-axis does not contain any information but only used for more clear representation of the braking rate values.

Figure 8-14 indicates that the average braking value ranges from 0.24 m/s² to 0.11 m/s², which is much lower than the train's braking rate setting in OpenTrack (0.66m/s²). The reason for this phenomenon is that in reality, trains will not always keep braking in the braking segment, but do a combined braking/cruising instead. This actual behavior is exemplified in Figure 8-15:

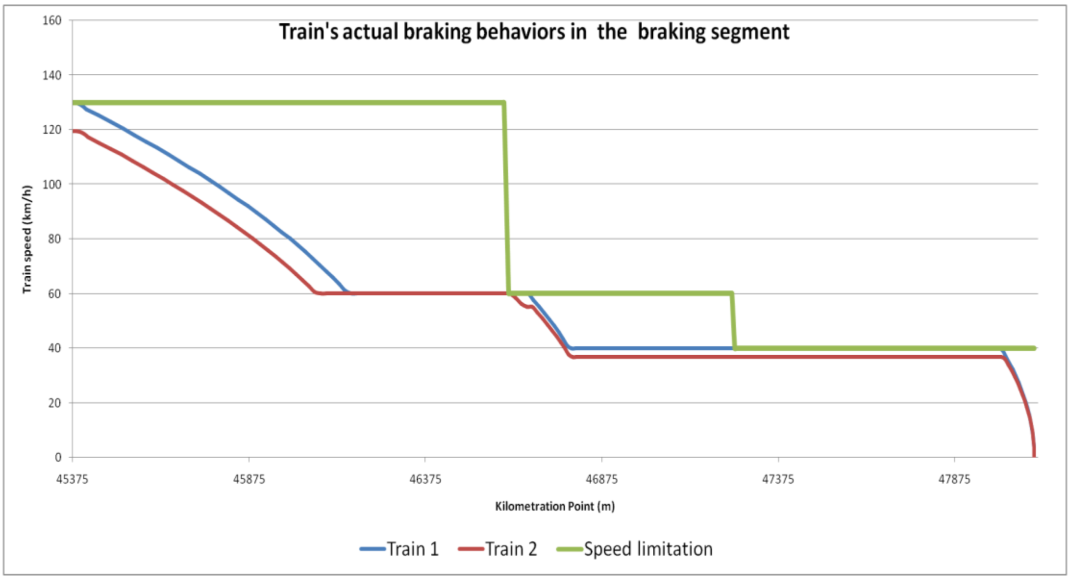


Figure 8-15 Train's actual braking behavior in the braking segment

In Figure 8-15, two train runs are shown.These two train runs are created by OpenTrack: train 1's performance factor is set as 100%, while train 2's is set as 92%. Both of them show a combined braking/cruising driving behavior during the study area.

Therefore, in this study's average braking rate calculation, the length for a train to do actual cruising will strongly influence the average braking rate value. But as the objectives of using this indicator is to compare different train's braking rates, this low value estimation would not influence the result of driving behavior analysis.

Finally, 800 series trains' average braking rate distributions are given in Figure 8-16:

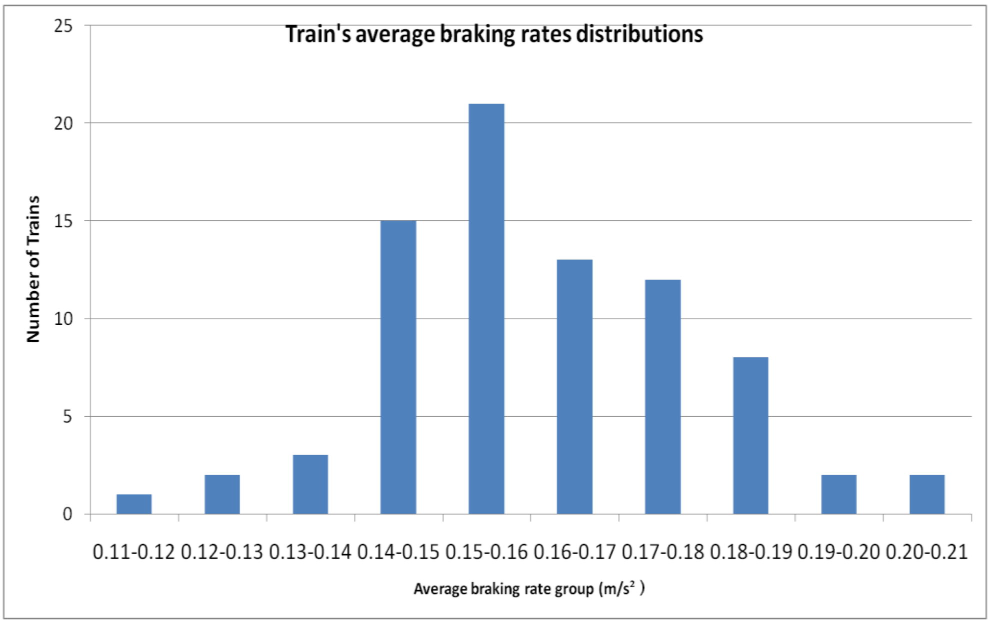


Figure 8-16 800 series train's average braking rate distribution

Figure 8-16 indicates the average braking rate distribution for 800 series trains. The average braking rate group with a range of [0.15,0.16] contains highest number of trains, and most of trains' average braking rate are located in the range of [0.14, 0.19] [m/s^2].

For 3500 series trains, their average braking rates are shown in Figure 8-17:

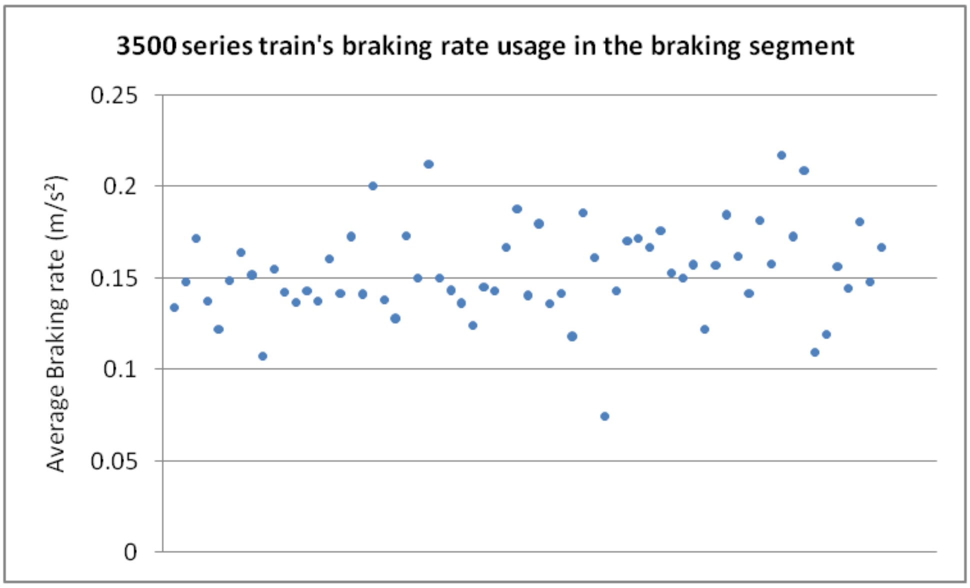


Figure 8-17 3500 series trains' average braking rates in the braking segment

One reminder in Figure 8-17 is that the x-axis does not contain any information but only used for more clear representation of the braking rate values.

Figure 8-17 indicates that the average braking rates of 3500 series trains vary in the same range as that of the 800 series trains. Although 800 series trains and 3500 series trains use different routes, they have the same running distance (this study assumes the halting points for both train series in Den Bosch station is kilometration point 48100), therefore, these two train series' speed profiles should be comparable.

Additionally, 800 series trains' average braking rate distributions are given in Figure 8-18:

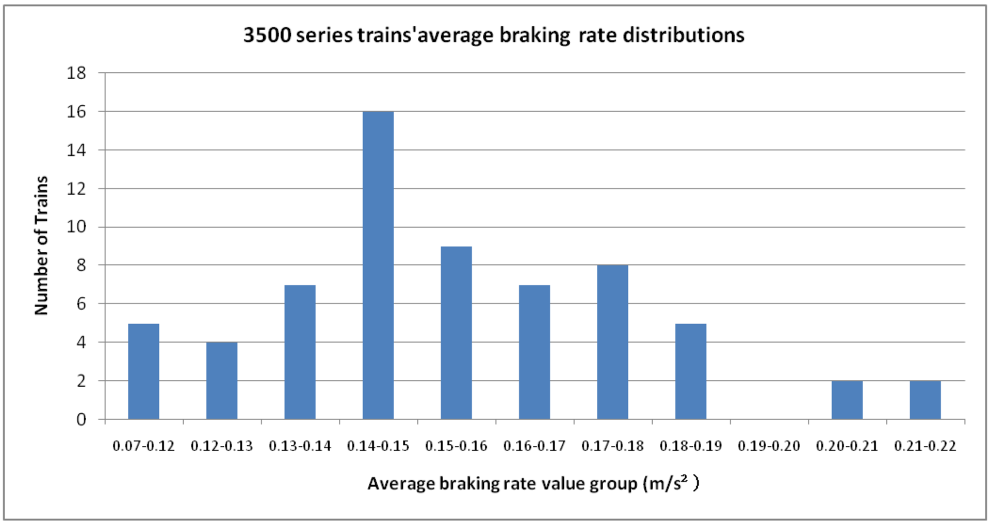


Figure 8-18 3500 series trains' average braking rate distribution

Figure 8-18 illustrates that for 3500 series trains, the average braking rate group of [0.14, 0.15] m/s^2 has the highest number of trains. Most of trains are located with the area of [0.12, 0.19].

8.4.2. *Train's acceleration rate usage in the direction of Den Bosch to Geldermalsen*

For the running direction of Den Bosch to Geldermalsen, because the Utrecht station is out of selected study area, only trains' acceleration rate usage status can be investigated.

An indicator of "average acceleration rate" is introduced to represented a train's acceleration rate usage in the acceleration segment. This average acceleration rate $a_{average}$ is given in formula (8-3):

$$a_{average} = L_{brakingsegment} * 2 / t^2_{brakingsegment} \qquad (8\text{-}3)$$

Where $L_{brakingsegment}$ is the length of the acceleration segment area [m];

$t_{brakingsegment}$ is train's running time in the acceleration segment

Formula (8-3) indicates that in this method, trains are assumed to do constant acceleration in the acceleration segment. But chapter 2 and chapter 4 have explained that this constant acceleration assumption is not realistic, because a train's traction force is also determined by train's speed at the moment.

For 800&3500 series trains, their average braking rates are shown in Figure 8-19:

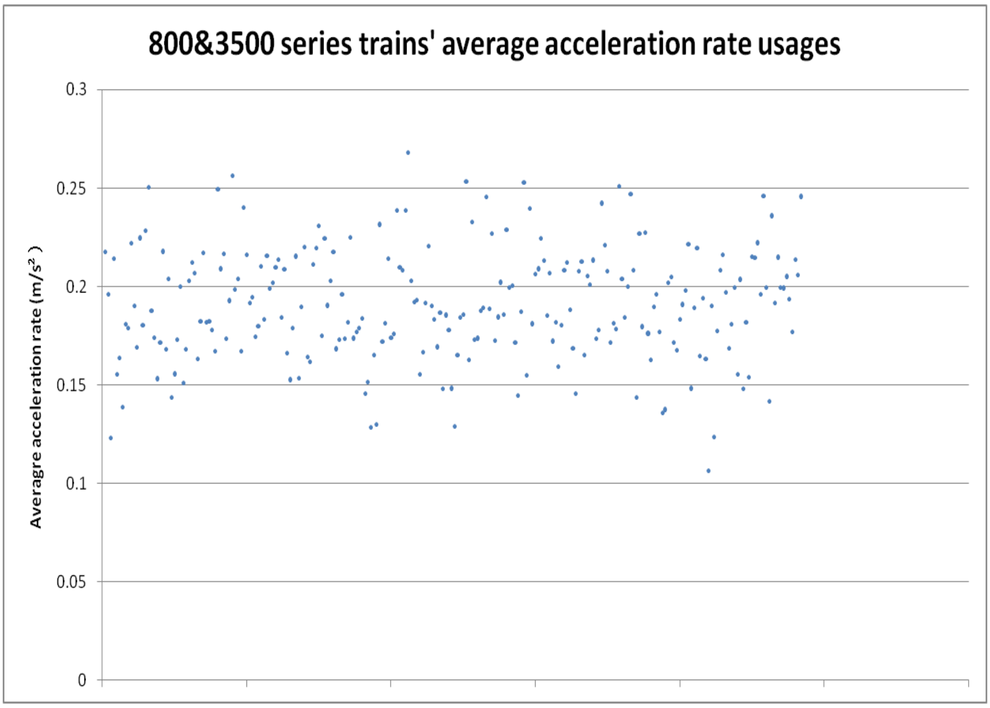


Figure 8-19 800&3500 series trains' average acceleration rate

Figure 8-19 illustrates that for 800&3500 train series running from Den Bosch to Geldermalsen, their average acceleration rates vary from 0.1m/s^2 to 0.27 m/s^2. As in normal cases, a train's maximum acceleration rate can never be higher than 0.5 m/s^2, the average acceleration rate derived from this study is very low, and the reason for this phenomenon is similar to has been discussed in section 8.4.1.

A train's actual speed profile in the acceleration segment can be exemplified in Figure 8-20:

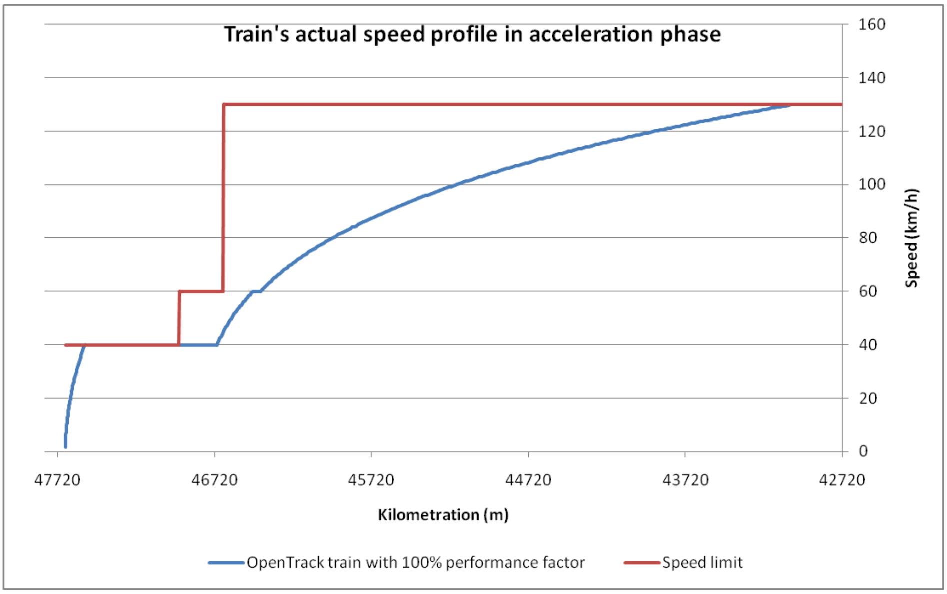


Figure 8-20 Train's actual braking behavior in the braking segment

Figure 8-20 shows that the method used for train's average acceleration rate calculation does not exclude train's cruising behavior, therefore the derived value is quite low comparing to a train's normal maximum acceleration rate.

800&3500 series trains' average acceleration rate distributions is given in Figure 8-21:

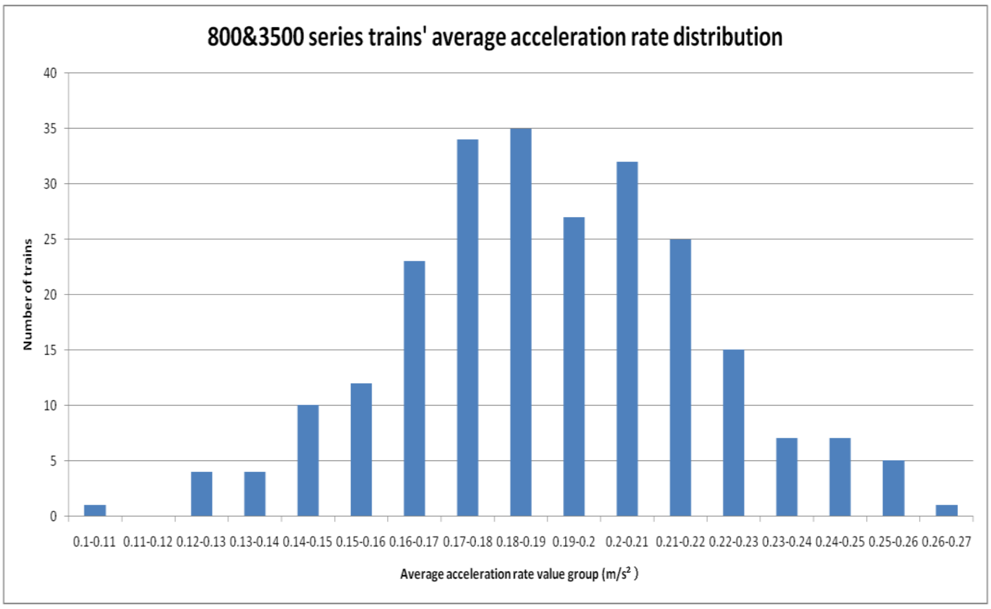


Figure 8-21 800&3500 series trains' average acceleration rate distribution

Figure 8-21 illustrates that for 800& 3500 series trains, the average acceleration rates generally shows a normal distribution, and most of train's average acceleration rate are between 0.15m/s^2 and 0.23 m/s^2. For the value group between 0.18 m/s^2 and 0.19 m/s^2, it has the highest number of trains.

8.5 Train's cruising speed usage

8.5.1. *Train's cruising speed usage in the direction of Geldermalsen to Den Bosch*

For 800 series trains, their cruising speed distribution in the cruising phase is given in Figure 8-22:

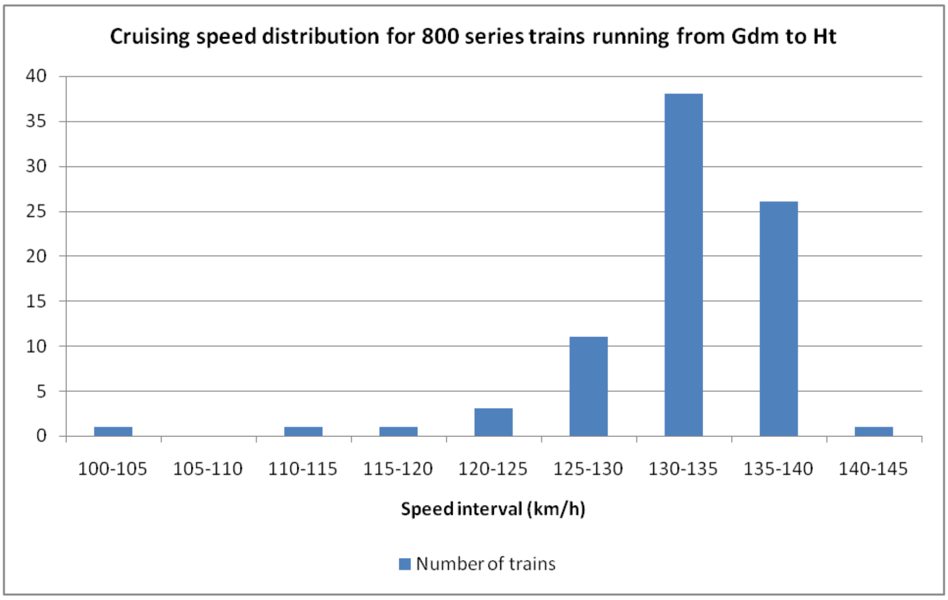


Figure 8-22 Cruising speed distribution for 800 series trains running from Gdm to Ht

Figure 8-22 indicates that most of trains' cruising speeds are within the interval of [120km/h, 140km/h]. One reminder is that the cruising speeds higer than 130/km is mainly due to the deviations of the Trento data.

The Relationship between a train's cruising speed in the cruising phase and braking rate in the braking phase is given in Figure 8-23:

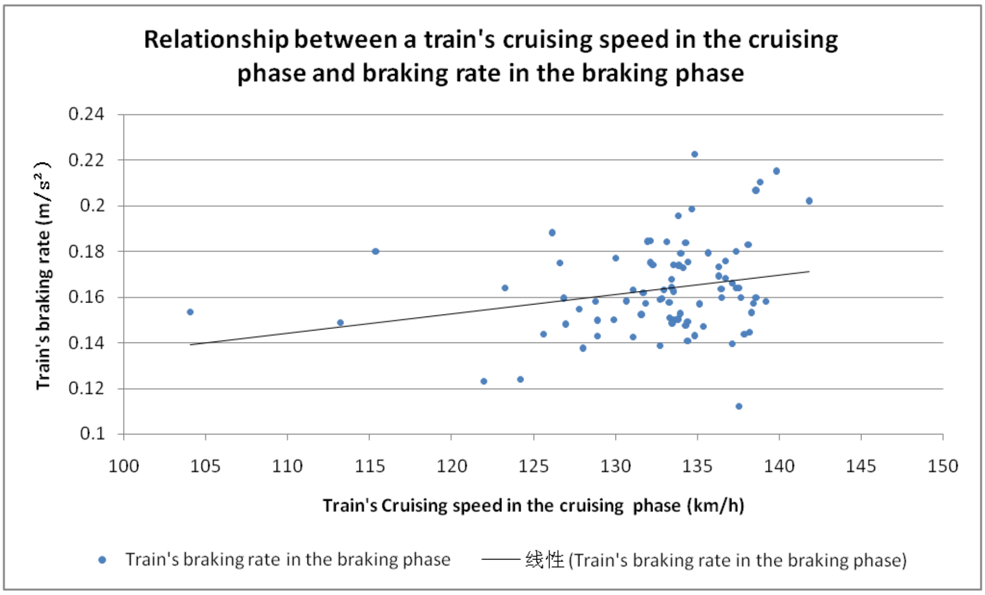


Figure 8-23 The Relationship between a train's cruising speed in the cruising phase and braking rate in the braking phase

Additionally, the regression analysis result of this relationship is given in Table 8-6:

Table 8-6 Regression analysis result of the Relationship between a train's cruising speed in the cruising phase and braking rate in the braking phase

Multiple R	0.24956
R Square	0.06228
Adjusted R Square	0.050559
Standard Deviations	0.019609
Samples	82

Figure 8-23 and Table 8-6 indicate that the relationship between a train's cruising speed in the cruising phase and braking rate in the braking phase is lacking because the R Square value is low. This result is also found in the 3500 train series group running from Geldermalsen to Den Bosch.

8.5.2. *Train's cruising speed usage in the direction of Geldermalsen to Den Bosch*

For all the trains running from Den Bosch to Geldermalsen, their cruising speed distribution in the cruising phase is given in Figure 8-24:

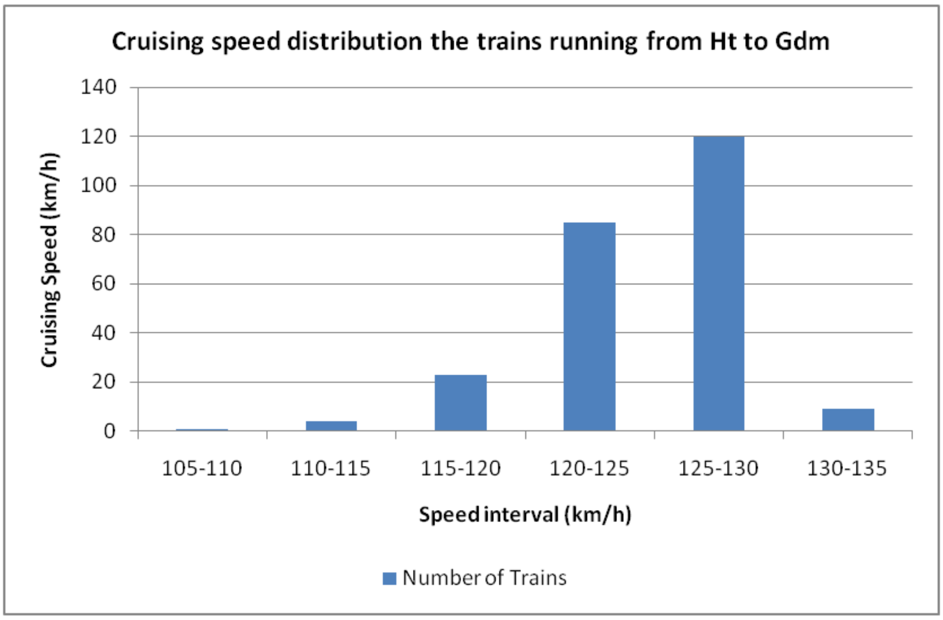


Figure 8-24 Cruising speed distribution the trains running from Ht to Gdm

Figure 8-24 indicates that most of trains' cruising speeds are within the interval of [115km/h, 130km/h], which is lower than the trains running in the other direction. One reminder is that the cruising speeds higer than 130/km is mainly due to the deviations of the Trento data.

The Relationship between a train's cruising speed in the cruising phase and acceleration rate in the acceleration phase is given in Figure 8-25:

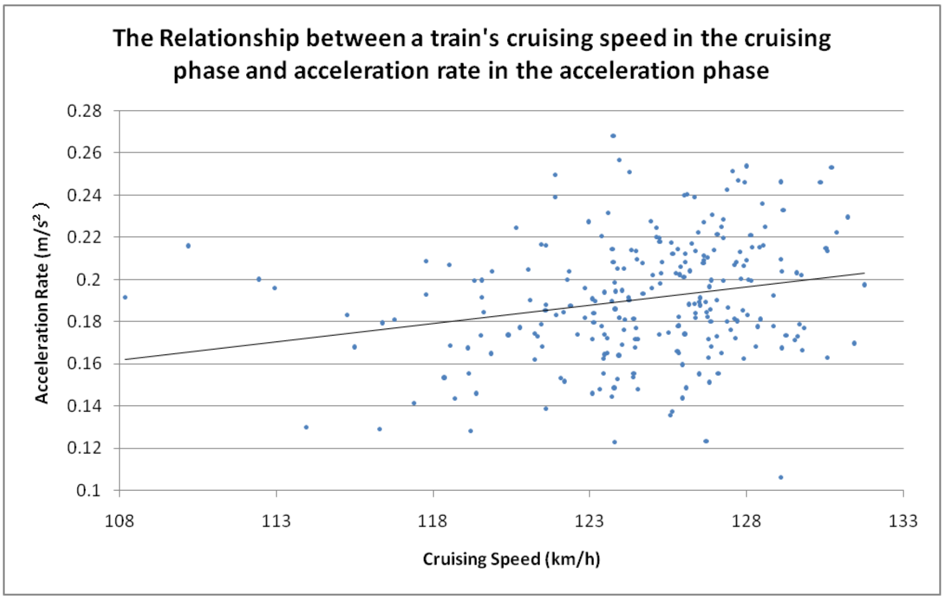


Figure 8-25 The Relationship between a train's cruising speed in the cruising phase and acceleration rate in the acceleration phase

Additionally, the regression analysis result of this relationship is given in Table 8-7:

Table 8-7 Regression analysis result of the Relationship between a train's cruising speed in the cruising phase and braking rate in the braking phase

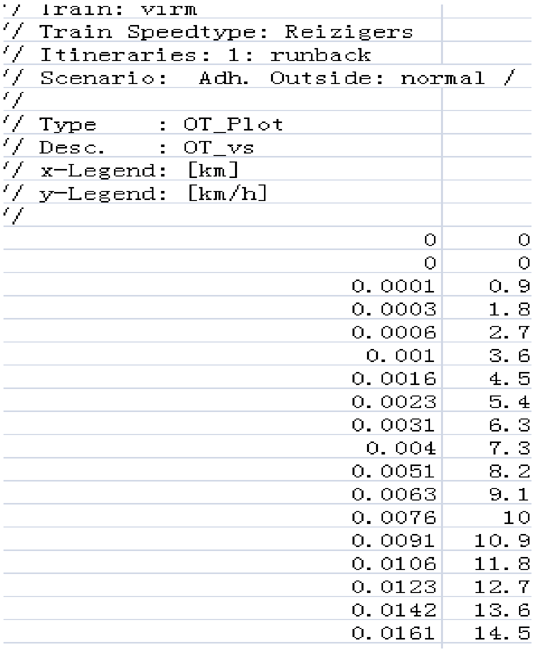
Multiple R	0.227151
R Square	0.051598
Adjusted R Square	0.047646
Standard Deviations	0.027925
Samples	242

Figure 8-25 and Table 8-7 indicate that the Relationship between a train's cruising speed in the cruising phase and acceleration rate in the acceleration phase is lacking, because the R Square value is low.

9 OpenTrack modelling abilities

In this section, OpenTrack’s modelling ability in modelling a train’s coasting, acceleration and braking behavior will be discussed separately.

In OpenTrack, the simulation output can be presented in many formats, for example, time/distance files and speed/distance files. In this study, the speed /distance output format (vs file) is selected, because this type of output can be directly used to construct the speed profile. The format of vs file is exemplified in Figure 9-1:



```
'/ Train: virm
'/ Train Speedtype: Reizigers
'/ Itineraries: 1: runback
'/ Scenario:  Adh. Outside: normal /
'/
'/ Type     : OT_Plot
'/ Desc.    : OT_vs
'/ x-Legend: [km]
'/ y-Legend: [km/h]
'/
                                     0      0
                                     0      0
                                0.0001    0.9
                                0.0003    1.8
                                0.0006    2.7
                                 0.001    3.6
                                0.0016    4.5
                                0.0023    5.4
                                0.0031    6.3
                                 0.004    7.3
                                0.0051    8.2
                                0.0063    9.1
                                0.0076     10
                                0.0091   10.9
                                0.0106   11.8
                                0.0123   12.7
                                0.0142   13.6
                                0.0161   14.5
```

Figure 9-1 Example of vs files

Figure 9-1 illustrates that in a vs file, value records of speed and distance are based on the simulation steps, for example, 1 or 0.5 second. Therefore, the user could not “request” OpenTrack to record the train’s speed on a specific kilometration point. Furthermore, in a vs file, it will not show an indication of the section, because the recorded distance is only represented in length values instead of sections.

In order to make a fair comparison between the Trento data and OpenTrack result, the OpenTrack result is also grouped by the grouping method, which is introduced in chapter 6.

9.1 OpenTrack modelling ability for coasting behavior

Section 8.3 has proved that among the investigated train runs of this study, there is only one train doing coasting - train 881 on 5th October.

In OpenTrack, a coasting board (signal) can be used to force the simulated train to do coasting after this coasting board. When a train has passed this board, it can not use the engine for acceleration and cruising again. Therefore, a coasting board is set at the beginning of the cruising/possible coasting segment. But one reminder is that as a coasting board has been put in the infrastructure model, it will affect all the trains encountering this board.

The speed profiles for OpenTrack coasting trains and train 881 on 10-5 is given in Figure 9-2:

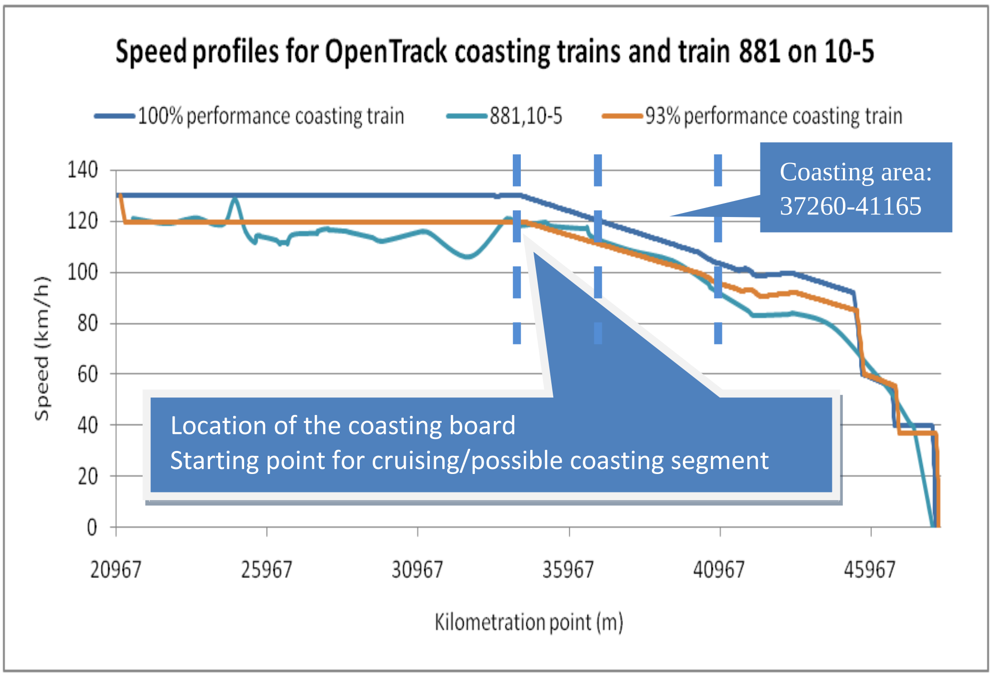


Figure 9-2 Speed profiles for OpenTrack coasting trains and train 881 on 10-5

One reminder in Figure 9-2 is that for the OpenTrack data, the grouping method is not applied. But as proved in section 6.7, the grouping method will not influce the speed profiles a lot in the area indicated in Figure 9-2.

In OpenTrack simulation, two train runs are simulated, with a performance factor value of 100% and 93%, respectively. The purpose of setting a 93% performance coasting train is to make sure that it has the same entrance speed as train 881 at the entrance point of the cruising/possible coasting segment.

The yellow area indicates the coasting section for train 881.Figure 9-2 illustrates that the constructed OpenTrack model has a good ability in modelling a train’s coasting behavior itself. However, this coasting board setting is not suitable in practical because the coasting board will influence all trains that encounter this board, while in reality, not all trains will do coasting.

9.2 OpenTrack modelling ability for acceleration/braking

After the OpenTrack result is also grouped by the grouping method used for the Trento data, the modelled train's average acceleration rate and braking rate could also be derived by using same calculation method. Their values' comparison with the Trento data is given in Figure 9-3, Figure 9-4and Figure 9-5 separately:

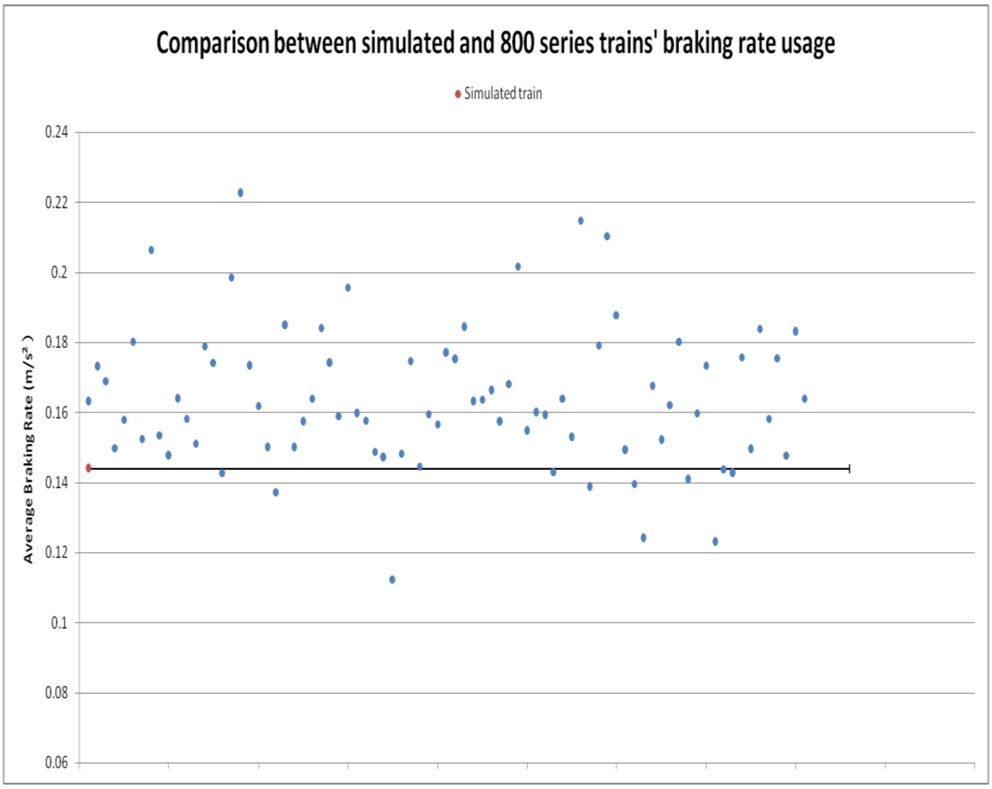


Figure 9-3 Comparison between simulated and 800 series trains' braking rate usage

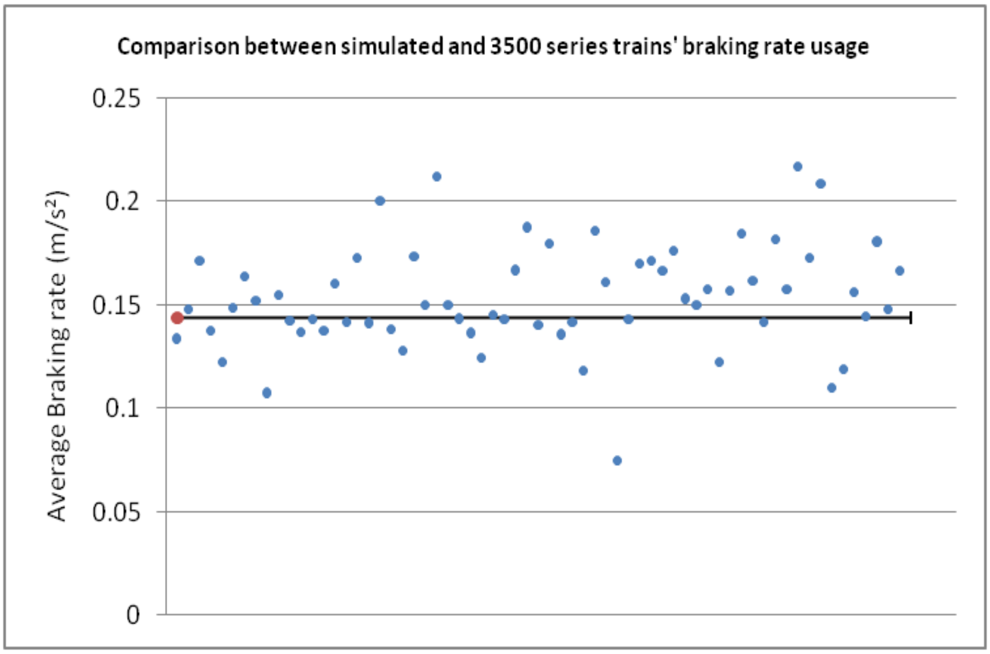


Figure 9-4 Comparison between simulated and 3500 series trains' braking rate usage

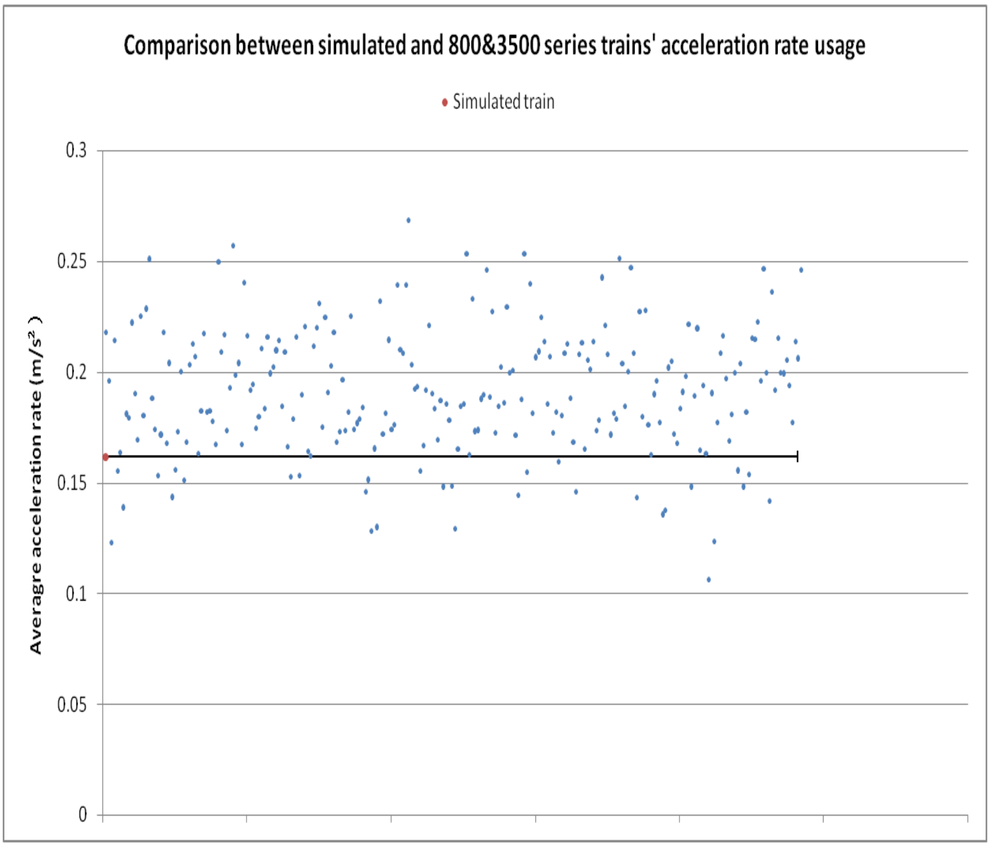


Figure 9-5 Comparison between simulated and 800&3500 series trains' acceleration rate usage

One reminder in Figure 9-3, Figure 9-4 and Figure 9-5 is that the x-axis does not contain any information but only used for more clear representation of the braking rate values.

Figure 9-3, Figure 9-4 and Figure 9-5 illustrates that the simulated train's acceleration/braking rate values derived from the model in chapter 7 are lower than the medium values of that in the Trento data.

For the braking situation, there could be two reasons to this phenomenon: firstly, in reality, train drivers could like to use higher braking rate than the setting of OpenTrack; secondly, train drivers would like to use a lower braking rate and avoid cruising in the braking phase. In this case, train drivers still spent less time than the combined braking/cruising behavior and the derived average braking rates are higher than that of OpenTrack model.

For the acceleration situation, there is only one reason that in reality, train drivers would like to use a higher acceleration rate than the setting of the OpenTrack model.

Finally, Comparisions of the speed profiles derived from the Trento Data and OpenTrack (for both directions) are given as Figure 9-6 and Figure 9-7 respectively:

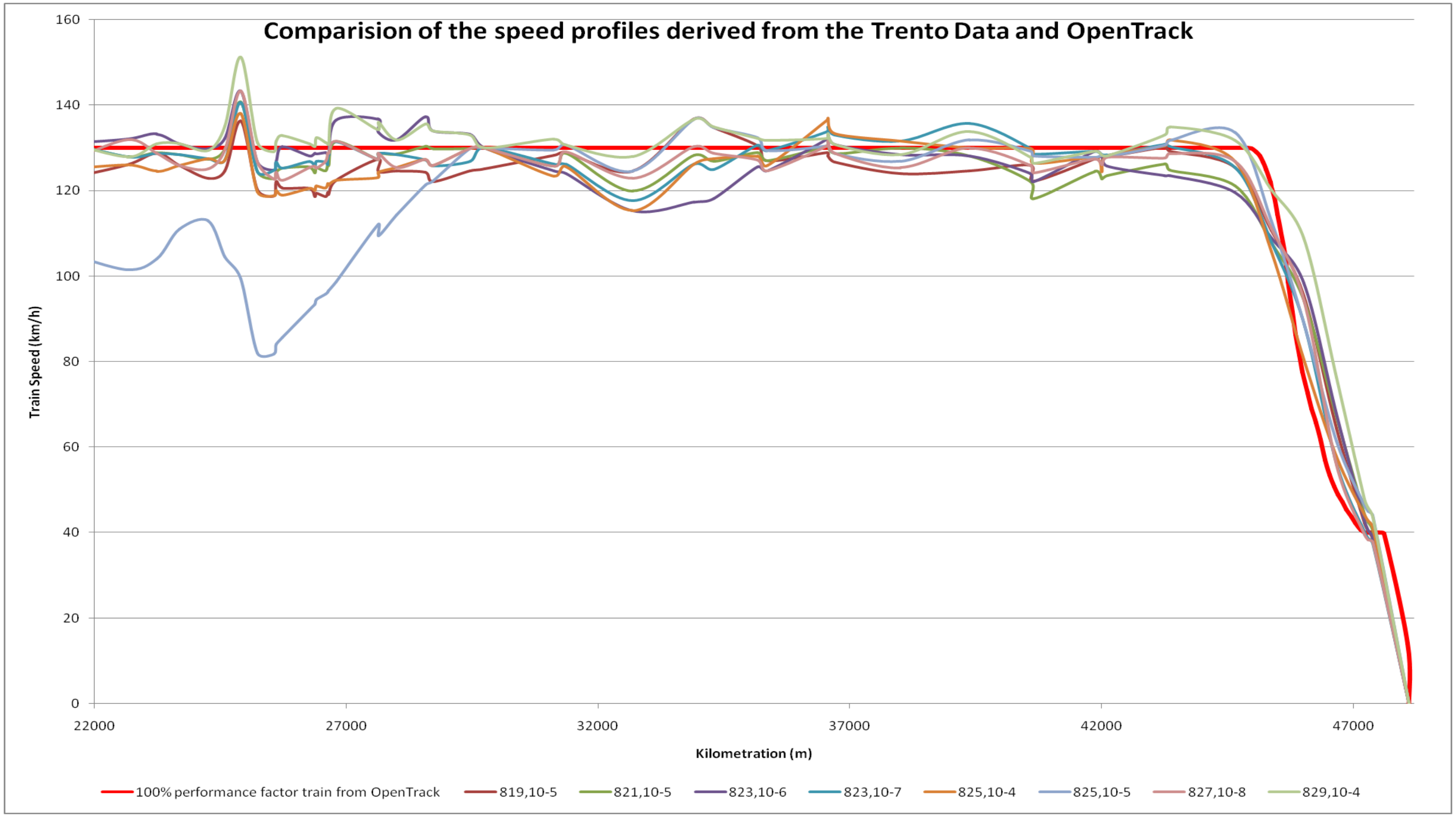


Figure 9-6 Comparision of the speed profiles derived from the Trento Data and OpenTrack for the direction of Gdm to Ht

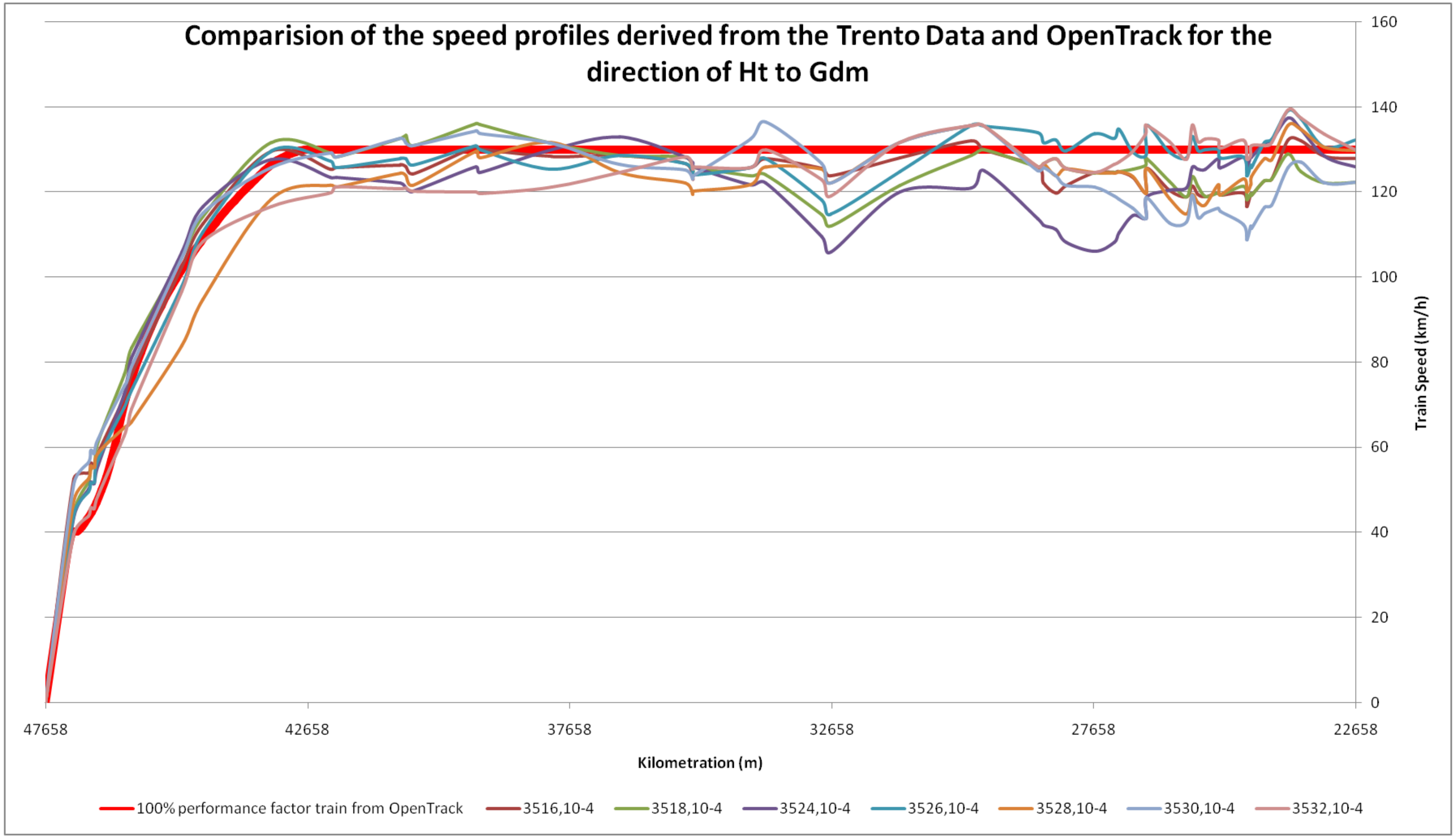


Figure 9-7 Comparision of the speed profiles derived from the Trento Data and OpenTrack for the direction of Ht to Gdm

Figure 9-6 and Figure 9-7 indicates that, the OpenTrack's overall modeling ability can be proved by the Trento data. One reminder is that for the OpenTrack train's data used in Figure 9-6 and Figure 9-7, they have been grouped by the grouping method.

10 Conclusions of this study and recommendations for OpenTrack model

10.1 Conclusions of this study

The conclusions of this study are formulated as follows:

a) For the Trento data selection and Grouping method aspects:
1. Although the Trento data is derived from the TROTS log files, their accuracy is not the same. In other words, Trento (version 1.3) does not extract the more accurate "event time" from the TROTS log files, but uses "final recording times" instead (section 6.4 and 6.5).

   This conclusion makes that the current Trento data does not have enough accuracy to construct a reliable speed profile in accordance with Newtonian mechanics.

2. The crude speed profile construction method proposed in this study has been proved to be a proper method to construct speed profiles from the Trento data, with an acceptable accuracy level with respect to the resulting running time.

3. The section grouping method is effective in removing the deviations from the Trento data. After section grouping, newly derived speed profiles are more realistic than the speed profiles that are derived from the original Trento data.

   Additionally, it is recommended to group the sections based on length instead of number of sections, because the average length of a section in station areas is much shorter than that of a section on open lines. Thus, grouping sections based on length achieves a better balance in keeping a realistic speed profile's satisfying the given accuracy.

4. If the section grouping method based on length is selected, it is recommended to define minimum grouping length for a section group as 900 meters, and mimimum overlap length between two section groups as 800 meters.

b) For the data analysis of realization data aspect:

5. For all the three train series groups in this study, there is no evidence that the initial delay is an important influencing factor for a driver to make his driving strategy, i.e., a clear relationship between a train's initial delay and its corresponding running time is lacking.

6. For all the three train series groups in this study, the data analysis reveals that train drivers seldom do coasting during their driving times. Only one train run is proved to be a coasting train, while the number of train runs considered in this study is 388.

7. For all the three train series groups in this study, there is no evidence of a relationship between a train's acceleration/deceleration rate usage and cruising

speed. In other words, it is not likely for a train driver to choose a higher cruising speed if he lost some time in the acceleration/deceleration phase.

8. For each train series group in this study, it is shown that the acceleration/deceleration rate used in one train series group generally shows an approximate symmetric distribution.

c) For the OpenTrack Modelling ability validation aspect:

9. This study shows that compared to the nominal train that is created in the OpenTrack model (in chapter 7), most of the acceleration/deceleration rates derived from the Trento data are higher.

10. It is shown that OpenTrack has a limited ability to model coasting behavior by using coasting boards since the setting of coasting board is not suitable for practical cases.

10.2 Recommendations for OpenTrack

1. As mentioned in chapter 7, the performance factor in the OpenTrack settings is used for determining cruising speed, acceleration and deceleration together, however, it is has been observed from this study that train drivers will apply different “performance factors” in cruising and acceleration/deceleration phases. Therefore, different performance factors are recommended to be set in OpenTrack settings in the future.

2. Although the coasting board in OpenTrack has been proved to be a good setting for simulating coasting trains, it will force all trains to do coasting when they pass the coasting board, which is not useful for multi-train simulations. This study recommends OpenTrack to implement coasting behavior settings to trains instead of the infrastructure. This means that some coasting indicators can be assigned to trains so that these trains can start coasting at some points, while the other trains can still do normal driving in the simulated infrastructure.

10.3 Next steps

This study also had many limitations that can be developed in the future:

1. A better data processing method is needed to analyze the data from train describers, and the new method should be faster and provide more reliable results to construct speed profiles.

2. Although the relationship between a train’s initial delay and running time is not clear in this study, more studies have to be done to investigate this relationship. Because drivers are taught that if their trains have initial delays, they should drive as soon aspossible to recover these delays.

3. More studies have to be done to investigate a train’s acceleraton/deceleration rate usage in both reality and OpenTrack settings.

4. It is essential to propose a solution to make more drivers choose to do coasting in proper cases, i.e. when they have enough slack time during their driving.

Reference

Agricola, M. *Validatie Micro-Simulatiemodellen voor Railverkeer Kalibratie en Validatie van OpenTrack.* TU Delft, 2009.

Albrecht, T. "Energy efficient train operation." In *Railway Timetable & Traffic*, by I.A. Hansen and J. Pachl, 83-105. Hamburg: Eurailpress, 2008.

Albrecht, T., and S. Oettich. "A New Integrated Approach to Dynamic Schedule Synchronization and Energy-saving Train Control." *Computers in Railways Ⅷ.* Southampton: WIT Press, 2002.

Albrecht, T., C. Gassel, A. Binder, and J. van Luipen. "Dealing With Operational Constraints In Energy Efficient Driving." *Railway Traction Systems (RTS 2010).* Birmingham, 2010.

Albrecht, T., C. Gassel, J. Knijff, and J. van Luipen. "Analysis of Energy Consumption and Traffic Flow by Means of Track Occupation Data." *Railway Traction Systems (RTS 2010).* Birmingham, 2010.

Albrecht, T., R.M.P. Goverde, V.A. Weeda, and J. van Luipen. "Reconstruction of train trajectories from track occupation data to determine the effects of a Driver Information System." *Computers in Railways X,.* Southampton: WIT Press, 2006. 207-216.

Brunger, O., and E. Dahlhaus. "Running time estimation." In *Railway Timetable & Traffic*, by I.A. Hansen and J. Pachl, 58-80. Hamburg: Eurailpress, 2008.

Chevrier, R., G. Marliere, B. Vulturescu, and J. Rodriguez. "Multi-objective Evolutionary Algorithm for Speed Tuning Optimization with Energy Saving in Railway: Application and Case Study." *RailRome 2011.* Rome, 2011.

Coleman, D., P. Howlett, P. Pudney, X. Vu, and R. Yee. “Coasting boards vs optimalcontrol.” *Railway Traction Systems (RTS 2010).* Birmingham, 2010.

Goverde, R.M.P, and I.A. Hansen. "TNV-Prepare: Analysis of Dutch Railway Operations Based on Train detection Data." *Computers in Railways VII.* Southampton: WIT Press, 2000. 779-788.

Goverde, R.M.P. “Delay estimation and filtering of train detection data.” *Proceedings TRAIL 6th annual congress2000.* Delft: TRAIL Research School, 2000.
Goverde, R.M.P., W. Daamen, and I.A. Hansen. ““Automatic Identification of Route Conflict Occurrences and Their Consequences.” *Computers in Railways XI.* Southampton: WIT Press, 2008. 473-482.

Hürlimann, D. *OpenTrack, Simulation of Railway Networks, Release Notes Version 1.6.* ETH Zürich, 2010.

Marinov, M., and J. Viegas. “A Mesoscopic Simulation Modelling Methodology for Analyzing and Evaluating Freight Train Operations in a Rail Network.” *Simulation Modelling Practice and Theory,* 2011: 516-539.

Medeossi, G., S. De Fabris, and G. Longo. "An implementation of stochastic blocking times to support timetable planning." *4th International Seminar on Railway Operations Modelling and Analysis (RailRome 2011).* Rome, 2011.

NedTrain Consulting. "Materieelparameters en - Karakteristieken bij 1500V/4KA Versie 3.0." 2001, 27-29.

Profillidis, V.A. *Railway Management and Engineering (Third Edition).* Hampshire: Ashgate, 2006.

Radtke, A. "Infrastructure modelling." In *Railway Timetable & Traffic*, by I.A. Hansen and J. Pachl, 43-57. Hamburg: Eurailpress, 2008.

Appendix 1 Figure list

Appendix 2 Table list

Appendix 3 OBE drawings list

Drawing Name	Belonging Line name	Starting Kilometration	End Kilometration	Version	Series	Version Date
Culemborg	Ut-Btl	13500	24100	L	OBE-blad 2	22-07-2010
Geldermalsen	Ut-Btl	24100	26900	U	OBE-blad 1	22-07-2010
Meteren Aansl	Ut-Btl	26900	30800	N	OBE-blad 5	16-02-2010
Meteren Aansl - Zaltbommel	Ut-Btl	30800	34100	F	OBE-blad 6	06-12-2010
Zaltbommel	Ut-Btl	34100	39100	O	OBE-blad 4	12-10-2009
Hedel	Ut-Btl	39100	46500	Q	OBE-blad 8	06-12-2010
’s-Hertogenbosch	Ut-Btl	46500	48000	U	OBE-blad 1	26-02-2010
’s-Hertogenbosch	Ut-Btl	48000	49300	N	OBE-blad 2	16-02-2010

Appendix 4 OS drawings list

Drawing Name	Belonging Line name	Starting Kilometration	End Kilometration	Version	Series	Version Date	Direction
Culemborg	Ut-Btl	14700	22500	B	OS-blad 1322	06-12-2010	Btl-Ut
Geldermalsen	Ut-Btl	20000	26900	H	OS-blad 0321	06-12-2010	Btl-Ut
Meteren Aansl	Ut-Btl	26900	30800	G	OS-blad 0181	06-12-2010	Btl-Ut
Zaltbommel	Ut-Btl	34100	39100	E	OS-blad 0323	06-12-2010	Btl-Ut
Hedel	Ut-Btl	39000	42000	C	OS-blad 0325	06-12-2010	Btl-Ut
Hedel - ’s-Hertogenbosch	Ut-Btl	42000	46500	D	OS-blad 0208	06-12-2010	Btl-Ut
’s-Hertogenbosch	Ut-Btl	-	-	C	OS-blad 0763	15-01-2008	Btl-Ut
’s-Hertogenbosch	Ut-Btl	-	-	B	OS-blad 0764	08-01-2003	Btl-Ut
Culemborg	Ut-Btl	14700	22500	B	OS-blad 1321	06-12-2010	Ut-Btl
Geldermalsen	Ut-Btl	22000	27000	K	OS-blad 0320	06-12-2010	Ut-Btl
Meteren Aansl	Ut-Btl	26900	30800	G	OS-blad 0180	06-12-2010	Ut-Btl
Zaltbommel	Ut-Btl	34100	39100	D	OS-blad 0322	06-12-2010	Ut-Btl
Hedel	Ut-Btl	39000	42000	E	OS-blad 0324	06-12-2010	Ut-Btl
Hedel - ’s-Hertogenbosch	Ut-Btl	42000	46500	D	OS-blad 0207	06-12-2010	Ut-Btl
’s-Hertogenbosch	Ut-Btl	-	-	C	OS-blad 0761	15-01-2008	Ut-Btl
’s-Hertogenbosch	Ut-Btl	-	-	B	OS-blad 0762	08-01-2003	Ut-Btl

Appendix 5 Original Trento data sheets list

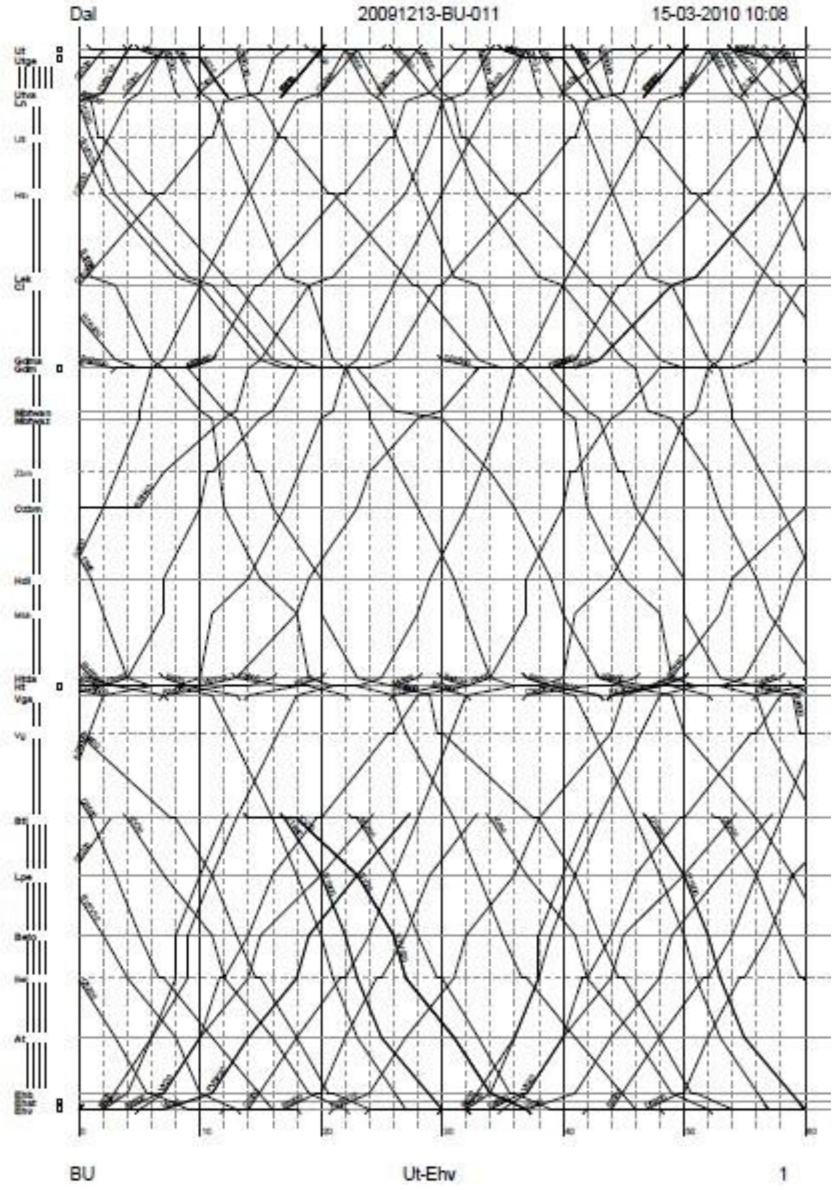
1) Section occupancy data

File name	Recording Date
routes_2010-08-23	23-08-2010
routes_2010-08-24	24-08-2010
routes_2010-08-25	25-08-2010
routes_2010-08-26	26-08-2010
routes_2010-08-27	27-08-2010
routes_2010-10-04	04-10-2010
routes_2010-10-05	05-10-2010
routes_2010-10-06	06-10-2010
routes_2010-10-07	07-10-2010
routes_2010-10-08	08-10-2010

2) Signal aspect data

File name	Recording Date	Sheet Name
Signal_04-10-2010	04-10-2010	route setting time
Signal_05-10-2010	05-10-2010	
Signal_06-10-2010	06-10-2010	
Signal_07-10-2010	07-10-2010	
Signal_08-10-2010	08-10-2010	

Appendix 6 The VPT timetable used in this study



Appendix 7 The excel files constructed for this study

Excel file name	Function
04-3500-neg.xlsx	Studied train selection
04-3500-pos.xlsx	Studied train selection
04-800-neg.xlsx	Studied train selection
04-800-pos.xlsx	Studied train selection
05-3500-neg.xlsx	Studied train selection
05-3500-pos.xlsx	Studied train selection
05-800-neg.xlsx	Studied train selection
05-800-pos.xlsx	Studied train selection
06-3500-neg.xlsx	Studied train selection
06-3500-pos.xlsx	Studied train selection
06-800-neg.xlsx	Studied train selection
06-800-pos.xlsx	Studied train selection
07-3500-neg.xlsx	Studied train selection
07-3500-pos.xlsx	Studied train selection
07-800-neg.xlsx	Studied train selection
07-800-pos.xlsx	Studied train selection
08-3500-neg.xlsx	Studied train selection
08-3500-pos.xlsx	Studied train selection
08-800-neg.xlsx	Studied train selection
08-800-pos.xlsx	Studied train selection
23-3500-neg.xlsx	Studied train selection
23-3500-pos.xlsx	Studied train selection
23-800-neg.xlsx	Studied train selection
23-800-pos.xlsx	Studied train selection
24-3500-neg.xlsx	Studied train selection
24-3500-pos.xlsx	Studied train selection
24-800-neg.xlsx	Studied train selection
24-800-pos.xlsx	Studied train selection
25-3500-neg.xlsx	Studied train selection
25-3500-pos.xlsx	Studied train selection
25-800-neg.xlsx	Studied train selection
25-800-pos.xlsx	Studied train selection

26-3500-neg.xlsx	Studied train selection
26-3500-pos.xlsx	Studied train selection
26-800-neg.xlsx	Studied train selection
26-800-pos.xlsx	Studied train selection
27-3500-neg.xlsx	Studied train selection
27-3500-pos.xlsx	Studied train selection
27-800-neg.xlsx	Studied train selection
27-800-pos.xlsx	Studied train selection
All the studied trains summary.xlsx	Studied train selection
Excel tool for coverting Trento data.xlsx	Studied train selection
Grouping method test based on length for neg train.xlsm	Grouping method investigation
Grouping method test based on length for pos train.xlsm	Grouping method investigation
Grouping method test based on number of sections for neg on 04-10-2010.xlsm	Grouping method investigation
Coasting detection for 800 from Gdm to Ht.xlsm	Coasting train detection
3500 series from Gdm to Ht.xlsm	Data analysis based on different groups
Coasting train validation.xlsx	Data analysis based on different groups
Initial delay analysis for 800 from Gdm to Ht.xlsm	Data analysis based on different groups
Neg trains from Ht to Gdm.xlsm	Data analysis based on different groups
Other studies for 800series from Gdm to Ht.xlsm	Data analysis based on different groups