Stochastic Modelling of Train Delays and Delay Propagation in Stations

Jianxin Yuan

Stochastic Modelling of Train Delays and Delay Propagation in Stations

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op woensdag 18 oktober 2006 om 10:00 uur

door

Jianxin YUAN

Master of Science in Engineering, UNESCO-IHE, Delft

geboren te Shanxi, China

Dit proefschrift is goedgekeurd door de promotor:

Prof.Dr.-Ing. I.A. Hansen

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.Dr.-Ing. I.A. Hansen	Technische Universiteit Delft, promotor
Prof.dr.ir. P.H.L. Bovy	Technische Universiteit Delft
Prof.dr. F.M. Dekking	Technische Universiteit Delft
Prof.dr. L.G. Kroon	Erasmus Universiteit Rotterdam
Prof.dr. M. Carey	Queen's University Belfast, United Kingdom
Prof.Dr.-Ing. E. Wendler	RWTH Aachen University, Germany

Preface

This PhD thesis is a result of six years of research work carried out at the Department of Transport and Planning, Delft University of Technology. I hereby express my special thanks to my promoter Prof. Ingo Hansen, who gave me the opportunity to start my PhD research in October 2000 and has provided his continuous support over all these years. His encouragement and supervision all the way and the detailed comments regarding the structure and texts of this thesis are highly appreciated.

My sincere gratitude also goes to all colleagues from the Transport and Planning Department and the TRAIL research school, who helped me a lot during my PhD research. In particular, I want to thank Ir. Paul Wiggenraad and Dr. Rob Goverde for their patient tutoring when I studied some subjects of railway operations and learned how to use several statistical software packages. I am also very thankful for the daily discussions with them during these years and their comments and suggestions during the finalization of this thesis. In addition, the assistance I obtained from Dr. Winnie Daamen in the use of Scientific Word for the edition of this thesis are especially acknowledged.

I am grateful to Dr. Geoff Rose at Monash University, Australia, for correcting the English texts of a part of my draft thesis during his sabbatical visit at the Department of Transport and Planning. I sincerely thank the members of the promotion committee for reading the entire manuscript carefully and giving their valuable comments. Finally, I thank all my family members and all my friends for their support during the period of my PhD research and in particular: my wife Bei and my daughter Qiaochu.

Jianxin Yuan

September 2006

Contents

List of Figures

List of Tables

Notations

The main symbols, variables, parameters and abbreviations that are used in this thesis are presented as follows:

Statistical symbols and parameters

P($\leq$60s)	Percentage of delays less than $60\,\mathrm{s}$
P($\leq$180s)	Percentage of delays less than $180\,\mathrm{s}$
SD	Standard deviation
H_0	Null hypothesis
H_1	Alternative hypothesis
$F_X(x)$	Cumulative distribution function of a random variable X
M_{X_i}	Median of a random variable X_i
T	Wilcoxon rank sum test statistic
α	Significance level
p	p-value of a statistical test
$K(u)$	Kernel
$f_{n,h}(t)$	Kernel density estimate (n is the sample size and h is the bandwidth)
χ^2	Chi-square test statistic
$\chi^2_{k-1,1-\alpha}$	Upper $1-\alpha$ critical point of a chi-square distribution with $k-1$ degrees of freedom
D_n	Kolmogorov-Smirnov test statistic
$d_{n,1-\alpha}$	Upper $1-\alpha$ critical point of the distribution of D_n
$S_n(x)$	Empirical distribution function
$f(x)$	Probability density function
L	Likelihood function
Λ	Logarithmic likelihood function
θ_i	Distribution parameter
μ	Mean
σ	Standard deviation
c	Shift parameter
k	Shape parameter
λ	Inverse of scale parameter
R_l	Rank of log-normal model among the candidate distributions
R_w	Rank of Weibull model among the candidate distributions
m_r	The r th moment about the origin
Γ	Gamma function
cv	Variation coefficient

Symbols, variables and parameters used in the delay propagation model

sa_i^{p}	Scheduled arrival time of train i
sd_i^{p}	Scheduled departure time of train i
$sc_{i,j}^{\mathrm{r}}$	Scheduled release time of the route of train j by train i
A_i^{p}	Arrival time of train i
A_i^{as}	Arrival time of train i at the approach signal of the station home signal
D_i^{p}	Departure time of train i
$C_{i,j}^{\mathrm{r}}$	Release time of the route of train j by train i
C_i^{r}	Clearance time of the route of train i by all the conflicting trains
t^{as}	Sight and reaction time
t^{hs}	Time duration for a hindered train to react to clearance of the home signal and to proceed to it in case of an extra stop before this signal
$E_{j,k}$	Typical event k ($k = 1, 2$ and 3) defined for an approaching train j
$T_j^{\mathrm{as\text{-}p}(1)}$	Running time of train j from the approach signal to the stop position on the platform track in case of event $E_{j,1}$
$T_j^{\mathrm{as\text{-}p}(2)}$	Running time of train j from the approach signal to the stop position on the platform track in case of event $E_{j,2}$
$T_j^{\mathrm{hs\text{-}p}(3)}$	Running time of train j from the home signal to the stop position on the platform in case of event $E_{j,3}$
$T_i^{\mathrm{as\text{-}rp}}$	Occupancy time of a track section by an inbound train i from the approach signal of the home signal to the release position
$T_i^{\mathrm{p\text{-}rp}}$	Occupancy time of a track section by an outbound train i from the stop position of the train on the platform track to the track release position
$\widehat{\mathcal{T}}_j^{\mathrm{a}}$	Set of the conflicting trains of train j that are approaching the station
$\widehat{\mathcal{T}}_j^{\mathrm{d}}$	Set of the conflicting trains of train j that are departing from the station
Δ_j^{as}	Time difference between clearance of the inbound route and the arrival of train j at the approach signal of the home signal
$\Delta_{i,j}^{\mathrm{as}}$	Time difference between release of the inbound route of train j by i and the arrival of train j at the approach signal of the home signal
$F_{X_1} * F_{X_2}(x)$	Convolution of independent distributions $F_{X_1}(x)$ and $F_{X_2}(x)$
L_j^{aknock}	Total knock-on delay of train j suffered before arriving at the station
a_j^{as}	Realization of A_j^{as}
$E_{j,3}(a_j^{\mathrm{as}})$	Typical event 3 defined for an approaching train j in case of a realization a_j^{as} of A_j^{as}
S_j^{p}	Free dwell time of train j
$S_j^{\mathrm{p}}(a_j^{\mathrm{p}})$	Free dwell time of train j in case of a realization a_j^{p} of A_j^{p}
$A_j^{\mathrm{p+}}$	Later one of the scheduled and the actual arrival times of train j
$A_j^{\mathrm{p,l}}$	Arrival time of a late train j
$S_j^{\mathrm{p,e}}$	Free dwell time of an early arriving train j
$S_j^{\mathrm{p,l}}$	Free dwell time of a late arriving train j
$S_j^{\mathrm{p,e*}}$	Free dwell time excluding the early arrival time of an early train j
$\widehat{\mathcal{T}}_j$	Set of the feeder trains of train j
$t_{k,j}$	Minimum transfer time of all the passengers from train k to j

t^{p}	Dispatching time of a train at platform
L_j^{dknock}	Total knock-on delay of a departing train j
$L_{i,j}^{\mathrm{dknock}}$	Knock-on delay of a departing train j caused by a conflicting train i
$L_{k,j}^{\mathrm{dknock}}$	Knock-on delay of a departing train j caused by a feeder train k
A_i^{us}	Arrival time of train i at the conflict detecting point upstream a route node
a_j^{us}	Realization of A_j^{us}
$a_i^{\mathrm{us}**}$	Critical arrival time of the preceding train i at the conflict detecting point, which is used to determine the alteration of the passing order of a pair of scheduled trains i and j at a route node
$\widetilde{C}_{i,j}^{\mathrm{r}}$	(Virtual) release time of the route of the second scheduled train j by the first scheduled train i in case the alteration of train orders is considered
$\widetilde{\Delta}_{i,j}^{\mathrm{as}}$	Time difference between $\widetilde{C}_{i,j}^{\mathrm{r}}$ and A_j^{as}
$\widetilde{C}_{j,i}^{\mathrm{r}}$	(Virtual) release time of the route of the first scheduled train i by the second scheduled train j in case the alteration of train orders is considered
$\widetilde{\Delta}_{j,i}^{\mathrm{as}}$	Time difference between $\widetilde{C}_{j,i}^{\mathrm{r}}$ and A_i^{as}
$T_i^{\mathrm{us\text{-}rp}}$	Occupancy time of the track section between the conflict detecting position upstream a route node and the release position of this node by train i
$T_j^{\mathrm{us\text{-}as}}$	Running time of train j between the conflict detecting position upstream a route node and the approach signal of this node
$\widetilde{A}_k^{\mathrm{p}}$	(Virtual) arrival time of a feeder train k, which is used to model the (in)dependence of the actual departure of train j upon the arrival of train k in case the cancellation of scheduled transfer connections is taken into account
$l_{j,k}$	Synchronization control time, i.e., maximum acceptable waiting time of train j for train k

General abbreviations

AR	Agglo/Regional (local) train series
CADANS	Combinatorisch-Algebraïsch DienstregelingAlgoritme voor de Nederlandse Spoorwegen
DONS	Designer Of Network Schedules, Dutch timetable design system
FCFS	First Come First Served
HST	High Speed Train series
IC	InterCity train series
INT	INTernational train series
IR	InterRegional train series
K-S	Kolmogorov-Smirnov goodness-of-fit test
MLE	Maximum Likelihood Estimator
NSR	Nederlandse Spoorwegen Reizigers, passenger train division of Dutch Railways
TNV	TreinNummerVolgsysteem, Dutch train describer system

Station abbreviations

Asd	Amsterdam CS	Hrl	Heerlen
Brusc	Brussel CS	Ledn	Leiden CS
Bd	Breda	Rsd	Roosendaal
Ddr	Dordrecht	Vl	Venlo
Gv	The Hague HS	Vs	Vlissingen
Gvc	The Hague CS		

Chapter 1

Introduction

1.1 Research background

Reliability and punctuality are critical measures of the quality of scheduled train operations and both are important to operators and passengers. Harris (1992) indicates that the punctuality of scheduled train operations is often perceived to be worse than it actually is, since passengers have a selective memory tending to remember poor performances at the expense of good ones. To keep the existing customers and attract more passengers, scheduled train services must get the punctuality at a high level. Many European governments and railway authorities have established striving goals for the reliability and punctuality of train operations (Hansen (2001)). For instance, the punctuality of the Dutch Railways is required to increase from 81% in 2002 to 89% in 2007.

Infrastructure capacity is another important issue regarding railway operations, which limits the volume of train services. Facing continuous growth of traffic demand and needed train services, most railway infrastructure managers are not only extending and upgrading the track network and improving the signalling systems to create additional transport capacity, but also striving to utilize the existing capacity more efficiently. However, a trade-off exists between the degree of utilization of network capacity and the quality of train services, e.g. the reliability and punctuality of train operations.

Due to random disturbances at the nodes and links of railway networks, running time supplements and buffer times are added, respectively, to the minimum process times and minimum headway between train paths in timetables in order to absorb smaller train delays and assure some degree of robustness of the timetable and the punctuality of train operations. However, large running time supplements and buffer times result in longer planned travel times for passengers, lower operating speed, higher operating cost and less efficient infrastructure capacity utilization. Therefore, the desired level-of-service may affect the expected operating cost and revenues and the efficiency of capacity utilization.

The optimization of running time supplements and their allocation with respect to the expected operating cost and performance of train operations has received much attention in the literature (Schwanhäußer (1974), Carey & Kwiecinski (1995), Carey (1998), Rudolph (2003), Vromans (2005)). However, the trade-off between efficient utilization of network

capacity and a higher level of the reliability and punctuality of train operations has been less studied and a generic approach to determine optimality is still to be developed.

In practice, the utilization of track capacity is generally estimated by virtually compressing timetable train paths up to the minimum headway according to their timetable order, without inserting any buffer time. The recommended maximal infrastructure utilization is defined by UIC (2004) at 75% for peak hours and 60% for the day in the case of mixed traffic lines. Furthermore, the robustness of timetables against perturbations can be estimated by means of queueing or simulation models by comparing the estimated waiting time and delay, respectively, with a certain maximum waiting time which was derived heuristically and considered to represent the desired level of operations quality (Hansen (2004)).

The optimization of capacity utilization and timetable design requires the prediction of the reliability and punctuality level of train operations, which is determined by the delays of trains and delay propagation. Actually, the propagation of train delays mainly occurs during the approach or departure of trains at stations, since the crossing or merging of lines and platform tracks are in most cases bottlenecks in highly used railway networks. To predict the punctuality level of train operations, the distributions of arrival and departure delays must be estimated as realistically as possible by taking into account the impact of knock-on delays due to headway and route conflicts as well as late arriving of feeder trains.

Modelling the propagation of train delays always focuses on a specific track layout, signalling and train protection system and timetable design. Based on the input delays at the boundary of and primary delays within the investigated area, the distributions of knock-on delays, arrival and departure delays can be estimated by stochastic models. The distribution of input delays and that of primary delays are often assumed by rules of thumb or based on rather limited empirical studies. In recent years, more comprehensive statistical analyses of train delays have been performed in the Netherlands by means of mining and compilation of standard track occupancy and release data (Goverde et al. (2001), Yuan et al. (2002)). This enables evaluation of the suitability of the distribution models applied in the literature for stochastic train event and process times on the basis of empirical traffic data.

The delay propagation models developed in North America usually predict the delay propagation caused by the meets and overtakes of trains on a single-track line with two-way traffic (Petersen (1974), Chen & Harker (1990)). In Europe, trains are mostly running on double-track lines with one-way traffic. The train movements here seem to be simpler. However, the European railway networks are more complex due to the mixed operation of links by different lines and trains. Furthermore, many large stations exist that have multiple intersecting inbound routes and outbound routes connected to multiple parallel or sequential stopping platforms. It is also common that the large stations serve several hundred, or even more than a thousand trains per day (Carey & Carville (2003)). In addition, the train arrivals and departures of different lines are often well coordinated at the railway nodes due to the clockface timetable. To reflect the impact of complex infrastructure networks, scheduled transfer connections and heterogeneous train traffic with high frequency, advanced models are needed for predicting the stochastic propagation of train delays.

Well proven microscopic simulation tools such as RailSys (Radtke & Hauptmann (2004)) and OpenTrack (Nash & Huerlimann (2004)) can be used to analyze the propagation of train delays in large complex railway networks. However, they do not take into account stochastic variation of train running times due to e.g. driver behavior and cannot hence accurately reflect the interrelation among all the factors affecting the dynamic process of conflict resolution. In addition, working with these models may require enormous amounts of preparation time and computing time, and simulation offers generally less insight into structural relations between input and output (Huisman & Boucherie (2001)).

Existing analytical probability models of delay propagation focus either on a single link or network corridor. In some of the models (Weigand (1981), Mühlhans (1990)), knock-on delays caused by route conflicts between scheduled trains are not explicitly taken into account. The probability delay propagation models proposed by Carey (1994b) and Higgins & Kozan (1998) explicitly incorporate the knock-on delays due to tight headway and route conflicts. However, they also assume that hindered trains follow at deterministic minimum headway when route conflicts occur. This is not realistic especially when the hindered trains stop in front of block signals due to a red signal aspect because of occupied route sections or platform tracks further ahead. Until now, there is no existing model that accurately estimates the delays of trains due to deceleration and acceleration in case of route conflicts and different reaction times of train drivers and traffic controllers.

A few stochastic optimization timetable models have been developed recently, which optimize timetable design by incorporating the estimation of train punctuality at stations. Vromans (2005) presented a two-stage stochastic optimization model that combines perturbation simulation and timetable optimization. There, knock-on delays of trains caused by route conflicts are simply considered by applying predefined minimal headway times as well.

In heavily occupied complex railway networks, systematic interdependences exist between train movements of different lines, which may result in multi-source knock-on delays. A train suffering knock-on delay may cause further knock-on delays to other trains, which is called 'dynamic delay propagation' in this thesis. In case of big perturbations in real operations, rescheduling trains, e.g. altering the order of trains at critical track sections or turning short earlier than the scheduled terminal stations is applied to reduce the propagation of train delays. However, such aspects are insufficiently reflected in existing delay propagation models. A generic analytical probability model is needed to accurately estimate multi-source knock-on delays and the resulting punctuality of trains by incorporating the impact of dynamic delay propagation and rescheduling. This is a prerequisite for improving capacity utilization, timetable design and rescheduling.

As the bottlenecks of a railway network are generally situated at stations (Hermann (1996), Hansen (2000), Carey & Carville (2000)), the real dwell times of trains often exceed the scheduled dwell times due to prolonged boarding and alighting caused by more passengers waiting at the platform for the delayed trains. The arriving and departing trains often suffer from knock-on delays due to the late release of platform tracks or junctions by other trains. In addition, 'ready to depart' trains at stations wait for late arriving feeder trains to secure the scheduled transfer connections. Hence, this thesis focuses on the analysis of empirical data from train describer systems and develops the stochastic models of train delays and delay propagation in stations.

1.2   Research objectives

This thesis aims mainly at modelling train delays and delay propagation at stations to support the improvement of capacity utilization and timetable design at a desired reliability and punctuality level of train operations. To this end, the following objectives are derived:

1. Getting more insight into the stochastic characteristics of train delays, delay propagation and real capacity utilization;

2. Modelling probability distributions of the input delays of trains at the boundary of the modelled station area and of the primary delays of trains within this area;

3. Developing an analytical probability model to predict the knock-on delays caused by route conflicts and late transfer connections including the impact on the punctuality of train operations more realistically than the existing models;

4. Validating the developed delay propagation model with empirical train detection data;

5. Demonstrating the applicability of the developed delay propagation model for maximizing the capacity utilization and improving the quality of timetable design.

1.3   Research relevance

Analysis and modelling of train delays and delay propagation in stations is an important research topic in the field of railway operations research. The relevance is summarized in both scientific and practical aspects.

Scientific relevance

The statistical analysis and modelling of train operations on the basis of a data mining tool called TNV-Prepare (Goverde & Hansen (2000)) is an extension of the previous study by Goverde et al. (2001). We evaluate several distribution models applied in the literature for stochastic train event and process times using empirical traffic data. The distribution evaluation results not only confirm certain findings in the literature (Schwanhäußer (1974), Goverde et al. (2001), Wendler & Naehrig (2004)), but also fill in a number of blank spots in the existing knowledge.

The developed delay propagation model is a blocking time theory based analytical probability model which estimates the knock-on delays caused by route conflicts and late transfer connections including the impact on the punctuality of trains. Reflecting the constraints of the signalling system and train protection operations rules in complicated station and interlocking areas and the impact of rescheduling in real operations, this model extends and refines the existing models developed by Weigand (1981), Mühlhans (1990), Carey (1994*b*) and Higgins & Kozan (1998). As a result, the model enables accurate predictions of the knock-on delays of trains suffered at critical track sections, e.g. platform tracks and junctions in complex station and interlocking areas and of the resulting punctuality of train arrivals and departures at the station.

Practical relevance

The empirical analysis of train delays and delay propagation reveals a number of main factors that affect the knock-on delays at platform tracks and junctions and the punctuality of train arrivals and departures at stations. Without detailed analysis, the proposed models of train delays and delay propagation in stations cannot be developed successfully. Such an analysis can support railway managers, timetable designers and train operators to adopt appropriate strategies to reduce knock-on delays and increase the punctuality level. The evaluation results with respect to the distributions of train event and process times can be used by timetable designers for predicting the propagation of train delays and the punctuality of train operations, especially when empirical data is not available, e.g. when timetables for a new infrastructure project are designed and evaluated.

The delay propagation model enables the maximization of capacity utilization for critical track sections within a complex station network and the improvement of detailed timetable design assuring a maximally acceptable knock-on delay at a certain confidence level and a desired punctuality level simultaneously. These reliability and punctuality measures can be specified more easily in practice than the total (or mean) waiting time estimations derived by queueing models (Wakob (1985), Schwanhäußer (1994), Huisman et al. (2002)), while representing the timetable constraints explicitly. Incorporating the impact of alteration of train order on delay propagation, this model also supports traffic controllers to on-line make appropriate rescheduling decisions in case of big operational disturbances.

When the precise position and delay of each train on its route is transmitted by advanced telecommunication technology, e.g. GSM-R (GSM for Railways), continuously to the area traffic control center, the developed delay propagation model could be incorporated in a computerized decision support tool which enables an on-line prediction of the expected arrival time of each train at downstream conflicting points accurately. Such a decision support tool would enable the determination of the appropriate speed, acceleration or deceleration rate for each involved train to assure the appropriate arrival time at the next conflict point, thus avoiding knock-on delays and increasing the punctuality of train operations.

1.4   Thesis contributions

This thesis contributes to a better understanding of the stochastic characteristics of train operations and to an improved modelling of probability distributions for train event and process times based on empirical data. The most important scientific contribution of this thesis is an innovative analytical probability model that accurately predicts the knock-on delays of trains including the impact on the punctuality at stations on the basis of an extension of the blocking time theory (Pachl (2002)) of railway operations to stochastic phenomena.

New insight into the stochastic characteristics of train operations is based on a comprehensive and detailed statistical analysis of real-world train detection data recorded at a major Dutch railway station. The statistical analysis has been performed not only for

train delays, but also for the propagation of train delays between train pairs and the estimation of the real capacity utilization of critical track sections, i.e., platform tracks and track junctions in the station and adjacent interlocking areas. One of the most remarkable findings is that the blocking times of the preceding open track signal block of the station by some frequently hindered train series are even exceeding those of the station block including the platform track. In case of hinder due to route conflicts, the bottleneck of the station and interlocking areas is shifted upstream. Statistical analysis results in the distribution of knock-on delays for each individual train series at the critical track sections and reveals the factors leading to knock-on delays as well as the impact on the punctuality of trains.

We have proposed a new method for fine-tune the parameters of theoretical distribution models on the basis of an initial estimate obtained by applying the maximum likelihood method. Several distribution models commonly applied in the literature for train event and process times are then evaluated by means of the Kolmogorov-Smirnov (K-S) goodness-of-fit test (Ross (2004)). The evaluation of distribution models is performed not only for the arrival and departure times of trains at the station, but also for the arrival times at the boundary of the local station area and the process times on the track sections within the area. In particular, the conditional distributions of train running and track occupancy times in case of (no) hinder from other trains and the distribution of the dwell times of trains in the absence of hindrance are evaluated, which so far hardly appears in the literature.

The key contribution of this thesis is an analytical probability model that enables much more realistic estimations of the knock-on delays of trains than known by existing models, while predicting its impact on the train punctuality at stations. Some of the innovative aspects of this model are specified as follows:

1. The concept of conditional distributions is adopted for the first time to reflect the stochastic variation of the running times of trains due to speed fluctuations in case of different aspects of the station home signal and the approach signal, respectively;

2. Possible dependence of dwell times of trains at stations on the corresponding arrival times is taken into account by conditional dwell time distributions;

3. The stochastic interdependences between train movements of different lines in complicated station and interlocking areas are explicitly expressed;

4. The dynamic delay propagation between trains in highly utilized station areas and the alteration of the order of pairs of trains passing the critical track sections are considered.

The validated delay propagation model can be used to estimate the maximum capacity utilization of the critical track sections within a complex station and interlocking area and to optimally allocate the buffer times between scheduled train paths at these track sections while assuring a maximally acceptable knock-on delay at a certain confidence level and a desired punctuality level. A generic approach for optimizing the capacity utilization and timetable design at stations with a certain reliability and punctuality of train operations is achieved.

1.5   Thesis outline

This section provides an outline of this thesis. The thesis structure is shown in Figure 1.1.

```
┌─────────────────────────────────────┐
│        Chapter 1. Introduction       │
└─────────────────────────────────────┘
                   │
                   ▼
┌─────────────────────────────────────┐
│   Chapter 2. State-of-the-art of     │
│   railway operations and scheduling  │
└─────────────────────────────────────┘
                   │
                   ▼
┌─────────────────────────────────────┐
│  Chapter 3. Statistical analysis of  │
│            train operations          │
└─────────────────────────────────────┘
          │                   │
          ▼                   ▼
┌──────────────────────┐  ┌──────────────────────┐
│ Chapter 4. Statistical│  │ Chapter 5. Probability│
│ distributions of train│  │  model of train delay │
│ event and process     │  │     propagation       │
│ times                 │  │                       │
└──────────────────────┘  └──────────────────────┘
          │                   │
          ▼                   ▼
┌─────────────────────────────────────┐
│  Chapter 6. Improving station        │
│  capacity utilization and timetable  │
│  design                              │
└─────────────────────────────────────┘
                   │
                   ▼
┌─────────────────────────────────────┐
│    Chapter 7. Conclusions and        │
│          recommendations             │
└─────────────────────────────────────┘
```

Figure 1.1: Thesis structure

Chapter 2 presents an overview of the state-of-the-art in the fields of railway operations and scheduling after a general introduction given in this chapter. We discuss, in particular, the advantages and drawbacks of commonly used analysis and modelling approaches of train delays and existing probability models of delay propagation. Chapter 3 deals with a detailed statistical analysis of real-world train detection data recorded in the area of The Hague HS station, The Netherlands. This provides a solid basis for the modelling of train delays and delay propagation. Next, Chapter 4 reports on the results of statistical modelling of the distributions of train event and process times, which will be used as input data for estimating knock-on delays including the impact on the punctuality of trains to

improve station capacity utilization and timetable design. In Chapter 5, a new analytical probability model that enables accurate predictions of the knock-on delays of trains including the impact on the punctuality at stations is first developed. This model is then solved on the basis of a numerical approximation of the convolution of individual distributions and validated using empirical train detection data. Chapter 6 further demonstrates an application of the proposed delay propagation model for maximizing station capacity utilization and optimizing timetable design while assuring a maximally acceptable knock-on delay at a certain confidence level and a desired punctuality of trains. Some modelling results of statistical distributions of train event and process times are herein applied. Finally, the main conclusions are drawn and further research aspects are recommended in Chapter 7.

Chapter 2

State-of-the-art of railway operations and scheduling

2.1 Introduction

An overview of the state-of-the-art in the fields of railway operations and scheduling is given in this chapter. Although the modelling of train delays and delay propagation is our main concern, this literature overview also covers the basic concepts of operations reliability and capacity of railway networks in Section 2.2. The propagation of train delays depends, among others, on the utilization of railway network capacity and timetable design. In addition, the analysis of train delays and modelling of delay propagation help to assess the quality of designed timetables regarding the reliability and punctuality of the scheduled trains and to achieve a more efficient utilization of existing network capacity. Therefore, the existing approaches and models of timetable design and evaluation are reviewed in Section 2.3. Next, we introduce the commonly used statistical analysis approaches of train delays and summarize a number of analytical probability delay propagation models presented in the literature in Section 2.4. Finally, conclusions are drawn in Section 2.5.

2.2 Capacity utilization and operations reliability

Operations quality and capacity are two basic characteristics of public transport. Reliability and punctuality are critical measures of the quality of scheduled train operations and are being paid much attention at present by customers and public authorities. Railway infrastructure capacity and capacity utilization affect the volume of train services and the reliability and punctuality of real operations, respectively.

2.2.1 Railway capacity and capacity utilization

Railway capacity is studied to validate the feasibility of scheduled timetables in an existing railway network, to evaluate the necessity of extending the track infrastructure and

modernizing the signalling system, and to provide assistance for designing a new railway network in order to satisfy the expected future traffic demand (Pachl (2002)). Another motivation of railway capacity research results from the separation of railway infrastructure management and train operations, and the introduction of railway capacity charges to train operators. To ensure fair track access and congestion charges, the analysis and modelling of the relationship between capacity utilization and train performance is now being paid much attention to in European countries (Gibson et al. (2002), UIC (2004)).

While the term capacity is well known, it is not easily defined or quantified. The carrying capacity of a railway transit system expressed by the number of passenger spaces per time period is determined by the transport capacity of the railway infrastructure in terms of train throughput, the number of carriages in a train and the capacity of the carriages. In this thesis, we focus on the railway capacity, which depends on the physical characteristics of the infrastructure itself and on the timetable. These characteristics include the number of tracks between stations, one-way or two-way operations on the tracks, the number of platforms within stations, the number of level crossings and mergings/divergings between different railway lines, alignment and slope of the tracks, train heterogeneity on the railway route, and connections of trains in stations.

The *maximum capacity* expressed in terms of the maximum number of trains in a railway line can only be achieved in ideal conditions. In this case, the scheduled headway between each pair of successive trains would need to be equal to the minimum headway, which is quite unrealistic. The minimum headway on a line depends on the so-called *blocking times* (Pachl (2002)) of any scheduled pair of trains. The blocking time is the time interval in which a section of a track (usually a block section) is exclusively allocated to a train at its scheduled speed and blocked for any other train. Thus, the blocking time lasts from issuing a movement authority (i.e. by clearing a block signal) until the possibility of issuing a movement authority to another train to enter the same section. The blocking time of a signal block is usually much longer than the time the train occupies the block physically. In a territory with lineside fixed signals, the blocking time of a block section consists of the following time intervals (Figure 2.1):

- the time for clearing the signal,

- a certain time for the driver to view the clear aspect of the signal that gives the approach indication to the signal at the entrance of the block section (this can be the preceding block signal or a separate approach signal),

- the approaching time between the signal that provides the approach indication and the signal at the entrance of the block section,

- the running time between the block signals,

- the clearing time to clear the block section and, if required, the overlap with the full length of the train,

- the release time to 'unlock' the blocking system.

Drawing the blocking times of all block sections that a train passes into a time-distance diagram yields the so-called *blocking time stairway* (Figure 2.2). The blocking times
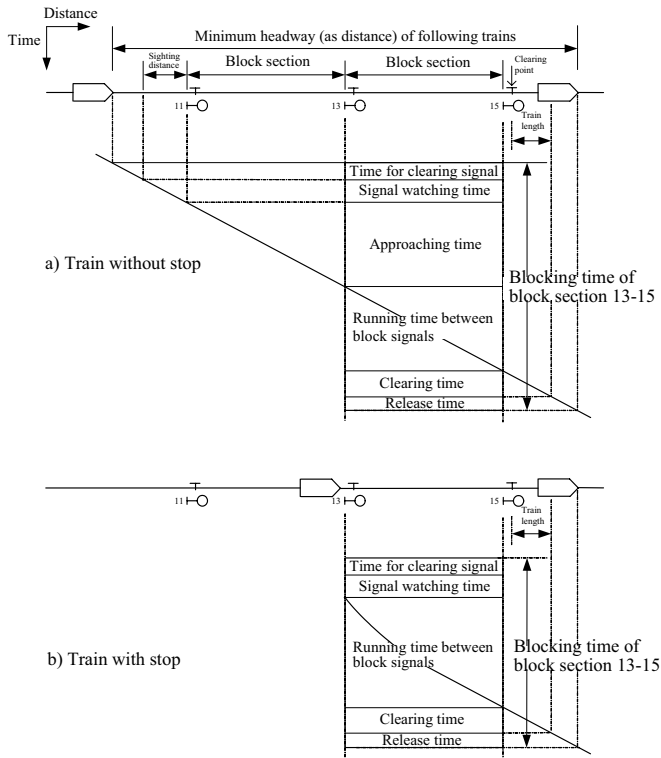
Figure 2.1: Blocking time of a block section (Source: Pachl (2002))

directly establish the signal headway as the time interval between two successive trains in each block section. The minimum line headway is the headway between two trains not only considering one block section but the whole blocking time stairways of the line. In this case the blocking time stairways of two following trains just touch each other in at least one block section and at this critical block section, there is no *buffer time* available between the blocking time stairways of the two successive trains. The buffer time is defined as the time difference between the end of the blocking time of the first train and the start of the blocking time of the second train at the critical block section.

Railway capacity depends on the timetable, too. To achieve the maximum capacity of a railway route, the speed of all trains must be scheduled equally. However, such a schedule can only be found on routes dedicated to no more than one type of traffic, e.g. metro lines. On railway routes, a variety of train types with different stopping patterns and running speeds are generally operated. This has led to the terminology of *timetable capacity* (Pachl (2002)), the maximum number of heterogeneous train paths (time-distance curves) that could be scheduled without considering any buffer time. The difference between the timetable capacity and the maximum capacity of a line is a reserve capacity in order to quickly reduce queues in case of perturbations by temporarily equalizing the speed differences to enable a higher output than scheduled. Adopting such a reserve helps to prevent scheduled operations from severe congestion.
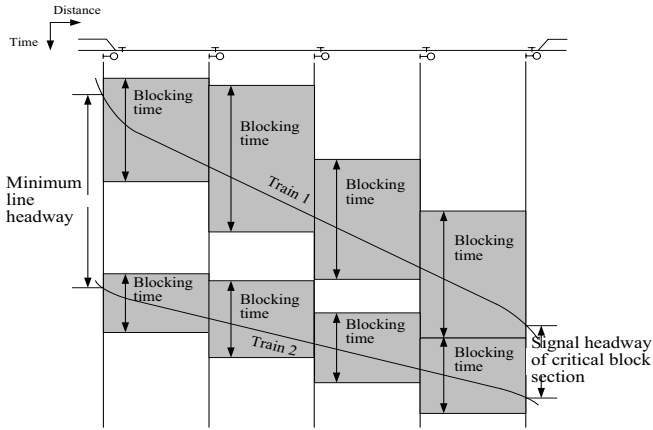
Figure 2.2: Blocking time stairways, signal headways and minimum line headway (Source: Pachl (2002))

To investigate the capacity of a complex railway network, it is necessary to determine capacity measures separately for different parts of the network. A network has to be decomposed into terminal tracks, lines between terminals including intermediate passing tracks, and interlocking areas. An interlocking is an arrangement of switches and signals interconnected in a way that each train movement follows the other in a proper and safe sequence. By analyzing the capacity for each part especially for the weakest one, i.e., critical nodes, the *network capacity* can be estimated. When analyzing the capacity for interlocking areas, the interdependences between the decomposed infrastructure elements must be taken into account.

To estimate the maximum capacity of a railway network, a standard set of train paths would need to be defined, which rarely exists in real operations like single lines on dedicated routes. To estimate the timetable capacity of a railway network, a variety of path mixes would have to be considered. It is not so easy to convert every train path to a standard path without affecting the overall capacity level (Burdett & Kozan (2006)). However, the level of capacity utilization for scheduled train operations can be calculated more easily for any given timetable of a particular railway infrastructure.

Capacity utilization is generally analyzed within a line section by virtually compressing the timetable train paths in a predefined time window. The time window is fixed by reference to the purpose of the analysis, which can, for example, be a peak period (at least one hour) or 24 hours daily total (representative weekday). The compression of train paths is done by pushing all single train paths together up to the minimum headway according to their timetable order without inserting any buffer time (UIC (2004)).

The capacity utilization rate (Pachl (2002)) of a line can be formulated as

$$\eta = \frac{n \cdot t_{\text{h}}}{t_{\text{p}}}, \tag{2.1}$$

where $\eta \in [0, 1]$ represents the capacity utilization rate, n denotes the number of scheduled

trains in the considered time window [-], t_h is the average minimum line headway [min], and t_p stands for the time window [min].

Capacity utilization of a line is the result of the compression process and measured at the beginning of the first block section within the line. The remaining capacity that equals the chosen time window minus the capacity utilization and total buffer time is called *unused capacity*. If necessary, additional train paths could be incorporated into the original timetable and a new estimation of capacity utilization would then be required. Usually, the capacity utilization rate is recommended to be smaller than a predefined threshold value, which is determined based on rules of thumb or the expected reliability and punctuality of train operations.

In order to investigate the capacity utilization of a complex network, the capacity utilization rate should be calculated separately for each part. The highest rate of capacity utilization of a certain network part determines the capacity utilization in the whole railway network of interconnected lines.

2.2.2   Determination of maximal capacity utilization

Facing a continuous growth of traffic demand and consequently needed additional train services, most railway infrastructure managers strive to utilize the existing capacity as much as possible. However, if the utilization rate of railway capacity is designed too high, the train traffic becomes easily perturbed, leading to a lower level of reliability and punctuality of operations. A compromise must be found to utilize more efficiently the existing capacity of railway networks on the one hand, and improve the level of reliability and punctuality of operations on the other hand.

The determination of the maximal capacity utilization is not simple, as a large number of relevant factors must be taken into account. Up to now, the capacity utilization applied for the planning of train services, has mainly been based on rules of thumb. The recommended maximal capacity utilization is defined by UIC (2004) at 75 % for peak hours and 60 % for the entire day in the case of mixed traffic lines.

The German Railways use the total waiting time of trains as a benchmark to determine the maximally acceptable capacity utilization, which is based on the maximal queue length per day defined by Schwanhäußer (1974). This queue length is characterized by a simple exponential equation as a function of the share of passenger trains:

$$L_{mq} = \alpha e^{-\beta p^p}, \tag{2.2}$$

where L_{mq} represents the queue length [-], p^p the share of passenger trains, and α, β coefficients that were calibrated as 0.257 and 1.3, respectively, for the German Railways.

Three different levels of quality of operations (0.5=very good, 1.0=satisfactory, 1.5=unsatisfactory) are determined by the estimated length of the queue divided by the maximally acceptable length of the queue. The maximal waiting time tolerated by the German Railways is between 130 up to 300 min per station per day depending on the share of passenger trains. These threshold values of the maximal waiting time, however, look somewhat arbitrary and lack sufficient empirical and theoretical evidence.

Hertel (1992) developed a theoretical approach for the determination of the maximal capacity utilization, situated between the minimal relative sensitivity of the waiting time (partial derivative of mean waiting time to traffic flow) and the maximal traffic energy (product of train density and the square of the speed). Based on simulation experiments, the range of optimal use of track occupation as a function of the waiting time, speed and traffic energy, in general, was determined to be between 150 to 200 trains per day, while the waiting time per train would be up to about 10 min. Further verification of Hertel's approach is still pending.

The utilization of railway capacity has a considerable impact on the reliability and punctuality of train operations. Running and dwell time supplements, which reduce the impact of primary delays of individual trains, decrease capacity use. Buffer times between scheduled train paths reduce the propagation of delays from one train to the next, but lead to a lower use of capacity.

2.2.3   Operations reliability and punctuality

Reliability is the ability of an item to perform a required function, under given environmental and operational conditions and for a stated period of time (ISO (1994)). It is also defined as the probability that an item will perform a required function under stated conditions for a stated period of time (Smith (1988)). This terminology has been adopted in the field of transportation research for a few decades. However, there is hardly any literature that gives a specific definition for the reliability of railway operations.

Punctuality is a commonly used measure of the reliability of railway operations. Usually, it is defined as the percentage of trains that do not arrive at (depart from, pass) a location later than a certain time in minutes. For most European railway companies, trains arriving less than 5 min late are generally not considered to be delayed. In the Netherlands, the threshold value for defining the punctuality of train operations is 3 min. In Japan, this value is even defined as 10 to 15 s. The punctuality of train operations may be measured not only at terminal stations, but also at intermediate stations and timetable reference points. However, no standard definition exists yet for measuring punctuality in larger networks with a mix of lines (Hansen (2001)).

Punctuality is an important measure of the reliability of railway operations, but it gives only limited information about train delays. The large number of smaller train delays that are not considered as delays according to public reports on punctuality have a considerable impact on the quality of train operations. A comprehensive punctuality definition should be more accurate and include not only the level of punctuality, but also the mean and standard deviation. Additionally, the reliability of train operations is reflected by the percentage of cancelled trains and percentage of realized train connections.

The measures of the reliability and punctuality of scheduled train services are useful for railway infrastructure managers and train operators in planning, management and marketing of these services, and for customers in making travel choices. For railway regulators these measures are also needed to check whether train operators together with infrastructure managers are providing the promised or contracted quality of services to customers. Therefore, most operators record detailed train traffic data and regularly publish the realized reliability and punctuality of operations as required by law or regulation. Measuring

the reliability and punctuality of train operations can be done by a detailed ex-post statistical analysis of train operations data.

A more difficult and important work is to predict the reliability and punctuality of scheduled train services in advance. That capability would help railway infrastructure managers to maximize the capacity utilization and operators to minimize the operating cost, while simultaneously assuring a desired level of reliability and punctuality of train operations. For the prediction of the reliability and punctuality of train operations according to a certain timetable, it is necessary to distinguish two types of delays, primary delays and knock-on delays. Primary delays of trains may be due to technical failures, running at lower than scheduled speed, prolonged alighting and boarding times of passengers, and bad weather conditions. When a train is delayed, it may hinder other trains by still occupying (part of) the scheduled route and preventing other trains from passing (crossing). The delays to other trains are called knock-on delays. Another source of knock-on delays for departing trains are transfer connections at large stations, where scheduled train arrivals and departures on different lines are coordinated. The amount and survival time of knock-on delays of trains reflect the stability of timetables and reliability of train operations in networks and depend on the size and distribution of buffer times.

Modelling the propagation of train delays in a railway network requires the input delays at the boundary and primary delays within the network. Based on the distributions of the input delays and primary delays, the distribution of knock-on delays suffered within the network and the distributions of the arrival delays and departure delays at the stations included in the network can be estimated by a stochastic delay propagation model. The input delays brought from the upstream stations and open tracks are recorded at cordons of the network and they may include primary delays as well as knock-on delays. In general, the real distribution of the input delays at the boundary of the studied network and that of the primary delays within the network can be obtained from a statistical analysis of existing empirical data (track occupancy and release records or on-board computers).

Knock-on delays should be distinguished clearly from the scheduled waiting times devised first by Schwanhäußer (1974). The scheduled waiting times result from headway and route conflicts between requested random train paths during the scheduling process and they can be estimated on the basis of train frequencies, inter-arrival times between trains, and minimum headways at the corresponding track section. Knock-on delays, however, are a part of real train delays generated in train operations. Both the amount of the scheduled waiting times and knock-on delays are applied in the German Railways' timetable standards to determine the maximally acceptable degree of network capacity utilization.

Carey (1999) introduces a number of ex ante heuristic reliability measures for scheduled public transport services, which were classified into two groups. Some of the heuristic measures are expressed by probabilities of delays, which are own delays, one-stage knock-on delays, own delays plus one-stage knock-on delays and multi-stage knock-ons. The others are just based on scheduled headways, which include minimum headway percentiles, (weighted) headway spread, weighted average headway and heuristics for solving line conflicts. Such heuristic reliability measures can give crude estimations of the reliability of timetables and punctuality of scheduled train operations, but do not reflect dispatching rules and dynamic behavior of trains in case of route conflicts.

2.3   Timetable design and evaluation

The main task of timetable design is to determine the arrival and departure times of trains at successive stations and the routes of trains through the network. In most European countries, passenger train services are basically scheduled according to a periodic (cyclic) timetable, repeating the same arrival and departure times each hour, with the exception of additional trains operated in peak hours and some international trains scheduled only in a part of the day. Routing trains through stations without hindrance, while assuring good connections, is a complicated issue in the process of timetable design.

The stability of timetables is very important. To provide high quality train services to customers, it is necessary to evaluate the quality of planned timetables. An alternative is to develop stochastic optimization timetable models that explicitly take into account the reliability and punctuality of train operations.

2.3.1   Timetabling methods

The existing railway timetabling models and approaches are based on deterministic running, dwell and headway times between stations. These times are usually rounded to minutes and refer to virtual stopping points at the stations. Small variations of the process times are compensated by standard running and dwell time supplements, as well as buffer times between train paths. The amount of time supplements and buffer times applied for scheduling, however, are mainly based on rules of thumb, sometimes checked by simulation, and only seldom derived from statistical analysis of real-world operations data (Hansen (2004), Goverde (2005)).

A running time supplement is usually added to the running time as a percentage of the minimal process time, which depends on the track infrastructure and dynamic characteristics of the train. This time supplement is generally 3-7% on European railways and 6-8% for passenger trains on North American railways (Pachl (2002)). The scheduled dwell time of a train at stations depends mainly on the type of the train and the volume of boarding and alighting passengers. Synchronization time may be included in the scheduled dwell time depending on the network timetable characteristics (Goverde (2005)). On some railways, the time supplement is evenly spread over the train path while other railways prefer to concentrate the supplement at the end of the path or at large intermediate stations. At large stations, the supplement may not be added to the running time in the section leading to the station but to the dwell time in the station (Pachl (2002)).

In case of conventional 'rules of thumb' timetabling, a consistent calculation of blocking time stairways is not performed and the minimal headway times are determined in a simplified way (Pachl (2002)). The track blocking times are considered by adding a general extra time, e.g. 1 min to the running time. Furthermore, a buffer time is added that usually lies within the range of 1 to 3 min. The detailed assignment of the buffer times between pairs of trains depends on the priority of trains and most railways use basic rules as follows:

- large buffer time when the following train has a higher priority than the preceding train,

- small buffer time when the preceding train has a higher priority than the following train,

- middle buffer time when both trains have the same priority.

Traditionally, timetable planners use two types of graphs as the main design tools, which are known as the time-distance diagram and the platform occupation chart. In a time-distance diagram, train movements are represented by train paths and the detailed use of the tracks within stations is not displayed. The platform occupation chart graphically zooms into the details within a station.

The estimation of blocking time stairways enables a clear recognition of the conflicts between train paths for a given timetable. By displacing, bending, or changing the train paths and/or changing the train order step by step, a conflict-free timetable can be achieved. However, the calculation of blocking time stairways for a timetable requires detailed information on the track infrastructure, train characteristics and the signalling system. The computation is therefore rather time-consuming in case of large railway networks.

Schrijver & Steenbeek (1994) compute a feasible periodic network timetable using a combinatorial model, which is based on the Periodic Event Scheduling Problem (Serafini & Ukovich (1989)). They also developed a powerful constraint propagation algorithm, called CADANS (Combinatorisch-Algebraïsch DienstregelingAlgoritme voor de Nederlandse Spoorwegen), which is capable of computing feasible timetables (if they exist) for the national Dutch railway network in a reasonable short time (several seconds to 15 min depending on the constraints). The performance of CADANS led to the development of the Dutch railway timetable design system DONS (Designer Of Network Schedules) by NSR (Nederlandse Spoorwegen Reizigers) and Railned (ProRail) (Hooghiemstra et al. (1999), Goverde (2005)).

It should be mentioned that CADANS considers the railway infrastructure only from a global point of view, neglects the detailed layout and signalling of the network within the railway stations, and applies predefined minimal headway times as constraints. Thus, the feasibility of the computed network timetable with respect to local infrastructure needs to be investigated in more detail. Kroon et al. (1997) consider the complexity of routing trains in stations and formulate it as a fixed interval scheduling problem. Zwaneveld et al. (1996) and Zwaneveld et al. (2001) developed an algorithm for solving the routing problem in a station to optimality, which is based on preprocessing, valid inequalities, and a branch-and-cut approach. This algorithm has been implemented in the module STATIONS as part of DONS (Zwaneveld et al. (2001)).

Several integrated scheduling and routing models have been presented in the literature. Carey (1994*a*) considers a rail corridor, but assumes that the stations along the corridor have unlimited capacity so that finding time slots and platforms is not a problem. In addition, a mathematical programming branch-and-bound approach was adopted, with heuristics to guide the branching and bounding, to restrict the search domain and speed up the search. More recently, Carey & Carville (2003) and Carey & Crawford (2005) use a heuristic approach without mathematical programming. The former focuses on the scheduling and platforming for a single busy complex station and the latter extends the algorithms to a series of complex stations linked by multiple one-way lines in each direction. The heuristic approach has the advantage that it can easily be extended to handle

additional information or objectives, making the scheduling choices more understandable and acceptable to timetable planners.

2.3.2 Timetable quality assessment

The timetabling methods based on deterministic process times do not sufficiently take into account the expected reliability and punctuality of the scheduled trains. As a result, the stability of the generated timetables still needs to be evaluated. In addition, the trade-off between improving the reliability and punctuality of train operations and making full use of the existing network capacity has to be solved by optimization of the running and dwell time supplements of scheduled trains and of the buffer times between scheduled train paths. The distribution of the total supplement time over a train line and assignment of the total buffer time within a cyclic timetable period are important for improving the robustness of a timetable and overall operations punctuality.

The scheduled waiting time devised first by Schwanhäußer (1974) is used by the German Railways to express the expected reliability of the scheduled train operations. Adopting queueing theory from operations research, Schwanhäußer (1974) predicts the average scheduled waiting time of trains on open tracks by assuming heterogeneous train traffic with random train orders, exponential interarrival times, and priority rules between different train types. Wakob (1985) extends Schwanhäußer's method to the computation of scheduled waiting time at railway stations and adjacent junctions. In addition, he assumes the interarrival times between trains to be gamma distributed. However, the route nodes are considered to be independent and the interaction between train movements over a sequence of route nodes is neglected. Wendler (1999) developed several extensions to the queueing models of Schwanhäußer and Wakob, based on conflicts between train triples and new approximation algorithms, which improve the estimation of waiting times at route nodes and the interactions between trains.

Huisman et al. (2002) presented an analytically tractable queueing network model for railway networks, which are divided into stations, junctions and sections, and transformed into a so-called product form queueing network. Closed form expressions for average waiting times are obtained and different network designs, traffic scenarios and capacity expansions may be evaluated. However, a few simplifying and modifying assumptions are made. At first, stations are modelled in a rather simple way, with the assumption that every train may use every platform track. Second, it is assumed that trains arrive according to a Poisson process, while the occupation times and minimum headway times are exponentially distributed. Another important assumption is the existence of artificial intermediate queues between the components of the network with unlimited capacities. In fact, this is not realistic, and therefore, the model results cannot be accurate for track sections with limited capacity, i.e., bottlenecks of the network.

Queueing models estimate the waiting time of trains and are applied in the course of strategic planning to evaluate the impact of increasing train frequencies and modifying infrastructure and train characteristics on the waiting time. The waiting times estimated by queueing models may not be true if the distributions of both interarrival times and minimal headway times are not stochastically independent as in highly utilized networks with clockface timetables (Hansen (2004)). In addition, queueing models are timetable-free, i.e., only train frequencies are defined and neither specific scheduled arrival and

departure times nor the impact of speed variations are considered. Thus, the punctuality of train operations for a given railway network and timetable cannot be predicted with a high level of confidence by means of queueing models.

Simulation models are used to estimate delay propagation in railway networks for given timetables. Macroscopic simulation models replicate real train operations on the basis of a macroscopic network and associated stochastic running and dwell times rather than the constraints of the signalling system which would be included in a microscopic infrastructure network and associated random track occupancy and release times. An example of this type of models is SIMONE (Middelkoop & Bouwman (2001)). Due to lack of application of detailed safety rules imposed by the signalling system, the delays estimated by macroscopic simulation models may be not sufficient for accurately evaluating the reliability and punctuality of train operations in highly utilized railway networks with many scheduled interconnections.

Microscopic simulation models, e.g. RailSys (Radtke & Hauptmann (2004)) and Open-Track (Nash & Huerlimann (2004)), intend to model real train operations according to the scheduled arrival and departure times, dispatching rules, track configuration, train dynamic characteristics, and signalling system. These models replicate the operation of each individual train during a user-defined time step and then repeat the process for the entire simulation period by means of synchronous simulations. Microscopic simulation models can be used to estimate the distributions of knock-on delays and the punctuality level of any timetable and operational disturbance during different time periods not only for stations, but also for timetable reference points, lines, and even whole networks. In addition, users can predefine the boundaries of the modelled area arbitrarily.

Applying RailSys, Kaminsky (2001) computes an 'optimal' buffer time between two successive trains at the conflicting track section on the basis of the performance criterion that a given percentage, e.g. 80%, of the train delays do not propagate to the following train. However, the punctual arrival of only the following train line is considered here. Rudolph (2003) presented a 'buffer train' approach for assigning the buffer time between the scheduled paths of two successive trains at each location of a line on the basis of the expected delay of the preceding train at the local position. It is also suggested to insert timetable slack predominantly as part of the scheduled dwell time rather than as part of the scheduled running time if station capacity suffices.

Microscopic simulation models are quite powerful, but they require extensive work to enter the detailed infrastructure topology, train characteristics, signal berths and timetables. Working with these models also requires enormous amounts of computing time. The existing micro-simulation models alone cannot be used for optimization of train throughput and the allocation of buffer times between scheduled train paths with respect to a desired reliability and punctuality level of train operations.

Scheduled train operations can be modelled analytically as a Discrete Event Dynamic System (Goverde (2005)). Given necessary data such as the timetable, train routes, and connections between train lines, Petri net theory enables the representation of scheduled train operations by Timed Event Graphs. An algebraic state representation of such a timed event graph is given by the associated max-plus linear system (Goverde & Koelemeijer (2000)). To analyze and quantify the stability of large network timetables, Goverde & Odijk (2002) developed a max-plus algebraic tool named PETER. The max-plus algebra

models determine the network timetable slack and clearly display the delay propagation in railway networks by assuming deterministic train process times and headways according to schedule. These models are suitable for evaluating the overall stability of timetables of interconnected lines, but cannot be used to estimate the distributions of knock-on delays and the punctuality level of the scheduled trains, as they are still based on a deterministic modelling approach.

A few stochastic optimization timetable models that explicitly incorporate the reliability and punctuality of the scheduled trains have been proposed until now. Carey (1994*b*) introduces a linear function of costs of travel times and punctuality and then formulates algorithms to minimize the sum of the costs for all involved trains at successive stations. The mean arrival and departure delays are estimated by an analytical probability model. Finally, optimal scheduled running and dwell times, or arrival and departure times are achieved by recursive substitutions. In this approach, the supplement times are hidden in the scheduled running and dwell times, while the buffer times are included in the scheduled headway times. Furthermore, Carey (1998) states that the process times often tend to last longer if more time is allocated and addresses the impact of behavioral response on optimizing the scheduled times. He also formulates a nonlinear function of costs of travel times and lateness. Clearly, the modelling principles are straightforward, but the calibration of the model based on empirical data is still pending.

Goverde (1999) proposes an objective of a weighted sum of separate convex cost functions of the recovery times for all relevant train processes to optimize timetable design. The cost of a transfer connection is the mean transfer time including the risk of a missed connection, which depends on several parameters such as mean arrival delay, frequency, and waiting time regulations. The convexity of the cost functions guarantees that a local optimal solution is also the global optimum. However, the impact of the safety and signalling system on train operations are taken into account only by the minimum headway times of trains.

Vromans (2005) recently presented a stochastic optimization model that distributes the running and dwell time supplements and the buffer times between train pairs such that the average delay of trains is minimal and the overall punctuality of trains is maximal under stochastic disturbances. In this model, the total of running and dwell time supplements and buffer time are fixed in advance and stochastic disturbances of train operations are simulated. The optimization model, however, neglects the impact of dynamic headway and route conflicts on train speed and track occupation by applying predefined minimal headway times.

Given certain timetables, actual process times of train operations deviate more or less from the scheduled ones due to random disturbances. As a result, route conflicts often occur at station tracks and junctions in highly utilized stations and interlocking areas and trains frequently wait for late arriving feeder trains at stations. In case of large perturbations, on-line rescheduling would need to be based on a comprehensive evaluation of a variety of control strategies with respect to the punctuality of the involved trains and optimal network capacity utilization. To assess the quality of timetables and rescheduling, it is necessary to develop an analytical probability model that enables accurate estimations of the propagation of train delays and punctuality of train operations especially at critical nodes and track junctions.

2.4 Analysis and modelling of train delays and delay propagation

Analysis and modelling of train delays and delay propagation enable the prediction of reliability and punctuality of the scheduled trains and optimization of a variety of rescheduling strategies. Statistical analysis of train delays not only provides insight into the delay characteristics, but it also helps to identify the sources of delays. Usually, the analysis of the sources of delays is done by multiple variable regression. Modelling the distribution of train delays is another important research topic. Given the input delay distribution of train arrivals at the boundary of a railway network and the primary delay distribution of the trains within this network, the distributions of knock-on delays and resulting exit delays can be estimated.

Some literature exists that models the propagation of train delays on single-track routes with two-way traffic. For example, Chen & Harker (1990) estimate the mean and variance of stochastic travel times of trains by taking into account the actual departure delays, random meeting/overtaking at sidings, and the priorities of trains. The modelling of delay propagation on single-track railway routes is not of main concern in Europe as the main railway networks are characterized by double or even four tracks with, in general, one-way traffic per track. However, this does not mean that the European railway networks are quite simple. On the contrary, the railway networks in Europe are more complex due to the mixed operation of the lines with different scheduled speeds as well as large stations that often have multiple intersecting routes around the station. In addition, the train arrivals and departures of different lines may be synchronized at the busy complex stations to facilitate the transfer of passengers between trains. In the following section, we will summarize the probability models that have been developed in recent years and can be used to predict the propagation of train delays in the complex European railway networks.

2.4.1 Multiple variable regression analysis

The literature on multiple variable regression analysis is rather limited. This may be due to the difficulty of obtaining relevant data from railway companies. Harris (1992) uses least-squares multiple linear regression to get a better understanding of the punctuality of train services in the UK and the Netherlands. He chose several variables and analyzed the dependence of the punctuality of trains. The variables include length of train in terms of the number of carriages, previous number of station stops, previous distance covered, age of the motive power unit, and track occupation. The regression results reveal that the distance covered and the length of the train (in number of carriages) are statistically significant in determining the punctuality of long- and short-distance services, respectively. Thus, Harris states that the appropriate management response is to ensure sufficient time supplements for long-distance services, and to enhance demand management for short-distance services. In addition, he also mentions a critical impact of the track layout in influencing the operational performance of train services.

Olsson & Haugland (2004) calculate the correlation coefficients of a number of selected factors for the punctuality of trains for the Norwegian Railways. The purpose was to

identify the critical factors that affect the punctuality level and further to prepare corresponding strategies for the improvements. The selected factors include the number of passengers, occupancy ratio (passengers/seats), infrastructure capacity utilization, number of cancellations, and temporary speed reductions. The correlation coefficient was found to be significant at the 0.01 level between arrival punctuality and the number of passengers, occupancy ratio (passengers/seats) and departure punctuality. Arrival punctuality is reduced as the number of passengers increases, and departure punctuality is decreased as the occupancy ratio (passengers/seats) of trains rises. They concluded that the key success factor for punctuality on local and regional trains in congested areas is the management of boarding and alighting passengers. The capacity utilization rate of a scheduled timetable has a big impact on the punctuality of train operations, but it alone cannot explain all variations in punctuality during a day.

Lin & Wilson (1992) establish dwell time regression models for one- and two-car light rail operations on the basis of data gathered on the Green Line of the Massachusetts Bay Transportation Authority. They assume that the cumulative dwell time may represent a significant proportion of the total train trip time and can contribute a lot to headway variability, which in turn affects passenger service quality. The resulting models show that both the numbers of passengers boarding and alighting and the level of passenger crowding on-board the trains significantly affect the dwell times. Evidence was also found for a non-linear increase of the marginal delay with the number of standees. Important differences exist between dwell time estimation models for one- and two-car trains as a result of typically uneven distributions of passenger movements and passenger loads between cars in a two-car train.

Wiggenraad (2001) analyzes the impact of a number of variables on dwell times of trains at several Dutch railway stations on the basis of on-site observations. He states that the length of actual dwell times is determined by the length of the planned dwell times, the numbers of alighting and boarding passengers, train and infrastructure characteristics, and the arrival and departure process of trains. Measured dwell times, especially of intercity trains, in general, are longer than scheduled. Dwell times during peak and off-peak hours are about the same, although the number of passengers boarding and alighting is quite different. Trains with wider doors (3 passenger lane) show about 10% shorter typical alighting and boarding times and with narrower doors (1 passenger lane) about 10% longer typical times than trains with standard doors (2 passenger lane). The total dwell time is split into: alighting and boarding time, unused time, and dispatching time. The alighting and boarding time is composed of two parts: the first part is alighting and boarding in a cluster; the second part is individual alighting and boarding. The clustered alighting and boarding times vary from 20 to 60% of the total dwell time. The unused and dispatching times contribute about 20% of the total dwell time. In addition, there are clear concentrations of waiting and boarding passengers in the vicinity of platform accesses.

Multiple variable analysis regression is an explicit approach to reflect the impact of selected factors on train delays and the punctuality of train operations. However, empirical analysis alone cannot provide quantitative predictions and regression analysis models may ignore some important factors affecting train delays or indicate seeming but not real correlations. In addition, the data collection and quantification of the impact of other factors like the behavior of train drivers and conductors, as well as weather conditions, still remains a problem.

2.4.2   Probability distributions of train delays

The distribution of train delays and its parameters may depend on the types and routes of trains, differ from location to location, and vary over time. It appears difficult to find a standard distribution type to be applicable everywhere. To analyze the robustness of timetables and the reliability and punctuality of train operations in a railway network, the distribution of input delays at the network boundaries and the distribution of primary delays within the network are often assumed based on the experiences from real train operations. Limited literature exists with respect to the statistical inference of the probability distributions of train delays based on empirical data.

The negative exponential distribution is often considered to be a valid model for headways and delays in railway systems. Schwanhäußer (1974) is probably the first who concludes that the non-negative arrival delays of trains at stations fit well to the negative exponential distribution. This is confirmed by later studies (Hermann (1996), Goverde et al. (2001), Yuan et al. (2002), Wendler & Naehrig (2004)). The exponential distribution of non-negative arrival delays is widely used as input data for some delay propagation models (Weigand (1981), Mühlhans (1990), RMcon (2004)). Goverde et al. (2001) also fit the departure delays of trains at a station to the exponential distribution.

Carey & Kwiecinski (1994*a*) assume a shifted exponential distribution and a uniform distribution for the free running time of each train on an open track section to simulate the effect of headways on knock-on delays of trains. The free running time of a train is defined as the sum of the minimum running time and a random disturbance. The difference between the free running time and the scheduled running time is the primary delay of this train. Carey & Carville (2000) use uniform and beta distributions for the input delays of train arrivals and the primary delays during the dwell process to test the timetable reliability for train stations. Huisman et al. (2002) presented a queueing network model for estimating the average waiting time of trains, where exponential distributions are assumed for inter-arrival times and minimal headway times.

For the arrival times (delays) of trains at stations or at signal berths, different distribution models, e.g. the normal, Erlang, Weibull and log-normal distributions have been adopted to capture the stochasticity (Hermann (1996), Higgins & Kozan (1998), Bruinsma et al. (1999)). For the primary delays during train running or dwell processes, it appears that the distributions cannot be derived directly from track occupation and release records, which can only show the total delays and may include knock-on delays (Yuan (2002), Wendler & Naehrig (2004)). Therefore, data filtering is necessary to model the distribution of primary delays. Some statistical analyses reveal that the primary delays at stations and open track sections fit well to an exponential distribution (Schwanhäußer (1974), Ferreira & Higgins (1996)), but other studies deliver partly different distributions (Steckel (1991), Yuan et al. (2002)). To the best of our knowledge, there is no publication that discusses the conditional distributions of the running and dwell times of trains in case of (no) hinder due to route conflicts or late transfer connections on the basis of empirical data.

In view of the above, a further detailed statistical analysis of track occupation and release data is required. It is, in particular, necessary to evaluate the distribution models that have been frequently adopted in the literature. The distribution evaluation should be performed not only for the arrival and departure times (delays) of trains at stations, but also for the arrival times at the boundaries of railway networks and the process times at stations and

relevant track sections within the networks. Furthermore, the conditional running and dwell time distributions need to be analyzed on the basis of empirical data.

2.4.3   Probability models of delay propagation

In recent decades, several analytical probability models of delay propagation have been proposed. Analytical probability models reflect explicitly the impact of various factors on the propagation of train delays and enable direct prediction of knock-on train delays and possibly of the punctuality level of scheduled train operations.

Weigand (1981) applies probability theory in analyzing the evolution of delays in long-distance railway networks. He describes the variation of (negative, zero and positive) delays by a probability and an exponential distribution of the non-negative delays. Based on the primary and additional delay distributions, as well as the available running time supplement, Weigand derives a model for the distribution of the exit delays on a line and further in long-distance railway networks. Mühlhans (1990) generalizes the analytical probability model developed by Weigand. He considers a generic cumulative probability distribution of delays rather than the specific exponential distribution and then computes the exit delay distribution by a convolution of the primary and additional delay distributions. He solves his model with a numerical approach. The knock-on delays caused by route conflicts are only included implicitly in the additional delays and the systematic interdependences between scheduled train pairs at critical track sections, e.g. stations and junctions, are not explicitly modelled.

Carey & Kwiecinski (1994*a*) estimate the knock-on delay occurring on a single link due to tight headway and speed variations by non-linear regression and heuristic approximations. In the simulation experiments for three-aspect signalling, they assume a simple speed reduction to a fraction of 0.6 in case a train meets a yellow signal. In addition, the downstream station capacity is considered to be higher than the number of trains arriving, thus excluding any waiting before entering the station. In fact, station capacity may be less than the upstream capacity demand in practice during some time periods and incidents. Moreover, trains generally may arrive at a fixed platform track unless a preceding train occupies the track for too long.

Considering the knock-on delays of trains due to the occupancy of on-route critical track sections by the preceding trains, Carey (1994*b*) computes the distributions of arrival and departure times of trains at successive stations with recursive substitutions starting from known departure times for each train at the initial station. However, these models do not take into account the interrelationship between multiple lines in more complex networks and scheduled transfer connections between trains.

A general stochastic model of scheduled transport formulated by Carey & Kwiecinski (1995) represents the random deviations of actual time moments, intervals from the corresponding scheduled timings and various forms of interdependence between the timings of different transport units. In principle, this model could be applied to railway systems, but the detailed interdependences between different lines in railway networks are not specified. In addition, the delay distributions of scheduled services and their impact on the level of punctuality are not discussed.

Higgins & Kozan (1998) presented an analytical model to quantify the expected positive delay for individual passenger trains and signal blocks in an urban rail network, by considering the knock-on delays caused by tight headways, speed variation of trains, and late transfer connections. However, the increase of minimal headway times due to speed fluctuation in case of route conflicts, alteration of the scheduled order of trains passing a route node and cancellation of scheduled transfer connections in real operations are not considered.

In view of the above, most existing delay propagation models assume that hindered trains occupy the next block section following the preceding trains at scheduled minimal time headway in case of route conflicts. These models do not consider the stochastic variation of track blocking times caused by different speeds and dwell times of the involved trains. In reality, primary delays due to longer than scheduled times for alighting and boarding and lower than scheduled train speeds can be observed frequently and those perturbations generally occur in and nearby station tracks. In case of route conflicts, hindered departing trains just extend the dwell process until clearance of the departure signal, while approaching trains regularly decelerate first and even may stop in front of the home signal of the station (junction), and then accelerate once the stop signal clears. There exists no analytical model that accurately represents the delays of trains due to acceleration and deceleration time loss and that forecasts the knock-on delays and exit delays at stations (junctions) reliably.

In addition, the systematic interdependences between train movements of different lines in complicated stations and interlocking areas are seldom analyzed and there is no existing analytical probability model yet that estimates knock-on train delays resulting from multiple sources. Moreover, the impact of dynamic delay propagation and rescheduling is hardly considered. Thus, a generic analytical probability delay propagation model is still needed to accurately estimate knock-on delays caused by route conflicts and late transfer connections in complicated stations and interlocking areas including the impact on the punctuality of train operations, by incorporating the impact of dynamic delay propagation and rescheduling.

2.5 Conclusions

In summary, the trade-off between increasing the utilization of railway networks on the one hand and improving the reliability and punctuality of train operations on the other hand has been widely recognized in the literature. However, a more precise analytical approach is still needed, which reflects the dynamic interaction between stochastic train movements, infrastructure and signalling constraints at highly utilized tracks and routes in stations and interlocking areas. To increase the capacity utilization and improve the quality of timetable design and rescheduling in densely occupied railway networks with interconnected lines, it is necessary to formulate a generic analytical probability model that is capable of accurately predicting knock-on delays and exit delays especially at station tracks and junctions based on certain input and primary delay distributions.

To this end, a detailed statistical analysis of empirical railway traffic data is a prerequisite for creating a deeper insight into the stochastic characteristics of train operations. This

will constitute a solid base for developing a more realistic delay propagation model to estimate the size of knock-on delays consistently and with a considerably higher accuracy than the currently available one(s). In addition, a detailed statistical analysis will also contribute to the determination of the distribution models for train event and process times, which can be used as the input data of delay propagation models for estimating the distribution of knock-on delays and train punctuality.

Chapter 3

Statistical analysis of train operations

3.1 Introduction

To model train delays and the propagation of train delays realistically and to achieve a maximal utilization of track capacity at a desired level of punctuality, it is important to obtain a real-world image of stochastic train event and process times and the interdependences between trains. Goverde et al. (2001) carried out a statistical analysis of real train traffic data recorded at Eindhoven station, the Netherlands. However, the effect of route conflicts in complicated station areas on the delays of trains and the real use of track capacity was not discussed. Hence, we have performed a more comprehensive statistical analysis of train operations, particularly focusing on the propagation of train delays and the use of track capacity on the basis of train traffic data recorded at The Hague Holland Spoor (The Hague HS) station, the Netherlands. In total, nearly 10000 trains proved to be recorded during the whole month of September 1999.

The data used for our statistical analysis is based on track occupancy and release records of train operations. In the Dutch Railways, the train describers, i.e., TNV systems (Trein-NummerVolgsystemen) keep track of the progress of trains at discrete steps over their routes. The position of a train number is based on TNV-windows. A TNV-window is a route between two (not necessarily adjacent) signals. At the start of a train journey the assigned train number is inserted into the TNV-system at the TNV-window of the departure platform. The TNV-window in which a train number is located is called its train number position or TNV-position. The movement and direction of a train is deduced from received safety and signalling information of signal controls, switch detection, and track circuits. The next TNV-position is determined by the set route after the destination signal of the current TNV-position. A TNV-step of the train number to the next TNV-position is triggered when the train enters the first associated track section. Hence, the TNV-position of a train corresponds with the track circuit occupancy but moves in larger steps. The Dutch railway network has been partitioned into 13 TNV traffic control areas, each having a separate TNV-system. If a train approaches an adjacent traffic control area, its number is transmitted to the associated TNV-system. When the train enters the next traffic control area the train number is cleared from the current TNV-system and inserted into the next. See Goverde (2005) for a more detailed description about the Dutch TNV-systems.

A TNV-system keeps a real-time record of received inputs and TNV-events. These TNV-logfiles of about 25 MB ASCII-format per day per TNV-system contain chronological information about all signalling and interlocking events in a traffic control area. The TNV logfiles give an accurate description of train movements with a maximal error of 1 s, but track section messages are not matched to individual train numbers. By posterior analysis of the TNV logfiles, Goverde & Hansen (2000) developed the data mining tool TNV-Prepare that couples events of infra elements to train numbers, such as section occupations and releases on open tracks and along routes in railway stations, triggered signal changes, and switch position lockings. The TNV-Prepare output consists of TNV-tables per train line service and (sub)route. For each individual train, event times along the route are given, including train description steps, section entries and clearances, signals, and point switches. From this information other interesting process times such as running times, blocking times, and headways can be derived easily.

The actual arrival and departure times of a train at a platform stop are in general not recorded in the TNV records and the platform track section borders are often more or less distant from the train stop position at the platform track. Therefore, the tool TNV-Filter (Goverde (2000)) was developed that estimates the arrival and departure times of trains at platforms based on data generated by TNV-Prepare. In detail, the speed trajectory of each train entering or departing a station is estimated on the basis of section (track circuit) occupation and clearance times and scheduled train characteristics. Estimation errors due to rounding-off (passing times are given in seconds) and deceleration (acceleration) variations during the approach (departure) are filtered by means of a nonlinear least-squares method taking into account speed limits at, e.g. signals and switches. Subsequently, the running times at the platform section before and after standstill are estimated from the filtered inbound and outbound speed profiles, known standard deceleration and acceleration characteristics per type of train, and stop location on the platform section (depending on train length). The arrival and departure delays of each individual train are then calculated by comparison with the scheduled times, see also Goverde (2005).

This chapter is structured as follows. Section 3.2 starts with an introduction to The Hague HS station, where the empirical data was collected. Section 3.3 then presents the main descriptive analysis results regarding arrival delays, dwell times and departure delays. In Section 3.4, the statistical difference of train delays between the days of a week and the periods of a day is illustrated. Next, Section 3.5 shows the knock-on delays caused by route conflicts in the area of the station and their impact on the punctuality of train operations at the station. In Section 3.6, the capacity utilization of The Hague HS station is analyzed. Finally, conclusions are drawn in Section 3.7.

3.2   The Hague HS station

The Hague HS (Figure 3.1) is one of the two main railway stations in the city of The Hague, the Netherlands. This station is a through station of two lines: one between Amsterdam/Leiden and Rotterdam and the other between The Hague Central Station (The Hague CS) and Rotterdam. The routes of the trains departing from The Hague HS and leading to The Hague CS and the routes of the trains coming from Amsterdam/Leiden cross at the junctions situated North of The Hague HS. The detailed track layout, switches

and signals in the area of The Hague HS station are shown in Figure 3.2 (those stabling tracks and the tracks connecting Voorburg/Utrecht and used for cargo train operations at night are discarded since they do not affect this analysis).
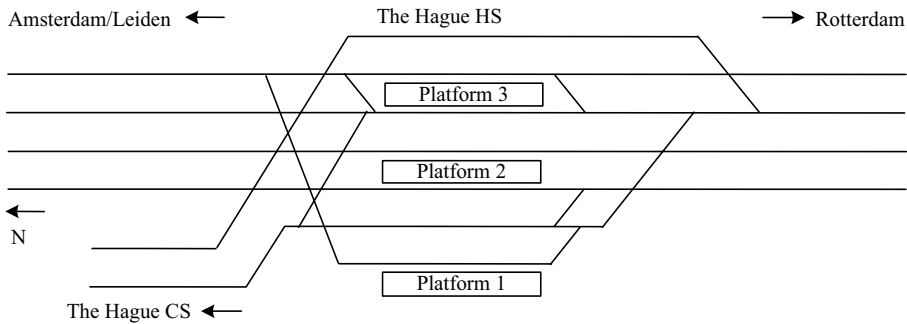
Figure 3.1: Scheme of railway network of The Hague HS station
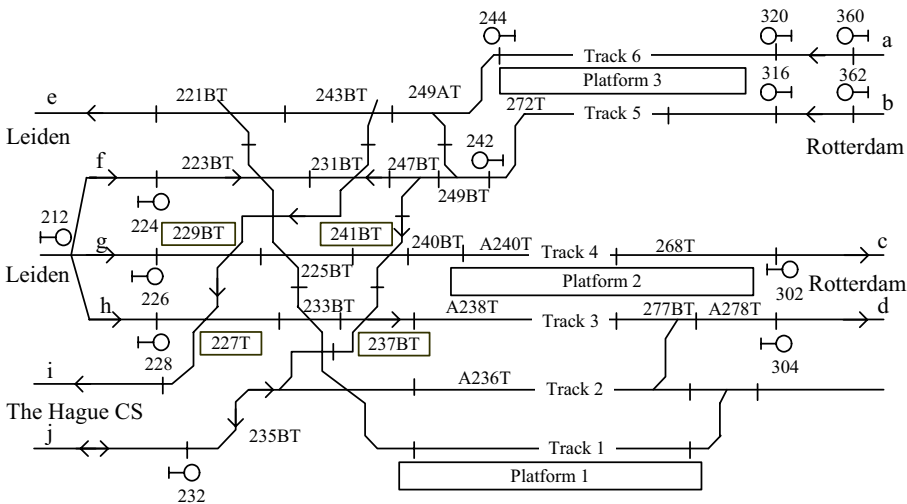
Figure 3.2: Track layout, switches and signals in the area of The Hague HS station

During the time period of data recording, 24 passenger trains arrived and departed at The Hague HS per hour corresponding to 9 different train series, i.e., 1 High Speed Train (HST) series, 1 INTernational (INT) series, 4 InterCity (IC) series, 1 InterRegional (IR) series and 2 Agglo/Regional (AR) series in both southbound and northbound directions (NSR (1999)). The HST series applies high speed rolling stock, but it was operated on the Dutch tracks only at the same maximal speed as IC-trains. The train routes for each train series per direction have been confirmed by means of TNV-Prepare. The scheduled arrival and departure times at the station together with the confirmed routes around the station are displayed in Table 3.1 for each train series per direction. The seven standard routes

used by these train series at most days within the investigated period are schematically illustrated in Figure 3.3 (infrequent routes used at a few days are neglected).

Table 3.1: Scheduled arrival and departure times and train route for each train series per direction at The Hague HS station (September 1999)

Train series	Orig.-Dest.	Southbound(S)			Northbound(N)		
		Arr.	Dep.	Route	Arr.	Dep.	Route
AR5000	Ledn-Ddr	12	14	h→d, f→d*	18	20	b→e
		42	44	"	48	50	"
AR5100	Gvc-Rsd	00	03	j→d	29	30	b→j
		30	33	"	59	00	"
IR2200	Asd-Bd	28	29	g→c, f→c*	05	06	a→e
		58	59	"	35	36	"
IC1900	Gvc-Vl	24	25	j→d	07	08	b→i
IC2100	Asd-Vs	45	46	g→c, f→c*	17	18	a→e
IC2400	Asd-Ddr	15	16	"	47	48	"
IC2500	Gvc-Hrl	54	55	j→d	37	38	b→i
INT600	Asd-Brusc	04	06	g→c, f→c*	58	59	a→e
HST9300	Asd-Paris	34	36	"	28	30	"

 * Infrequent routes used at a few days within the investigated period
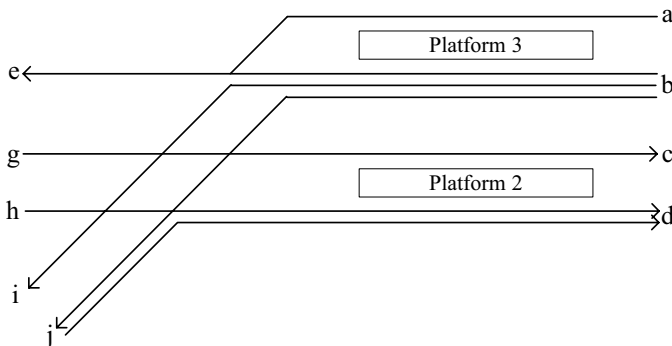
Figure 3.3: Schematic illustration of standard train routes in the area of The Hague HS station (September 1999)

3.3   Arrival delays, dwell times and departure delays

Train series have different operations histories upstream and distinct routes in the station area. The corresponding trains may also have different dynamic characteristics. These factors eventually lead to different statistics of train delays. Hence, our statistical analysis has been performed of the arrival delays, dwell times and departure delays per individual train series in both southbound (S) and northbound (N) directions.

3.3.1 Arrival delays

The arrival delay of a train at a station usually is the result of a late departure at the upstream station, low speed and longer running time from the upstream station, or route conflicts particularly due to the occupancy of a junction on the route or the platform track by other trains. In this thesis, we define the arrival delay of a train to be the difference between the actual and the scheduled arrival time at a station. It includes non-negative arrival delay (punctual and late arrival) and negative arrival delay (early arrival). Early arrivals do not affect the level of punctuality, but they may lead to a less efficient use of the station capacity.

Punctuality of train services, in general, is expressed as the percentage of trains passing, arriving or departing at given locations of the railway network not later than a certain time in minutes. In the Netherlands, the punctuality of a train line is defined officially at the threshold of 3 min. The punctuality reflects only the overall performance of train operations. To get a more comprehensive impression of the quality of train services, it is necessary to know other statistics of train delays, at least the mean and standard deviation.

Table 3.2: Statistics of arrival delays

Train series	Sample size	Mean [s]	SD [s]	Median [s]	Late [%]	P($\leq$60s) [%]	P($\leq$180s) [%]
AR5000S	557	82	106	55	88	55	89
AR5100S	950	69	88	47	98	64	94
IR2200S	975	64	146	27	66	65	89
IC1900S	415	98	104	79	100	31	93
IC2100S	495	40	121	9	54	72	91
IC2400S	482	58	183	8	55	71	89
IC2500S	434	106	129	76	100	34	88
INT600S	452	78	199	31	68	67	87
HST9300S	143	1	130	-21	38	89	97
AR5000N	547	51	140	3	51	71	89
AR5100N	884	-8	157	-56	26	86	94
IR2200N	984	-5	138	-44	24	87	94
IC1900N	414	75	250	14	56	67	89
IC2100N	508	27	194	-34	39	76	87
IC2400N	509	-14	227	-66	27	85	94
IC2500N	397	86	291	11	57	73	89
INT600N	430	37	276	-31	34	79	90
HST9300N	144	66	378	-49	28	80	88
Total	9720						

S: southbound; N: northbound

Table 3.2 displays a summary of the main statistics of the arrival delays for each train series per direction. The mean and standard deviation are also shown in Figure 3.4. Although the mean value is less than 2 min for each case, the biggest standard deviation is even beyond 6 min in case of the high speed trains HST9300N. This indicates that the scheduled timetable has been designed reasonably well without resulting in large structural delays, but there exist big variations of train operations.

The mean arrival delay of southbound trains is generally bigger than that of the corresponding northbound trains. This is partly because the southbound trains coming from Amsterdam/Leiden suffer from route conflicts before arriving at The Hague HS due to occupancy of on-route level crossings, e.g. track junctions 229BT, 241BT, 227T and 237BT (see Figure 3.2) situated North of the station by the trains leading to The Hague CS. Although there is no conflicting point on the running route, the southbound trains originating from The Hague CS have a mean arrival delay of 1 to 2 min. This arrival lateness is caused by late departure at the origin station and lower running speed between the two distant at 3 km stations than scheduled (Nie & Hansen (2005)). The arrival delays of the northbound intercity trains IC1900N and IC2500N originating from Venlo and Heerlen, which are located nearby the south-east boundary of the Netherlands, the northbound international trains INT600N from Brussels and the northbound high speed trains HST9300N from Paris have a standard deviation bigger than 4 min. This can be attributed to the longer journeys since the travel times of long-distance trains, in general, have larger variations than short-distance trains.
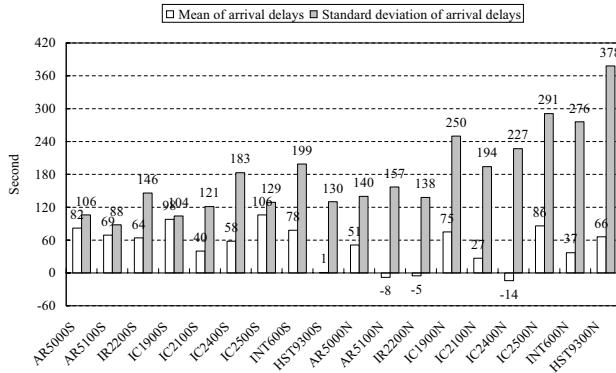
Figure 3.4: Mean and standard deviation of arrival delays
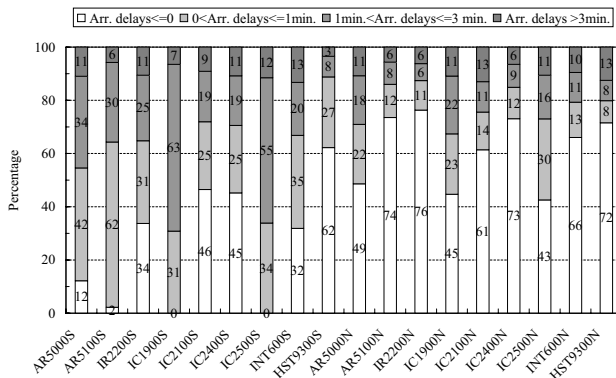
Figure 3.5: Distribution of arrival delays

Figure 3.5 shows the distribution of arrival delays. According to the official definition of the Dutch Railways, the punctuality of train arrivals at the station is between 87% and 97% for each series per direction. However, if punctuality was defined at the level of 1 min, the percentage of trains that arrived punctually would reduce to between 31% and 89%. The percentage of trains that arrived earlier than the scheduled time is between 0% and 76%. The southbound trains AR5100S, IC1900S and IC2500S, which originate from neighbor station The Hague CS, arrived at The Hague HS later than the corresponding scheduled arrival time with an extreme high percentage of more than 98%.

3.3.2   Dwell times

The dwell time of a train at a station is the difference between the arrival and departure time. The scheduled dwell time of a train series per direction is generally designed to be a constant value of 1, 2 or 3 min. This depends on the type of train series, station size and scheduled transfer connections and so on. The actual dwell times per train line vary randomly, affected by the scheduled dwell time, number of passengers alighting and boarding, train characteristics, platform design, and the behavior of train drivers and conductors. In addition, the dwell time of a train can be prolonged due to route conflicts or late transfer connections.

Table 3.3: Statistics of dwell times

Train		All trains		Early arrivals		Late arrivals	
series	Plan	Mean	SD	Mean	SD	Mean	SD
	[s]	[s]	[s]	[s]	[s]	[s]	[s]
AR5000S	120	108	53	164	39	100	49
AR5100S	180	150	59	206	65	148	58
IR2200S	60	96	54	123	48	82	52
IC1900S	60	75	32	-	-	75	32
IC2100S	60	108	51	130	57	88	34
IC2400S	60	98	52	126	54	75	35
IC2500S	60	77	39	-	-	77	39
INT600S	120	138	56	191	40	113	44
HST9300S	120	186	49	206	44	152	39
AR5000N	120	173	76	198	65	150	79
AR5100N	60	168	74	194	64	96	50
IR2200N	60	136	58	149	49	92	63
IC1900N	60	101	40	126	31	82	35
IC2100N	60	143	66	170	56	99	58
IC2400N	60	154	62	177	51	89	39
IC2500N	60	104	46	126	43	88	42
INT600N	60	143	55	162	46	106	52
HST9300N	120	185	67	212	58	117	27

Table 3.3 lists a summary of the main statistics of the dwell times for each train series per direction. A train is not allowed to depart earlier than the scheduled departure time. Consequently, the mean dwell time of early arriving trains is 38 to 98 s longer than that of

late arriving trains for all relevant train series. We hereafter take the part of the observed dwell time after the scheduled arrival time as the dwell time for an early arriving train. To investigate the development of train delays at a station, it is important to compare the actual dwell times and the scheduled dwell times. We show, in Figure 3.6, the difference between the mean dwell time and the associated scheduled dwell time per train series in each direction, by considering the early and late arriving trains separately. This time difference is actually the mean additional delay of trains at the station.
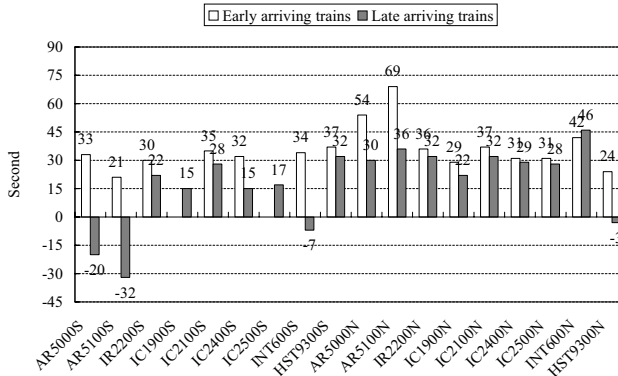
Figure 3.6: Difference between mean dwell time and scheduled dwell time

The mean dwell time of early arriving trains is 20 to 70 s longer than scheduled for all the train series even when the part of the dwell times due to the early arrivals is excluded. The mean dwell time of late arriving trains is 15 to 45 s longer than scheduled except for the international and two regional (local) train series in southbound direction and the north-bound high speed trains. An investigation at several Dutch railway stations (Wiggenraad (2001)) shows that about 20% of the observed dwell times is unused even during peak periods for dwelling at stations. The additional train delays at The Hague HS station might partly be caused, too, by a lack of strict departure behavior. The exceptional trains have a relatively longer scheduled dwell time of 2 or 3 min, which compensates for the dwell disturbances in the station. The mean additional delay of late arriving trains, which is negative in these cases, differs significantly from that of early arriving trains of the same train series.

We have noticed that the mean additional delay of northbound trains at the station is longer than that of the southbound trains for most train series, e.g. AR5000 and AR5100. This is mainly caused by late setting of outbound routes for the northbound trains due to occupancy of on-route mergings by trains departing for Leiden/Amsterdam in the same direction and of level crossings by trains coming from Amsterdam/Leiden in the opposite direction, respectively. An exception is the high speed train series HST9300. The mean additional delay of the northbound trains is much smaller than that of the southbound trains. This is because less people board the high speed train going to Schiphol Airport and Amsterdam from The Hague HS station.

3.3.3   Departure delays

The departure delay of a train is the difference between the actual and the scheduled departure time and it is generally non-negative. Departure delays may be caused by a late arrival or a prolonged dwell process. Just like the dwell time, the departure time of a train may be postponed due to route conflicts in the outbound direction or due to waiting for connections. A summary of the main statistics of the departure delays is listed in Table 3.4 for each train series per direction.

Table 3.4: Statistics of departure delays

Train series	Mean [s]	SD [s]	Median [s]	P($\leq$60s) [%]	P($\leq$180s) [%]
AR5000S	70	96	43	59	91
AR5100S	39	80	14	82	96
IR2200S	104	181	50	56	85
IC1900S	113	108	94	21	90
IC2100S	88	115	51	55	88
IC2400S	96	177	41	61	87
IC2500S	122	137	93	23	85
INT600S	96	181	41	67	89
HST9300S	67	117	40	72	94
AR5000N	104	140	57	52	84
AR5100N	100	142	62	49	89
IR2200N	71	130	31	69	92
IC1900N	116	246	46	58	87
IC2100N	110	179	48	60	85
IC2400N	80	209	28	70	92
IC2500N	130	291	43	60	85
INT600N	120	262	47	60	87
HST9300N	131	351	23	70	86
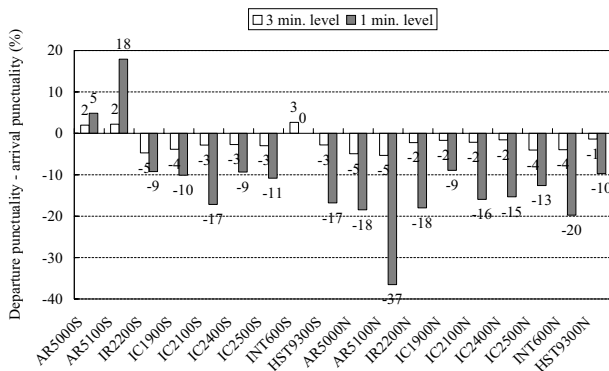
Figure 3.7: Difference between the arrival and departure punctuality of trains

Figure 3.7 shows the difference between the arrival and departure punctuality defined at the levels of 3 and 1 min. Except for the international and two local train series in southbound direction, the officially defined punctuality at the level of 3 min is reduced between 1% and 5% from arrival to departure while the punctuality defined at the level of 1 min is even reduced between 9% and 37%. This is due to the additional delays during the dwell process of trains. The larger percentage of additional delays of the northbound international trains INT600N is caused by a shorter scheduled dwell time, i.e., 1 min and the relatively longer alighting and boarding times of passengers. The larger percentage of additional delays of the northbound local trains AR5000N and AR5100N is caused mainly by route conflicts in the outbound direction.

3.4   Train delays by type and period of day

The arrival delays, dwell times and departure delays of trains may vary among different days of a week and between peak and off-peak periods in a day because of differences in traffic flows and an uneven distribution of passengers alighting and boarding trains. The impact of type of day and period of day on the delays of trains is studied for two typical train series: the northbound intercity trains IC1900N and the southbound interregional trains IR2200S.

In case of the intercity train series, whose trip ends at The Hague CS, there is a nearly fixed pattern of passenger flows boarding the trains at The Hague HS station. For the interregional train series, the number of passengers getting on and off the trains at The Hague HS station varies a lot even during a day because of the commuting characteristic.

We investigate the median difference for the train delays and the dwell times of the chosen train series by means of the Wilcoxon rank sum test (A detailed description of this test can be found in Appendix A.1), assuming the same shape of the delay distribution by type and period of the day. This assumption can be partially justified by the facts that a train series runs on the same route and has the same stopping pattern in different days of a week and during different periods of a day.

The decision to accept the null hypothesis of a statistical test or to reject it in favour of the alternative hypothesis depends on the choice of a *significance level* α, which is the maximally allowed probability of rejecting the null hypothesis when it is actually true. In the sequel, we use a significance level $\alpha = 0.05$ for the Wilcoxon rank sum tests to be performed. The statistical analysis tool S-Plus (MathSoft (1999)) is used to calculate the test statistics and the corresponding p-values. If the p-value is smaller than the significance level, the null hypothesis is rejected, otherwise it is accepted.

3.4.1   Train delays by type of the day

We separated the arrival and departure delays of trains and the dwell times per train series by type of the day: Sunday, Monday, weekdays (Tuesday, Wednesday and Thursday), Friday and Saturday. The dwell times are further distinguished between early and late arriving trains, respectively. It should be mentioned that the observed dwell times due

to early arrivals were excluded. The corresponding sample sizes are given in Table 3.5. Table 3.6 lists the median values for the delays and the dwell times of trains split by the five types of the day for both IC1900N and IR2200S.

We compare the median of the arrival delays between the five types of the day using the Wilcoxon rank sum test. The p-values of the test statistic between the split arrival delays are listed in Table 3.7 for both IC1900N and IR2200S (within parentheses those for IR2200S). There is a significant difference between the median of the arrival delays of the intercity train series IC1900N on Sunday and that on weekdays and Friday. In case of the interregional train series IR2200S, the median of the arrival delays on Sunday differs from that on Monday and weekdays significantly. Similar results have also been obtained for the departure delays in both cases.

Table 3.5: Sample sizes of IC1900N and IR2200S by type of the day

Type of day	IC1900N			IR2200S		
	All	Early	Late	All	Early	Late
Sunday	59	34	25	130	56	74
Monday	56	25	31	136	45	91
weekdays	182	73	109	449	136	313
Friday	63	24	39	132	47	85
Saturday	54	26	28	128	45	83
Total	414	182	232	975	329	646

Table 3.6: Medians of delays and dwell times of IC1900N and IR2200S by type of the day [s]

Type of day	IC1900N				IR2200S			
	Arr. delays	Dwell times		Dep. delays	Arr. delays	Dwell times		Dep. delays
		Early	Late			Early	Late	
Sunday	-12	78	75	31	11	79	68	33
Monday	8	81	73	37	35	77	69	61
weekdays	24	85	70	49	33	77	70	53
Friday	34	85	81	69	27	79	67	57
Saturday	3	83	79	45	24	76	62	43
Total	14	83	73	46	27	77	68	50

Table 3.7: p-values of the Wilcoxon rank sum test for arrival delays IC1900N(IR2200S) by type of the day

	Monday	weekdays	Friday	Saturday
Sunday	0.2889	0.0197	0.0166	0.2981
	(0.0030)	(0.0015)	(0.0600)	(0.1285)
Monday		0.2748	0.1326	0.9642
		(0.5093)	(0.2434)	(0.1625)
weekdays			0.4211	0.3562
			(0.4727)	(0.2327)
Friday				0.1504
				(0.7477)

We also investigate the median difference for the dwell times of trains by type of the day. In case of the intercity train series IC1900N, the median of the dwell times of early arriving trains on Sunday differs significantly from that on weekdays, Friday and Saturday. The median of the dwell times of late arriving trains on Friday is significantly different from that on weekdays. There is, however, no significant difference between the median of the dwell times of trains by type of the day for the interregional train series IR2200S. The p-values of the test statistic between the split dwell times are given in Tables 3.8 and 3.9 for both IC1900N and IR2200S, respectively (within parentheses those for the late arriving trains).

Table 3.8: p-values of the Wilcoxon rank sum test for the dwell times of early(late) arriving trains IC1900N by type of the day

	Monday	weekdays	Friday	Saturday
Sunday	0.4802	0.0107	0.0449	0.0366
	(0.7919)	(0.4490)	(0.3528)	(0.6303)
Monday		0.2168	0.2711	0.3087
		(0.6051)	(0.2926)	(0.4079)
weekdays			0.9201	0.9556
			(0.0412)	(0.1556)
Friday				0.7930
				(0.8091)

Table 3.9: p-values of the Wilcoxon rank sum test for the dwell times of early(late) arriving trains IR2200S by type of the day

	Monday	weekdays	Friday	Saturday
Sunday	0.4943	0.9067	0.6692	0.8967
	(0.7382)	(0.2067)	(0.1989)	(0.8839)
Monday		0.5177	0.3135	0.5747
		(0.3661)	(0.3494)	(0.7482)
weekdays			0.5992	0.8851
			(0.8278)	(0.1744)
Friday				0.6225
				(0.2254)

The significant difference of the arrival and departure delays and of the dwell times on Sunday can be attributed to the following facts. At first, there are less passengers who take trains on Sunday, leading to less primary delays. In addition, the capacity utilization on Sunday is lower than that on Monday, weekdays and Friday, resulting in less knock-on delays. The higher median of the dwell times for late arriving trains IC1900N on Friday might be due to longer alighting times of accumulated passengers that boarded the trains at the upstream stations in the case of larger arrival delays.

3.4.2   Train delays by period of the day

A similar analysis is carried out of the delays and dwell times of trains for the chosen train series IC1900N and IR2200S by period of the day: morning peak (scheduled arrival time 7:00 till 9:00 o'clock), evening peak (scheduled arrival time 16:30 till 18:30 o'clock) and off-peak (the remaining period of a day). The corresponding sample sizes are given in Table 3.10. Table 3.11 lists the median values of the arrival delays, departure delays and dwell times of trains by the three periods of the day for both IC1900N and IR2200S.

Table 3.10: Sample sizes of IC1900N and IR2200S by period of the day at The Hague HS

Period	IC1900N			IR2200S		
of day	All	Early	Late	All	Early	Late
morning peak	38	14	24	107	21	86
evening peak	51	29	22	113	40	73
off-peak	325	139	186	755	268	487
Total	414	182	232	975	329	646

Table 3.11: Medians of delays and dwell times of IC1900N and IR2200S by period of the day [s]

Period	IC1900N				IR2200S			
of	Arr.	Dwell times		Dep.	Arr.	Dwell times		Dep.
day	delays	Early	Late	delays	delays	Early	Late	delays
morning peak	30	90	82	53	63	92	81	86
evening peak	-13	82	72	48	25	79	75	48
off-peak	14	82	73	34	22	76	66	46
Total	14	83	73	46	27	77	68	50

We compare the median of the arrival delays corresponding to the three periods of the day with each other for both train series IC1900N and IR2200S, respectively. The p-values of the Wilcoxon rank sum test statistic between the split arrival delays are not smaller than 0.05 in case of the intercity train series IC1900N. We conclude that there is no significant difference between the median of the arrival delays by period of the day. In case of the interregional train series IR2200S, however, the median of the arrival delays during the morning peak period differs significantly from those during the evening peak and off-peak periods. The corresponding p-values of the Wilcoxon rank sum test are 0.011 and 0.000, respectively. In fact, similar results have also been obtained for the departure delays of both train series.

We also compare the median of the dwell times of trains corresponding to the three periods of the day using the Wilcoxon rank sum test. It has been found that the median of the dwell times for early arriving trains during the morning peak period is significantly different from that in the off-peak period for both selected train series. The corresponding p-values of the Wilcoxon rank sum test are 0.0259 and 0.011, respectively. This median difference may be because the conductors of trains would wait longer for late arriving passengers in the morning peak period if the train arrived at the station earlier than the scheduled time.

The Wilcoxon rank sum test shows that the median of the dwell times for late arriving trains during the off-peak period is significantly different from those in the morning and evening peak periods for the interregional train series IR2200S. The corresponding p-values of the Wilcoxon rank sum test are 0.0155 and 0.0105, respectively. These differences can also be found in the separate histograms shown in Figure 3.8, where the corresponding kernel density estimates (A detailed description of kernel density estimates can be found in Appendix A.2) are also added. The revealed dwell time differences may be because there are less passengers boarding and alighting during the off-peak period than in the morning and evening peak periods at the station. For the intercity train series IC1900N, however, there is no significant difference between the median of the dwell times for late arriving trains by period of the day. The corresponding p-values of the Wilcoxon rank sum test are all bigger than 0.05.
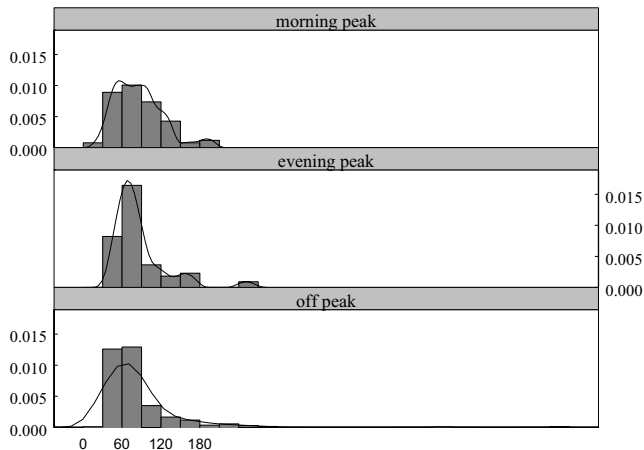
Figure 3.8: Histograms and kernel density estimates of the dwell times of late arriving trains IR2200S by period of the day

3.5 Route conflicts, knock-on delays and the impact on the punctuality of train operations

Route conflicts and the resulting knock-on delays often occur at busy stations and junctions, which may affect the punctuality of train arrivals and departures at the stations and the use of network capacity. The knock-on delay of a train is determined by the buffer time between the scheduled path of this train and that of the preceding train as well as the delays of the two successive trains. The buffer time is defined as the time added to the minimum line headway to avoid the transmission of smaller delays and it is the smallest slot between the blocking time stairways of trains (Pachl (2002)). We herein analyze route

conflicts, knock-on delays of trains and the impact on the punctuality of train operations by means of a case study of three pairs of train series at The Hague HS station.

The first pair of train series consists of the southbound intercity IC2100S and the interregional IR2200S, both of which originate from Amsterdam and approach The Hague HS station along the route $g \rightarrow c$ (Figures 3.2 and 3.3). The buffer time between the scheduled train paths is nearly 10 min at platform track 4. The second pair of train series is the southbound local AR5100S coming from The Hague CS along the route $j \rightarrow d$ and the northbound local AR5100N leading to The Hague CS along the route $b \rightarrow j$. The buffer time between the scheduled train paths is only 3 s at track junction 237BT. The third pair of train series is the northbound local AR5100N leading to The Hague CS and the southbound international INT600S coming from Amsterdam along the route $g \rightarrow c$. The buffer time between the scheduled train paths is 68 s at track junction 241BT (see Figure 3.9) .
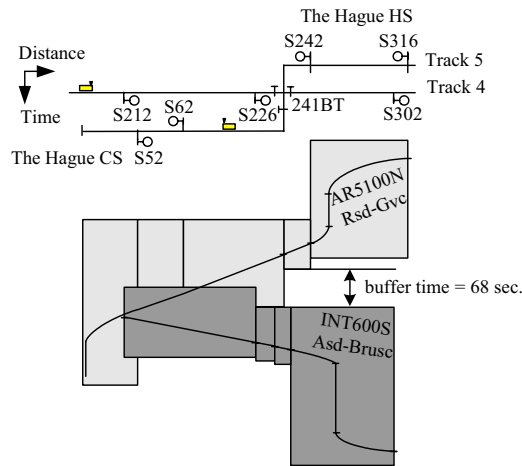
Figure 3.9: Buffer time between scheduled paths of AR5100N and INT600S

3.5.1   Route conflicts and knock-on delays

To analyze the route conflict between a pair of trains passing a signal block or a track junction in real operations, we hereby introduce a 'time lag' between the actual track release by the preceding train and the start of the virtual occupancy by the following train. In case of a train approaching a station, the virtual occupancy of the corresponding station block starts when the train arrives at sight distance of the approach signal of the home signal. For a departing train at a station, the virtual occupancy of the first signal block in the outbound direction starts at the later one of the so-called 'ready to depart' time (Carey & Kwiecinski (1995)) of the train at the platform track and the scheduled departure time. The 'ready to depart' time of a train at a station cannot be determined only on the basis of train detection data. Therefore, it is here assumed that a train is 'ready to depart' from The Hague HS station after the train has stopped at the platform track during the scheduled dwell time.

When the introduced time lag is negative, a route conflict occurs between a pair of trains. It is important to obtain the probability that the time lag is negative, which represents the chance of a route conflict. As we know, the order of two successive trains at a track junction may be altered in case of a large delay of the preceding one in real operations, but the chance is rather limited according to our analysis and these events have therefore been excluded.

According to our analysis, the probability of a route conflict is 0.4%, 80% and 62% for the three pairs of train series, respectively. The larger the buffer time between the scheduled train paths, the smaller the probability of a route conflict between a pair of train series. Figure 3.10 shows the empirical distribution function curve of the time lag between the release of junction 241BT by the northbound local AR5100N and the start of the virtual occupancy by the southbound international INT600S. The probability that the time lag is negative, which equals 62%, represents the probability that the international trains have a route conflict at this junction.
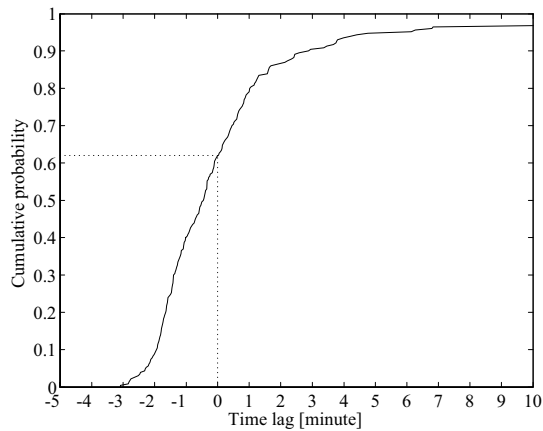
Figure 3.10: Empirical distribution function curve of the time lag between the release of track junction 241BT by AR5100N and the start of the virtual occupation by INT600S

In case of an inbound route conflict, an approaching train is forced to slow down from the approach signal of the home signal and even to make a temporary stop in front of the home signal of the station. Figure 3.11 displays the kernel density estimates of the average speeds in the open track signal block before the home signal for the international trains INT600S. The kernel estimate curve in case of late route setting due to route conflicts is shifted to the left compared to on-time route setting. The speeds of hindered trains in the open track block are on average 40 km/h less than the free running speed. This significant speed difference results in a bimodal kernel estimate of the train speeds in the open track signal block if all the international trains are considered. Note that the speeds of free running trains are, on average, around 20 km/h less than the scheduled one. This may be because of coasting of the trains when they are approaching the station.

The deceleration and even temporary stop of a hindered approaching train results in a longer running time in the open track signal block before the home signal of The Hague HS station. It is found that the running times of hindered trains increase by 70% on
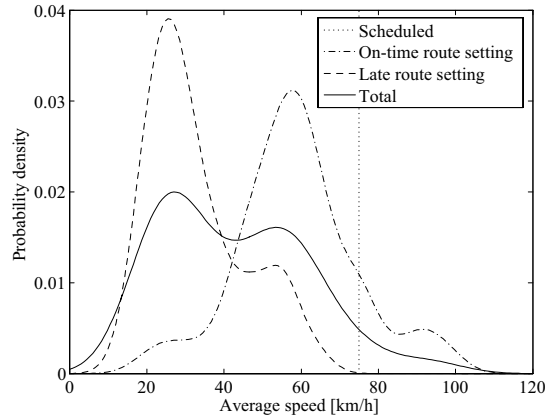
Figure 3.11: Kernel density estimates of the average speed of INT600S in the open track signal block before the home signal of The Hague HS

average compared to the running times of trains that are not hindered, while the 90th percentile of the running times increases by 115% in case of hinder. The mean knock-on delay for the hindered trains is about 1 min. The significant knock-on delays in case of route conflicts also result in a bimodal probability density curve for the running times of all the international trains in the open track signal block.

In case of an outbound route conflict, a departing train has to stop at the platform track for a longer time, resulting in a prolonged dwell time and a postponed departure. We have investigated the actual dwell times of late arriving trains for the northbound local train series AR5100N, whose scheduled dwell time is 1 min. It is found that the dwell times of hindered trains in case of route conflicts increase by 82% on average compared to the dwell times of the trains which depart from the station without any hindrance due to route conflicts, while the 90th percentile of the dwell times increases by 57% in case of hinder. The mean knock-on delay for the hindered departing trains at the station is about 1 min.

3.5.2   Impact on the punctuality of train operations

The overall impact of knock-on train delays caused by route conflicts on the punctuality of train operations is analyzed below by a case study of the third pair of train series, i.e., the northbound local AR5100N and the southbound international INT600S. Figure 3.12 shows a detailed classification of the outbound route setting and the departure delays of AR5100N, where the occurring percentage of a relevant event is an absolute percentage. Although the late setting of the train route due to the late release of track junction 237BT by the preceding train series AR5100S causes a considerable increase of the dwell times for the northbound trains AR5100N at The Hague HS, it leads to a relatively big share of moderate departure delays only. The probability that AR5100N suffers a route conflict and has a departure delay of less than 3 min is 76% and the conditional probability of this smaller delay in case of late route setting is 95%.

We recognized that the southbound local trains AR5100S which start the trip at The Hague
CS (3 km away from The Hague HS) have only small delays when releasing track junc-
tion 237BT. The departure punctuality of the northbound trains AR5100N leading to The
Hague CS is hence not very low although the trains suffer from knock-on delays caused
by a high probability of 80% of a late release of the conflicting junction.

A similar classification of the inbound route setting and the arrival delays of the south-
bound international INT600S has also been done. The percentage that the international
trains suffer a route conflict and have an arrival delay of less than 3 min is 61%, and the
conditional probability of this smaller delay in case of late route setting is more than 98%.
The international trains may arrive early at the approach signal of The Hague HS since
the last station Schiphol Airport is about 30 km away and a supplement time of about 7%
of the minimal running time is included in the scheduled running time. Thus, the route
conflicts of the international trains at track junction 241BT may be subject to early arrivals
of the trains themselves.
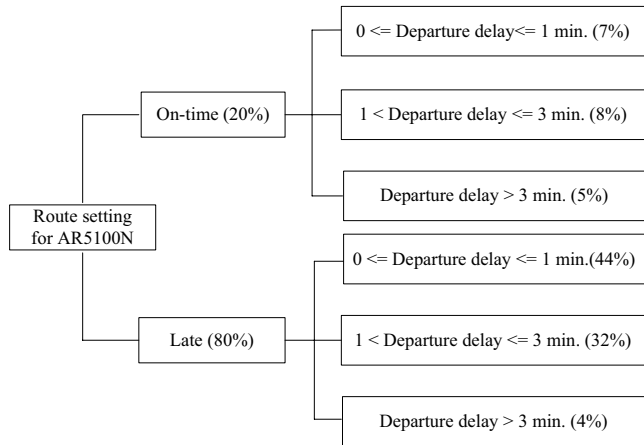
Figure 3.12: Classification of the route setting for and the departure delay of AR5100N
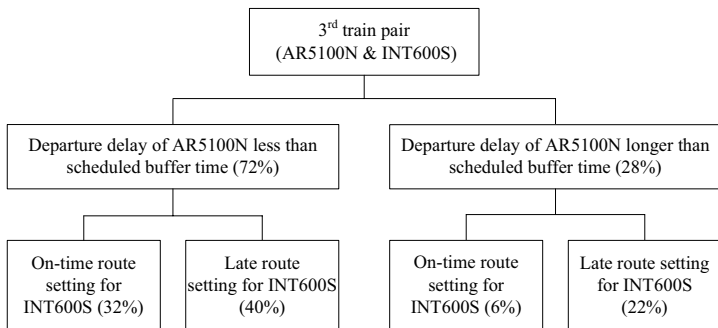
Figure 3.13: Classification of the departure delay of AR5100N and the setting of the inbound
route for INT600S

Figure 3.13 shows a classification of the departure delays of AR5100N and the setting of inbound routes for INT600S. Around two third of the late route setting due to route conflicts occur when the departure delays of the local trains are shorter than the buffer time between the scheduled paths of AR5100N and INT600S. This part of the route conflicts is mainly caused by early arrivals of the international trains at the approach signal of the home signal. Apparently, the remaining one third of the route conflicts is caused by departure delays of the preceding local trains. As we know, when a preceding local train has a large arrival delay at The Hague HS station, the following international train might get priority to first pass the conflicting track junction. Thus, the arrival punctuality of the international trains would be rather well, even in case of more anticipated route conflicts caused by the departure delays of the preceding local trains.

We have noticed that for this pair of train series passing a conflict point consecutively, the probability of the following one experiencing a delay larger than 3 min at The Hague HS is even higher in case of on-time route setting than in case of route conflicts. The larger train delays are mainly due to input delays generated at the upstream stations.

3.6   Station capacity utilization

The track occupation of railway networks operated according to a timetable has been studied frequently in other countries (Barter (2004), Höllmüller & Klahn (2005), Wahlborg (2005)). Most of the research with regard to station track capacity is either analytical or by simulation models. The analytical estimation of scheduled waiting times is based on queueing theory approaches (Wakob (1985), Schwanhäußer (1994), Wendler (1999)), whereas simulation models estimate the amount of total delays by generating random primary delays from a certain distribution and computing knock-on delays (Carey & Carville (2003), Carey & Crawford (2005)). In the following, we first introduce an analytical capacity research approach used for our analysis. Then, we discuss the main results regarding real track capacity utilization of the station and compare them with the estimation of the scheduled one.

3.6.1   Analytical estimation of station capacity

To perform an analytical capacity study of a complicated railway station, it is necessary to determine the capacity for different parts of the network. The analysis of station capacity, platform tracks and interlocking routes is done separately. The interlocking area is subdivided into smallest track elements called route nodes (Figure 3.14), which can be occupied by only one train at the same time. Every route node, thus, works as a single server (Schwanhäußer (1994)).

Based on the partitioning of the station tracks, we can estimate analytically the utilization of the station capacity by means of blocking times at two levels: 1) platform tracks and route nodes; 2) signal blocks and routes. The signal blocks include station blocks that are situated between a home signal and a departure signal of the station, as well as the open track signal blocks that are next to the station blocks in the inbound direction. The station blocks, in general, have a higher capacity utilization than other blocks because of longer
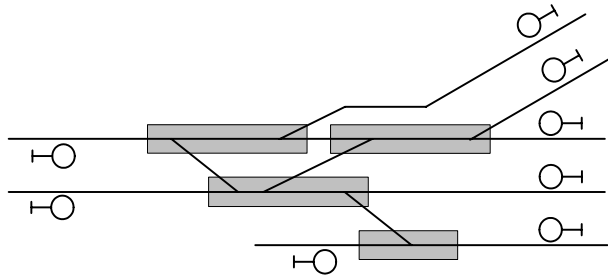
Figure 3.14: Dividing an interlocking area into route nodes (Source: Pachl (2002))

dwell times of trains for passengers' alighting and boarding and even indirect occupancy by trains. The indirect occupancy of a station block means the block is occupied indirectly by other trains that cross only a part of the track sections of the route.

For the blocks next to a station block in the inbound direction, the realized utilization of the track capacity may also be higher than scheduled due to a lower actual train speed, stochastic occupancy of the platform tracks or route nodes in the station area and the resulting knock-on delays of trains in front of the home signals of the station. The capacity utilization of a route is determined by the sequence of trains, headway times, speed differences, and train and block section lengths within the route.

The estimation of scheduled track capacity utilization is carried out based on the scheduled train paths and by compressing virtually the timetable train paths (Pachl (2002)) for a certain time period, e.g. a peak hour or a whole day. The capacity utilization rate is expressed as the ratio (in %) of the total track blocking time, which is the result of the compression process, and the considered time period. UIC (2004) recommends a capacity utilization rate of at most 75% during peak hours and 60% within a daily period for open tracks, but no threshold value has been determined for station tracks.

The real utilization of track capacity is, however, determined by stochastic instead of deterministic blocking times, which depend on variations of train speeds, acceleration/deceleration of trains and so on. The actual blocking time of a signal block (route node) by a train starts at the route setting for this train. In case of a route conflict, the route of a train is set later than the start of the virtual occupancy time. In this situation, the start of the track blocking time by the train depends on the time of track release by the preceding train.

On the other hand, the route of a train is often set slightly earlier than the start of the virtual track blocking because of automatic route setting according to the timetable and automatic detection of trains in the approach. Once the route has been set up, it is locked by signals and flank protection to prevent simultaneous occupation by other trains. Including the time of setting up the new route for the trains when it is different from the preceding one, this may result in a considerable increase of the estimated station track occupancy. The impact of stochastic delays and delay propagation to other trains on station track occupancy will be taken into account in order to enable a true estimation of the station track capacity as a function of a certain level of punctuality (Chapter 6).

3.6.2   Analysis of real capacity utilization of The Hague HS station

To analyze the real capacity utilization of The Hague HS station, we consider the standard routes confirmed by means of TNV-Prepare: a→e, b→e, b→i, b→j, g→c, h→d and j→d (Figures 3.2 and 3.3). It is very important to recognize the following facts:

1. The merging of the route (a→e) of the trains departing at platform track 6 and the route (b→e) of the local trains departing at platform track 5 and leading to Leiden in the first outbound block;

2. The crossing of the routes (b→i and b→j) of the trains departing at platform track 5 and leading to The Hague CS with the routes (g→c and h→d) of the trains approaching platform track 3 and 4;

3. The sharing of part of the route (b→j) of the local trains departing at platform track 5 and leading to The Hague CS with part of the route (j→d) of the trains originating from The Hague CS;

4. The merging of the route (h→d) of the local trains originating from Leiden and the route (j→d) of the trains originating from The Hague CS at platform track 3.

Considering the dwell process of trains and the above facts, we estimate the capacity utilization of The Hague HS station by decomposing this part of the station into four different platform tracks, i.e. track 3 to 6, and four route nodes, i.e. 229BT, 241BT, 227T and 237BT (Figure 3.2). The routes including and trains passing each of the decomposed infra elements are shown in Table 3.12.

Table 3.12: Routes including and trains passing each of the decomposed infra elements in The Hague HS

Infra. element	Passing routes	Passing trains	Trains /h
Track 3	h→d, j→d	AR5000S, AR5100S, IC1900S, IC2500S	6
Track 4	g→c	IR2200S, IC2100S, IC2400S, INT600S, HST9300S	6
Track 5	b→e, b→i, b→j	AR5000N, AR5100N, IC1900N, IC2500N	6
Track 6	a→e	IR2200N, IC2100N, IC2400N, INT600N, HST9300N	6
229BT	b→i, g→c	IR2200S, IC2100S, IC2400S, INT600S, HST9300S, IC1900N, IC2500N	8
241BT	b→j, g→c	IR2200S, IC2100S, IC2400S, INT600S, HST9300S, AR5100N	8
227T	b→i, h→d	AR5000S, IC1900N, IC2500N	4
237BT	b→j, h→d, j→d	AR5000S, AR5100S, IC1900S, IC2500S, AR5100N	8

The estimation results regarding the scheduled and realized utilization of the platform tracks and route nodes are shown in Figure 3.15. The scheduled capacity utilization rate

varies between 35% and 60% for the different platform tracks. This rate is considerably higher than that of the adjacent route nodes being 15% to 35%. This is due to the fact that the dwell times of trains increase the blocking times of platform tracks. A relatively higher capacity utilization at route node 237BT and platform track 3 is related to a low design speed of 40 km/ h between The Hague CS and HS station and a longer scheduled dwell time of in total 12 min for the corresponding platform track.
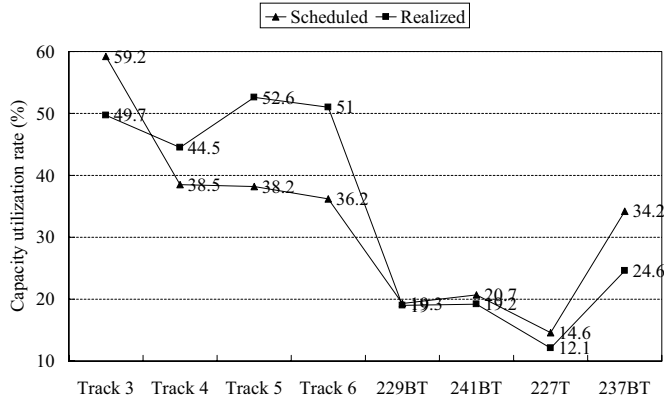
Figure 3.15: Scheduled and real capacity utilization of platform tracks and route nodes in The Hague HS

The realized track capacity utilization is between 6% and 15% higher than the scheduled one at platform tracks 4, 5 and 6. This is due to lower than scheduled running speeds and prolonged dwell times during train operations. As we have shown, the dwell times of trains are prolonged not only because of the variation of the dwell process for passengers' alighting and boarding, but also due to early arrivals of a big percentage of the trains in both directions and the outbound route conflicts for the northbound trains leading to The Hague CS and Leiden/Amsterdam.

Surprisingly, at platform track 3 and each of the route nodes, the realized track capacity utilization is between 0.3% and 10% lower than scheduled, although route conflicts often occur due to short headway and buffer times between train paths, as well as stochastic train delays. In case of a route conflict between a pair of trains at a platform track or route node, the distribution of the virtual blocking time of the track section by the following train overlaps with the physical occupancy by the preceding one. The actual track blocking by the following train can, in fact, only start after the preceding train releases the infrastructure. Before this happens, the following train approaches the home signal of the station and may have stopped in front of this block signal. As there remains only a short approaching distance and time, the actual track blocking time of the following train may be shorter than the scheduled one. This partially explains why the realized track capacity utilization of the critical route nodes is somewhat lower than the scheduled one. In addition, the station home signal S232 for the trains that originate from The Hague CS may be cleared late even in case of no train occupying platform track 3. This may contribute to the lower real capacity utilization of platform track 3 and route node 237BT, as well as the late departure of trains from The Hague CS station.

The relevant station blocks for our analysis include S320-244, S316-242, S226-302, S228-304 and S232-304 (Figure 3.2). The routes passing, infra elements included in, and trains utilizing each of the blocks are given in Table 3.13. The scheduled and realized capacity utilization of the five station blocks is displayed in Figure 3.16. Although the same number of trains pass platform tracks 3, 4, 5 and 6, the scheduled capacity utilization of station blocks S228-304, S232-304 and S226-302 is 10% to 25% higher than that of station blocks S316-242 and S320-244 because of the infrastructure and timetable characteristics. The high scheduled capacity utilization of station blocks S228-304 and S232-304 of about 60% is mainly caused by a low design speed between The Hague CS and HS stations and a longer total dwell time at platform track 3. The relatively higher scheduled capacity utilization of station block S226-302 compared to station blocks S316-242 and S320-244 is mainly because of indirect occupancy of the station block S226-302 by the four trains per hour departing from platform track 5 and leading to The Hague CS.

Table 3.13: Routes passing, infra elements included in and trains utilizing each of the station blocks in The Hague HS

Station block	Passing routes	Infra. elements	Trains utilizing the block	Trains /h
S320-244	a→e	Track 6	IR2200N, IC2100N, IC2400N, INT600N, HST9300N	6
S316-242	b→e, b→i, b→j	Track 5	AR5000N, AR5100N, IC1900N, IC2500N	6
S226-302	g→c	Track 4 229BT 241BT	IR2200S, IC2100S, IC2400S, INT600S, HST9300S IC1900N*, IC2500N*, AR5100N*	10
S228-304	h→d	Track 3 227T 237BT	AR5000S AR5100S*, IC1900S*, IC2500S*, IC1900N*, IC2500N*, AR5100N*	10
S232-304	j→d	Track 3 237BT	AR5100S, IC1900S, IC2500S AR5000S*, AR5100N*	8

* Indirect utilization
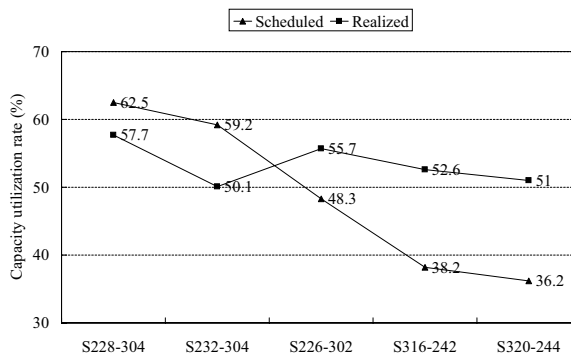
Figure 3.16: Scheduled and real capacity utilization of station blocks in The Hague HS

The real capacity utilization is nearly 15% higher than the scheduled one for station blocks S316-242 and S320-244, and 7.4% higher in case of station block S226-302. The real capacity utilization of station blocks S228-304 and S232-304 is, however, 4.8% and 9.1% lower than the scheduled one. The realized lower and slightly higher than the scheduled capacity utilization at station blocks S228-304, S232-304 and S226-302 can be explained just like for the four route nodes and platform 3.

In case of a route conflict between a pair of trains at a platform track or an inbound route node included in the corresponding station block, the actual occupancy of the critical track infrastructure by the following train can be very short and even shorter than the scheduled blocking time. However, if we consider the open track signal block in front of the home signal of the track infrastructure, the real blocking time becomes very long and generally much longer than the scheduled blocking time. This is illustrated by the kernel density estimates of the track blocking times of the signal block before the home signal of the station by the southbound international trains INT600S (Figure 3.17).
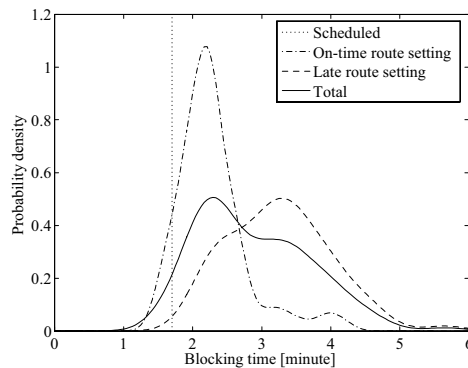
Figure 3.17: Kernel density estimates of the track blocking times of the signal block before the home signal of The Hague HS by INT600S

The above analysis has shown the high occupation of the station blocks and of the open track signal blocks before a home signal of the station. As we know, platform tracks and junctions are, in general, the bottlenecks in a station area. This is because of the longer dwell times of trains at the station, speed limits of trains at the switches, deceleration and acceleration of trains during approaching and departing the station and higher frequency of trains passing the junctions where different train routes merge or cross. To confirm the existence of bottlenecks of The Hague HS station in real-time train operations, we show the empirical distribution function curve of the differences between the station block occupancy times and the blocking times of the preceding open track signal block by INT600S in Figure 3.18.

The cumulative probability of the blocking time differences being negative, i.e., the probability that the blocking time of the preceding signal block is longer than the station block occupancy time is nearly 10%. This implies that a new bottleneck has been created just before the interlocking of the station tracks. It should be mentioned that the scheduled dwell time of the southbound international INT600S is 2 min. If the scheduled dwell time changes to only 1 min, this new bottleneck will be more critical in the area of the station.
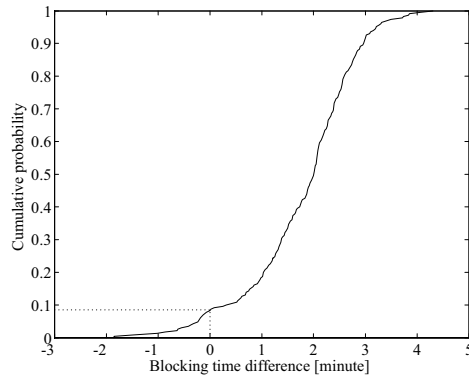
Figure 3.18: Empirical distribution function curve of the differences between the station block occupancy times and the blocking times of the preceding signal block by INT600S

3.7   Conclusions

The main research results obtained from a detailed statistical analysis of train delays and delay propagation at The Hague HS station have been presented. According to the official definition of the Dutch Railways, the arrival punctuality of each individual train series at this station is between 87% and 97%. If punctuality was defined at the level of 1 min, the percentage of trains that arrived punctually at the station would drop to between 31% and 89%. The delays of trains increase generally from arrival to departure at the station. The mean dwell time of early arriving trains is 20 to 70 s longer than the scheduled one for all the train series even when the part of dwell times due to early arrivals is excluded, while the mean dwell time of late arriving trains is 15 to 45 s longer than the scheduled one for most of the train series.

Train delays are not significantly different between working days and Saturday, but the delays of trains on Sunday differ significantly from those from Monday till Friday. For a selected intercity train series in northbound direction, the train delays do not vary much during different time periods of the day, whereas the arrival delays, departure delays and dwell times of an interregional train series in southbound direction differ significantly between peak hour periods and off-peak period. It can be concluded that the delays of commuter trains depend more on the period of the day.

Route conflicts between a pair of train series cause considerable knock-on delays and reduce the punctuality of the following train series. To reduce the probability of knock-on delays for a train series, it is necessary to decrease not only the delays of the preceding train series, but also to avoid early arrivals of the train series itself.

The real capacity utilization of most track infrastructure elements of the station is 8 to 15% higher than the scheduled one due to the train delays and delay propagation in the area of the station. The estimated real capacity utilization of one platform track and the route nodes in the area of the station is surprisingly lower than the scheduled one due to rather frequent route conflicts at these track sections and the following reduction of the approaching time after the release of the track sections.

However, in case of route conflicts, the average blocking time of the open track signal block before the home signal of the station is significantly longer than the scheduled one and that in case of on-time route setting. As a result, the blocking times of the preceding signal block by an investigated international train series in southbound direction become even longer than the station block occupancy times with a probability of nearly 10%. This implies that in case of hinder the bottleneck may be shifted upstream just before the interlocking area of the station tracks.

The statistical analysis of train operations establishes a solid base for developing a more realistic delay propagation model to estimate knock-on delays and the impact on the punctuality of trains with a higher accuracy than the existing one(s). The train detection data adopted for this analysis will be used again for the assessment of the distribution models of train event and process times and for the validation of the developed delay propagation model in the next chapters.

Chapter 4

Statistical distributions of train event and process times

4.1   Introduction

In recent decades, a variety of theoretical distributions such as the normal, uniform, exponential, gamma, beta, Weibull, and log-normal distributions (Bury (1999)) have been adopted in the literature to model the stochasticity of train event and process times. However, limited literature exists for statistical modelling of the distributions based on empirical data. This chapter deals with an evaluation of those seven candidate distribution models on the basis of the train detection data used for the statistical analysis of train operations in the previous chapter. The evaluation of the distribution models is to be performed not only for the arrival and departure times (delays) of trains at the station, but also for the arrival times of trains at the boundary of the local railway network and the train running, dwell and track occupancy times within the local network. Furthermore, the distribution models will be evaluated for the running and dwell times as well as track occupancy times of trains in case of (no) hinder caused by other trains.

4.2   Common approaches to distribution evaluation

For evaluating the 'quality' of fitted distributions, both heuristic procedures and goodness-of-fit hypothesis tests are generally applied. Law & Kelton (2000) presented five main heuristic or graphical procedures for comparing fitted distributions with the true underlying distribution. These procedures include density/histogram overplot, frequency comparison, distribution function differences plot, probability-probability plot and quantile-quantile plot.

A goodness-of-fit test is used to assess formally whether a data set comes from a postulated particular distribution. Before proceeding with a description of common goodness-of-fit tests, we feel necessary to mention several properties of these tests. First, failure to reject the null hypothesis H_0 (the probability distribution function of the observed random variable is $F_0(x)$) should not be interpreted as 'accepting H_0 as being true' (Law &

Kelton (2000)). This property is also applicable to other statistical hypothesis tests. In addition, goodness-of-fit tests are often not very powerful for small to moderate sample sizes; that is, they are not very sensitive to subtle disagreements between the data and the fitted distribution. On the other hand, if the sample size n is very large, then these tests will almost always reject H_0 (Gibbons (1985)). This is an unfortunate property of these tests, since it is usually sufficient to have a distribution which is nearly correct.

The oldest goodness-of-fit test is the chi-square test (Gibbons (1985), Law & Kelton (2000)), which can be thought of as a more formal comparison of a histogram with the fitted density function. First, suppose that all parameters of the hypothesized distribution have been specified without making use of the corresponding data. In this case, if the null hypothesis H_0 is true, the chi-square test statistic χ^2 converges in distribution (as $n \to \infty$) to a chi-square distribution with $k - 1$ degrees of freedom (k is the number of grouped categories). Thus, for large n, a test with a significance level α is obtained approximately by rejecting H_0 if $\chi^2 > \chi^2_{k-1,1-\alpha}$, where $\chi^2_{k-1,1-\alpha}$ is the upper $1 - \alpha$ critical point for a chi-square distribution with $k - 1$ degrees of freedom.

Second, suppose that the parameters of the hypothesized distribution have been estimated from the data. If H_0 is true, then as $n \to \infty$ the distribution function of χ^2 converges to a distribution function that lies between the distribution functions of chi-square distributions with $k - 1$ and $k - m - 1$ degrees of freedom (m represents the number of estimated parameters). If we let $\chi^2_{1-\alpha}$ be the upper $1 - \alpha$ critical point of the asymptotic distribution of χ^2, then

$$\chi^2_{k-m-1,1-\alpha} \leq \chi^2_{1-\alpha} \leq \chi^2_{k-1,1-\alpha}. \tag{4.1}$$

It is often recommended that we reject H_0 only if $\chi^2 > \chi^2_{k-1,1-\alpha}$, since this is conservative (the rejection area is smaller than necessary). This choice, however, will entail loss of power (probability of rejecting a false H_0) of the test.

Another commonly used goodness-of-fit test is the Kolmogorov-Smirnov (K-S) test (Gibbons (1985), Law & Kelton (2000)), which compares an empirical distribution function with the hypothesized distribution function. A detailed description of this test can be found in Appendix A.3. The form of the K-S test is to reject the null hypothesis H_0 if the statistic D_n exceeds some constant $d_{n,1-\alpha}$, which is the upper $1 - \alpha$ critical point of the distribution of D_n.

If all the parameters of the hypothesized distribution are specified without making use of the data, a single table of values for $d_{n,1-\alpha}$ will suffice for all continuous distribution forms. This all-parameters-known case is the original form of the K-S test (Law & Kelton (2000)). If the distribution parameters have been estimated from the data, the outcome of the K-S test in the original form is conservative. More recently, the K-S test has been extended to allow for estimation of the parameters from the data in the cases of normal (log-normal), exponential, and Weibull distributions by modifying the original critical point $d_{n,1-\alpha}$, respectively.

Both the chi-square test and the K-S test have been widely used to determine goodness-of-fit, but they have their own advantages and drawbacks. Perhaps the main advantage of the chi-square test is that when unknown parameters must be estimated from the data, a generic correction can be introduced in the statistic distribution by reducing the number of degrees of freedom. If parameters must be estimated to apply the K-S test, no general

adjustment is available for the upper $1 - \alpha$ critical point of the statistic distribution. However, the K-S test also has several advantages over the chi-square test. The K-S test does not require us to group the data in any way, so no information is lost; this also eliminates the troublesome problem of interval specification, which occurs in case of the chi-square test. Another main advantage of the K-S test is that this test is exact for any sample size in the all-parameters-known case, whereas the chi-square test is valid only in an asymptotic sense. Therefore, for all continuous population distributions, if the parameters of the hypothesized distribution are specified without making use of the data, the K-S test is applied preferably.

4.3 Distribution evaluation based on a parameter fine-tuning method

The parameters of the hypothesized distribution of a goodness-of-fit test may be specified by experience only. In this case, the null hypothesis can often be rejected due to a gross estimation of the parameters. Therefore, we herewith propose a new method for fine-tuning the parameters of theoretical distributions. The fitted different forms of distributions with fine-tuned parameters are then evaluated by the K-S goodness-of-fit test.

To fit a distribution to a data set (Ω) of train event or process times, we first split the data set into two parts (Ω_1 and Ω_2) randomly, e.g., by assigning the chronologically recorded times into the data subsets alternately. Then, using the split data subset Ω_1, an initial estimate of the parameters of the distribution is obtained. Furthermore, we fine-tune the parameters of the distribution to locally optimize the goodness-of-fit of the K-S test where the empirical distribution is obtained from the data subset Ω_2. The sample size of these data subsets has been halved, but the original form of the K-S test, applicable to all distribution types, can be used accurately. In addition, by randomly splitting the original data set into two parts, we are sure that the obtained distribution fit to the data subset Ω_2 is the distribution we need for modelling the stochasticity of the original data set.

On the basis of empirical data, the parameters of a given distribution can be estimated using the moment method, maximum likelihood method, and Bayesian estimation (Ross (2004)). Estimates by the moment method sometimes fail to take into account all relevant information in the sample; i.e., they are sometimes not sufficient statistics. The Bayesian estimation is generally used when there is prior information of distribution parameters. The maximum likelihood method is widely used in practice because a Maximum Likelihood Estimator (MLE) is the minimum variance unbiased as the sample size increases (Dekking et al. (2005)). A detailed description of this parameter estimation approach can be found in Appendix A.4.

The estimation of distribution parameters can be influenced by outliers. A outlier is a data point which deviates from the bulk of the data (Goverde et al. (2001)). In case of train event and process times, the outliers are mostly some very large values, as has been confirmed by our empirical data analysis. The impact of an outlier depends on the sample size and the spread of the data. However, there is no (exact) standard method to detect outliers. A new heuristic procedure to deal with outliers is incorporated in the following approach for fine-tuning the parameters of a distribution model of a data set (Ω):

1. An initial estimate of the distribution parameters is obtained by using the MLE based on the split data subset Ω_1.

2. The large delays in the data subset Ω_1 are omitted iteratively one by one estimating the distribution parameters correspondingly using the MLE. In each iteration, we compute the p-value of the K-S test where the empirical distribution is, however, obtained from the split data subset Ω_2. The iterative procedure terminates if the p-value cannot be increased any more.

3. After the iterative procedure, we fine-tune the parameter estimate again to further improve the estimation by maximizing the p-value of the K-S test, where of course the empirical distribution remains unchanged. In detail, a neighborhood ± 1.0 and a fine-tuning step 0.1 are considered for the shape parameters of the gamma, beta and Weibull distributions as well as the parameters of the log-normal distribution. In case of other distribution parameters, a neighborhood ± 10 and a fine-tuning step 1.0 are considered. Herein, the unit of time is in seconds.

The parameters of a distribution model can be classified, on the basis of their physical or geometric interpretation, as being one of three basic types: location, scale, or shape parameters. They are represented respectively by c, $1/\lambda$, and k in this thesis. If the location parameter of a distribution model is the lower endpoint of the distribution's range and is nonzero, then the location parameter is also called shift parameter and the distribution is called a location-shifted distribution. Train event and process time distributions generally have a lower bound which equals the earliest or the minimal time. Therefore, a location-shifted distribution is mostly applied. It should be mentioned that we take the shift parameter as the minimum value of the data observations.

To assess the candidate distribution models, the K-S tests are performed at a commonly adopted significance level $\alpha = 0.05$. Each candidate distribution is ranked according to the p-value of the K-S test where the hypothesized distribution is specified with the fine-tuned parameters. In detail, we give rank 1 to the distribution model with the biggest p-value, rank 2 to the distribution with the second biggest p-value and so on. The distribution model ranked 1 is considered to be the 'best' among the candidate distributions. The quality of the distribution fitting for the event and process times of trains is also visualized by comparing the fitted distribution density curve with the kernel density estimate and the empirical histogram and by applying the distribution differences plot for the fitted distribution and the empirical one.

4.4 Results of the distribution evaluation

Based on the empirical data recorded in The Hague HS station, we evaluate the candidate distributions for train event and process times. Train event and process time distributions may depend on types and routes of the trains. We hence assess the candidate distributions for the event and process times of trains per train series in each direction (southbound or northbound). A train series in one direction is hereafter referred to as a studied case.

4.4.1 Arrival times

The modelling of the train arrival time distribution is a prerequisite for predicting the propagation of train delays at stations. To incorporate the impact of knock-on delays caused by route conflicts in a delay propagation model, we need to distinguish the arrival times of trains at the station platform from those at the approach signal of the station home signal. Early arriving trains generally have a longer dwell time than late arriving trains. To model the distribution of departure times more realistically by distinguishing the dwell times of late arriving trains from those of early arriving trains, the distribution of non-negative arrival delays is also needed (Chapter 5).

We have evaluated the candidate distributions for both the arrival times of trains at the approach signal of the station home signal and at the platform track. For the arrival times of trains at the approach signal, the log-normal distribution gives the best fit in 9 out of 14 studied cases, and the gamma, Weibull, or normal distribution is the best among the candidate distributions in the other 5 cases. In addition, the log-normal distribution ranks second or third in 4 out of the 14 cases. For the arrival times of trains at the platform track, the log-normal distribution is the best model among the candidate distributions in 11 out of 14 studied cases and it ranks second or third in the other 3 cases, where the gamma, Weibull or normal distribution gives the best fit to the applied data subsets.

Table 4.1: Fine-tuned parameters and the K-S test result of the fitted log-normal distribution as well as the sample size of the original data set of the arrival times of trains in The Hague HS station (the reference time is defined as the scheduled arrival time at the platform track per studied case)

Train series	n	At approach signal					At platform track				
		c [s]	μ [s]	σ [s]	p	R_1	c [s]	μ [s]	σ [s]	p	R_1
IR2200S	772	-190	4.6	0.6	0.27	1	-108	4.8	0.6	0.48	1
IC2100S	373	-184	4.4	0.7	0.83	1	-98	4.6	0.7	0.86	1
IC2400S	388	-177	4.2	0.8	0.43	1	-79	4.3	0.9	0.74	1
INT600S	361	-279	4.9	0.8	0.68	2	-126	5.0	0.5	0.61	1
HST9300S	103	-369	4.7	0.8	0.74	3	-141	4.9	0.5	0.54	2
AR5000N	545	-219	4.7	0.4	0.00	3	-99	4.8	0.5	0.02	1
AR5100N	802	-255	4.5	0.5	0.00	1	-152	4.6	0.5	0.01	1
IR2200N	974	-248	4.6	0.4	0.00	2	-136	4.5	0.5	0.01	2
IC1900N	395	-251	4.9	0.5	0.87	1	-143	5.0	0.5	0.44	1
IC2100N	506	-285	5.0	0.7	0.70	1	-182	5.0	0.7	0.78	1
IC2400N	503	-290	4.8	0.6	0.88	1	-197	4.9	0.6	0.99	1
IC2500N	391	-242	4.8	0.6	0.44	1	-117	4.9	0.5	0.38	1
INT600N	429	-266	4.7	0.6	0.28	1	-162	4.9	0.7	0.33	1
HST9300N	143	-259	4.6	1.4	0.00	6	-146	4.6	1.4	0.01	3

Note: p-values have been rounded to two decimals

Table 4.1 lists the fine-tuned parameters and the K-S test result of the fitted log-normal distribution as well as the sample size (n) of the original data set for both the arrival times of trains at the approach signal of the station home signal and at the platform track per train series per direction. Herewith, μ and σ represent the mean and standard deviation

of the underlying normal distribution, respectively, and c denotes the shift parameter; R_1 denotes the rank of the log-normal fit among the candidate distributions. It should be mentioned that the sample size of the data subsets used for parameter specification and the goodness-of-fit test has been halved in each studied case. The p-values of the K-S test are larger than 0.05, thus the fitted log-normal distributions have been accepted in most cases.

Figure 4.1 shows the log-normal and gamma density fits, kernel density estimate and empirical histogram for the arrival delays of the northbound intercity trains IC2100N at the platform track. The distribution differences plots are also given in Figure 4.2. Both density fits match with the kernel density estimate and the empirical histogram rather well since the log-normal and gamma distributions feature shape flexibility. It appears that the log-normal fit is more attractive than the gamma fit. The differences between the log-normal distribution and the empirical distribution are overall smaller than the differences between the gamma distribution and the empirical one, too. The log-normal model appears better than the gamma model for fitting a distribution to a date set with a positively(right) skewed density estimate and empirical histogram including a long tail.
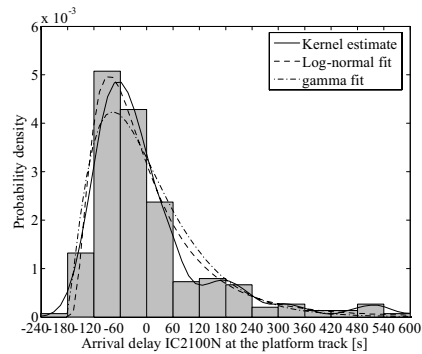
Figure 4.1: Log-normal and gamma density fits, kernel density estimate and histogram for the arrival delays of IC2100N at the platform track

Figure 4.2: Distribution function difference plot for the log-normal fit and the arrival delays of IC2100N at the platform track and the plot for the gamma fit and the empirical delays

We have also evaluated the candidate distributions for non-negative arrival delays at the platform track. The evaluation results reveal that the Weibull model gives the best fit among the candidate distributions in 12 out of 18 studied cases and it ranks second or third in the other 6 cases. In these cases, either the gamma or log-normal fit is the best among the candidate distributions. Table 4.2 lists the fine-tuned parameters (k and $1/\lambda$) and the K-S test result of the fitted Weibull distribution as well as the sample size of the original data set of the non-negative arrival delays per train series in each direction. R_{w} is hereafter used to represent the rank of the Weibull fit among the candidate distributions. The p-values are larger than 0.05, the Weibull fits have been hence accepted by the K-S test in all the cases. In addition, it has also been found that the shape parameter of the Weibull fit to non-negative arrival delays is not larger than 1.0 except for the southbound trains originating from The Hague CS station.

Table 4.2: Fine-tuned parameters and the K-S test result of the fitted Weibull distribution as well as the sample size of the original data set of non-negative arrival delays

Train series	n	k	$1/\lambda$ [s]	p	R_{w}
AR5000S	380	1.0	90	0.99	1
AR5100S	818	1.5	60	0.10	2
IR2200S	465	0.9	89	0.71	1
IC1900S	411	1.7	78	0.80	3
IC2100S	162	0.9	89	0.97	1
IC2400S	184	0.7	111	0.76	1
IC2500S	210	1.8	86	0.93	1
INT600S	237	0.9	85	0.99	1
HST9300S	35	1.0	56	0.95	1
AR5000N	279	1.0	106	0.98	1
AR5100N	199	0.9	89	0.44	2
IR2200N	228	0.7	114	0.94	1
IC1900N	219	1.0	103	0.95	1
IC2100N	194	0.8	137	0.08	3
IC2400N	134	1.0	91	0.71	1
IC2500N	222	0.9	80	0.61	1
INT600N	146	0.7	144	0.75	2
HST9300N	40	0.5	146	0.85	2

Note: p-values have been rounded to two decimals

Figure 4.3 shows the Weibull and gamma density fits as well as the empirical histogram for the non-negative arrival delays of the northbound interregional trains IR2200N. It appears that the Weibull density fit matches with the empirical histogram better than the gamma fit. Figure 4.4 displays the distribution differences plots. The differences between the Weibull distribution and the empirical distribution are overall smaller than the differences between the gamma distribution and the empirical one, too.

In conclusion, a location-shifted log-normal distribution can be considered as the best model among the candidate distributions for both the arrival times of trains at the platform track and at the approach signal of the station home signal. The earliest arrival time

of trains corresponds to the shift parameter of the distribution model. The Weibull distribution with a shape parameter not larger than 1.0, i.e., with a decreasing density curve is generally the best model among the candidate distributions for non-negative arrival delays. If the shape parameter can be fine-tuned to be nearly 1.0, the exponential distribution as a special type of the Weibull distribution, will be used to approximate the stochasticity of non-negative arrival delays. This is in line with Schwanhäußer (1974), Goverde et al. (2001), and Wendler & Naehrig (2004).
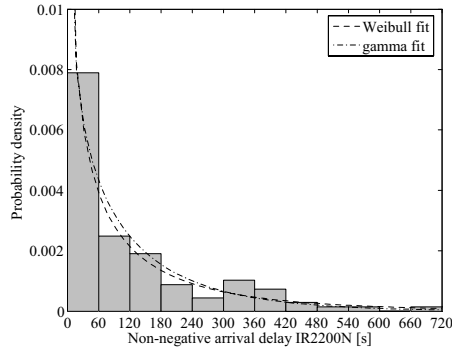
Figure 4.3: Fitted Weibull and gamma density curves and histogram of the non-negative arrival delays of IR2200N
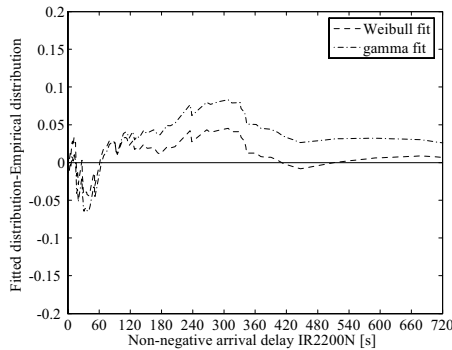
Figure 4.4: Distribution differences plot for the Weibull fit and the non-negative arrival delays of IR2200N and the plot for the gamma fit and the empirical delays

4.4.2   Departure delays

Departure delays are generally non-negative since passenger trains are not allowed to depart from the station earlier than the scheduled departure time. The distribution of departure delays can be used to predict the distribution of outbound track release times and the distribution of train arrival times at the following stations.

It has been found that the Weibull distribution is the best model among the candidate distributions for the departure delays of trains in 11 out of 18 studied cases and it ranks second or third in the other 7 cases. In these cases, the gamma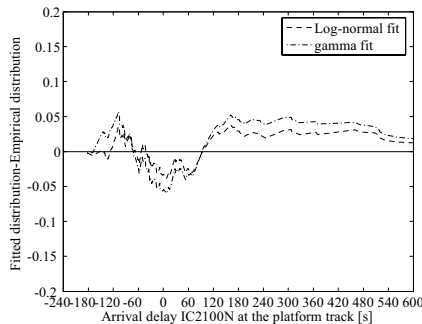, log-normal or beta distribution is the best model among the candidate distributions. Table 4.3 lists the fine-tuned parameters and the K-S test result of the Weibull fit as well as the sample size of the original data set of the departure delays per train series per direction. The p-values are larger than 0.05, the Weibull distribution fits have been hence accepted by the K-S test in most cases. In addition, we have found that the shape parameter of the Weibull fit is not larger than 1.0 in 11 out of the 18 cases.

Table 4.3: Fine-tuned parameters and the K-S test result of the Weibull fit as well as the sample size of the original data set of departure delays

Train series	n	k	$1/\lambda$ [s]	p	R_w
AR5000S	443	0.5	43	0.00	3
AR5100S	836	0.5	29	0.00	2
IR2200S	772	0.9	68	0.18	1
IC1900S	411	2.1	111	0.66	2
IC2100S	373	1.0	69	0.29	1
IC2400S	388	0.8	58	0.28	1
IC2500S	210	1.9	113	0.98	1
INT600S	361	1.4	59	0.02	3
HST9300S	103	1.3	67	0.88	2
AR5000N	545	1.0	93	0.42	1
AR5100N	802	1.4	85	0.13	2
IR2200N	974	0.9	49	0.18	1
IC1900N	395	1.1	72	0.40	1
IC2100N	506	0.9	72	0.18	1
IC2400N	503	0.9	43	0.25	1
IC2500N	391	1.0	71	0.34	1
INT600N	429	1.2	75	0.07	2
HST9300N	143	0.5	38	0.19	1

Note: p-values have been rounded to two decimals

Figures 4.5 and 4.6 visualize the goodness-of-fit of the Weibull and gamma distributions for the departure delays of the southbound intercity trains IC2400S. It appears that the Weibull density fit matches with the empirical histogram better than the gamma fit. Moreover, the maximum difference between the Weibull distribution and the empirical distribution are smaller than the maximum difference between the gamma distribution and the empirical one. Based on these graphical illustrations and the results of the K-S goodness-of-fit test, we conclude that the Weibull distribution can generally be considered as the best model among the candidate distributions for departure delays. However, the distribution shape parameter can be either smaller or larger than 1.0, which is subject to the dwell times at platform. Just like for non-negative arrival delays, the exponential distribution will be used to approximate the distribution of departure delays if the shape parameter can be fine-tuned to be nearly 1.0.
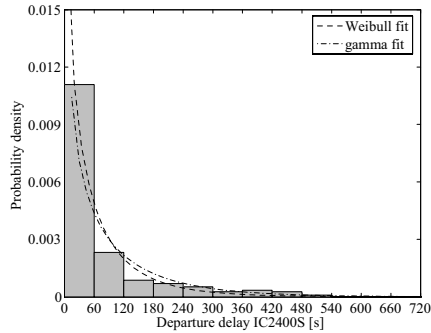
Figure 4.5: Fitted Weibull and gamma density curves and histogram of the departure delays of IC2400S
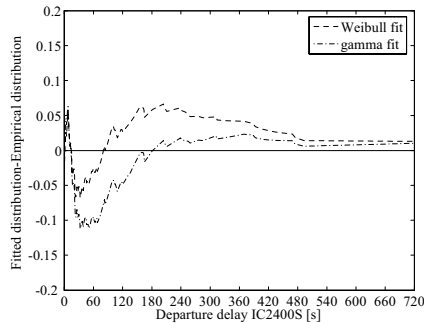
Figure 4.6: Distribution differences plot for the Weibull fit and the departure delays of IC2400S and the plot for the gamma fit and the empirical delays

During the above analysis, we found that the empirical density curves of stochastic arrival and departure times often have a long tail. It appears that these event times are not a simple random variable. As we know, train arrival and departure times are subject to the alteration of train orders in real-time operations, which may contribute to the long tails of the empirical probability density curves. A mixture distribution, which is applied for modelling the combined effect of a number of random variables, might be an alternative distribution model which can explicitly incorporate the impact of the alteration of train orders. However, finite mixtures will introduce additional complexity because they have more parameters, whose estimation usually requires solving a complex system of non-linear equations. Moreover, it appears that most people prefer to use a simpler model provided that it can approximate the reality reasonably well. Thus, fitting a mixture distribution to the observed arrival and departure times, respectively, is not studied further.

4.4.3 Free dwell times

The dwell times of trains at a station are the difference between the arrival and departure times. To estimate the knock-on delays and the departure delays of trains at a station, it

is necessary to know the free dwell times. In this thesis, we define the *free dwell time* of a train as the necessary dwell time for passenger alighting and boarding in the absence of hindrance from other trains. We realize that it is hardly possible to measure the free dwell time for an individual train which is hindered at the platform track due to the occupancy of the next track block only based on the track occupancy and release records, i.e., TNV data. Therefore, we are going to obtain the distribution of the free dwell times of trains per train series per direction, which will be used to estimate the knock-on delay distribution and the departure delay distribution. The measurement of the free dwell time for every involved train is not our main concern.

The free dwell time of a train at a station is subject to the arrival delay. To accurately estimate the knock on delay distribution and the departure delay distribution, it is necessary to adopt the conditional free dwell time distributions in case of different sizes of arrival delays (see Chapter 5). For a shorter computation time, we will only use the conditional free dwell time distributions in case of early and late arrivals. According to our empirical data analysis, the free dwell time of an early arriving train is influenced by the arrival earliness significantly. Therefore, we will estimate the knock-on delay and the departure delay for an early arriving train on the basis of the scheduled arrival time and the free dwell time excluding the arrival earliness. We hereafter take the part of the free dwell time as the free dwell time for an early train. In this case, the minimum free dwell time equals the scheduled dwell time. However, the minimum free dwell time can be smaller than the scheduled dwell time for late arriving trains, since they may dwell at the station for a shorter time to compensate for the arrival delays.

We fit the free dwell time distribution for early and late arriving trains per train series per direction using the dwell time observations of the trains that respectively satisfy

$$A_i^p < sa_i^p \text{ and } C_i^r < sd_i^p, \tag{4.2}$$

and

$$A_i^p \geq sa_i^p \text{ and } C_i^r < A_i^p + 30, \tag{4.3}$$

where A_i^p and sa_i^p denote the actual and scheduled arrival times of train i at the platform track, sd_i^p stands for the scheduled departure time and C_i^r represents the clearance time of the outbound route of this train. The unit of these times is in seconds. In case of early arriving trains, the chosen ones are not hindered by other trains at the station after the scheduled departure time. In case of late trains, the selected ones are not hindered by other trains after a minimal dwell time of 30 s assumed on the basis of on-site observations (Wiggenraad (2001)). Synchronized transfer connections between different train lines at The Hague HS station are rare and the impact is hence neglected. Some trains which do not satisfy Inequalities (4.2) and (4.3) may not be hindered at the station, too, because of their longer free dwell times. This means that the trains satisfying Inequality (4.2)((4.3)) is a subset of the early (late) arriving trains which are not hindered at the station. In case of a considerably large scheduled headway, each one of a train series cannot be hindered by the preceding train of a different train series. In this case, the free dwell times of trains equal the observed dwell times.

We have evaluated the candidate distributions for the free dwell times of both early and late arriving trains per train series per direction. It has been found that for the free dwell

times of early arriving trains, the Weibull model gives the best distribution fit in 8 out of 15 studied cases and the normal, gamma or log-normal distribution gives the best fit in the other 7 cases. In addition, the Weibull fit ranks second among the candidate distributions in 5 out of the 15 cases. For the free dwell times of late arriving trains, the Weibull distribution is the best model among the candidate distributions in 7 out of 18 studied cases and the normal, gamma or log-normal distribution is the best model in the other 11 cases. In addition, the Weibull fit ranks second among the candidate distributions in 10 out of the 18 cases.

Table 4.4 lists the fine-tuned parameters and the K-S test result of the fitted Weibull distribution as well as the sample size of the original data set for the free dwell times of both early and late arriving trains per train series in each direction. The p-values of the K-S test are larger than 0.05 except for the free dwell times of early arriving trains AR5000S, therefore the Weibull distribution fits have been accepted in most cases.

Table 4.4: Fine-tuned parameters and the K-S test result of the fitted Weibull distribution as well as the sample size of the original data set of the free dwell times for both early and late arriving trains

Train series	Early						Late					
	n	c [s]	k	$1/\lambda$ [s]	p	R_{w}	n	c [s]	k	$1/\lambda$ [s]	p	R_{w}
AR5000S	60	120	0.3	14	0.00	3	118	22	1.3	52	0.47	2
AR5100S	-	-	-	-	-	-	49	20	1.5	43	0.84	2
IR2200S	280	60	1.4	23	0.69	1	299	20	2.4	59	0.54	2
IC1900S	-	-	-	-	-	-	398	20	2.2	41	0.80	1
IC2100S	195	60	1.3	25	0.85	1	119	37	1.8	52	0.67	2
IC2400S	181	60	1.4	21	0.81	1	117	24	2.1	42	0.99	2
IC2500S	-	-	-	-	-	-	190	28	2.5	53	0.91	1
INT600S	114	120	2.2	34	0.81	2	101	52	1.4	35	0.76	3
HST9300S	67	120	2.1	40	0.98	1	29	108	2.3	42	0.92	2
AR5000N	146	120	1.0	31	0.49	2	53	33	1.6	51	0.50	2
AR5100N	103	60	1.7	23	0.94	1	68	23	2.1	41	0.92	2
IR2200N	720	60	1.0	28	0.13	1	132	25	1.8	63	0.98	1
IC1900N	168	60	1.5	30	0.50	2	155	35	2.4	38	0.62	2
IC2100N	307	60	1.2	35	0.65	1	145	36	2.0	53	0.98	1
IC2400N	357	60	1.1	26	0.45	1	100	37	1.8	43	0.80	1
IC2500N	153	60	0.9	23	0.22	2	144	32	2.5	54	0.76	1
INT600N	273	60	1.9	40	0.27	3	133	33	1.6	41	0.86	1
HST9300N	99	120	1.0	23	0.43	2	26	61	2.3	47	0.98	2

Note: p-values have been rounded to two decimals

Figure 4.7 shows the Weibull and normal density fits, kernel density estimate and empirical histogram for the free dwell times of late arriving trains IR2200N. The fitted Weibull distribution with a shape parameter of 1.8 matches with the kernel estimate for the empirical data much better than the normal fit with respect to overall shape and the tails of the distribution. The better fitting of the Weibull model is also illustrated by the distribution differences plots given in Figure 4.8.
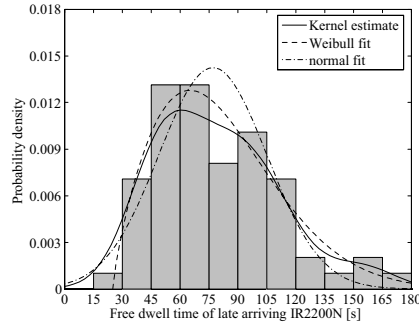
Figure 4.7: Weibull and normal fits, kernel estimate and histogram for the free dwell times of late arriving trains of IR2200N
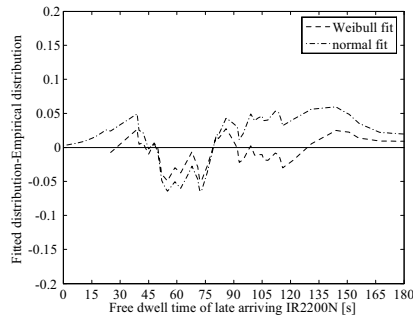
Figure 4.8: Distribution differences plots for the Weibull and normal fits and the free dwell times of late arriving trains of IR2200N

It has been found that the density estimate and empirical histogram of free dwell times are generally positively skewed. The free dwell times of early arriving trains cannot be shorter than the scheduled dwell time. The free dwell times of late arriving trains may be shorter than the scheduled time, but they are longer than the necessary time for opening the doors of a train, announcing the departure, and closing the doors. Taking into account the positive skewness and the lower bound, a location-shifted Weibull distribution better captures the stochasticity of free dwell times than a normal distribution.

In case of an early arrival, the train driver may not feel so hurry and wait for some late passengers boarding the train. In case of a late arrival, the train stops at the platform longer than the minimum dwell time due to an increased number of passengers boarding the train. Obviously, the probability of large free dwell times is very small. Therefore, the probability density curve of free dwell times is generally non-decreasing and unimodal. The shape parameter of the Weibull distribution fit is larger than 1.0 in most cases. If the scheduled dwell time is long enough compared to the necessary times for passenger alighting and boarding, the shape parameter of the Weibull distribution fit to the free dwell times of early arriving trains (excluding the arrival earliness) can be smaller than 1.0.

4.4.4   Running times and track occupancy times

To incorporate knock-on delays in the modelling of delay propagation in a railway network, the distributions of train running times and track occupancy times are required. Herewith, we focus on the statistical distribution of the running times of trains on the preceding block of The Hague HS station and that of the occupancy times of adjacent junctions around this station.

Considering an approaching train, if the inbound route is released earlier than the train arrives at sight distance of the approach signal of the home signal, the train proceeds freely to the platform. Otherwise, the train is hindered and has to decelerate and even to stop before the home signal. To accurately estimate the knock-on delays caused by route conflicts in a station area, it is necessary to know the conditional distributions of train running and track occupancy times in case of (no) hinder (Chapter 5).

Carey & Kwiecinski (1994a) adopted the exponential and uniform distributions to model the stochasticity of the free running time of a train on an open track. For simplicity, they assume the running time of a hindered train on an open track section to be the free running time divided by a deterministic coefficient of 0.6. In real operations, the increase of the running time of a hindered train compared to the corresponding free running time is affected not only by the signalling system and the operational rules, but also by the behavior of the train driver, who may react to a change of signal aspects early or late. It is necessary to model the conditional distributions of train running and track occupancy times using a statistical analysis of the empirical data.

To obtain those conditional distributions, the first step is to classify the data. By comparing the arrival time of each train at the approach signal to the clearance time of the inbound route, we have extracted a data set suited for fitting the free running time distribution in each studied case (the arrival time of each of the trains at sight distance of the approach signal must be later than the inbound route clearance time). A hindered approaching train will pass the home signal with a reduced speed if it may proceed to this signal. Otherwise, the hindered train will pass the home signal with an accelerating speed after stopping before this signal. Since the standstill of a train on a track is not recorded, we cannot directly identify whether or not a hindered train stops before the home signal based on track occupancy and release records.

However, we have realized that the difference between the arrival time of a hindered train at the approach signal and the clearance time of the inbound route is generally longer in case the train stops before the home signal. On the other hand, the difference between the route clearance time and the passing time of the train at the home signal is generally shorter in case the train stops before the home signal than if it may proceed. Having adopted the k-means data clustering (see Appendix A.5) routine included in the statistical analysis tool S-Plus (MathSoft (1999)), we split the data sample of hindered trains for each studied train series operating on routes $a \rightarrow e$, $b \rightarrow e$ and $g \rightarrow c$ (Figure 3.3) into two separate parts (Figure 4.9) which correspond approximately to the aforementioned two cases. For the hindered trains that stop before the home signal, it is still unknown when these trains stop. Therefore, the running times of these trains on the preceding block of the station are lacking. Having applied the k-means data clustering method, the data sets suited for fitting the conditional distributions of the inbound occupancy times of route node 241BT (Figure 3.2) by each relevant passing train series were also extracted.
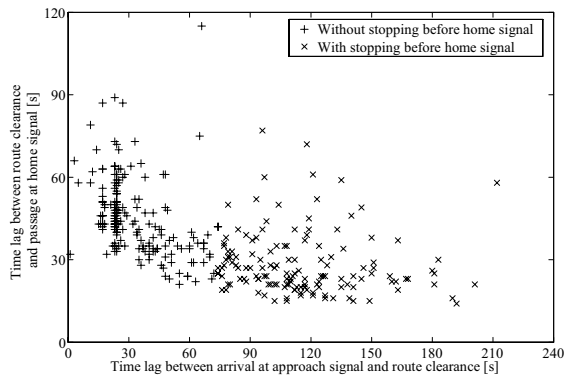
Figure 4.9: Data sample split of hindered trains INT600S

In case of a northbound departing train, if it is hindered due to outbound route conflicts, it dwells at the station for a longer time. However, it will not be hindered again on the next track sections. Thus, the conditional distributions are not applicable to the running times of trains on the outbound track sections and outbound track occupancy times.

We have evaluated the candidate distributions for inbound train running times and track occupancy times in case of different conditions and for outbound track occupancy times. For the free running times of trains on the preceding block of the station, both the Weibull and normal distributions have not been rejected by the K-S test in 9 out of 13 studied cases. In addition, each of these distributions is the best model among the candidate distributions in 5 out of the 13 cases. For inbound occupancy times of route node 241BT by freely passing trains, the Weibull distribution is the best fit in 3 out of 4 studied cases and has not been rejected by the K-S test. The Weibull and normal distributions are the best distribution models for outbound track occupancy times of route nodes 227T and 237BT in 2 and 1 out of 3 studied cases, respectively. Furthermore, both distributions have not been rejected by the K-S test in 2 out of the 3 cases. The goodness-of-fit of the Weibull and normal distributions for the aforementioned process times is shown in Figures 4.10 and 4.11, respectively.

In case of hindered approaching trains which do not stop before the home signal, both the Weibull and normal distributions have not been rejected by the K-S test for the running times on the preceding block of the station in 6 studied cases. In addition, each of these distributions is the best model among the candidate distributions in 3 out of the 6 cases, respectively. For inbound occupancy times of route node 241BT by hindered trains, both the Weibull and normal distributions have not been rejected by the K-S test in 5 studied cases and these distributions are the best distribution model in 1 and 2 out of the 5 cases, respectively. In the other 2 cases, either the gamma or beta distribution is the best distribution model. In case of hindered trains which stop before the home signal, the normal distribution fits best to the inbound track occupancy times and has not been rejected by the K-S test in 2 studied cases. The goodness-of-fit of the Weibull and normal distributions for the aforementioned train process times is given in Figures 4.12 and 4.13, respectively.
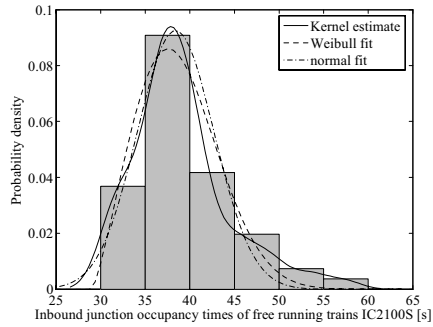
Figure 4.10: Weibull and normal density fits, kernel density estimate and empirical histogram for inbound occupancy times of 241BT by freely running trains IC2100S
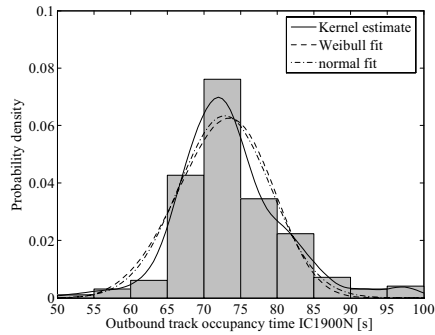
Figure 4.11: Weibull and normal density fits, kernel density estimate and empirical histogram for outbound occupancy times of 227T by IC1900N
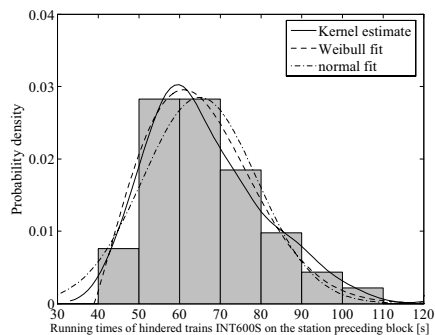
Figure 4.12: Weibull and normal density fits, kernel density estimate and empirical histogram for inbound running times of hindered INT600S that do not stop before the home signal
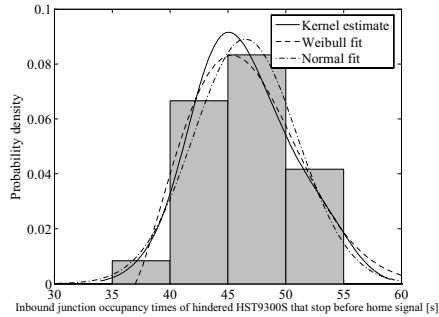
Figure 4.13: Weibull and normal density fits, kernel density estimate and empirical histogram for inbound occupancy times of 241BT by hindered HST9300S that stop before the home signal

A good generic distribution has not been found for the conditional train running and track occupancy times in case of (no) hinder. This might be caused by the big variation of train speeds on the short track sections in the complicated station and interlocking area. The data classification and the further data split, which reduce the size of the data sample, also affect the determination of a generic distribution for the conditional train running and track occupancy times. However, it has been found that the conditional distribution fits for inbound train running or track occupancy times vary considerably. The average running time of hindered trains in the preceding block of the station is higher by 17% to 135% than the average running time of freely passing trains for the 6 studied cases. The average inbound occupancy time of route node 241BT by the trains which are hindered but proceed to the home signal with a reduced speed is higher by 26% and 29% than that by freely passing trains for the two studied cases, respectively. The increased percentages are 50% and 41%, correspondingly, if the trains which are hindered and forced to stop before the home signal are considered. These results have been obtained for the station and interlocking area, where the maximum speed of approaching trains is decreased to 80 km/ h. In case of a junction on an open track, the differences between the conditional distribution fits for train running or track occupancy times would be even more significant.

4.5   Conclusions

In this chapter, we have presented a new method for fine-tuning the parameters of distribution models. Based on this parameter fine-tuning approach, we have compared several commonly applied distribution models for train event and process times by the K-S goodness-of-fit test. It has been found that a location-shifted log-normal distribution can be considered as the best model among the candidate distributions for both the arrival times of trains at the platform track and at the approach signal of the station home signal. In both situations, the earliest arrival time corresponds to the shift parameter of the distribution model. The Weibull distribution can generally be considered as the best distribution model for non-negative arrival delays, departure delays and the free dwell times of trains. The shape parameter of a Weibull distribution fit to non-negative arrival delays is mostly less than 1.0. The shape parameter of a Weibull distribution fit to the free dwell times

of trains is generally larger than 1.0. The presented parameter fine-tuning method is an engineering approach for locally optimize the goodness-of-fit of the K-S test. No probability distribution can fit empirical data exactly. The evaluation results of the distribution models for train event and process times will be used as the input of the delay propagation model to be presented. By estimating knock-on delay distributions and the impact on the punctuality of trains, we will improve station capacity utilization while assuring a desired punctuality level of train operations.

Chapter 5

Probability model of train delay propagation

5.1 Introduction

This chapter presents a new analytical probability model for estimating the propagation of train delays in a station. The model reflects the constraints of the signalling system and the train protection operations rules. Railway signalling systems can be categorized as a one-block, multiple-block or moving block signalling system (Bailey (1995), Pachl (2002)). In one-block signalling, a block signal can only give information about the block section behind the signal but no approach information for the next signal. Therefore, every block signal must have a special distant (approach) signal whose only purpose is to give the required approach information. In multiple-block signalling, a block signal gives information about two or more following block sections. A common example of multiple-block signalling is two-block signalling in which the approach information is integrated in the aspect of the preceding block signal. Because in two-block signalling in its simplest form, a block signal shows three different aspects, i.e., clear (green), approach (yellow) and stop (red), it is also called 'three-aspect signalling'. A moving block system would yield the best possible headway in the sense of capacity use, but it would be only practical with cab signalling and automatic train control. The proposed model is designed for the three-aspect two-block signalling system, which is widely used on the main lines of many railways.

This chapter is structured as follows. Section 5.2 first describes the modelling methodology adopted in the proposed delay propagation model. The composite trapezoidal rule is then applied in Section 5.3 to obtain a numerical approximation of the model. Furthermore, the model is validated in Section 5.4 with the empirical train detection data recorded at The Hague HS railway station. Finally, the main conclusions are drawn in Section 5.5.

5.2   Modelling methodology

The proposed model focuses on the propagation of train delays in a station. In case of route conflicts, hindered approaching trains decelerate first and even may stop in front of the home signal of the station, and then accelerate once the stop signal clears, while departing trains just prolong the dwell process until clearance of the departure signal. Transfer connections are another source of knock-on delays of departing trains at large stations, where scheduled train services on different lines are well coordinated. Therefore, we describe the modelling of delay propagation to an approaching and departing train in Sections 5.2.1 and 5.2.2, respectively. In addition, a train suffering knock-on delay may cause knock-on delays to other trains. To decrease knock-on delay of a train during real-time operations, the prespecified order of this train and the scheduled preceding train at the conflict route node may be altered in case of a large delay of the preceding one. A scheduled transfer connection may also be canceled if the feeder train comes too late. The dynamic delay propagation in a complicated station area and the impact of rescheduling on the propagation of train delays are discussed in Section 5.2.3.

5.2.1   Delay propagation to an approaching train

To focus on the modelling of delay propagation in a complicated station and interlocking area, we assume that an approaching train never stops on the open track signal block before the approach signal of the home signal. This assumption can be partially justified by small probability that a chain of trains are hindered on the same route before arriving at a station. It implies that when a train arrives at sight distance of the approach signal of the home signal of a station, the aspect of this signal is either green (G) or yellow (Y).

If the approach signal shows a green aspect, the train proceeds freely to the platform track and does not suffer a knock-on delay. A yellow aspect of the approach signal indicates occupancy of the platform track (Figure 5.1) or of a route node (Figure 5.2) included in the station block by a conflicting train and forces the approaching train to decelerate and even to stop if the following block signal (the home signal) is not released in time.

When the hindered approaching train proceeds to the home signal and this signal shows a green or yellow aspect, the train proceeds to the platform without an extra stop. If this signal remains red (R), the train has to stop first in front of this signal and accelerates again after a reaction time following signal clearance and finally proceeds to the platform. It should be mentioned that when a train stops before the home signal of a station, there is generally a small distance offset.
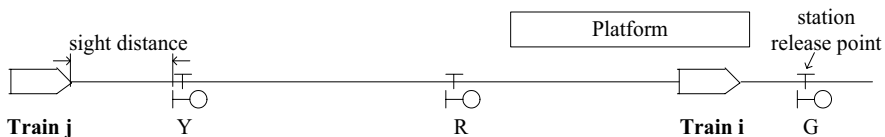
Figure 5.1: Hindrance of an approaching train j due to occupancy of the platform track by a departing train i
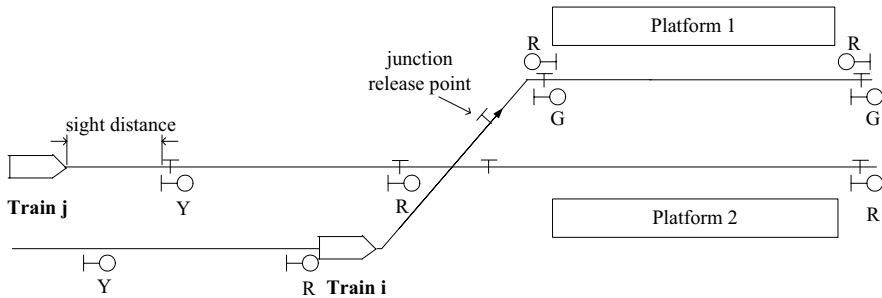
Figure 5.2: Hindrance of an approaching train j due to occupancy of a route node included in the station block by an approaching train i

In this section, it is also assumed that the prespecified order of trains must be adhered to. We model the delay propagation to an approaching train j depending on one of the following typical events:

$E_{j,1}$: *When train j arrives at sight distance of the approach signal of a station, this signal shows a green aspect.*

$E_{j,2}$: *When train j arrives at sight distance of the approach signal of the station, this signal shows a yellow aspect (the aspect of the home signal is red), however, when the train proceeds to the home signal, the latter signal changes to a green or yellow aspect.*

$E_{j,3}$: *When train j arrives at sight distance of the approach signal of the station, this signal shows a yellow aspect and when the train proceeds to the home signal, the latter signal still has a red aspect.*

These random events are formulated respectively by

$$
\begin{aligned}
E_{j,1}: \quad & C_j^{\mathrm{r}} \leq A_j^{\mathrm{as}} - t^{\mathrm{as}}, \\
E_{j,2}: \quad & A_j^{\mathrm{as}} - t^{\mathrm{as}} < C_j^{\mathrm{r}} \leq A_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}}, \\
E_{j,3}: \quad & C_j^{\mathrm{r}} > A_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}},
\end{aligned}
\tag{5.1}
$$

where A_j^{as} denotes the arrival time of train j at the approach signal of the station, C_j^{r} the clearance time of the route of train j, $T_j^{\mathrm{as\text{-}hs}(2)}$ the running time of train j from the approach signal to the home signal in case of event $E_{j,2}$, and t^{as} and t^{hs} are sight and reaction time of an approaching train before an approach signal of a station and the time needed for a hindered approaching train to react to the clearance of the home signal and to pass by it in case of an extra stop before the home signal of a station, respectively. Note that the possibility of a late setting of a train route is neglected.

In the proposed model, those main event and process times related to stochastic train movements and track occupancy and release are modelled by a random variable and represented by a capital letter. The sight and reaction time of a train, the time needed for an approaching train which is hindered and even has an extra stop before the home signal to pass by this signal after clearance, and the dispatching time of a train at platform are considered to be deterministic parameters. The realizations of the random variables and

the deterministic parameters are denoted by a small letter. The standard time for sight of and reaction to fixed signals is 12 s in case of a train running on open tracks at a design speed of 140-160 km/ h (Hansen et al. (1999)). This standard time consists of a reaction time of 3-5 s and the running time of about 8 s on a track section of about 300 m long before the signals. For an approaching train which is hindered and even forced to stop before the home signal, it has been found from the TNV data recorded in The Hague HS station that the time needed for the train to react to the clearance of the home signal and pass by it is around 25 s.

The clearance time of the route of train j is bounded by the release of the platform track and the route nodes on this route by other trains. If we consider a conflicting train i that hinders train j by occupying the platform track or a track junction on the route, the release time of the critical track section by train i can be estimated by

$$C_{i,j}^{\mathrm{r}} = \left\{ \begin{array}{ll} A_i^{\mathrm{as}} + T_i^{\mathrm{as\text{-}rp}} & \text{for train } i \text{ approaching the station} \\ D_i^{\mathrm{p}} + T_i^{\mathrm{p\text{-}rp}} & \text{for train } i \text{ departing from the station.} \end{array} \right. \tag{5.2}$$

In the above, A_i^{as} and $T_i^{\mathrm{as\text{-}rp}}$ represent the arrival time of approaching train i at the approach signal of the critical track section and the occupancy time of the track section between the approach signal and the release position by train i, respectively (Figure 5.2); D_i^{p} (the estimation will be discussed in Section 5.2.2) and $T_i^{\mathrm{p\text{-}rp}}$ denote the actual departure time of departing train i at the platform track and the occupancy time of the track section between the stop position on the platform track and the release point of the critical track section by train i, respectively (Figures 5.1 and 5.3). The track occupancy time here consists of the time for the train running on the corresponding track section, clearing time of the full length of the train at the release point of the track section and the time needed to set up a new route for the following train j.
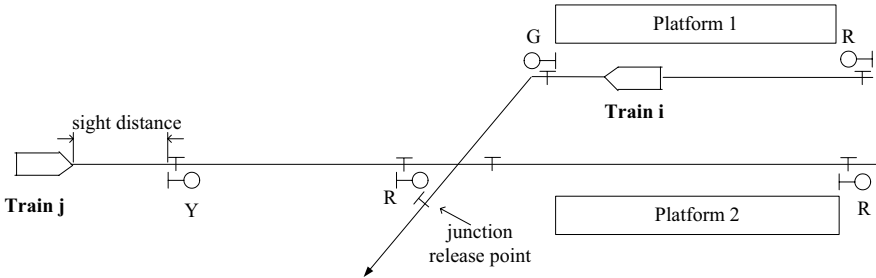
Figure 5.3: Hindrance of an approaching train j due to occupancy of a track junction on the route by a departing train i

In many cases, the knock-on delay of a train comes from multiple sources. The clearance time of the route of this train is the latest of the release times of the critical track sections on the route by a number of conflicting trains. Let $\widehat{\mathcal{T}}_j$ denote the set of all the conflicting trains of train j. Then the clearance time of the route of train j is expressed by

$$C_j^{\mathrm{r}} = \max_{i \in \widehat{\mathcal{T}}_j}(C_{i,j}^{\mathrm{r}}). \tag{5.3}$$

We represent the arrival time of train j at the stop location on the platform track by

$$A_j^{\text{p}} = \begin{cases} A_j^{\text{as}} + T_j^{\text{as-p}(1)} & \text{for } E_{j,1}, \\ A_j^{\text{as}} + T_j^{\text{as-p}(2)} & \text{for } E_{j,2}, \\ C_j^{\text{r}} + t^{\text{hs}} + T_j^{\text{hs-p}(3)} & \text{for } E_{j,3}, \end{cases} \tag{5.4}$$

where $T_j^{\text{as-p}(1)}$ and $T_j^{\text{as-p}(2)}$ denote the running time of train j from the approach signal to the stop location on the platform track in case of event $E_{j,1}$ and $E_{j,2}$, respectively, and $T_j^{\text{hs-p}(3)}$ the running time of train j from the home signal to the stop location on the platform track in case of event $E_{j,3}$.

Let $\Delta_j^{\text{as}} = C_j^{\text{r}} - A_j^{\text{as}}$. The total knock-on delay of approaching train j suffered before arriving at the platform track can be expressed by

$$L_j^{\text{aknock}} = \begin{cases} 0 & \text{for } E_{j,1}, \\ T_j^{\text{as-p}(2)} - T_j^{\text{as-p}(1)} & \text{for } E_{j,2}, \\ \Delta_j^{\text{as}} + T_j^{\text{hs-p}(3)} - T_j^{\text{as-p}(1)} + t^{\text{hs}} & \text{for } E_{j,3}. \end{cases} \tag{5.5}$$

Modelling the propagation of stochastic delays to an approaching train, it is important to predict the knock-on delay distribution and the distribution of the arrival time of this train at the stop location on the platform track. The former distribution reflects the capacity utilization of the critical track sections, i.e., platform track and route nodes on the route of this train. The latter distribution is used to calculate the arrival punctuality of the train. We estimate the knock-on delay distribution by the probability that each of the distinguished events occurs and the conditional probability of the knock-on delay in case of each of the events.

$$\begin{aligned} F_{L_j^{\text{aknock}}}(x) &= \sum_{k=1}^{3} \Pr[L_j^{\text{aknock}} \leq x | E_{j,k}] \cdot \Pr(E_{j,k}) \\ &= \Pr(E_{j,1}) + \sum_{k=2}^{3} \Pr[L_j^{\text{aknock}} \leq x | E_{j,k}] \cdot \Pr(E_{j,k}), \text{ for } x \geq 0. \end{aligned} \tag{5.6}$$

To obtain the probability that each of the distinguished events occurs, it is needed to know the distribution of an algebraic sum of the arrival time of train j at the approach signal of the station, the running time of train j from the approach signal to the home signal in case of event $E_{j,2}$ and the clearance time of the route of train j by conflicting trains, see Inequalities (5.1). If train j and a conflicting train i are running on different routes, the arrival time of train j at the approach signal is obviously independent of the release time of the critical track section by train i. This independence is still applicable when both trains i and j are running on the same route successively (Figure 5.1), provided that train j never meets a red aspect of the approach signal of the station. The running time of train j in case of event $E_{j,2}$ is determined by the signalling control system, length of the track section, dynamic characteristics of this train, and random behavior of the train driver. The distribution of an algebraic sum of these individual independent distributions is given by the so-called convolution (Law & Kelton (2000), Dekking et al. (2005)).

Let $F_{X_1}(x)$ and $F_{X_2}(x)$ be two independent cumulative probability distributions, and $F_{X_1} * F_{X_2}(x)$ denote the convolution, i.e., the distribution of the sum of the corresponding random variables X_1 and X_2. We have

$$F_{X_1+X_2}(x) = F_{X_1} * F_{X_2}(x) = \iint\limits_{u+v \leq x} dF_{X_2}(v)dF_{X_1}(u) = \int_{-\infty}^{+\infty} F_{X_2}(x-u)dF_{X_1}(u). \quad (5.7)$$

The convolution of n independent distributions $F_{X_1}(x)$, $F_{X_2}(x)$, ..., $F_{X_n}(x)$ can be estimated recursively by

$$F_{X_1+X_2+\cdots+X_n}(x) = F_{X_1} * F_{X_2} * \cdots * F_{X_n}(x) = (F_{X_1} * F_{X_2} * \cdots * F_{X_{n-1}}) * F_{X_n}(x). \quad (5.8)$$

The probability that each of the distinguished events occurs is then derived by

$$\begin{aligned}
\Pr(E_{j,1}) =\ & \Pr(C_j^r \leq A_j^{as} - t^{as}) = F_{\Delta_j^{as}}(-t^{as}), & (5.9)\\
\Pr(E_{j,2}) =\ & \Pr(A_j^{as} - t^{as} < C_j^r \leq A_j^{as} + T_j^{as\text{-}hs(2)} - t^{hs}) \\
=\ & \Pr(C_j^r \leq A_j^{as} + T_j^{as\text{-}hs(2)} - t^{hs}) - \Pr(C_j^r \leq A_j^{as} - t^{as}) & (5.10)\\
=\ & F_{\Delta_j^{as}} * F_{-T_j^{as\text{-}hs(2)}}(-t^{hs}) - F_{\Delta_j^{as}}(-t^{as}), \\
\Pr(E_{j,3}) =\ & \Pr(C_j^r > A_j^{as} + T_j^{as\text{-}hs(2)} - t^{hs}) \\
=\ & 1 - \Pr(C_j^r \leq A_j^{as} + T_j^{as\text{-}hs(2)} - t^{hs}) & (5.11)\\
=\ & 1 - F_{\Delta_j^{as}} * F_{-T_j^{as\text{-}hs(2)}}(-t^{hs}),
\end{aligned}$$

where

$$F_{\Delta_j^{as}}(x) = F_{C_j^r} * F_{-A_j^{as}}(x). \quad (5.12)$$

It should be mentioned that in this chapter we often apply the following equation:

$$\begin{aligned}
F_{-X}(x) =\ & \Pr(-X \leq x) = \Pr(X \geq -x) \\
=\ & 1 - F_X(-x), \text{ for any continuous random variable } X.
\end{aligned} \quad (5.13)$$

The conditional distributions of the knock-on delay of train j suffered at the preceding block of the station in the case of events $E_{j,2}$ and $E_{j,3}$ are also convolutions of several individual distributions. Eq. (5.6) is then expanded as

$$\begin{aligned}
F_{L_j^{\text{aknock}}}(x) =\ & \Pr(E_{j,1}) + F_{T_j^{as\text{-}p(2)}} * F_{-T_j^{as\text{-}p(1)}}(x) \cdot \Pr(E_{j,2}) \\
& + F_{\Delta_j^{as(3)}} * F_{T_j^{hs\text{-}p(3)}} * F_{-T_j^{as\text{-}p(1)}}(x - t^{hs}) \cdot \Pr(E_{j,3}), \text{ for } x \geq 0
\end{aligned} \quad (5.14)$$

where $F_{T_j^{\text{as-p(1)}}}$, $F_{T_j^{\text{as-p(2)}}}$ and $F_{T_j^{\text{hs-p(3)}}}$ denote the conditional running time distributions of train j on relevant track sections in the case of those distinguished events, respectively, which must be given. $F_{\Delta_j^{\text{as(3)}}}$ represents the conditional distribution of the time lag Δ_j^{as} in case of event $E_{j,3}$. If the running time of train j on the preceding block of the station is deterministic, the conditional distribution $F_{\Delta_j^{\text{as(3)}}}$ would be a left-truncated distribution of Δ_j^{as}. This is because the time lag Δ_j^{as} in the case of event $E_{j,3}$ must be longer than the running time of train j from the approach signal of the station to the extra stop location before the home signal. Taking into account the stochastic variation of $T_j^{\text{as-hs(2)}}$, $F_{\Delta_j^{\text{as(3)}}}$ can be derived by

$$
\begin{aligned}
F_{\Delta_j^{\text{as(3)}}}(x) &= \frac{\Pr[(\Delta_j^{\text{as}} \le x) \cap E_{j,3}]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr[(\Delta_j^{\text{as}} \le x) \cap (C_j^{\text{r}} > A_j^{\text{as}} + T_j^{\text{as-hs(2)}} - t^{\text{hs}})]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr(\Delta_j^{\text{as}} \le x) - \Pr[(\Delta_j^{\text{as}} \le x) \cap (C_j^{\text{r}} \le A_j^{\text{as}} + T_j^{\text{as-hs(2)}} - t^{\text{hs}})]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr(\Delta_j^{\text{as}} \le x) - \Pr[(\Delta_j^{\text{as}} \le x) \cap (\Delta_j^{\text{as}} - T_j^{\text{as-hs(2)}} \le -t^{\text{hs}})]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{F_{\Delta_j^{\text{as}}}(x) - \int_{-\infty}^{x} F_{-T_j^{\text{as-hs(2)}}}(-t^{\text{hs}} - u)\, dF_{\Delta_j^{\text{as}}}(u)}{1 - F_{\Delta_j^{\text{as}}} * F_{-T_j^{\text{as-hs(2)}}}(-t^{\text{hs}})}.
\end{aligned}
\tag{5.15}
$$

To estimate the knock-on delay distribution for an approaching train j, the distributions of A_j^{as} and C_j^{r} are needed. The former distribution $F_{A_j^{\text{as}}}(x)$ can be obtained on the basis of empirical data or estimated in real operations. We derive the distribution of C_j^{r} as follows:

$$
\begin{aligned}
F_{C_j^{\text{r}}}(x) &= \Pr(C_j^{\text{r}} \le x) = \Pr[\max_{i \in \widehat{T}_j}(C_{i,j}^{\text{r}}) \le x] \\[2mm]
&= \prod_{i \in \widehat{T}_j} F_{C_{i,j}^{\text{r}}}(x) = \prod_{i \in \widehat{T}_j^{\text{a}}} F_{C_{i,j}^{\text{r}}}(x) \cdot \prod_{i \in \widehat{T}_j^{\text{d}}} F_{C_{i,j}^{\text{r}}}(x) \\[2mm]
&= \prod_{i \in \widehat{T}_j^{\text{a}}} F_{A_i^{\text{as}}} * F_{T_i^{\text{as-rp}}}(x) \cdot \prod_{i \in \widehat{T}_j^{\text{d}}} F_{D_i^{\text{p}}} * F_{T_i^{\text{p-rp}}}(x),
\end{aligned}
\tag{5.16}
$$

where $\widehat{T}_j^{\text{a}}$ denotes the set of the conflicting trains of train j that are approaching the station and $\widehat{T}_j^{\text{d}}$ represents the set of the conflicting trains that are departing from the station. The estimation of $F_{D_i^{\text{p}}}(x)$ will be discussed later in Section 5.2.2.

In many cases, we are interested in investigating the knock-on delay of train j caused by an individual train i which occupies the platform track or a specific route node on the route of train j. The knock-on delay can be estimated by replacing C_j^{r} with $C_{i,j}^{\text{r}}$ in Inequalities (5.1) and Δ_j^{as} with $\Delta_{i,j}^{\text{as}} = C_{i,j}^{\text{r}} - A_j^{\text{as}}$ in Eq. (5.5). Similarly, we can obtain the distribution, which reflects the capacity utilization of the specific track section.

In the above, we have derived the distribution of the knock-on delay of an approaching train j suffered before arriving at the platform. In a similar way, Yuan & Hansen (2006) represent the arrival time of the train at the platform as

$$
\begin{aligned}
F_{A_j^{\mathrm{p}}}(x) = & \sum_{k=1}^{3} \Pr[A_j^{\mathrm{p}} \leq x | E_{j,k}] \cdot \Pr(E_{j,k}) \\
= & F_{A_j^{\mathrm{as}(1)}} * F_{T_j^{\mathrm{as\text{-}p}(1)}}(x) \cdot \Pr(E_{j,1}) + F_{A_j^{\mathrm{as}(2)}} * F_{T_j^{\mathrm{as\text{-}p}(2)}}(x) \cdot \Pr(E_{j,2}) \\
& + F_{C_j^{\mathrm{r}(3)}} * F_{T_j^{\mathrm{hs\text{-}p}(3)}}(x - t^{\mathrm{hs}}) \cdot \Pr(E_{j,3}),
\end{aligned}
\tag{5.17}
$$

where $F_{A_j^{\mathrm{as}(1)}}$ and $F_{A_j^{\mathrm{as}(2)}}$ represent the conditional distributions of the arrival time of train j at the approach signal given events $E_{j,1}$ and $E_{j,2}$, respectively; $F_{C_j^{\mathrm{r}(3)}}$ stands for the conditional distribution of the route clearance time given event $E_{j,3}$.

If all trains were operated randomly without a scheduled timetable, then the conditional distributions $F_{A_j^{\mathrm{as}(1)}}$ and $F_{A_j^{\mathrm{as}(2)}}$ would equal the distribution of A_j^{as}, and the conditional distribution $F_{C_j^{\mathrm{r}(3)}}$ would be the same as the distribution of C_j^{r}. For scheduled train operations, let us consider a pair of trains i and j passing a track junction included in a station block successively, see Figure 5.3. If the approaching train j arrives at sight distance of the approach signal of the home signal earlier than the scheduled arrival time and the earliness is beyond the buffer time between the scheduled paths of this pair of trains at the route node, it will be hindered due to the occupancy of the conflict route node by the scheduled preceding train i which departs from the station even on time. Given the distinguished event $E_{j,1}$ occurring, it is clear that train j could not arrive too early at sight distance of the approach signal. However, train j may arrive early or late in case of event $E_{j,2}$. This implies that the conditional distributions $F_{A_j^{\mathrm{as}(1)}}$ and $F_{A_j^{\mathrm{as}(2)}}$ are different. In addition, we have no evidence that the conditional distribution $F_{C_j^{\mathrm{r}(3)}}$ is the same as the distribution of C_j^{r} for scheduled train operations. Therefore, we must not simplify Eq. (5.17) by replacing $F_{A_j^{\mathrm{as}(1)}}$ and $F_{A_j^{\mathrm{as}(2)}}$ with $F_{A_j^{\mathrm{as}}}$ and $F_{C_j^{\mathrm{r}(3)}}$ with $F_{C_j^{\mathrm{r}}}$. These conditional distributions can be derived respectively by

$$
\begin{aligned}
F_{A_j^{\mathrm{as}(1)}}(x) & = \frac{\Pr[(A_j^{\mathrm{as}} \leq x) \cap E_{j,1}]}{\Pr(E_{j,1})} \\
& = \frac{\Pr[(A_j^{\mathrm{as}} \leq x) \cap (C_j^{\mathrm{r}} \leq A_j^{\mathrm{as}} - t^{\mathrm{as}})]}{\Pr(E_{j,1})} \\
& = \frac{\Pr(A_j^{\mathrm{as}} \leq x) - \Pr[(A_j^{\mathrm{as}} \leq x) \cap (C_j^{\mathrm{r}} > A_j^{\mathrm{as}} - t^{\mathrm{as}})]}{\Pr(E_{j,1})} \\
& = \frac{\Pr(A_j^{\mathrm{as}} \leq x) - \Pr[(A_j^{\mathrm{as}} \leq x) \cap (A_j^{\mathrm{as}} - C_j^{\mathrm{r}} < t^{\mathrm{as}})]}{\Pr(E_{j,1})} \\
& = \frac{F_{A_j^{\mathrm{as}}}(x) - \int_{-\infty}^{x} F_{-C_j^{\mathrm{r}}}(t^{\mathrm{as}} - u) dF_{A_j^{\mathrm{as}}}(u)}{F_{\Delta_j^{\mathrm{as}}}(-t^{\mathrm{as}})},
\end{aligned}
\tag{5.18}
$$

$$\begin{aligned}
F_{A_j^{\text{as}(2)}}(x) &= \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap E_{j,2}\right]}{\Pr(E_{j,2})} \\[2mm]
&= \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap (A_j^{\text{as}} - t^{\text{as}} < C_j^{\text{r}} \le A_j^{\text{as}} + T_j^{\text{as-hs}(2)} - t^{\text{hs}})\right]}{\Pr(E_{j,2})} \\[2mm]
&= \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap (A_j^{\text{as}} - t^{\text{as}} < C_j^{\text{r}})\right]}{\Pr(E_{j,2})} \\[2mm]
&\quad - \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap (C_j^{\text{r}} > A_j^{\text{as}} + T_j^{\text{as-hs}(2)} - t^{\text{hs}})\right]}{\Pr(E_{j,2})} \\[2mm]
&= \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap (A_j^{\text{as}} - C_j^{\text{r}} < t^{\text{as}})\right]}{\Pr(E_{j,2})} \\[2mm]
&\quad - \frac{\Pr\left[(A_j^{\text{as}} \le x) \cap (A_j^{\text{as}} + T_j^{\text{as-hs}(2)} - C_j^{\text{r}} < t^{\text{hs}})\right]}{\Pr(E_{j,2})} \\[2mm]
&= \frac{\int_{-\infty}^{x} \left[F_{-C_j^{\text{r}}}(t^{\text{as}} - u) - F_{T_j^{\text{as-hs}(2)}} * F_{-C_j^{\text{r}}}(t^{\text{hs}} - u)\right] dF_{A_j^{\text{as}}}(u)}{F_{\Delta_j^{\text{as}}} * F_{-T_j^{\text{as-hs}(2)}}(-t^{\text{hs}}) - F_{\Delta_j^{\text{as}}}(-t^{\text{as}})}, \qquad (5.19)
\end{aligned}$$

$$\begin{aligned}
F_{C_j^{\text{r}(3)}}(x) &= \frac{\Pr\left[(C_j^{\text{r}} \le x) \cap E_{j,3}\right]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr\left[(C_j^{\text{r}} \le x) \cap (C_j^{\text{r}} > A_j^{\text{as}} + T_j^{\text{as-hs}(2)} - t^{\text{hs}})\right]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr(C_j^{\text{r}} \le x) - \Pr\left[(C_j^{\text{r}} \le x) \cap (C_j^{\text{r}} \le A_j^{\text{as}} + T_j^{\text{as-hs}(2)} - t^{\text{hs}})\right]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{\Pr(C_j^{\text{r}} \le x) - \Pr\left[(C_j^{\text{r}} \le x) \cap (C_j^{\text{r}} - A_j^{\text{as}} - T_j^{\text{as-hs}(2)} \le -t^{\text{hs}})\right]}{\Pr(E_{j,3})} \\[2mm]
&= \frac{F_{C_j^{\text{r}}}(x) - \int_{-\infty}^{x} F_{-A_j^{\text{as}}} * F_{-T_j^{\text{as-hs}(2)}}(-t^{\text{hs}} - u) dF_{C_j^{\text{r}}}(u)}{1 - F_{\Delta_j^{\text{as}}} * F_{-T_j^{\text{as-hs}(2)}}(-t^{\text{hs}})}. \qquad (5.20)
\end{aligned}$$

We have realized that the conditional distribution of C_j^{r} in case of event $E_{j,3}$ might be subject to the stochastic variation of A_j^{as} particularly under the assumption that the pre-specified order of trains must be adhered to. For instance, if train j arrives at the home signal of a station later than the scheduled arrival time while this signal shows a red aspect, then it is clear that the inbound route of train j is released much later than scheduled by the conflicting trains (a late route setting is neglected). If train j arrives at the home signal of the station earlier than the scheduled arrival time and this signal shows a red aspect, the inbound route of train j may be released on time by the conflicting trains. Let a_j^{as} be a realization of A_j^{as}. By considering the possible dependence of the conditional distribution $F_{C_j^{\text{r}(3)}}$ on A_j^{as}, we can expand Eq. (5.17) into:

$$
\begin{aligned}
F_{A_j^{\mathrm{p}}}(x) = \ & F_{A_j^{\mathrm{as}(1)}} * F_{T_j^{\mathrm{as\text{-}p}(1)}}(x) \cdot \Pr(E_{j,1}) + F_{A_j^{\mathrm{as}(2)}} * F_{T_j^{\mathrm{as\text{-}p}(2)}}(x) \cdot \Pr(E_{j,2}) \\
& + \int F_{C_j^{\mathrm{r}(3)}(a_j^{\mathrm{as}})} * F_{T_j^{\mathrm{hs\text{-}p}(3)}}(x - t^{\mathrm{hs}}) \cdot \Pr[E_{j,3}(a_j^{\mathrm{as}})] dF_{A_j^{\mathrm{as}}}(a_j^{\mathrm{as}}),
\end{aligned}
\tag{5.21}
$$

where $E_{j,3}(a_j^{\mathrm{as}})$ represents the distinguished event $E_{j,3}$ in case of the realization a_j^{as} of A_j^{as}; $F_{C_j^{\mathrm{r}(3)}(a_j^{\mathrm{as}})}$ denotes the conditional distribution of the route clearance time C_j^{r} given event $E_{j,3}(a_j^{\mathrm{as}})$. We have

$$
\begin{aligned}
\Pr[E_{j,3}(a_j^{\mathrm{as}})] &= \Pr(C_j^{\mathrm{r}} > a_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}}) \\
&= 1 - \Pr(C_j^{\mathrm{r}} \le a_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}}) \\
&= 1 - F_{C_j^{\mathrm{r}}} * F_{-T_j^{\mathrm{as\text{-}hs}(2)}}(a_j^{\mathrm{as}} - t^{\mathrm{hs}}),
\end{aligned}
\tag{5.22}
$$

and

$$
\begin{aligned}
F_{C_j^{\mathrm{r}(3)}(a_j^{\mathrm{as}})}(x) &= \frac{\Pr[(C_j^{\mathrm{r}} \le x) \cap E_{j,3}(a_j^{\mathrm{as}})]}{\Pr[E_{j,3}(a_j^{\mathrm{as}})]} \\
&= \frac{\Pr[(C_j^{\mathrm{r}} \le x) \cap (C_j^{\mathrm{r}} > a_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}})]}{\Pr[E_{j,3}(a_j^{\mathrm{as}})]} \\
&= \frac{\Pr(C_j^{\mathrm{r}} \le x) - \Pr[(C_j^{\mathrm{r}} \le x) \cap (C_j^{\mathrm{r}} \le a_j^{\mathrm{as}} + T_j^{\mathrm{as\text{-}hs}(2)} - t^{\mathrm{hs}})]}{\Pr[E_{j,3}(a_j^{\mathrm{as}})]} \\
&= \frac{\Pr(C_j^{\mathrm{r}} \le x) - \Pr[(C_j^{\mathrm{r}} \le x) \cap (C_j^{\mathrm{r}} - T_j^{\mathrm{as\text{-}hs}(2)} \le a_j^{\mathrm{as}} - t^{\mathrm{hs}})]}{\Pr[E_{j,3}(a_j^{\mathrm{as}})]} \\
&= \frac{F_{C_j^{\mathrm{r}}}(x) - \int_{-\infty}^{x} F_{-T_j^{\mathrm{as\text{-}hs}(2)}}(a_j^{\mathrm{as}} - t^{\mathrm{hs}} - u) dF_{C_j^{\mathrm{r}}}(u)}{1 - F_{C_j^{\mathrm{r}}} * F_{-T_j^{\mathrm{as\text{-}hs}(2)}}(a_j^{\mathrm{as}} - t^{\mathrm{hs}})}.
\end{aligned}
\tag{5.23}
$$

In the above, we have described the conditional modelling of the delay propagation to an approaching train and of the impact on the punctuality of the arrival time at the platform. To estimate the inbound knock-on delay distribution and the distribution of the arrival time at the platform, the distribution of the arrival time at the approach signal of the station block, the conditional distributions of the running times on the relevant track sections in case of the distinguished events, and the distribution of the inbound route clearance time are required.

It should be mentioned that if the scheduled headway time between a pair of trains i and j passing a conflict route node is long enough compared to the input delays, the probability of the distinguished events $E_{j,2}$ and $E_{j,3}$ occurring is very small for the following train j. In this case, the conditional modelling of delay propagation and of the impact on the

arrival punctuality at the platform can be simplified by neglecting the second and third items in Eq.s (5.14) and (5.17).

Based on the derived distribution of the arrival time of a train at the platform, we can easily calculate the statistics such as the mean, standard deviation and punctuality levels at different threshold values defined. These estimated statistics can be used to determine the capacity utilization of inbound track junctions and platform tracks at a desired punctuality of train arrivals.

5.2.2   Delay propagation to a departing train

The actual departure of a train at a station is determined by a number of factors. The most important factor is the free dwell time introduced in the previous chapter. The free dwell time of a train is the necessary dwell time in the absence of hindrance due to route conflicts and late transfer connections. In case of a short scheduled headway, a departing train may be hindered due to the occupancy of on-route junctions by preceding trains that are approaching or have departed from the station (Figures 5.4 and 5.5). After the route nodes on the outbound route have been released and the departure signal clears, the train conductor may close the doors and the train departs from the station.
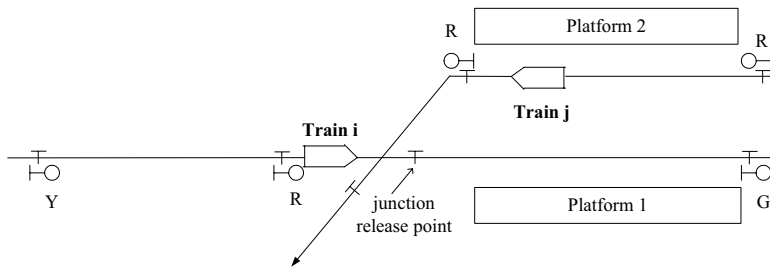
Figure 5.4: Hindrance of a departing train j due to occupancy of a route node on the route by an approaching train i
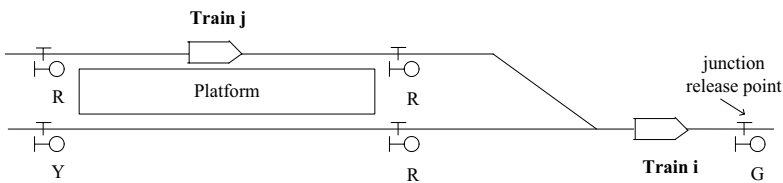
Figure 5.5: Hindrance of a departing train j due to occupancy of a route node on the route by a train i having just departed

The actual departure time of a train is also subject to the scheduled transfer connections. Small arrival delays of feeder trains can usually be absorbed by transfer buffer times. If the arrival delay of a feeder train is larger, a connecting train may wait to secure the transfer connection. In this section, it is assumed that the scheduled transfer connections

from one train to another are absolutely secured and the prespecified order of a pair of trains passing a route node must be adhered to.

Let $\overrightarrow{\mathcal{T}}_j$ denote the set of feeder trains, the actual departure time of a train j at a station is given by

$$D_j^{\mathrm{p}} = \max\left[A_j^{\mathrm{p}} + S_j^{\mathrm{p}}, sd_j^{\mathrm{p}}, C_j^{\mathrm{r}} + t^{\mathrm{p}}, \max_{k \in \overrightarrow{\mathcal{T}}_j}(A_k^{\mathrm{p}} + t_{k,j})\right] \qquad (5.24)$$

In the above, A_j^{p} and D_j^{p} represent the actual arrival and departure times of train j at the station, respectively, S_j^{p} the free dwell time, sd_j^{p} the scheduled departure time, C_j^{r} the clearance time of the outbound route, t^{p} the dispatching time, i.e., the time offset used to watch the departure signal and close the doors of a train at a station, which requires about 20 s according to on-site observations in some Dutch railway stations (Wiggenraad (2001)); A_k^{p} denotes the actual arrival time of a feeder train k at the platform track, $t_{k,j}$ the minimum transfer time of all passengers from train k to j. A diagram illustrating the modelling of the actual departure time of train j hindered by a conflicting train i and a feeder train k is shown in Figure 5.6.
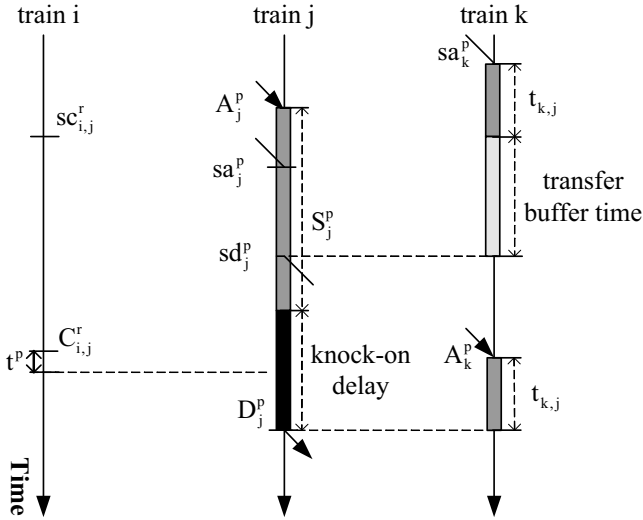
Figure 5.6: Illustration diagram of the actual departure procedure of train j hindered by a conflicting train i and a feeder train k (sa_j^{p} and sa_k^{p} represent the scheduled arrival times of trains j and k, respectively; $sc_{i,j}^{\mathrm{r}}$ denotes the scheduled release time of the critical track section on the route of train j by train i)

Next, the delay propagation to a departing train j whose route is different from those of the conflicting trains (Figures 5.4 and 5.5) is analyzed. Unlike the actual departure time, the arrival and free dwell times of train j are independent of the time when the route is released by the conflicting trains. The inbound route of train j is usually different from that of each feeder train k and the routes neither merge nor cross. In this case, the arrival

times of trains j and k are independent. Therefore, the distribution of the actual departure time of train j is formulated by

$$
F_{D_j^p}(x) = \begin{cases} 0 & \text{if } x < sd_j^p \\ \Pr(A_j^p + S_j^p \leq x) \cdot F_{C_j^r}(x - t^p) \cdot \prod_{k \in \overrightarrow{\mathcal{T}}_j} F_{A_k^p}(x - t_{k,j}) & \text{if } x \geq sd_j^p, \end{cases} \quad (5.25)
$$

In practice, the free dwell time of a train at a station is subject to the arrival time. Early arriving trains generally have a longer dwell time than late trains, because earlier departure than scheduled is not allowed. Furthermore, when a train arrives at a station earlier than the scheduled arrival time, the crew of the train may not feel so hurry and hence often wait for some late arriving passengers. However, when a train arrives late, it may dwell at the station for a shorter time to compensate for the arrival delay if there are few waiting passengers to alight and board the train. Taking into account the above-mentioned stochastic dependence, we can estimate the item $\Pr(A_j^p + S_j^p \leq x)$ in Eq. (5.25) as follows:

$$
\begin{aligned}
\Pr(A_j^p + S_j^p \leq x) &= \int \Pr[a_j^p + S_j^p(a_j^p) \leq x] dF_{A_j^p}(a_j^p) \\
&= \int F_{S_j^p(a_j^p)}(x - a_j^p) dF_{A_j^p}(a_j^p),
\end{aligned} \quad (5.26)
$$

where a_j^p is a realization of A_j^p; $S_j^p(a_j^p)$ implies that S_j^p depends on a_j^p; $F_{S_j^p(a_j^p)}$ denotes the conditional distribution of S_j^p in case of the realization a_j^p of A_j^p. Eq. (5.25) can then be expanded as

$$
F_{D_j^p}(x) = \begin{cases} 0 & \text{if } x < sd_j^p \\ \int F_{S_j^p(a_j^p)}(x - a_j^p) dF_{A_j^p}(a_j^p) & \\ \cdot F_{C_j^r}(x - t^p) \cdot \prod_{k \in \overrightarrow{\mathcal{T}}_j} F_{A_k^p}(x - t_{k,j}) & \text{if } x \geq sd_j^p. \end{cases} \quad (5.27)
$$

Apparently, it is ideal for us to estimate the departure time distribution of a train by Eq. (5.27), since this is a complete conditional modelling approach. To consider the dependence of the free dwell time S_j^p of train j upon the arrival time A_j^p on the one hand and to avoid requesting a great number of conditional free dwell time distributions and spending long computational time on the other hand, Yuan & Hansen (2006) express the departure time distribution by

$$
F_{D_j^p}(x) = \begin{cases} 0 & \text{if } x < sd_j^p \\ F_{A_j^{p+}} * F_{S_j^{p,l}}(x) \cdot F_{C_j^r}(x - t^p) \cdot \prod_{k \in \overrightarrow{\mathcal{T}}_j} F_{A_k^p}(x - t_{k,j}) & \text{if } x \geq sd_j^p, \end{cases} \quad (5.28)
$$

where A_j^{p+} represents the later one of the actual and scheduled arrival times of train j; $S_j^{p,l}$ denotes the free dwell time of train j in case of late arrival. This is in accordance with the modelling approach of the simulation tool RailSys (RMcon (2004)).

It should be mentioned that Eq. (5.28) is based on the assumption that the distribution of the free dwell times excluding the early arrival times in case of early arriving trains is the same as the distribution of the free dwell times of late arriving trains. To achieve a better estimate of the departure time distribution than that by Eq. (5.28), we relax the assumption as below. Let $S_j^{p,e*}$ denote the free dwell time excluding the early arrival time for an early arriving train j and $A_j^{p,l}$ represent the arrival time of a late train j. Herewith, we use $S_j^{p,e*}$ to distinguish it from $S_j^{p,e}$, which represents the free dwell time of a train in case of early arrival. We then obtain

$$
\begin{aligned}
\Pr(A_j^p + S_j^p \leq x) = \ & \Pr[sa_j^p + S_j^{p,e*} \leq x] \cdot F_{A_j^p}(sa_j^p) \\
& + \Pr[A_j^{p,l} + S_j^{p,l} \leq x] \cdot [1 - F_{A_j^p}(sa_j^p)] \\
= \ & F_{S_j^{p,e*}}(x - sa_j^p)F_{A_j^p}(sa_j^p) + F_{A_j^{p,l}} * F_{S_j^{p,l}}(x) \cdot [1 - F_{A_j^p}(sa_j^p)].
\end{aligned}
\tag{5.29}
$$

Therefore, the departure time distribution of train j can be estimated by

$$
F_{D_j^p}(x) = \begin{cases}
0 & \text{if } x < sd_j^p \\
\{F_{S_j^{p,e*}}(x - sa_j^p)F_{A_j^p}(sa_j^p) + F_{A_j^{p,l}} * F_{S_j^{p,l}}(x) \\
\cdot [1 - F_{A_j^p}(sa_j^p)]\} \cdot F_{C_j^r}(x - t^p) \cdot \prod_{k \in \overrightarrow{T}_j} F_{A_k^p}(x - t_{k,j}) & \text{if } x \geq sd_j^p.
\end{cases}
\tag{5.30}
$$

Let L_j^{dknock} denote the total knock-on delay of train j suffered at the station. We have

$$
L_j^{dknock} = \max\{\max[C_j^r + t^p, \max_{k \in \overrightarrow{T}_j}(A_k^p + t_{k,j})] - \max(A_j^p + S_j^p, sd_j^p), 0\}.
\tag{5.31}
$$

The above equation assures a nonnegative knock-on delay. Let

$$
D_j^{p1} = \max(A_j^p + S_j^p, sd_j^p),
\tag{5.32}
$$

and

$$
D_j^{p2} = \max[C_j^r + t^p, \max_{k \in \overrightarrow{T}_j}(A_k^p + t_{k,j})].
\tag{5.33}
$$

We can then express the distribution of L_j^{dknock} by

$$
F_{L_j^{dknock}}(x) = \begin{cases}
0 & \text{if } x < 0 \\
F_{D_j^{p2}} * F_{-D_j^{p1}}(x) & \text{if } x \geq 0.
\end{cases}
\tag{5.34}
$$

The distributions of D_j^{p1} and D_j^{p2} are expressed respectively by

$$
\begin{aligned}
F_{D_j^{\mathrm{p1}}}(x) &= \Pr[\max(A_j^{\mathrm{p}} + S_j^{\mathrm{p}}, sd_j^{\mathrm{p}}) \le x] \\[1em]
&= \begin{cases} 0 & \text{if } x < sd_j^{\mathrm{p}} \\ \Pr(A_j^{\mathrm{p}} + S_j^{\mathrm{p}} \le x) & \text{if } x \ge sd_j^{\mathrm{p}} \end{cases} \\[1em]
&= \begin{cases} 0 & \text{if } x < sd_j^{\mathrm{p}} \\ F_{S_j^{\mathrm{p,e*}}}(x - sa_j^{\mathrm{p}}) F_{A_j^{\mathrm{p}}}(sa_j^{\mathrm{p}}) + F_{A_j^{\mathrm{p,l}}} * F_{S_j^{\mathrm{p,l}}}(x) \\ \quad \cdot [1 - F_{A_j^{\mathrm{p}}}(sa_j^{\mathrm{p}})] & \text{if } x \ge sd_j^{\mathrm{p}}, \end{cases}
\end{aligned}
\tag{5.35}
$$

$$
\begin{aligned}
F_{D_j^{\mathrm{p2}}}(x) &= \Pr\left(\max\left[C_j^{\mathrm{r}} + t^{\mathrm{p}}, \max_{k \in \overrightarrow{\mathcal{T}}_j}(A_k^{\mathrm{p}} + t_{k,j})\right] \le x\right) \\[1em]
&= \Pr(C_j^{\mathrm{r}} + t^{\mathrm{p}} \le x) \cdot \prod_{k \in \overrightarrow{\mathcal{T}}_j} \Pr(A_k^{\mathrm{p}} + t_{k,j} \le x) \\[1em]
&= F_{C_j^{\mathrm{r}}}(x - t^{\mathrm{p}}) \cdot \prod_{k \in \overrightarrow{\mathcal{T}}_j} F_{A_k^{\mathrm{p}}}(x - t_{k,j}).
\end{aligned}
\tag{5.36}
$$

Let $L_{i,j}^{\mathrm{dknock}}$ and $L_{k,j}^{\mathrm{dknock}}$ denote the knock-on delay of train j caused by individual conflicting train i and feeder train k respectively. We have

$$
L_{i,j}^{\mathrm{dknock}} = \max(C_{i,j}^{\mathrm{r}} + t^{\mathrm{p}} - D_j^{\mathrm{p1}}, 0),
\tag{5.37}
$$

and

$$
L_{k,j}^{\mathrm{dknock}} = \max(A_k^{\mathrm{p}} + t_{k,j} - D_j^{\mathrm{p1}}, 0)
\tag{5.38}
$$

The distributions of $L_{i,j}^{\mathrm{dknock}}$ and $L_{k,j}^{\mathrm{dknock}}$ are formulated respectively by

$$
F_{L_{i,j}^{\mathrm{dknock}}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{C_{i,j}^{\mathrm{r}}} * F_{-D_j^{\mathrm{p1}}}(x - t^{\mathrm{p}}) & \text{if } x \ge 0 \end{cases}
\tag{5.39}
$$

and

$$
F_{L_{k,j}^{\mathrm{dknock}}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{A_k^{\mathrm{p}}} * F_{-D_j^{\mathrm{p1}}}(x - t_{k,j}) & \text{if } x \ge 0 \end{cases}
\tag{5.40}
$$

To estimate the outbound knock-on delay distributions and the distribution of the departure times of a train, the arrival time distribution and the conditional free dwell time distributions of this train, the distribution of the outbound route clearance times, and the arrival time distribution of each of the feeder trains are required. Based on the derived departure time distribution, we can calculate a variety of departure punctuality statistics. The distribution of knock-on delays due to route conflicts can be applied to determine the

capacity utilization of the corresponding route node in the outbound direction. The distribution of knock-on delays due to late connections may enable improved transfer buffer times of the feeder trains.

5.2.3   Dynamic delay propagation and impact of rescheduling

In a complicated station area with busy traffic, the critical track sections including the platform tracks and adjacent track junctions are highly utilized. Consequently, a train suffering knock-on delays caused by the preceding trains may in turn hinder the following trains. The propagation of train delays is a dynamic process. Carey (1999) presented a number of ex ante heuristic measures of schedule reliability including the so-called 'multi-stage knock-on delays', which are actually the results of dynamic delay propagation.

To estimate the distributions of knock-on delays, and arrival and departure times for all the trains during a certain time period, we use recursive substitutions. For all the trains passing a station during a day, we model the dynamic delay propagation starting from the initial train in the morning. The recursive substitutions are repeated step by step. When the headway of a train is scheduled sufficiently, the knock-on delay of this train will be smaller than a prespecified small value which thus can be neglected and results in the end of the recursive process. It is well known that most passenger trains are operated according to a periodic timetable. Therefore, the recursive process would be finished within the cycle of the operating period, e.g. one hour in the Netherlands if there is sufficient buffer time available for fading out the delay. Otherwise, the propagation of delays may extend to more than one cycle.

To realistically estimate the propagation of train delays in real operations, the impact of rescheduling has to be incorporated, too. Altering the prespecified order of trains is a common rescheduling approach for resolving route conflicts (Carey & Kwiecinski (1994b)). In addition, most railways apply certain guidelines for traffic controllers with respect to the synchronization control times of trains (Goverde (1998)), which are the maximally acceptable waiting times of trains for the feeder trains. Therefore, the assumptions that the prespecified order of trains must be adhered to and the planned transfer connections must be secured need to be relaxed in order to accurately predict on-line the propagation of train delays in real operations.

To support traffic controllers to make the most effective decision in real-time conflict-resolving, a number of optimization rescheduling models and computerized tools have been developed (Carey & Kwiecinski (1994b), Ho et al. (1997), D'Ariano & Pranzo (2004)). The quality of decision-making for resolving route conflicts is subject to the detection of conflicts (Makkinga & Zigterman (2002), Jacobs (2004)). To achieve a reliable detection of route conflicts and further efficient train rescheduling, the propagation of train delays in real operations should be estimated accurately by taking into account the stochasticity of train speeds and track occupancy and release times. An accurate estimation of real-time delay propagation will also contribute to the determination of optimal synchronization control times of trains. In the following, we will discuss the impact of altering the scheduled order of trains and canceling the scheduled transfer connections on the propagation of train delays, respectively.

Altering the scheduled order of trains

For a pair of trains passing a route node successively according to a given timetable, if the first scheduled train has a large delay upstream while the second one does not approach the route node even late, train controllers may on-line alter the prespecified passing order of this pair of trains to resolve an anticipated route conflict. The determination of the passing order of a pair of scheduled trains at a route node may be based on the simple criterion First Come First Served (FCFS), which solves the conflict by first assigning a route for the train first calling it. However, the priorities of trains, which are related to the type of the train and the number of passengers on board, must also be taken into consideration during an on-line train rescheduling. Ideally, the weighted total delay of all the involved trains and passengers respectively should be minimized.

In practice, train controllers usually make a on-line decision regarding the alteration of the scheduled order of a pair of trains passing the conflicting route node on the basis of a critical passing time (delay) of the first train and a critical time lag between the passage of the train pair at upstream stations (timing points). Note that the actual passing times of trains are recorded more or less distant from the route node, since a certain time is required for conflict recognition, on-line decision-making of train rescheduling and setting-up of a new route. If this pair of trains come along different railway lines, the conflict detecting positions are situated at different stations (timing points). The rescheduling guideline actually contains the FCFS criterion and it also takes into account the priorities of trains implicitly by specifying different threshold values of the passing time of the first train and time lag between the train pair at the conflict detecting positions. In the following, we will formulate the propagation of train delays provided that this train rescheduling guideline is adopted.

Considering a pair of trains i and j scheduled to pass a route node successively, it is assumed in the first instance that we only know the probability distribution of the actual passing time of the first scheduled train i, but we know the exact passing time of the second scheduled train j at their conflict detecting positions. Let A_i^{us} and a_j^{us} denote the passing times of the train pair i and j, respectively. The prespecified order of this pair of trains at the conflicting route node is altered during real operations if and only if

$$A_i^{\text{us}} > \max(a_i^{\text{us}*}, a_j^{\text{us}} + \delta_{j,i}^{\text{us}*}) = a_i^{\text{us}**}, \tag{5.41}$$

where $a_i^{\text{us}*}$ and $\delta_{j,i}^{\text{us}*}$ denote the predefined critical passing time of train i and the time lag between the passing of this pair of trains at the conflict detecting positions, respectively. Clearly, $a_i^{\text{us}**}$ depends on a_j^{us}.

Let $T_i^{\text{us-rp}}$ denote the occupancy time of the track section between the conflict detecting position upstream the route node and the release position of the route node by the first scheduled train i and

$$\widetilde{C}_{i,j}^{\text{r}} = \begin{cases} A_i^{\text{us}} + T_i^{\text{us-rp}} & \text{if } A_i^{\text{us}} \leq a_i^{\text{us}**} \\ -\infty & \text{if } A_i^{\text{us}} > a_i^{\text{us}**}. \end{cases} \tag{5.42}$$

When $A_i^{\text{us}} \leq a_i^{\text{us}**}$, $\widetilde{C}_{i,j}^{\text{r}}$ represents then the real release time of the route of the second scheduled train j by the first scheduled train i. When $A_i^{\text{us}} > a_i^{\text{us}**}$, we set a virtual route

release time $\widetilde{C}_{i,j}^{r} = -\infty$. This implies that the real operation of train j is not influenced by train i in case of an altered train order at the conflicting route node.

If the conflicting route node is situated on the inbound route of train j, let

$$\widetilde{\Delta}_{i,j}^{as} = \widetilde{C}_{i,j}^{r} - A_{j}^{as} = \widetilde{C}_{i,j}^{r} - a_{j}^{us} - T_{j}^{us\text{-}as}, \tag{5.43}$$

where $T_{j}^{us\text{-}as}$ represents a random running time of train j between the conflict detecting position upstream the route node and the approach signal of the conflicting route node.

Substituting $\widetilde{C}_{i,j}^{r}$ in Eq. (5.43) by Eq. (5.42), we have

$$\widetilde{\Delta}_{i,j}^{as} = \begin{cases} A_{i}^{us} + T_{i}^{us\text{-}rp} - a_{j}^{us} - T_{j}^{us\text{-}as} & \text{if } A_{i}^{us} \leq a_{i}^{us**}, \\ -\infty & \text{if } A_{i}^{us} > a_{i}^{us**}. \end{cases} \tag{5.44}$$

The distribution of $\widetilde{\Delta}_{i,j}^{as}$ can be derived by

$$F_{\widetilde{\Delta}_{i,j}^{as}}(x) = \begin{cases} 1 - F_{A_{i}^{us}}(a_{i}^{us**}) & \text{if } x = -\infty, \\ 1 - F_{A_{i}^{us}}(a_{i}^{us**}) + F_{\overline{A}_{i}^{us}} * F_{T_{i}^{us\text{-}rp}} * F_{-T_{j}^{us\text{-}as}}(x + a_{j}^{us}) \\ \cdot F_{A_{i}^{us}}(a_{i}^{us**}) & \text{if } x > -\infty, \end{cases} \tag{5.45}$$

where $F_{\overline{A}_{i}^{us}}(x)$ denotes the right-truncated distribution of A_{i}^{us} at $A_{i}^{us} = a_{i}^{us**}$, which is given by

$$F_{\overline{A}_{i}^{us}}(x) = \begin{cases} F_{A_{i}^{us}}(x)/F_{A_{i}^{us}}(a_{i}^{us**}) & \text{if } x \leq a_{i}^{us**} \\ 1 & \text{if } x > a_{i}^{us**} \end{cases} \tag{5.46}$$

Replacing Δ_{j}^{as} by $\widetilde{\Delta}_{i,j}^{as}$ in Eq. (5.9) to (5.11) and Eq. (5.14), we can obtain the distribution of the knock-on delay of an approaching train j caused by the scheduled preceding train i by considering the alteration of the scheduled order of this pair of trains in real operations. Similarly, we are able to estimate the distribution of the knock-on delay of train j resulting from multiple route conflicts in the inbound direction of the station and the distribution of the arrival time at the station by considering the alteration of the scheduled order for all the involved trains. In addition, the distribution of the knock-on delay of a departing train due to route conflicts in the outbound direction of the station and the distribution of the departure time at the station can be estimated by considering the alteration of the scheduled train order.

For a pair of trains passing a conflicting route node successively, altering the prespecified passing order reduces the knock-on delay of the second scheduled train. However, the probability that the first scheduled train suffers a knock-on delay from the second scheduled train becomes non-zero.

Let $T_{j}^{us\text{-}rp}$ denote the occupancy time of the track section between the conflict detecting position upstream the route node and the release position of the route node by the second scheduled train j and

$$\widetilde{C}_{j,i}^{\mathrm{r}} = \begin{cases} -\infty & \text{if } A_i^{\mathrm{us}} \leq a_i^{\mathrm{us**}} \\ a_j^{\mathrm{us}} + T_j^{\mathrm{us\text{-}rp}} & \text{if } A_i^{\mathrm{us}} > a_i^{\mathrm{us**}}. \end{cases} \tag{5.47}$$

When $A_i^{\mathrm{us}} > a_i^{\mathrm{us**}}$, $\widetilde{C}_{j,i}^{\mathrm{r}}$ represents then the real release time of the route of the first scheduled train i by the second scheduled train j. When $A_i^{\mathrm{us}} \leq a_i^{\mathrm{us**}}$, we set a virtual route release time $\widetilde{C}_{j,i}^{\mathrm{r}} = -\infty$. This indicates that the normal operation of train i is not influenced by the scheduled following train j.

If the conflicting route node is included within the inbound route of train i, let

$$\widetilde{\Delta}_{j,i}^{\mathrm{as}} = \widetilde{C}_{j,i}^{\mathrm{r}} - A_i^{\mathrm{as}} = \widetilde{C}_{j,i}^{\mathrm{r}} - A_i^{\mathrm{us}} - T_i^{\mathrm{us\text{-}as}}, \tag{5.48}$$

where $T_i^{\mathrm{us\text{-}as}}$ represents a random running time of train i between the conflict detecting position upstream the route node and the approach signal of the route node.

Substituting $\widetilde{C}_{j,i}^{\mathrm{r}}$ in Eq. (5.48) by Eq. (5.47), we obtain

$$\widetilde{\Delta}_{j,i}^{\mathrm{as}} = \begin{cases} -\infty & \text{if } A_i^{\mathrm{us}} \leq a_i^{\mathrm{us**}} \\ a_j^{\mathrm{us}} + T_j^{\mathrm{us\text{-}rp}} - A_i^{\mathrm{us}} - T_i^{\mathrm{us\text{-}as}} & \text{if } A_i^{\mathrm{us}} > a_i^{\mathrm{us**}} \end{cases} \tag{5.49}$$

The distribution of $\widetilde{\Delta}_{j,i}^{\mathrm{as}}$ can be derived by

$$F_{\widetilde{\Delta}_{j,i}^{\mathrm{as}}}(x) = \begin{cases} F_{A_i^{\mathrm{us}}}(a_i^{\mathrm{us**}}) & \text{if } x = -\infty \\ F_{A_i^{\mathrm{us}}}(a_i^{\mathrm{us**}}) + F_{T_j^{\mathrm{us\text{-}rp}}} * F_{-\underline{A}_i^{\mathrm{us}}} * F_{-T_i^{\mathrm{us\text{-}as}}}(x - a_j^{\mathrm{us}}) \\ \quad \cdot [1 - F_{A_i^{\mathrm{us}}}(a_i^{\mathrm{us**}})] & \text{if } x > -\infty \end{cases} \tag{5.50}$$

where $F_{\underline{A}_i^{\mathrm{us}}}(x)$ represents the left-truncated distribution of A_i^{us} at $A_i^{\mathrm{us}} = a_i^{\mathrm{us**}}$, which is given by

$$F_{\underline{A}_i^{\mathrm{us}}}(x) = \begin{cases} 0 & \text{if } x \leq a_i^{\mathrm{us**}} \\ [F_{A_i^{\mathrm{us}}}(x) - F_{A_i^{\mathrm{us}}}(a_i^{\mathrm{us**}})]/[1 - F_{A_i^{\mathrm{us}}}(a_i^{\mathrm{us**}})] & \text{if } x > a_i^{\mathrm{us**}} \end{cases} \tag{5.51}$$

Replacing the subscript j by i and Δ_j^{as} by $\widetilde{\Delta}_{j,i}^{\mathrm{as}}$ in Eq. (5.9) to (5.11) and Eq. (5.14), we are able to estimate the distribution of the knock-on delay of an approaching train i caused by the scheduled following train j at the conflicting route node by considering the alteration of the scheduled order of this pair of trains. Similarly, the distribution of the knock-on delay of a departing train from the scheduled following train at a conflicting route node in the outbound direction of the station can be estimated by considering the train order alteration.

In the above, it has been assumed that we know the exact passing time of the second scheduled train at the conflict detecting position upstream the conflicting route node between a train pair. In practical applications, we may only know the probability distribution. In this case, we are still able to predict the knock-on delay distribution by integrating the relevant equations over the passing time of the second train at the conflict detecting position.

Canceling scheduled transfer connections

Considering a departing train j and the scheduled feeder train $k \in \overrightarrow{\mathcal{T}}_j$ at a station, the transfer connection from train k to j is canceled if and only if

$$A_k^{\mathrm{p}} + t_{k,j} > sd_j^{\mathrm{p}} + l_{j,k}, \tag{5.52}$$

where $t_{k,j}$ denotes the minimum transfer time of all passengers from train k to j and $l_{j,k}$ is the maximum acceptable waiting time of train j for train k at the station.

Let

$$\widetilde{A}_k^{\mathrm{p}} = \begin{cases} A_k^{\mathrm{p}} & \text{if } A_k^{\mathrm{p}} + t_{k,j} \le sd_j^{\mathrm{p}} + l_{j,k} \\ -\infty & \text{if } A_k^{\mathrm{p}} + t_{k,j} > sd_j^{\mathrm{p}} + l_{j,k}. \end{cases} \tag{5.53}$$

When $A_k^{\mathrm{p}} + t_{k,j} > sd_j^{\mathrm{p}} + l_{j,k}$, we set a virtual arrival time $\widetilde{A}_k^{\mathrm{p}} = -\infty$. This is used to model the independence of the actual departure of train j upon the arrival of the feeder train k in case of cancellation of the scheduled transfer connection.

The distribution of the knock-on delay of train j caused by the feeder train k is then given by

$$F_{L_{k,j}^{\mathrm{dknock}}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{\widetilde{A}_k^{\mathrm{p}}} * F_{-D_j^{\mathrm{pl}}}(x - t_{k,j}) & \text{if } x \ge 0, \end{cases} \tag{5.54}$$

where

$$F_{\widetilde{A}_k^{\mathrm{p}}}(x) = \begin{cases} 1 - F_{A_k^{\mathrm{p}}}(sd_j^{\mathrm{p}} + l_{j,k} - t_{k,j}) & \text{if } x = -\infty \\ 1 - F_{A_k^{\mathrm{p}}}(sd_j^{\mathrm{p}} + l_{j,k} - t_{k,j}) + F_{A_k^{\mathrm{p}}}(x) & \text{if } -\infty < x \le sd_j^{\mathrm{p}} + l_{j,k} - t_{k,j} \\ 1 & \text{if } x > sd_j^{\mathrm{p}} + l_{j,k} - t_{k,j}. \end{cases} \tag{5.55}$$

Let $\widetilde{C}_j^{\mathrm{r}} = \max_{i \in \widetilde{\mathcal{T}}_j}(\widetilde{C}_{i,j}^{\mathrm{r}})$. By incorporating the cancellation of the scheduled transfer connections from one train to another and the alteration of the scheduled order of train pairs passing the conflicting route node in real operations, we can express the distribution of the actual departure time of a departing train j at a station as follows:

$$F_{D_j^{\mathrm{p}}}(x) = \begin{cases} 0 & \text{if } x < sd_j^{\mathrm{p}} \\ \{F_{S_j^{\mathrm{p,e*}}}(x - sa_j^{\mathrm{p}})F_{A_j^{\mathrm{p}}}(sa_j^{\mathrm{p}}) + F_{A_j^{\mathrm{p,l}}} * F_{S_j^{\mathrm{p,l}}}(x) \\ \cdot [1 - F_{A_j^{\mathrm{p}}}(sa_j^{\mathrm{p}})]\} \cdot F_{\widetilde{C}_j^{\mathrm{r}}}(x - t^{\mathrm{p}}) \cdot \prod\limits_{k \in \overrightarrow{\mathcal{T}}_j} F_{\widetilde{A}_k^{\mathrm{p}}}(x - t_{k,j}) & \text{if } x \ge sd_j^{\mathrm{p}}. \end{cases} \tag{5.56}$$

5.3   Numerical algorithm

Recalling the model description in the previous section, a key question is how to estimate several types of integrals. Among these integrals, the convolution of several independent distributions is frequently applied. In the field of railway operations research, different theoretical distributions can be used to model the stochastic event and process times and an analytical solution to the convolution often is not available. The empirical distributions of train event and proccess times may be applied. Moreover, the convolution estimation is integrated in the model to enable a computerized decision support for timetabling and rescheduling. Thus, we will calculate the convolution numerically. When no empirical data is available, the input distributions of the model may be assumed theoretically and then discretized to be used in the numerical scheme proposed as below.

There are a number of numerical schemes that can be used to approximate the convolution. Among others, the composite trapezoidal rule is a simple but robust approach and it can be easily implemented. Tortorella (1990) and Boehme et al. (1991) apply the composite trapezoidal rule for calculating the convolution distribution of two non-negative random variables. To estimate the convolution distribution of two generic random variables, let us find a big integer l satisfying

$$F_{X_1} * F_{X_2}(x) \approx \int_{-lh}^{lh} F_{X_2}(x - u) dF_{X_1}(u), \tag{5.57}$$

where h is the integral spacing of the composite trapezoidal rule. The numerical approximation of the above equation can be expressed as

$$Q(F_{X_1}, F_{X_2})(x_r) = \sum_{s=-l+1}^{l} \frac{F_{X_2}(x_r - x_s) + F_{X_2}(x_r - x_{s-1})}{2} [F_{X_1}(x_s) - F_{X_1}(x_{s-1})], \tag{5.58}$$

where $x_r = rh, x_s = sh; r, s = -l, -l+1, ..., l$.

The above equation can be rewritten as

$$Q(F_{X_1}, F_{X_2})(x_r) = \sum_{s=-l+1}^{l} \frac{F_{X_2}(x_{r-s}) + F_{X_2}(x_{r-s+1})}{2} [F_{X_1}(x_s) - F_{X_1}(x_{s-1})]. \tag{5.59}$$

Assuming that $F_{X_1}(x)$ and $F_{X_2}(x)$ have second continuous derivatives on $[-lh, lh]$, Tortorella (1990) proves the following error bound of the numerical approximation:

$$|\int_{-lh}^{lh} F_{X_2}(x_r - u) dF_{X_1}(u) - Q(F_{X_1}, F_{X_2})(x_r)| \leq \frac{lh^3}{6}(B_1^1 B_2^2 + B_1^2 B_2^1) \tag{5.60}$$

where B_i^j denotes $\sup\{|d^j F_{X_i}(u)|, u \in [-lh, lh]\}$.

The above inequality provides a rather conservative estimation regarding the approximation error of the composite trapezoidal scheme. In fact, we can apply a bigger integral spacing h than the value of h obtained on the basis of this conservative error bound. For the convolution of more than two independent distributions, the numerical approximation will be done via a recursive procedure.

To estimate the conditional distributions for the arrival time of a train at the approach signal, the clearance time of the route of this train as well as the time lag between these times in case of one of the distinguished events, we need to calculate the convolution but with a variable integral upper limit:

$$F_{X_1} * F_{X_2}(x, z) = \int\limits_{-\infty}^{z} F_{X_2}(x - u) dF_{X_1}(u). \qquad (5.61)$$

We approximate the above equation by

$$Q(F_{X_1}, F_{X_2})(x_r, x_p) = \sum_{s=-l+1}^{p} \frac{F_{X_2}(x_{r-s}) + F_{X_2}(x_{r-s+1})}{2}[F_{X_1}(x_s) - F_{X_1}(x_{s-1})], \qquad (5.62)$$

where x_p is defined just like x_r in the above.

Another type of integral to be calculated can be generalized as

$$F_{X_2}(x) = \int\limits_{0}^{+\infty} F_{X_2|X_1=u}(x) dF_{X_1}(u). \qquad (5.63)$$

It can be approximated by

$$Q(F_{X_2})(x_q) = \sum_{t=1}^{l} \frac{F_{X_2|X_1=x_t}(x_q) + F_{X_2|X_1=x_{t-1}}(x_q)}{2}[F_{X_1}(x_t) - F_{X_1}(x_{t-1})], \qquad (5.64)$$

where, $x_q = qh$, $x_t = th$; $q, t = 0,1,...,l$.

Adopting the introduced numerical schemes, we have discretized the proposed probability delay propagation model. Having coded a Matlab application program, we are able to estimate the knock-on delay distributions as well as the arrival and departure time (delay) distributions for each train series in a station. If the model is applied to improve the capacity utilization and timetabling, the input delay distributions may have a large standard deviation, the computation time for estimating the delay distributions of a train series may be up to 2 min when the probability of knock-on delays is very high. In case of real-time applications of the delay propagation model, we generally have more exact information with respect to the input delays and the computation time can thus be reduced to a few seconds.

5.4   Empirical validation

In this section, the developed model is validated using the local network of The Hague HS station, which is characterized by level crossings and mergings/divergings of different railway lines. A schematic track layout of the local network has been shown in Figure 3.1. Model validation is performed by comparing the delay estimates derived from the model with the empirical data recorded in September 1999 per train series in each direction. The scheduled arrival and departure times and operating routes of these train series have been given in Chapter 3. Because the knock-on delays of trains cannot be measured in practice, we validate the model estimates of the arrival and departure delays of trains at the station. Due to lack of information for the rescheduling rules applied in the operations during the period of data recording, we only consider the parts of the model described in Sections 5.2.1 and 5.2.2. The recorded data corresponding to alteration of the scheduled order of trains are hence not used.

To estimate the distributions of the arrival and departure delays at the station, the input distributions and conditional distributions as well as several model parameters are required. The empirical distribution of the arrival times of trains at the approach signal of the station can be easily calculated using the track occupancy and release time records. The distributions of the free dwell times of both early and late arriving trains as well as the conditional distributions of train running and track occupancy times in the case of different aspects of the approach signal and the home signal of the station are also calculated on the basis of the empirical data. To this end, the data must be classified beforehand, which has been discussed in Chapter 4. Those deterministic parameters of the model are assumed just like in the literature or on the basis of on-site observations and of the findings from the Dutch TNV data (see Sections 5.2.1 and 5.2.2).

Based on the empirical input distributions and conditional distributions as well as the assumed model parameters, we estimated the arrival and departure delay distributions for 14 and 18 cases (a train series in one direction is referred to as a case), respectively. To compare the estimated and empirical results, we use the scatter plot of delay statistics and the plot of distribution function differences. Figures 5.7 and 5.8 show the scatter plot for the estimated versus the empirical mean arrival delay and that for the estimated versus the empirical arrival punctuality at delays of 0, 1 and 3 min, respectively. The arrival delay distribution for each train series has been estimated by applying Eq. (5.21) (also referred to as 'new formula of arrival time distribution' in Figure 5.9), i.e., taking into account the dependence of the conditional distribution of the inbound route clearance times on the arrival times of trains at the approach signal of the station in case of the defined third typical event (see Section 5.2.1). The difference between the estimated and the empirical mean arrival delay is smaller than 9 s for all the 14 considered cases. The mean difference is even less than 2 s in 9 of the total 14 cases. For the arrival punctuality defined at delays of 0, 1 and 3 min, the maximum estimation error among all the cases is 4%, 3% and 2%, correspondingly. The punctuality estimation error is even smaller than 1% in 10, 11 and 13 of the total 14 cases, respectively, for those definitions.

It has been found that taking into account the dependence of the conditional distribution of the inbound route clearance times on the arrival times of trains at the approach signal of the station did not improve the distribution estimate by Yuan & Hansen (2006) for the arrival times of trains at The Hague HS station. In case of a sufficiently long buffer time
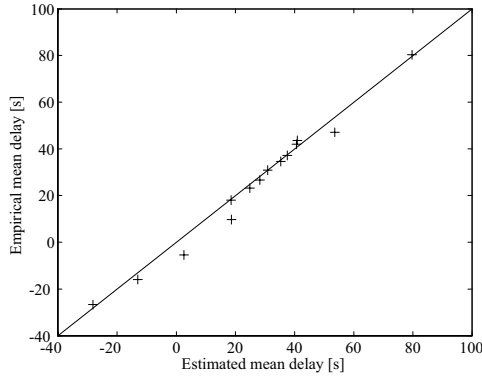
**Figure 5.7:** Scatter plot for mean arrival delay
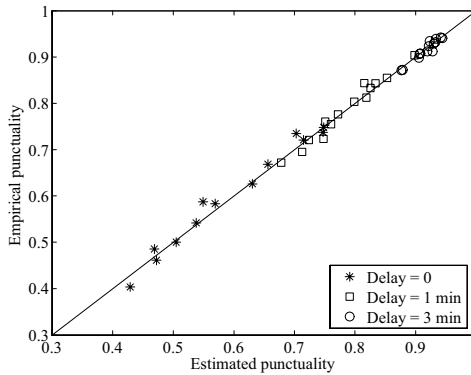
Figure 5.8: Scatter plot for the arrival punctuality of trains

between a pair of scheduled train paths, the following train has only a small probability of suffering a knock-on delay caused by the leading one. The probability of the third typical event defined is, of course, much smaller. In this case, the dependence consideration obviously cannot lead to much change with respect to the estimation of the arrival delay distribution.

Surprisingly, even when the scheduled buffer time is very short, e.g. about 1 min and the probability of the third typical event defined is quite high, the estimation of the arrival delay distribution for the southbound international trains INT600S could not be further improved by taking into account the stated distribution dependence (see Figure 5.9). This implies that the conditional distribution of the inbound route clearance times does not depend so much on the arrival times of trains at the approach signal of the station in this case. In fact, the knock-on delays of the international train series are mainly due to the early arrival itself, which has been revealed by the statistical analysis performed in Chapter 3. In case of larger delays of the preceding local train series AR5100N, the order of this train pair at the conflicting route node is altered because of a higher priority of the

international train series. These facts explains the weak dependence of the conditional distribution of the inbound route clearance times on the arrival times of the international trains at the approach signal of the station. The distribution estimation errors of less than 0.05 may be due to late setting of train routes that could not be reflected in the model. They are also subject to the approximation of the conditional distributions of the running times of hindered trains.
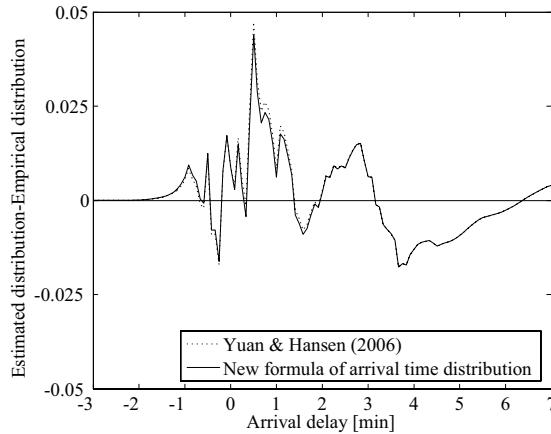
Figure 5.9: Distribution function differences plots for the estimated and empirical arrival delays INT600S

We have estimated the departure delay distribution for all the 18 train series that stop at the station by applying Eq. (5.30) (also referred to as 'new formula of departure time distribution' in Figure 5.12), respectively. This implies that the distribution of the free dwell times excluding the early arrival times for early arriving trains has been distinguished from the distribution of the free dwell times of late arriving trains per train series per direction. Figure 5.10 shows the scatter plot of the estimated versus the empirical mean departure delay. The difference between the estimated and the empirical mean departure delay is smaller than 9 s in all the cases. This mean difference is even less than 2 s in 10 of the total 18 cases. The scatter plot of the estimated versus the empirical departure punctuality at delays of 30 s, 1 and 3 min is given in Figure 5.11. The maximum estimation error among all the studied cases is 4%, 6% and 4% for the departure punctuality at these delays, respectively. The punctuality estimation errors are even smaller than 1% in 10, 10 and 12 of the total 18 cases, respectively, for the different definitions.

The distribution differences plots for the estimated and the empirical departure delays are shown in Figure 5.12 for the international trains INT600N. The dashed line corresponds to the departure delay estimate by applying Eq. (5.28), i.e., assuming that the distribution of the free dwell times excluding the early arrival times in case of early arriving trains is the same as the distribution of the free dwell times of late arriving trains per train series per direction (Yuan & Hansen (2006)). The solid line corresponds to the departure delay estimate by taking into account the distribution difference. In the former case, the maximum estimation error of the departure delay distribution is 0.11. In the latter case, we estimate the departure delay distribution with a maximum error of 0.02.

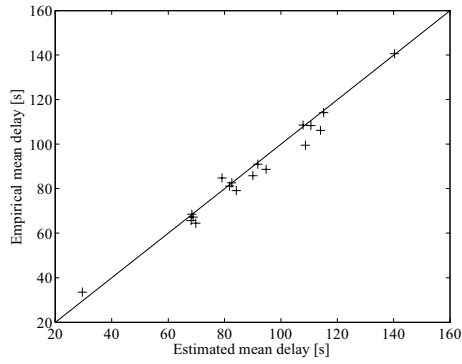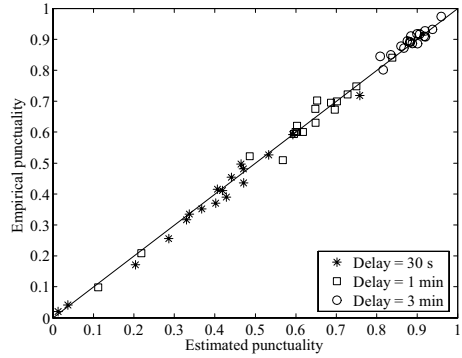**Figure 5.10:** Scatter plot for mean departure delay

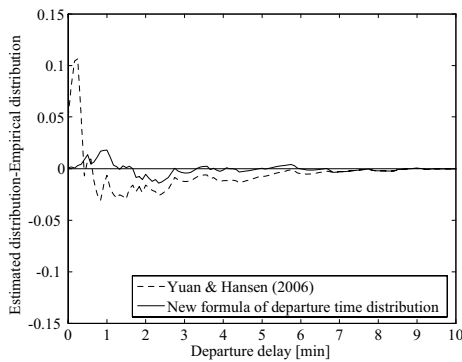**Figure 5.11:** Scatter plot for the departure punctuality of trains

Figure 5.12: Distribution function differences plots for the estimated and empirical departure delays of INT600N

For most of the studied cases, a complete distinction of the free dwell time distributions between early and late arriving trains leads to a much better estimation of the departure punctuality at smaller delays than that by Yuan & Hansen (2006). The maximum estimation error of the departure punctuality for delays of 30 s among the 18 cases decreased by 12%. However, the estimates of the departure punctuality at larger delays, e.g. 3 min as well as the estimated mean departure delay could not be improved. To further refine the departure delay estimation particularly in case of a longer scheduled dwell time, e.g. 3 min, a distinction of the arrival delays of trains into more intervals may be useful and the corresponding conditional distributions of the free dwell times of trains are to be applied.

The validation results reveal that the developed model estimates the arrival and departure distributions very well provided that the empirical input distributions are applied. This implies that the model is mathematically sound and is consistent with real train operations. However, a more challenging work is how to predict the propagation of train delays by the developed model. Having replaced the original free dwell times of both early and late arriving trains with the fitted Weibull distributions, we estimated the departure delay distribution for the northbound international trains INT600N. We compare, in Figure 5.13, the errors of the delay distribution estimated on the basis of the original free dwell times with those estimated by applying the fitted Weibull distributions. It appears that the estimation errors in the latter situation are overall bigger and the maximal estimation error has an increase of about 3%. In addition, the estimation error of the mean delay has an increase of 14 s. It should be mentioned that the parameters of the Weibull distributions have been specified by the proposed parameter fine-tuning method. If the parameters were specified with a lower level of accuracy, the distribution estimation would have a much bigger error. To predict the arrival and departure time distributions realistically, the parameters of the input distributions must be specified with a high accuracy. The number of passengers boarding and alighting the trains and the behavior of train crews during the processes of train running and dwelling would need to be modelled.
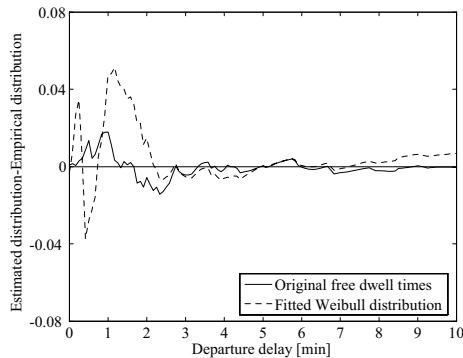
Figure 5.13: Error comparison between the departure delay distribution estimated on the basis of the original free dwell times and that estimated by adopting the fitted Weibull distributions for the northbound trains INT600N

5.5   Conclusions

In this chapter, we have presented our newly developed analytical stochastic model of train delay propagation in stations, which estimates the knock-on delays of trains caused by route conflicts and late transfer connections at a high precision. This model reflects the constraints of the signalling system and traffic management rules on the dynamic use of track capacity. The stochastic variations of track occupation times due to the fluctuation of train speed in case of different signal aspects are modelled by conditional probability distributions. The interdependences of arrival and departure times of different train lines in a complicated station area are taken into account. In addition, we model possible dependences of free dwell times of trains on the corresponding arrival times (delays) by conditional free dwell time distributions. The developed model is solved on the basis of a numerical approximation of the convolution of individual distributions and can be integrated into a larger computerized tool for timetable design and rescheduling.

Based on the local network of the Dutch railway station The Hague HS, we have validated the model using empirical train detection data. The validation results reveal that the developed model estimates the arrival and departure delay distributions very well provided that the empirical input distributions are applied. A reliable prediction of train arrival and departure time distributions can only be achieved by specifying the required input distributions including the parameters with a high accuracy. In case of other stations, the overall track layout and the interlocking of routes may be different. No matter how complex a station network is, it can always be decomposed into a number of platform tracks and conflicting route nodes (crossings and switches). A route conflict of a pair of trains in any station area may occur in the following standard conditions:

1. Consecutive arrival at a platform track (Figure 5.1);

2. Consecutive passing a common route node before arriving at different platform tracks (Figure 5.2);

3. Consecutive passing a common route node with the first train departing from and the second one approaching at the station (Figure 5.3) or vice versa (Figure 5.4);

4. Consecutive departure using the same outbound track sections (Figure 5.5).

The developed analytical probability model for estimating knock-on delays and the impact on the punctuality of trains is based on individual platform tracks and route nodes. Therefore, the developed model is of generic value. This presented model will be applied for improving station capacity utilization and timetable design by estimating knock-on delay distributions and the impact on the punctuality of train operations.

Chapter 6

Improving station capacity utilization and timetable design

6.1   Introduction

This chapter shows the applicability of the stochastic delay propagation model for improving station capacity utilization and timetable design by a case study of The Hague HS station. The level crossing (229BT) between the inbound route of the southbound trains from Amsterdam/Leiden and the outbound route of the northbound trains leading to The Hague CS is a bottleneck in the area of The Hague HS station (see Figure 3.2). If the capacity utilization of this route node is too high or the buffer time between scheduled train paths at this node is not properly distributed, the punctuality of train arrivals and departures at the station will be affected severely due to route conflicts and the resulting knock-on delays in real operations. By estimating the knock-on delays at this critical node using the delay propagation model, the maximal node capacity utilization is first determined through a maximally acceptable level of knock-on delays at a certain confidence level. The influences of the input and primary delays, buffer time assignment between scheduled train paths and the alteration of the train order in real operations are then discussed. Furthermore, the model application is extended to the whole station area for improving the station capacity utilization and timetable design at a desired punctuality level of train operations.

6.2   Model application to a critical route node

In this section, the delay propagation model developed in Chapter 5 is applied to the level crossing 229BT. For simplicity, we consider a number of trains with the same dynamic characteristics passing this route node in both southbound and northbound directions alternately. By assuming a standard running time supplement of 7% and a scheduled dwell time of 120 s for all trains in first instance, we obtain a maximal frequency of 13 pairs/ h in both directions based on the virtually compressed conflict-free blocking time stairways at the level crossing and at the relevant platform tracks (see Figure 6.1).
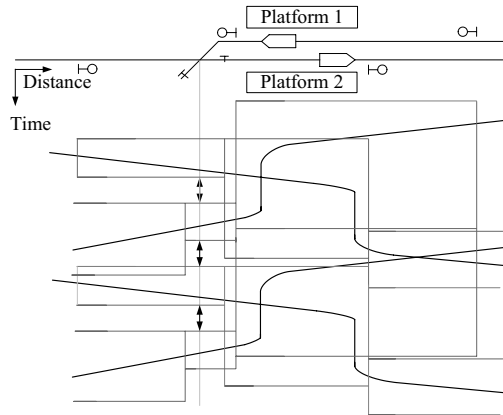
Figure 6.1: Virtually compressed conflict-free blocking time stairways at level crossing 229BT and at the relevant platform tracks in the area of The Hague HS station

The minimal buffer time between a pair of scheduled train paths is 48 s at the level crossing, 17 s at the platform track of the southbound trains from Amsterdam/Leiden and 0 s at the platform track of the northbound trains leading to The Hague CS. By calculating the buffer time between train paths given the frequency of scheduled trains passing the level crossing, it is found that the most critical bottleneck in the area of the station is the level crossing as far as the train frequency is lower than 10 pairs/ h. As soon as the train frequency is beyond 10 pairs/ h, the station track next to platform 1 becomes critical in the station area.

6.2.1 Maximizing route node capacity utilization with a maximally acceptable knock-on delay at a certain confidence level

The number of passenger train services is increasing with social progress and mobility development. Besides the extension of infrastructure, the existing capacity of railways needs to be utilized more efficiently, particularly in peak periods of a day. However, if the capacity utilization exceeds a certain level, knock-on delays of trains caused by other trains will occur frequently. In the following, we analyze the influence of capacity utilization of the critical level crossing 229BT on the knock-on delays of passing trains.

To estimate the knock-on delays of a train series within the local network, we focus in first instance on the impact of the preceding train series. It is here assumed that the trains which may suffer knock-on delays enter the network on time. Of course, the knock-on delays are also subject to the input delays of the train series itself. If a train arrives at the conflicting route node earlier than scheduled, the probability of suffering a knock-on delay would be higher than a train that is coming on time or later than scheduled. The expected knock-on delay of a train caused by its early arrival can be avoided by continuous automatic speed control in real operations. Therefore, it is paid less attention in the planning phase. The knock-on delay probability of a late train is less than that of a

punctual train. This partially justifies our assumption. The influence of the input delays of trains on the knock-on delays will be quantified in the next section.

Based on the statistical analysis results obtained in Chapter 4, we adopt a shifted log-normal model to represent the distribution for the arrival times of the preceding train series at the approach signal of the station signal block. The location parameter is assumed to be the earliest arrival time, e.g. 60 s earlier than the scheduled arrival time. The mean and standard deviation of the underlying normal distribution are derived using the moment method (Appendix A.6) on the basis of the assumed mean and standard deviation of the arrival times, e.g. 60 s and 90 s, respectively. For easier calculation, the primary delays of all involved trains during the running and dwell processes are neglected here.

In case of an approaching train hindered at the preceding block of the station, the running time on the track sections leading to the platform increases considerably. Of course, the running times of hindered trains may vary a lot due to the behavioral impact of train drivers, who may react to a change of signal aspects early or late. If a hindered train does not stop before the home signal of the station, the running time from the approach signal to the stop location on the platform is, for simplicity, assumed to be the scheduled time multiplied by a coefficient of 1.3, as is in accordance with Carey & Kwiecinski (1994a). If a hindered train stops before the home signal of the station, the running time from the home signal to the stop location at the platform is assumed to be the scheduled time multiplied by a coefficient of 1.4. Similar coefficients correspond, too, to a deterministic calculation according to the operational rules of the Dutch Railways.
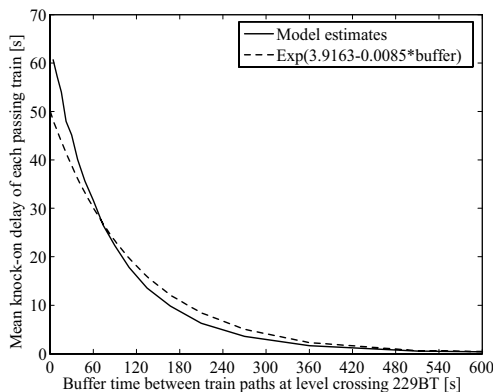
Figure 6.2: Mean knock-on delay as a function of buffer time between scheduled train paths at level crossing 229BT

Figure 6.2 shows the mean knock-on delay of each passing train at the route node 229BT as a function of the buffer time between the scheduled train paths at this node. It is clear that the mean knock-on delay increases with a decrease of the buffer time. In fact, an approximately exponential increase has been found by a statistical regression analysis. The mean knock-on delay of a train would be slightly larger than the mean delay of the preceding train when the buffer time between scheduled train paths is around 5 s. This mean delay difference is due to the necessary time for hindered trains to react to a change of the signal aspects, close the train doors and finally depart from the station, or run on the

track sections leading to the platform with slower speed due to forced deceleration and
acceleration in case of inbound route conflicts. The small buffer time of 5 s corresponds
to a capacity utilization rate of 95% and a maximal frequency, i.e., 19 pairs/ h of trains
passing this route node, in which the impact of train stop at the station is not considered.
Obviously, this maximal capacity utilization could not be feasible in real operations be-
cause the train stop for alighting and boarding of passengers at the station must be taken
into account.

For determining the scheduled railway capacity utilization, it is important to predict not
only the impact of the capacity utilization on the mean knock-on delay of each train series,
but also on the distribution of knock-on delays. Since the same number of homogeneous
trains has been assumed to pass the level crossing alternately in both southbound and
northbound directions, the capacity utilization of this node can be represented by the
frequency of the passing trains.

Figure 6.3 shows the estimated knock-on delay survival probability for a northbound de-
parting train as a function of the frequency of trains passing the level crossing. The
survival probability of a knock-on delay is here defined as the probability that a knock-on
delay is larger than this value and this probability equals 1.0 minus the cumulative dis-
tribution probability. Clearly, the knock-on delay survival probability develops with the
increase of the frequency of passing trains at the route node. When the train frequency is
1 pair/ h, the northbound departing train does not suffer any knock-on delay, whereas the
survival probability of a northbound departing train suffering knock-on delay becomes
about 42% when the train frequency increases to 13 pairs/ h.
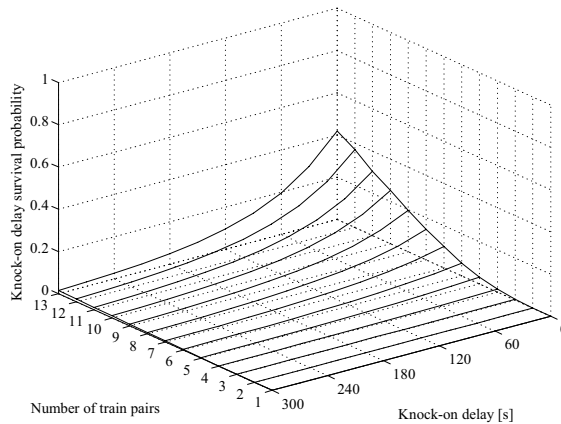
Figure 6.3: Knock-on delay survival probability for a northbound departing train as a function of
the frequency of trains passing level crossing 229BT.

Figure 6.4 shows the estimated knock-on delay distributions corresponding to different
frequencies of trains passing the level crossing. Given a maximally acceptable knock-on
delay, e.g. 60 s at a certain confidence level, e.g. 90%, we can determine the correspond-
ing maximal train frequency of 9 pairs/ h.
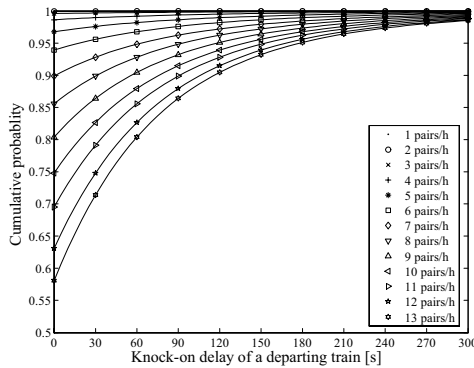
Figure 6.4: Knock-on delay distribution for a northbound departing train as a function of the frequency of trains passing level crossing 229BT.

6.2.2 Influence of input and primary delays

The knock-on delays of trains at a route node are affected considerably by the node capacity utilization. However, if there were no input delays at the boundary of the considered network and no primary delays within the network, no knock-on delays would occur. It is expected that the knock-on delays of a train series at the studied level crossing depend on the input delays and primary delays of the preceding train series that operates in the opposite direction and even of the train series itself. The input delays of trains at the boundary of the local station network consist of the primary delays and knock-on delays that occur in the upstream stations and on open tracks. The primary delays of trains within the local network are due to disturbances that occur during the acceleration and deceleration of trains in the station area and longer than scheduled dwell times at the platform track for alighting and boarding of passengers.

To investigate the impact of the input delays of preceding train series on the knock-on delays of a train series, we consider several scenarios of the mean and standard deviation for the input delays. The capacity utilization rate of the level crossing is estimated to be 65%, which corresponds to 13 pairs of trains per hour passing this route node. When the mean input delay becomes larger, the estimated knock-on delay survival probability overall increases significantly (Figure 6.5). When the standard deviation of the input delays becomes larger, the knock-on delay survival probability decreases at smaller delays, while it increases slightly at larger delays (Figure 6.6).

It is found that the statistics, e.g. mean and 90th percentile of the knock-on train delay increase not only with the mean, but also with the standard deviation of the input delay (see Figures 6.7 and 6.8, respectively). When the mean input delay increases by 30 s compared to the basic scenario of 60 s, the estimated mean and 90th percentile of the knock-on delay increase by 17 s and 36 s, respectively. When the input delay standard deviation increases by 30 s from the basic scenario of 90 s, the mean and 90th percentile of the knock-on delay increases by 7 s and 21 s, respectively. To reduce the knock-on delay of a train that is caused by the preceding one, it is necessary to decrease not only the mean, but also the standard deviation of the input delay.
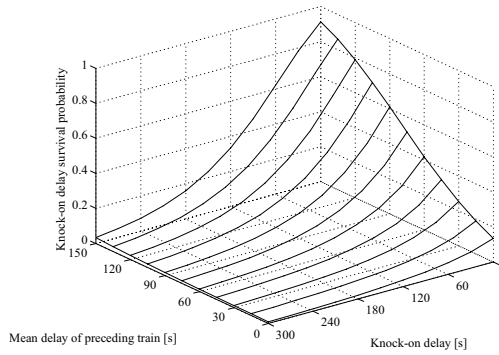
Figure 6.5: Knock-on delay survival probability for a northbound departing train as a function of the mean input delay of the preceding train in southbound direction.
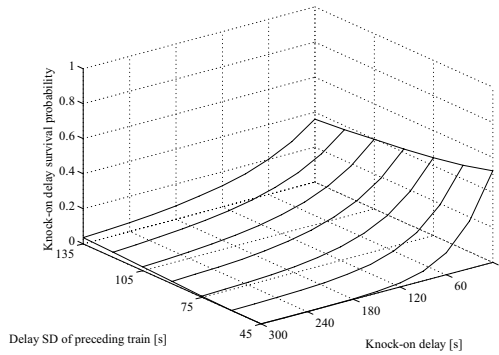
Figure 6.6: Knock-on delay survival probability for a northbound departing train as a function of the input delay standard deviation of the preceding train in southbound direction.
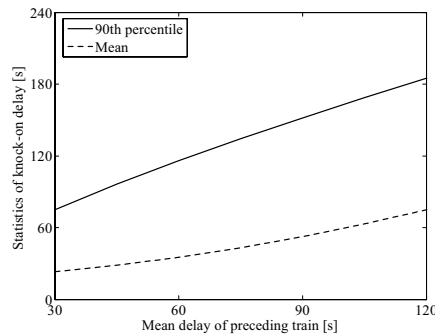
Figure 6.7: Knock-on delay statistics for a northbound departing train as a function of the mean input delay of the preceding train
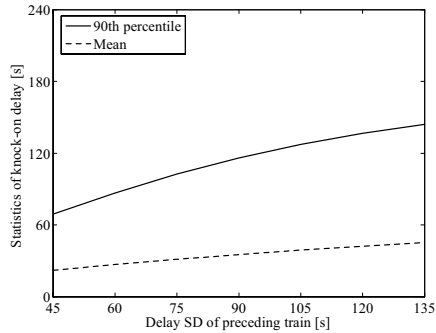
Figure 6.8: Knock-on delay statistics for a northbound departing train as a function of the standard deviation of the input delay of the preceding train

In the above analysis, the dwell times of trains are assumed to be equal to the scheduled ones. In real operations, they vary a lot because of the variation of the number of passengers alighting and boarding and behavior of train crews. The dwell times of trains are often longer in peak hours than in off-peak hours. Late arriving trains may have shorter dwell times than the trains that arrive at the station punctually. We analyze the impact of stochastic dwell times of the northbound departing trains at The Hague HS station on the knock-on delays of the southbound approaching trains at the level crossing. For simplicity, we do not distinguish the free dwell time distributions between early and late arriving trains. The distribution of the free dwell times is considered as a shifted Weibull distribution, according to the statistical modelling results obtained in Chapter 4. The location parameter of the distribution is assumed to be the minimal dwell time, e.g., 60 s. The shape and scale parameters are derived using the moment method (Appendix A.6) on the basis of the assumed mean and standard deviation of the free dwell times.

It is found that the knock-on delay of a southbound approaching train increases not only with the mean, but also with the standard deviation of the free dwell time of the preceding train. Let us consider a basic scenario of the free dwell time distribution with a mean of 120 s and a standard deviation of 30 s. If the mean dwell time increases by 30 s, the mean and 90th percentile of the knock-on delay increase by 16 s and 30 s, respectively, and if the mean dwell time is reduced by 30 s, the knock-on delay statistics decrease by 11 s and 28 s, respectively. Therefore, if the northbound preceding train is largely delayed, the dwell time at the station may be reduced as much as possible in order to decrease the delay propagation to the following train in southbound direction.

In addition, if the dwell time standard deviation increases by 30 s, then the mean and 90th percentile of the knock-on delay increase by 5 s and 25 s, respectively. The knock-on delay statistics decrease 4 s and 9 s, respectively, if the dwell time standard deviation reduces to 0 s, which corresponds to the scheduled dwell time assumed earlier. To reduce the knock-on delay of a southbound approaching train, the variation of the free dwell time of the preceding train at The Hague HS station can be decreased by not waiting for individual passengers who arrive late at the station.

To estimate the knock-on delay of a certain train, the input delay and primary delay of the train itself are not taken into account in the above experiments. However, the statistical

analysis results obtained in Chapter 3 reveal that the knock-on delay of a train at a route node is caused not only by late release of the preceding train, but also by an early arrival of the train itself. Applying the delay propagation model, we study the impact of stochastic arrival of a southbound approaching train at the approach signal of the station on the knock-on delay of the train itself. The distribution model for the arrival time is taken as a shifted log-normal distribution with the location parameter to be the earliest arrival time, which is, for instance, 60 s earlier than the scheduled arrival time. The input delay distribution for the preceding train in northbound direction is here assumed to be the same as in Section 6.2.1.

The impact of the stochastic arrival time of a southbound approaching train at the station approach signal on the knock-on delay is confirmed by the study results as follows. When the mean arrival time is assumed to be equal to the scheduled arrival time but the standard deviation is considered to be 90 s, the estimated mean and 90th percentile of the knock-on delay are 8 s and 17 s, respectively, larger than the estimated results in case of the basic scenario, where the arrival of the studied train is assumed to be 100% punctual. The increase of the knock-on delay statistics when considering the arrival time variation is because an early arrival of the train is taken into account in the case of a high utilization rate, i.e., 65% of the route node capacity.

We also investigate the impact of the mean arrival delay on the knock-on delay by fixing the standard deviation at 90 s. When the mean arrival delay reduces by 30 s, the estimated mean and 90th percentile of the knock-on delay increase by 16 s and 22 s, respectively. In real operations, speed control would reduce the probability, mean and variation of knock-on delays. On the contrary, when the mean arrival delay increases by 30 s, the estimated mean and 90th percentile of the knock-on delay reduce by 12 s and 25 s, respectively. The late arrival of a train reduces the probability and statistics for a knock-on delay of the train itself, but this may lead to a knock-on delay of the following train. In conclusion, deviation of the train operations from scheduled times need to be minimized in order to reduce knock-on delays.

6.2.3   Influence of buffer time assignment

When buffer times between scheduled train paths at a route node decrease, the mean knock-on delay of each passing train increases exponentially. The probability and size of knock-on train delays also depend on the amount of input and primary delays. If the input and primary delays vary between train series operating in a period of the timetable, the buffer time between each pair of scheduled train paths needs to be assigned optimally in order to enable a maximally tolerable knock-on delay at a certain confidence level. To assign buffer times between pairs of trains properly, the train priorities must be respected.

Figure 6.9 shows the 90th percentile of the knock-on delay for a northbound departing train as a function of the buffer time between the release of the level crossing 229BT by the southbound preceding train and the start of the blocking time by the studied train in the case of different mean input delays of the preceding one. Given a maximally acceptable knock-on delay at this confidence level, we are able to determine the minimal buffer times corresponding to the mean input delays. Similarly, we can determine the minimal buffer times between train pairs corresponding to the input or primary delays of trains to assure a
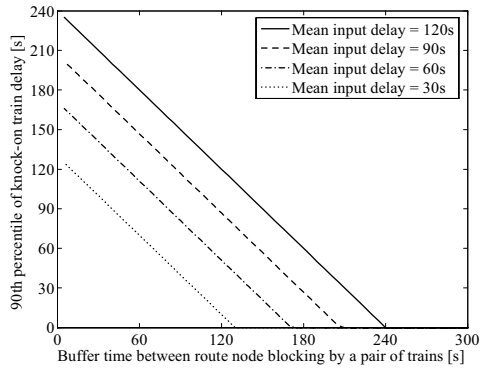
Figure 6.9: 90th percentile of knock-on delay of a northbound departing train as a function of the buffer time from the preceding train with different mean input delays

maximally acceptable knock-on delay at different confidence levels. Of course, in a cyclic timetable, increasing the buffer time between a pair of trains will decrease the buffer time somewhere else. To achieve an optimal buffer time assignment between all train pairs, both simulation and combinatorial optimization approaches may be adopted to minimize the weighted total knock-on delay of all the trains.

6.2.4   Influence of train order alteration

In some cases, the buffer time between train paths at route nodes cannot be assigned adequately because a lot of constraints have to be considered simultaneously for train operations in the whole railway network. Even though a timetable is well designed, knock-on delays may still occur due to route conflicts caused by stochastic train operations. To decrease the probability of knock-on delays and increase the punctuality of trains, traffic controllers may alter the scheduled order of a pair of trains that successively pass a route node or pass a station with passing tracks in case of a larger delay of the preceding train.

To investigate the influence of train order alteration, we take an example of a pair of train series that pass the critical level crossing north of The Hague HS station successively. The leading train comes from Leiden and the following one leads to The Hague CS. The scheduled buffer time between the release of the first scheduled train and the start of the blocking time of the following one at the route node is 48 s. It is assumed that the leading train arrives rather late at the station area whereas the following one is always ready to depart from The Hague HS station on time. Thus, the following train would frequently suffer knock-on delays from the leading one if the scheduled passing order is always applied in real operations. To decrease knock-on train delays, traffic controllers alter the order of this pair of trains passing the route node when the delay of the leading one becomes larger than a critical value.

Assuming that the train order is altered when the southbound leading train is delayed by 180 s, 240 s and 300 s, respectively, we estimate the knock-on delay distributions for the northbound following train in case of different mean input delays. The capacity utilization

rate of the level crossing is still considered to be 65%, corresponding to 13 pairs of trains per hour passing this route node. By considering the alteration of train order, the knock-on delay survival probability overall reduces. For a clear comparison of the impact of train order alteration and fixed order of trains on the estimated knock-on delay survival probability, see Figure 6.10.
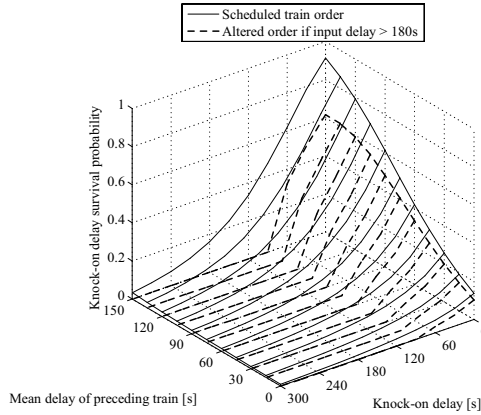
Figure 6.10: Impact of train order alteration on the knock-on delay survival probability for a northbound departing train

The impact of train order alteration on the statistics, e.g. mean and 90th percentile of the knock-on delay of a northbound departing train is shown in Figures 6.11 and 6.12, respectively. Clearly, alteration of the train order in case of a large input delay of the leading train reduces significantly the knock-on delay of the following one. If the train order is altered as long as the input delay is larger than 180 s, the mean knock-on delay of the following train series can be reduced by 69 s to 32 s in case the mean input delay of the leading train series is, e.g. 150 s. The 90th percentile of the knock-on delays can be reduced by 144 s to 73 s in this case.
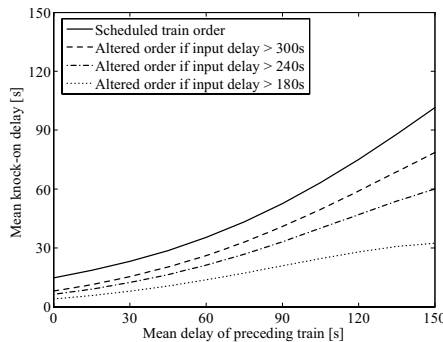
Figure 6.11: Impact of train order alteration on the mean knock-on delay for a northbound departing train
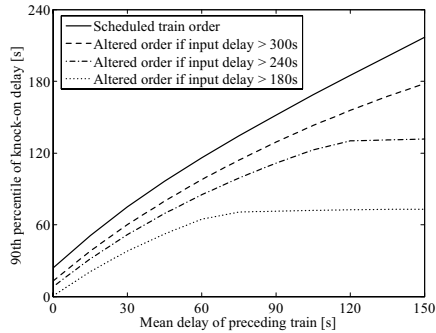
Figure 6.12: Impact of train order alteration on the 90th percentile of knock-on delay for a northbound departing train

Altering the prespecified order of a pair of trains that pass a route node consecutively reduces the knock-on delay of the second scheduled train, but this may result in a knock-on delay of the first scheduled train. To on-line determine the order of a pair of trains passing a route node, the priorities of trains, which are related to the type of the train and the number of passengers on board, must also be considered. Ideally, the weighted total delay of all the involved trains and passengers respectively should be minimized.

6.3   Extension of the model application to the whole station area

The optimization of capacity utilization for a critical route node in the station area has been discussed considering a maximally acceptable knock-on delay. It is well known that the punctuality of train operations is important not only to operators but also to passengers. The punctuality of trains at a station is determined by the input delays at the boundary of the station area and the primary delays as well as knock-on delays within the station area.

The tracks of a station are generally designed to minimize the number of mergings and crossings and maximize the flexibility of operations (Hansen (2000)). However, large stations in Europe typically have multiple intersecting in-routes and out-routes, connected to multiple parallel or sequential platforms, through and possibly stabling tracks of different lengths, some being dead-end and some being one or two-way through-platforms (Carey & Carville (2003)). The trains differ in their types, speeds, scheduled dwell times and headways, origins and destinations, and preferred or required routes and platforms. Moreover, the scheduled train arrivals and departures are often synchronized to facilitate the transfer of passengers between different lines. In addition, busy stations are frequently visited by several hundred, or over a thousand trains per day. As a result, the capacity of platform tracks and route nodes in a station area needs to be optimized in order to achieve a desired level of punctuality for train arrivals and departures.

6.3.1   Maximizing station capacity utilization at a desired punctuality level of train operations

To achieve a desired punctuality level of train operations, we make a distinction between approaching trains and departing trains. In case of a southbound approaching train, the knock-on delays that occur within the station area may be caused by occupancy of the platform track and route nodes on the inbound route. In case of a northbound departing train, knock-on delays are caused by occupancy of route nodes on the outbound route. For easier analysis of the multi-source knock-on delays and resulting punctuality at the platform, we assume that a southbound approaching train may be hindered before arrival at the home signal of the station due to occupancy of the platform track and a level-crossing on the inbound route (Figure 6.13).
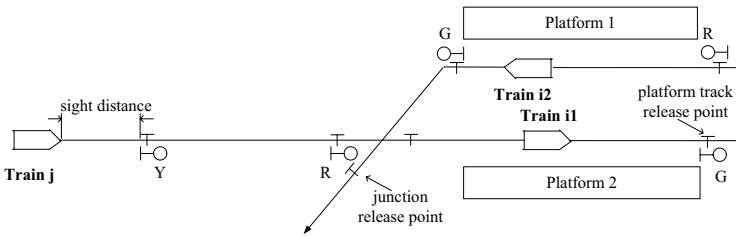
Figure 6.13: Hindrance of an approaching train j due to occupancy of the platform track and a route node on the inbound route

A northbound departing train may be hindered at the platform track due to occupancy of two level-crossings on the outbound route (Figure 6.14). We assume that the same number of trains pass the two route nodes in each direction.
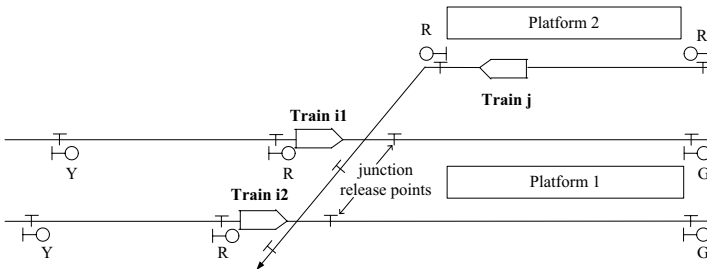
Figure 6.14: Hindrance of a departing train j due to occupancy of two route nodes on the outbound route

Most of the assumptions of the model remain the same as in Section 6.2.1. However, here we take into account the stochastic variation of the running times of hindered trains by a location-shifted Weibull distribution model. The location parameter of the distribution model for the running time of a hindered approaching train on the track sections leading to the platform is assumed to be the scheduled time. In addition, the mean running time of a hindered train is assumed to be the deterministic value used in Section 6.2.1 while

the standard deviation is assumed to be 10% of the mean value. The total knock-on delay of each studied train equals the arrival (departure) delay at the station, because the input delay and the primary delay of the train itself are not considered during the delay estimation.

The results of the model estimation prove that the probability of knock-on delays for a southbound approaching train due to the occupancy of the level crossing is higher compared to the occupancy of the platform track if the frequency of passing trains at the level crossing is less than 10 pairs/ h. As long as the train frequency is higher than this critical value, the knock-on delay caused by occupancy of the platform track becomes dominant. The arrival punctuality of an approaching train at a certain delay is lower than the cumulative probability of the same amount of knock-on delay generated at either the level crossing or at the platform track because of a joint effect of multiple stochastic knock-on delays (see Figure 6.15). The shape of the probability curves given in this figure are related to the conditional modelling of delay propagation and the assumed conditional distributions of the running time in case of different aspects of block signals. Knock-on delays less than 30 s hardly occur, which is due to the assumed average time difference of 45 s between free running and hindered running but without stop as well as an assumed small variation of the running times of hindered trains.
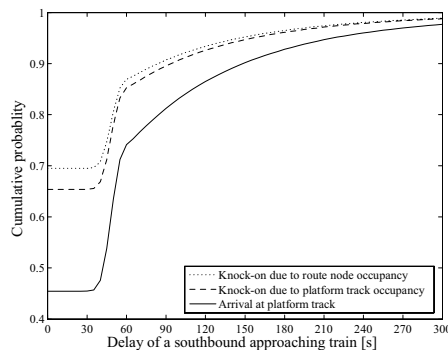
Figure 6.15: Cumulative probability of knock-on delays and arrival punctuality at the platform for a southbound train in case of a frequency of 11 pairs/h of trains passing level crossing 229BT

We also draw in Figure 6.16 the punctuality curves of the arrival time of a southbound approaching train at the platform track as a function of the frequency of trains passing the level crossing in both southbound and northbound directions. Given a desired arrival punctuality level, we can determine the maximally acceptable frequency of trains passing the level crossing. Of course, only 50% of the trains will arrive at the station after passing the level crossing, while the other 50% are departing from the station.

It is found that the probability of knock-on delays for a northbound departing train is higher than that for a southbound approaching train when the frequency of passing trains at the level crossings is lower than 10 pairs/ h. In this case, the buffer times between scheduled train paths at the route nodes are smaller than those at the platform tracks. Once the train frequency is higher than 10 pairs/ h, the probability of knock-on delays for a southbound approaching train becomes bigger because of the dominating bottleneck of the platform track.
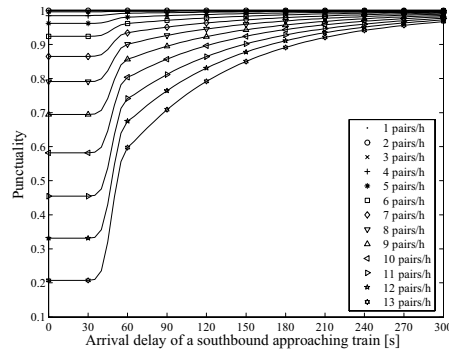
Figure 6.16: Arrival punctuality of a southbound train at the platform as a function of the frequency of trains passing the level crossing in the north of The Hague HS

In the above analysis, it has been assumed that the same number of trains pass each route node in both directions alternately. In practice, the frequency of trains passing route nodes are often different and the number of trains may also be different for each direction. The knock-on delays and the resulting punctuality of train arrivals and departures do not only depend on the capacity utilization rate of the route nodes, but also on the order of the trains and directly on the buffer time between train paths. The dependence of knock-on delays and punctuality of train operations on input delays and primary delays is obvious. In case of complicated operations within a complex station network, the requested train paths in an operating period could be scheduled sequentially by assigning optimal buffer times corresponding to the expected mean input delays and primary delays in order to assure the desired punctuality level.

6.3.2   Influence of train running and dwell time supplements

The punctuality of train operations is affected significantly by the capacity utilization of route nodes and platform tracks. Increasing the buffer time between train paths can reduce the knock-on delay of the following train, but this may not always be possible due to operational constraints in the whole network. Altering the scheduled order of trains can reduce knock-on delays, too. However, this solution is normally adopted in real operations when input delays and primary delays exceed a certain threshold. The running and dwell time supplements incorporated in timetables can reduce primary delays within the station area and compensate for input delays imported from upstream stations and links, as well as for knock-on delays caused locally by other trains.

We have analyzed the impact of running and dwell time supplements on the arrival and departure punctuality for the southbound approaching and northbound departing trains, respectively. It is assumed that the delays of these trains are caused exclusively by the hindrance from the preceding train at a critical level-crossing. Figure 6.17 shows a degressive increase of the departure punctuality for a northbound train when the dwell time supplement becomes larger. In this case, the preceding train is assumed to have a mean delay of 60 s when releasing the conflicting level crossing. If the scheduled dwell time

at the station is 120 s and the buffer time between the scheduled train paths is 60 s, the departure punctuality of the northbound train at a delay of 60 s is estimated to be 83%. If the desired departure punctuality at this delay is to be more than 90%, an extra dwell time supplement of 60 s would be necessary. If this is applied for each northbound train, the maximal capacity of the platform track for the northbound trains would drop from 13 to 11 trains/ h.
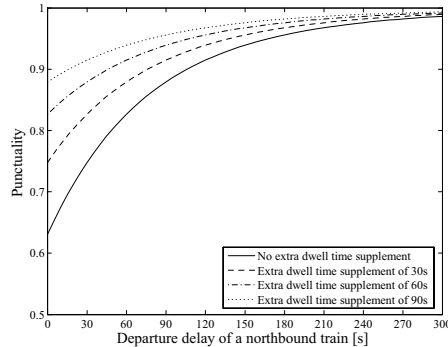
Figure 6.17: Impact of dwell time supplement on departure punctuality of a northbound train

In case of southbound trains, the arrival punctuality can be improved by increasing the running time supplement just before the arrival at the platform track. This would lead, however, to a drop of the maximal capacity of the platform track for southbound trains. Clearly, incorporating bigger running and dwell time supplements in station areas would result in longer planned travel times of passengers, too. But, this might be compensated by smaller running time supplements at open tracks. Using the stochastic delay propagation model, we can achieve an optimal assignment of the supplements for train running and dwell times corresponding to a certain level of punctuality.

6.4   Conclusions

The developed delay propagation model has been applied to optimize the station capacity utilization in a case study of The Hague HS station, characterized by different level crossings and mergings/divergings of railway lines. The mean knock-on delay of passing trains at a critical track section increases exponentially with a decrease of the scheduled buffer time between train paths. The model enables the determination of the maximal frequency of trains passing the critical track sections within the station areas for a maximally acceptable level of knock-on delays at a certain confidence level and a desired punctuality level. The distribution of knock-on delays for a train does not only depend on the mean and standard deviation of the input delay and primary delay of the preceding train, but also on those of the train itself. Given these delay statistics, the running and dwell time supplements and the buffer time between scheduled train paths can be minimized by means of the delay propagation model while assuring a maximally acceptable level of knock-on delay and a desired punctuality level. Although the model application has been

demonstrated only for a single railway station, it can be implemented similarly for other stations and interlocking areas with different track layouts. To estimate knock-on delays including the impact on the punctuality of trains successfully, it is required to carry out a systematic analysis of the interdependences between different train lines in advance.

Chapter 7

Conclusions

In this concluding chapter, we summarize the performed research and draw the main conclusions. Furthermore, some directions for follow up research are recommended.

7.1   Main results

Based on a detailed analysis of real-world train traffic data recorded at the Dutch railway station The Hague HS, this thesis deals mainly with statistical analysis of train delays and probability modelling of delay propagation in a complicated station and interlocking area. The developed models have been applied for maximizing station capacity utilization and improving timetable design while assuring a desired level of reliability and punctuality of train operations.

The ex-post analysis of train traffic data reveals a considerable evolution of train delays from arrival to departure at the case station The Hague HS. This analysis confirms the importance of analyzing the delays smaller than the current threshold value, i.e., 3 min, determined by the Dutch Railways for defining the punctuality of train operations. A more detailed analysis for two selected train series shows that the delays of trains on Sunday differ significantly from Monday till Friday because of less passengers alighting and boarding trains and lower capacity utilization of the railway infrastructure on Sunday. In addition, the delays of commuter trains depend a lot on the period of the day, confirming the research results obtained in previous studies.

In this thesis, the data analysis is not only performed for train delays themselves, but also for the propagation of delays between train pairs and the real track occupation of the critical sections in the station and adjacent interlocking area. It is clear that the headway and route conflicts between a pair of train series result in significant knock-on delays of the following one in case of insufficient buffer time between the scheduled train paths. However, the arrival punctuality of approaching trains degrades less than anticipated, because the knock-on delays at a route node are not only caused by a late release by a preceding train series, but also by an early arrival of the following one itself. In addition, running time supplements absorb, too, part of the knock-on train delays. To reduce the knock-on delays of approaching trains caused by route conflicts, it is required to decrease the

input and primary delays of preceding trains, as well as to avoid early arrivals of the approaching trains by means of dynamic speed control. The punctuality of train arrivals and departures at a station is determined predominantly by knock-on delays resulting from route conflicts in busy interlocking areas and input delays generated at upstream stations and open tracks.

It has been found that the real capacity utilization of most track sections in The Hague HS station is higher than scheduled. This is because of longer than scheduled train process times due to running and dwell perturbations as well as hindrance from preceding trains. The real capacity utilization of one platform track and its adjacent route nodes is surprisingly lower than scheduled. This is due to rather frequent route conflicts at these track sections and the following reduction of the approaching time after the release of the critical track sections. One of the most remarkable findings is that the blocking time of the signal block preceding a level crossing near the station by frequently hindered international trains is even longer than that of the station block itself. In case of hinder, the bottleneck on the operating route is shifted from the platform track to the preceding signal block just before the interlocking area in the inbound direction. This phenomenon corresponds to the well-known spill-back of vehicle queues on congested road sections.

We have evaluated several distribution models that are commonly applied in the literature for train event and process times on the basis of the empirical traffic data recorded at The Hague HS station. Before the distribution evaluation, a new engineering method for fine-tuning the parameters of distribution models was presented. It has been found that a location-shifted log-normal distribution can be considered as the best model among the candidate distributions for both the arrival times of trains at the platform track and at the approach signal of the station home signal. The distribution shift parameter corresponds to the earliest arrival time. The Weibull distribution is generally the best model among the candidate distributions for non-negative arrival delays, departure delays and the free dwell times of trains in the absence of hindrance from other trains. Obviously, the distribution location parameter is zero in case of both non-negative arrival delays and departure delays. For the free dwell times of trains, the location (shift) parameter of the Weibull distribution fit corresponds to the minimum free dwell time. The density curve of a Weibull fit to non-negative arrival delays is decreasing except for some special cases, where structural delays exist. This implies that the shape parameter of a Weibull fit to non-negative arrival delays generally cannot be larger than 1.0. It has also been found that the shape parameter of a Weibull fit to the free dwell times of trains is mostly beyond 1.0. The aforementioned distribution evaluation results can be adopted for estimating knock-on delays and the punctuality of trains at typical railway through stations with multiple platform tracks. In particular, if empirical operations data is lacking and the input delay and primary delay distributions can be reasonably estimated based on a certain theoretical distribution model with the mean and standard deviation of the delays assumed by experience.

The key contribution of this thesis is our innovative analytical probability model that estimates the knock-on delays of trains caused by route conflicts and late transfer connections including the impact on the punctuality of train operations. The model takes into account the interdependences of arrival and departure times of different train lines in a complicated station and interlocking area. We reflect the stochastic variations of track occupation times due to the fluctuation of the train speed in case of different signal as-

pects through conditional probability distributions. In addition, the possible dependence of the free dwell times of trains on the corresponding arrival delays is modelled, too, by conditional distributions. Furthermore, the impact of dynamic delay propagation between a chain of trains and of the alteration of train orders in real operations is incorporated. Consequently, the model enables an accurate estimation of the knock-on delays within a densely occupied station track network and the resulting punctuality of the train arrivals and departures at the station. Using empirical train detection data, we have validated the delay propagation model by means of a case study of The Hague HS station, which is characterized by different level crossings and mergings/divergings of railway lines. The validation results reveal that the model can estimate the knock-on delays of trains and resulting arrival and departure distributions very well, provided the distributions of the involved event and process times are known.

The delay propagation model has been applied for maximizing the capacity utilization of The Hague HS station and improving the allocation of buffer times between scheduled train paths at a critical track section. Assuming that a number of trains with the same dynamic characteristics pass a route node in two different directions alternately, we found that the mean knock-on delay of each passing train at the route node increases exponentially with a decrease of the buffer time between each pair of scheduled train paths. The delay propagation model enables the determination of the maximal frequency of trains passing critical track sections within the station area through a maximally acceptable knock-on delay at a certain confidence level and a desired punctuality level. These measures of level-of-service are much more straightforward than the estimated 'quality of operations' expressed by the total waiting time calculated by queueing models, as applied in practice at German Railways. We also found that the knock-on delay distribution for a train does not only depend on the mean input delay and mean primary delay of the preceding train and of the train itself, but also on the standard deviations. Therefore, in case the mean primary delay of a scheduled service cannot be reduced, a decrease of the variation by process management measures can limit the propagation of delays. Given input and primary delay distributions, the running and dwell time supplements and the buffer times between scheduled train paths can be minimized at a given maximally acceptable knock-on delay and a desired punctuality level by means of the delay propagation model.

Incorporating the impact of altering the scheduled order of trains in real operations, the delay propagation model can be used to improve the decision-making of rescheduling in case of operational disturbances. This model is solved on the basis of a numerical approximation of the convolution of the individual distributions and thus can be integrated into a larger computerized tool for timetable design and traffic management.

7.2   Recommendations

A detailed statistical analysis of train detection data has been carried out, which enables a deeper understanding of stochastic train operations including the identification of train event and process time distributions. This analysis is based on one-month track occupancy and release data recorded in the station of The Hague HS in September 1999. However, the distribution of train delays, especially the distribution parameters, may differ between stations depending on the travel distance from the origin of the trip, the timetable as well

as the track layout and signalling of the station. The distribution of train delays including the parameters may vary not only in space, but also in time. Furthermore, the behavior of train crews and traffic controllers may change depending on professional experience and process management rules implemented by the railway infrastructure manager and train operators in order to achieve a higher punctuality level of train operations. This may lead to a quite different quality of real-world train operations. Therefore, it is recommended that a detailed statistical analysis of railway infrastructure utilization and train operations at critical bottlenecks and links of railway networks is performed regularly.

Our delay propagation model estimates the knock-on delays caused by route conflicts and late transfer connections including the impact on the punctuality of trains more realistically than the existing models. This model has been validated by means of a case study for The Hague HS station, where synchronized transfer connections between different lines are rare. In addition, our model validation did not incorporate reordered train pairs and neglected the impact of the dynamic delay propagation between a chain of several trains. Therefore, it is still necessary to further validate the model by taking into account these aspects and for more stations.

The developed delay propagation model has been applied for maximizing the station capacity utilization and improving the timetable design while assuring a desired reliability and punctuality of train operations. To achieve an optimal buffer time assignment between all the train pairs in a cyclic timetable, heuristic, combinatorial optimization and simulation approaches could be adopted to minimize the weighted total knock-on delay of all the trains. The model can be further extended for improving the rescheduling measures in case of operational perturbations. Because on-line decision-making and route setting always need some time, the knock-on delays of trains at critical route nodes and the resulting punctuality of trains at stations need to be predicted on the basis of the latest input delays recorded some distance upstream the route nodes. The delay propagation model for stations would need to be extended to a network scale to maximize the network capacity utilization and improve the timetabling and rescheduling while assuring a desired reliability and punctuality of train operations in a larger railway network including successive stations and links.

To perform a continuously updated statistical analysis of train operations and further improve the capacity utilization, timetabling and rescheduling in a larger railway network, a generic computer-aided analysis, predicting and planning tool is needed. The tailor-made programs used for our statistical analysis of train operations and probability modelling of delay propagation would need to be extended and eventually combined with existing tools, i.e., TNV-Prepare and TNV-Filter. This could lead to a generic decision management information and support tool that enables dynamic reports and insight into the performance of train operations, infrastructure capacity utilization, timetabling quality and rescheduling effectiveness for larger railway networks.

Bibliography

Bailey, C., (ed.) (1995), *European Railway Signalling*, Institution of Railway Signal Engineers, A & C Black, London.

Barter, W. (2004), Forecasting robustness of timetables, *In:* J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, (eds.), *Computers in Railways IX*, WIT Press, Southampton, 563–572.

Boehme, T.K., W. Preuss & V. Van der Wall (1991), On a simple numerical method for computing Stieltjes integrals in reliability theory, *Probability in the Engineering and Informational Sciences* **5**, 113–128.

Bowman, A.W. & A. Azzalini (1997), *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, New York.

Bruinsma, F.R., P. Rietveld & D.J. Van Vuuren (1999), Unreliability in public transport chains, *In:* H. Meerman, E. Van de Voorde & W. Winkelmans, (eds.), *World Transport Research. Vol. 1: Transport Modes and Systems.*, Pergamon, Amsterdam, the Netherlands, 359–372.

Burdett, R.L. & E. Kozan (2006), Techniques for absolute capacity determination in railways, *Transportation Research Part B* **40**, 616–632.

Bury, K. (1999), *Statistical Distributions in Engineering*, Cambridge University Press, Cambridge.

Carey, M. (1994*a*), A model and strategy for train pathing with choice of lines, platforms, and routes, *Transportation Research B* **28**(6), 333–353.

Carey, M. (1994*b*), Reliability of interconnected scheduled services, *European Journal of Operational Research* **79**, 51–72.

Carey, M. (1998), Optimizing scheduled times, allowing for behavioural response, *Transportation Research Part B* **32**(5), 329–342.

Carey, M. (1999), Ex ante heuristic measures of schedule reliability, *Transportation Research Part B* **33**, 473–494.

Carey, M. & A. Kwiecinski (1994*a*), Stochastic approximation to the effects of headways on knock-on delays of trains, *Transportation Research Part B* **28**(4), 251–267.

Carey, M. & A. Kwiecinski (1994*b*), Swapping the order of scheduled services to minimize expected costs of delays, *Transportation Research Part B* **28**(6), 409–428.

Carey, M. & A. Kwiecinski (1995), Properties of expected costs and performance measures in stochastic models of scheduled transport, *European Journal of Operational Research* **83**, 182–199.

Carey, M. & I. Crawford (2005), Scheduling trains on a network of busy complex stations, *In:* I.A. Hansen, F.M.Dekking, R.M.P. Goverde, B. Heidergott & L.E.Meester, (eds.), *Proceedings of 1st International Seminar on Railway Operations Modelling and Analysis*, Delft, The Netherlands.

Carey, M. & S. Carville (2000), Testing schedule performance and reliability for train stations, *Journal of the Operational Research Society* **51**, 666–682.

Carey, M. & S. Carville (2003), Scheduling and platforming trains at busy complex stations, *Transportation Research Part A* **37**, 195–224.

Chen, B. & P.T. Harker (1990), Two moments estimation of the delay on single-track rail lines with scheduled traffic, *Tranportation Science* **24**(4), 261–275.

D'Ariano, A. & M. Pranzo (2004), A real-time train dispatching system based on the blocking time theory, *In:* P.H.L. Bovy, (ed.), *A World of Transport, Infrastructure and Logistics, Proceedings of 8th International TRAIL Congress 2004*, DUP Science, Delft, the Netherlands, 129–152.

Dekking, F.M., C. Kraaikamp, H.P. Lopuhaä & L.E. Meester (2005), *A Modern Introduction to Probability and Statistics*, Springer, London.

Ferreira, L. & A. Higgins (1996), Modelling reliability of train arrival times, *Journal of Transportation Engineering* **122**(6), 414–420.

Gibbons, J. D. (1985), *Nonparametric Methods for Quantitative Analysis*, American Sciences Press, INC., Columbus, Ohio.

Gibson, S., G. Cooper & B. Ball (2002), The evolution of capacity charges on the UK rail network, *Journal of Tranport Economics and Policy* **36**, 341–354.

Goverde, R.M.P. (1998), Synchronization control of scheduled train services to minimize passenger waiting times, *In:* P.H.L. Bovy, (ed.), *Competition, Innovation and Creativity, Proceedings of 4th TRAIL Annual Congress*, Delft University Press, Delft, the Netherlands, 1–34.

Goverde, R.M.P. (1999), Improving punctuality and transfer reliability by railway timetable optimization, *In:* P.H.L. Bovy, (ed.), *Proceedings of the 5th TRAIL Annual Congress*, TRAIL Research School, Delft, 1–30.

Goverde, R.M.P. (2000), Delay estimation and filtering of train detection data, *In:* P.H.L. Bovy, (ed.), *Proceedings of TRAIL 6th Annual Congress*, Delft University Press, Delft.

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

Goverde, R.M.P., G. Hooghiemstra & H. P. Lopuhaä (2001), *Statistical Analysis of Train Traffic*, DUP Science, Delft, the Netherlands.

Goverde, R.M.P. & G.S. Koelemeijer (2000), *Performance Evaluation of Periodic Railway Timetables: Theory and Algorithms*, Delft University Press, Delft, The Netherlands.

Goverde, R.M.P. & I.A. Hansen (2000), TNV-prepare: Analysis of dutch railway operations based on train detection data, *In:* J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, (eds.), *Computers in Railways VII*, WIT Press, Southampton, 777–788.

Goverde, R.M.P. & M.A. Odijk (2002), Performance evaluation of network timetables using PETER, *In:* J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, (eds.), *Compuers in Railways VIII*, WIT Press, Southampton, 731–740.

Hájek, J. & Z. Šidák (1967), *Theory of Rank Tests*, Academic Press Inc., New York.

Hansen, I. A. (2000), Station capacity and stability of train operations, *In:* J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S Sone, (eds.), *Computers in Railways VII*, WIT Press, Southampton, 809–816.

Hansen, I. A. (2001), Improving railway punctuality by automatic piloting, *In: 2001 IEEE Intelligent Transportation Systems Conference Proceedings*, Oakland (CA), USA, 792–797.

Hansen, I. A. (2004), Increase of capacity through optimised timetabling, *In:* J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, (eds.), *Computers in Railways IX*, WIT Press, Southampton, 529–538.

Hansen, I.A., A.F. De Kort & P.B.L. Wiggenraad (1999), *Transport Operation and Management*, Vol. A: Public Transport, Delft University of Technology.

Harris, N.G. (1992), *Planning Passenger Railways*, Transport Publishing Company, Derbyshire, chapter Punctuality and Performance, 130–142.

Hermann, U. (1996), *Untersuchung zur Verspätungsentwicklung von Fernreisezügen auf der Datengrundlage der Rechnerunterstützten Zugüberwachung Frankfurt am Main*, PhD thesis, Vom Fachbereich Bauingenieurwesen der Technischen Hochschule Darmstadt.

Hertel, G. (1992), Die maximale Verkehrsleistung und die minimale Fahrplanempfindlichkeit auf Eisenbahnstrecken, *Eisenbahntechnische Rundschau ETR* **41, H. 10**, 665–672.

Higgins, A. & E. Kozan (1998), Modeling train delays in urban networks, *Transportation Science* **32**(4), 346–357.

Ho, T.K, J.P. Norton & C.J. Goodman (1997), Optimal traffic control at railway junctions, *IEE Proc.-Electr. Power Appl.* **144**(2), 140–148.

Höllmüller, J. & V. Klahn (2005), Implementation of the UIC 406 capacity calculation at austrian railways (ÖBB), *In:* I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Heidergott & L.E. Meester, (eds.), *Proceedings of 1st International Seminar on Railway Operations Modelling and Analysis*, Delft, The Netherlands.

Hooghiemstra, J.S., L.G. Kroon, M.A. Odijk, M. Salomon & P.J. Zwaneveld (1999), Decision support systems support the search for win-win solutions in railway network design, *Interfaces* **29**(2), 15–32.

Huisman, T. & R.J. Boucherie (2001), Running times on railway sections with heterogeneous train traffic, *Transportation Research Part B* **35**, 271–292.

Huisman, T., R.J. Boucherie & N.M. van Dijk (2002), A solvable queueing network model for railway networks and its validation and applications for the netherlands, *European Journal of Operational Research* **142**, 30–51.

ISO (1994), *ISO 8402 - Quality Management and Quality Assurance - Vocabulary*, International Standards Organization.

Jacobs, J. (2004), Reducing delays by means of computer-aided 'on-the-spot' rescheduling, *In:* J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, (eds.), *Computers in Railways IX*, WIT Press, Southampton, 603–612.

Kaminsky, R. (2001), *Pufferzeiten in Netzen des spurgeführten Verkehrs in Abhängigkeit von Zugfolge und Infrastruktur*, PhD thesis, Wissenschaftl. Arbeiten Institut für Verkehrswesen, Eisenbahnbau und -betrieb Universität Hannover.

Kroon, L.G., H.E. Romeijn & P.J. Zwaneveld (1997), Routing trains through railway stations: complexity issues, *EJOR* **98**, 485–498.

Larose, D. T. (2005), *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Law, A. M. & W. D. Kelton (2000), *Simulation Modeling and Analysis*, McGraw-Hill Higher Education.

Lin, T.M. & H.M. Wilson (1992), Dwell time relationships for light rail systems, *Transportation Research Record* **1361**, 287–295.

Makkinga, F. & L. Zigterman (2002), Network control for improved performance- a new method for on-line scheduling and dispatching, *In:* J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, (eds.), *Computers in Railways VIII*, WIT Press, Southampton, 943–952.

MathSoft (1999), *S-Plus 2000, User's Guide*, Seattle, USA.

Middelkoop, D. & M. Bouwman (2001), SIMONE: Large scale train network simulations, *In:* B.A. Peters, J.S. Smith, D.J. Medeiros & M.W. Rohrer, (eds.), *Proceedings of the 2001 Winter Simulation Conference*, 1042–1047.

Mühlhans, E. (1990), Berechnung der Verspätungsentwicklung bei Zugfahrten, *Eisenbahntechnische Rundschau ETR* **39, H. 7/8**, 465–468.

Nash, A. & D. Huerlimann (2004), Railroad simulation using OpenTrack, *In:* J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, (eds.), *Computers in Railways IX*, WIT Press, Southampton, 45–54.

Nie, L. & I.A. Hansen (2005), System analysis of train operations and track occupancy at railway stations, *European Journal of Transport and Infrastructure Research* **5(3)**, 31–54.

NSR (1999), *Spoorboekje, 30 mei 1999 t/m 27 mei 2000*, NS Reizigers, Utrecht, Netherlands.

Olsson, N.O.E. & H. Haugland (2004), Influencing factors on train punctuality–results from some norwegian studies, *Transport Policy* **11**, 387–397.

Pachl, J. (2002), *Railway Operation and Control*, VTD Rail Publishing, Mountlake Terrace, USA.

Petersen, E.R. (1974), Over-the-road transit time for a single track railway, *Transportation Science* **8**(1), 65–74.

Radtke, A. & D. Hauptmann (2004), Automated planning of timetables in large railway networks using a microscopic data basis and railway simulation techniques, *In:* J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, (eds.), *Computers in Railways IX*, WIT Press, Southampton, 615–625.

RMcon (2004), *RailSys 3.0 Timetable and Infrastructure Management*, RMcon.

Ross, S.M. (2004), *Introduction to Probability and Statistics for Engineers and Scientists*, Elsevier Academic Press.

Rudolph, R. (2003), Allowances and margins in railway scheduling, *In: From birth to renaissance, Proceedings of World Congress on Railway Research 2003 (CD-ROM)*, 230–238.

Schrijver, A. & A. Steenbeek (1994), *Dienstregelingontwikkeling Voor Railned: Rapport CADANS 1.0*, CWI, Amsterdam.

Schwanhäußer, W. (1974), *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*, PhD thesis, Veröffentlichungen des verkehrswissenschaflichen Instituts RWTH Aachen.

Schwanhäußer, W. (1994), The status of german railway operations management in research and practice, *Transportation Research Part A* **28**(6), 495–500.

Serafini, P. & W. Ukovich (1989), A mathematical model for periodic scheduling problems, *SIAM Journal on Discrete Matheematics* **2(4)**, 550–581.

Smith, D.J. (1988), *Reliability and Maintainability in Perspective: Practical, Contractual, Commericial and Softwaree Aspects*, Macmillan Education, London.

Steckel, J. (1991), *Strategische Optionen für die zufällige Fahrzeit im Eisenbahnbetrieb*, PhD thesis, Hochschule für Verkehrswesen "Friedrich List" Dresden.

Tortorella, M. (1990), Closed Newton-Cotes quadrature rules for Stieltjes integrals and numerical convolution of life distributions, *SIAM J. Sci. Stat. Comput.* **11**, 732–748.

UIC (2004), *UIC Code 406, Capacity*, Union Internationale des Chemins de Fer (International Union of Railways).

Vromans, M. (2005), *Reliability of Railway Systems*, PhD thesis, Erasmus University Rotterdam.

Wahlborg, M. (2005), Application of the UIC 406 capacity leaflet at banverket, *In:* I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Heidergott & L.E. Meester, (eds.), *Proceedings of 1st International Seminar on Railway Operations Modelling and Analysis*, Delft, The Netherlands.

Wakob, H. (1985), *Ableitung eines generellen Wartemodells zur Ermittlung der planmässigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte der Leistungsfähigkeit und Anlagenbelastung*, PhD thesis, Veröffentlichungen des Verkehrswissenschaftlichen Institutes der RWTH Aachen.

Weigand, W. (1981), Verspätungsübertragungen in Fernverkehrsnetzen, *Eisenbahntechnische Rundschau ETR* **30**(12), 915–920.

Wendler, E. (1999), *Analytische Berechnung der planmässigen Wartezeiten bei asynchroner Fahrplankonstruktion*, PhD thesis, Veröffentl. verkehrswissenschaftl. Institut RWTH Aachen.

Wendler, E. & M. Naehrig (2004), Statistische Auswertung von Verspätungsdaten, *Eisenbahningenieurkalender EIK 2004* 321–331.

Wiggenraad, P.B.L. (2001), Alighting and boarding times of passengers at Dutch railway stations, *In:* I.A. Hansen, (ed.), *Train Delays at Stations and Network Stability, Proceedings of the TRAILWorkshop*.

Yuan, J. (2002), Train dwell times in stations, *In:* P.H.L. Bovy, (ed.), *TRAILblazing Into the Future! Proceedings of 7th TRAILCongress (CD-ROM)*, Delft University Press, Delft, the Netherlands.

Yuan, J. & I.A. Hansen (2006), Optimizing capacity utilization of stations by estimating knock-on train delays, *Transportation Research Part B* . In press.

Yuan, J., R.M.P. Goverde & I.A. Hansen (2002), Propagation of train delays in stations, *In:* J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, (eds.), *Computers in Railways VIII*, WIT Press, Southampton, 975–984.

Zwaneveld, P.J., L.G. Kroon, H.E. Romeijn, M. Salomon, S. Dauzère-Pérès, S.P.M. Van Hoesel & H.W. Ambergen (1996), Routing trains through railway stations: Model formulation and algorithms, *Transportation Science* **30**(3), 181–194.

Zwaneveld, P.J., L.G. Kroon & S.P.M. Van Hoesel (2001), Routing trains through a railway station based on a node packing model, *European Journal of Operational Research* **128**, 14–33.

Appendix A

Fundamentals of statistics

This appendix outlines fundamentals of mathematical statistics applied in this thesis.

A.1 The Wilcoxon rank sum test

The Wilcoxon rank sum test is a nonparametric test that is usually adopted to test whether the median differs between two sets of independent observations. Let $x_{1,1}, ..., x_{1,m}$ denote a sample data from a random variable X_1 and $x_{2,1},...,x_{2,n}$ be a sample data from another random variable X_2. The corresponding distribution functions are represented by $F_{X_1}(x)$ and $F_{X_2}(x)$, respectively. Then the null hypothesis of the Wilcoxon rank sum test is that these distributions are the same; that is

$$F_{X_1}(x) = F_{X_2}(x), \text{ for all } x. \tag{A.1}$$

Let M_{X_1} denote the median of X_1 and M_{X_2} the median of X_2, and assume that $F_{X_1}(x)$ and $F_{X_2}(x)$ are identical in form and shape, i.e.,

$$F_{X_2}(x) = F_{X_1}(x - c), \text{ for all } x, \tag{A.2}$$

where c is a constant representing the amount of the shift. Then the hypothesis set of the Wilcoxon rank sum test can be written as

$$\begin{aligned} H_0: & \quad M_{X_1} = M_{X_2}, \\ H_A: & \quad M_{X_1} \neq M_{X_2}. \end{aligned} \tag{A.3}$$

The Wilcoxon rank sum test calls initially for ranking, or ordering, the $m + n$ data values $x_{1,1}, ..., x_{1,m}, x_{2,1}, ..., x_{2,n}$. Give the smallest data value rank 1, the second smallest rank 2,..., and the $(m + n)$th smallest rank $m + n$.

For $i = 1, ..., m$, let

$$R_{1,i} = \text{rank of the data value } x_{1,i}. \tag{A.4}$$

The Wilcoxon rank sum test uses the test statistic

$$T = \sum_{i=1}^{m} R_{1,i}.$$

(A.5)

If the observed value of T is t, then H_0 should be rejected if either

$$\Pr(T \leq t | H_0) \leq \tfrac{\alpha}{2} \quad \text{or} \quad \Pr(T \geq t | H_0) \leq \tfrac{\alpha}{2}$$

(A.6)

where α denotes the significance level of the statistical test. The statistical package S-Plus (MathSoft (1999)) approximates the p-value of the Wilcoxon rank sum test. A more detailed description with respect to this test can be found in Hájek & Šidák (1967), Gibbons (1985), Ross (2004).

A.2 Kernel density estimates

Given a dataset sample of a random population, we can estimate the density function by using kernel smoothing methods. The basic ideas of kernel density estimation first appeared in the early 1950s, but this method has been become popular only since advanced computers with a high computational speed have been developed recently.

The idea behind the construction of a kernel density estimate of a dateset is to 'put a pile of sand' around each element of the dataset (Dekking et al. (2005)). At places where the elements accumulate, the sand will pile up. The kernel density estimate is constructed by choosing a *kernel K* and a *bandwidth h*. The kernel K reflects the shape of the piles of sand, whereas the bandwidth is a tuning parameter that determines how wide the piles of sand will be. Formally, a *kernel K* is a function $K\colon \mathbb{R} \to \mathbb{R}$. There are several well-known kernels, such as normal kernel, triangular kernel, and cosine kernel. A *kernel K* typically satisfies the following conditions:

1. K is a probability density, i.e., $K(u) \geq 0$ and $\int_{-\infty}^{+\infty} K(u)du = 1$;

2. K is symmetric around zero, i.e., $K(u) = K(-u)$;

3. $K(u) = 0$ for $|u| > 1$.

Let $f_{n,h}(t)$ denote a kernel density estimate of a dataset $x_1, x_2,..., x_n$. We have

$$f_{n,h}(t) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{t - x_i}{h})$$

(A.7)

Note that as a consequence of the first condition, $f_{n,h}(t)$ itself is a probability density:

$$f_{n,h}(t) \geq 0 \text{ and } \int\limits_{-\infty}^{+\infty} f_{n,h}(t)dt = 1$$

(A.8)

To construct a kernel density estimate, one has to choose a kernel K and a bandwidth h. The choice of the bandwidth h determines largely what the resulting kernel density estimate will look like. In practice one could do this by trial and error and continue until one obtains a reasonable picture. The choice of kernel is less important and a normal kernel is often applied. For a more detailed description of kernel density estimates, see Bowman & Azzalini (1997), Dekking et al. (2005).

A.3   The Kolmogorov-Smirnov goodness-of-fit test

The Kolmogorov-Smirnov (K-S) goodness-of-fit test is applied to test the goodness of fit between a set of sample observations and a theoretical distribution. Let $F_X(x)$ represent the real distribution of a random variable X and $F_0(x)$ denote the assumed distribution. The hypothesis set of the K-S test is

$$
\begin{aligned}
H_0\text{:} \quad & F_X(x) = F_0(x) \text{ for all } x \\
H_A\text{:} \quad & F_X(x) \neq F_0(x) \text{ for at least one value of } x.
\end{aligned}
\tag{A.9}
$$

The K-S statistic D_n is defined as

$$
D_n = \sup_x |S_n(x) - F_0(x)|
\tag{A.10}
$$

where $S_n(x)$ is the sample (or empirical) distribution function. For any value x, $S_n(x)$ is defined as the proportion of observations not exceeding x, where n denotes the number of observations.

Suppose there are r different values of observations for some $r \leq n$ and these values are sorted such that $x_1 < x_2 < \cdots < x_r$. The K-S test statistic D_n can be formulated by

$$
D_n = \max_{1 \leq i \leq r} \{ \max [|S_n(x_i) - F_0(x_i)|, |S_n(x_{i-1}) - F_0(x_i)|] \}.
\tag{A.11}
$$

If all observations are different, so that $r = n$, the last equation reduces to

$$
D_n = \max_{1 \leq i \leq n} \{ \max [|\frac{i}{n} - F_0(x_i)|, |\frac{i-1}{n} - F_0(x_i)|] \}.
\tag{A.12}
$$

The form of the K-S test is to reject the null hypothesis H_0 if D_n exceeds some constant $d_{n,1-\alpha}$, which is the upper $1 - \alpha$ critical point of the distribution of D_n. If none of the parameters of the hypothesized distribution are estimated from the data, a generic critical point $d_{n,1-\alpha}$ is available for all continuous distribution types. This is the so-called original form of the critical point $d_{n,1-\alpha}$ of the K-S test.

A.4   Maximum likelihood method

Given that X is a continuous random variable with $pdf = f(x; \theta_1, \theta_2, ..., \theta_k)$, we are going to estimate the k unknown parameters $\theta_1, \theta_2, ..., \theta_k$. If we have n independent

observations $x_1, x_2, ..., x_n$, the likelihood function is given by

$$L(x_1, x_2, ..., x_n | \theta_1, \theta_2, ..., \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, ..., \theta_k). \qquad (A.13)$$

The logarithmic likelihood function is,

$$\Lambda = \ln L = \sum_{i=1}^{n} \ln f(x_i; \theta_1, \theta_2, ..., \theta_k). \qquad (A.14)$$

The MLE of $\theta_1, \theta_2, ..., \theta_k$ may be obtained by maximizing L or Λ. By maximizing Λ, which is much easier to work with than L, the MLE of $\theta_1, \theta_2, ..., \theta_k$ is the simultaneous solution of k equations such that,

$$\frac{\partial(\Lambda)}{\partial \theta_j} = 0, \; j = 1, 2, ..., k. \qquad (A.15)$$

A.5 k-means clustering algorithm

k-means is one of the most well-known clustering methods of data. Cluster membership is determined by calculating the centroid (the multidimensional version of the mean) for each cluster. The clustering algorithm (Larose (2005)) proceeds as follows.

1. Ask the user how many clusters k the data set should be partitioned into;

2. Randomly assign k records to be the initial cluster center locations;

3. For each record, find the nearest cluster center. Thus, in a sense, each cluster center 'owns' a subset of the records, thereby representing a partition of the data set;

4. For each of the k clusters, find the cluster centroid, and update the location of each cluster center to the new value of the centroid;

5. Repeat steps 3 to 5 until convergence or termination.

A.6 Moment estimates of log-normal and Weibull distribution parameters

Log-normal distribution

Let X is a continuous random variable that has a log-normal distribution. If the mean and standard deviation of the underlying normal distribution is denoted by μ and σ, the r th moment of X about the origin is

$$m_r(X) = \exp(\mu r + \frac{r^2 \sigma^2}{2}), \qquad (A.16)$$

so that the mean and variance of X can be written as

$$\mu_x = m_1(X) = \exp(\mu + \frac{\sigma^2}{2}),$$
(A.17)

and

$$\sigma_x^2 = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1].$$
(A.18)

Given the mean μ_x and variance σ_x^2 of X, a moment estimate of σ can be derived by

$$\tilde{\sigma} = \sqrt{\ln(\sigma_x^2 + \mu_x^2) - 2\ln(\mu_x)}.$$
(A.19)

The corresponding moment estimate of μ then follows:

$$\tilde{\mu} = \ln(\mu_x) - \frac{\tilde{\sigma}^2}{2}.$$
(A.20)

Weibull distribution

Let X is a continuous random variable that has a two-parameter Weibull distribution. If the shape and the scale parameters are denoted by k and $1/\lambda$, respectively, the r th moment of X about the origin is

$$m_r(X) = \frac{1}{\lambda^r}\Gamma(1 + \frac{r}{k}),$$
(A.21)

where Γ is the gamma function (Bury (1999)).

Thus, the mean and variance can be expressed by

$$\mu_x = \frac{1}{\lambda}\Gamma(1 + \frac{1}{k}),$$
(A.22)

and

$$\sigma_x^2 = \frac{1}{\lambda^2}[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k})].$$
(A.23)

The coefficient of variation of X is therefore a function of k only:

$$cv_x = \frac{\sigma_x}{\mu_x} = \sqrt{\frac{\Gamma(1 + \frac{2}{k})}{\Gamma^2(1 + \frac{1}{k})} - 1}.$$
(A.24)

Given the mean μ_x and variance σ_x^2 of X, we can easily obtain the cv_x. The moment estimate $\tilde{k}$ can be obtained by solving the last equation with an iterative approach (Bury (1999)). The moment estimate of scale parameter $1/\lambda$ can then be derived from Equation A.22.

Summary

Facing continuous growth of transport demand, most railway infrastructure managers are not only extending and upgrading the track network and improving the signalling systems to create additional traffic capacity, but are also striving to utilize the existing capacity more efficiently. It is well known that the degree of utilization of network capacity affects the quality of train services, e.g. reliability and punctuality of train operations, which are important to passengers as well as operators. Trade-offs exist between the desired level-of-service for customers, the expected operating cost and revenues of train operators and efficient capacity utilization to be achieved by railway infrastructure managers, timetable designers, and train dispatchers. To maximize the capacity utilization and improve the timetable design at a desired reliability and punctuality level of train operations, it is necessary to analyze and model train delays and delay propagation. As the bottlenecks of double track railway networks are mostly situated at stations, this thesis focuses on the modelling of train delays and delay propagation in stations.

We have developed a blocking time theory based analytical probability model which estimates the knock-on delays caused by route conflicts and late transfer connections including the impact on the punctuality of trains at a station. To achieve a logic formulation of and a realistic estimation by this delay propagation model, a deeper insight into the stochastic characteristics of train operations is achieved by a detailed statistical analysis of real-world track occupation and release data recorded at the Dutch railway station The Hague HS, which contains various level crossings and mergings of railway lines. A number of existing theoretical distributions for train event and process times are assessed by the formal Kolmogorov-Smirnov (K-S) goodness-of-fit test, which may be taken as input of the delay propagation model. Finally, we apply the developed delay propagation model for maximizing the utilization of station capacity at a desired reliability and punctuality level of train operations.

The statistical analysis has been performed not only for train delays, but also for the propagation of train delays between train pairs and the estimation of the real capacity utilization of critical track sections, i.e., platform tracks and track junctions in the station and adjacent interlocking areas. The analysis results prove that route conflicts between a pair of train series cause considerable knock-on delays and reduce the punctuality of the following train series. To reduce the probability of knock-on delays for a train series, it is necessary to decrease not only the delays of the preceding train series, but also to avoid early arrivals of the train series itself. One of the most remarkable findings is that the blocking times at the open track signal block preceding the station of some frequently hindered train series are even exceeding those at the station block including the platform track. In case of hinder of following trains by delayed preceding trains at the station, the bottleneck of the station and interlocking areas may be shifted upstream.

To assess several commonly applied distribution models for train event and process times by the K-S goodness-of-fit test, we present a new method for fine-tuning the parameters of a distribution model. Then, we compare different distribution fits with fine-tuned parameters based on the K-S test results. It has been found that a location-shifted log-normal distribution with the shift parameter being the earliest arrival time can be generally considered as the best model among the candidate distributions for both the arrival times of trains at the platform track and at the approach signal of the station. The Weibull distribution can generally be considered as the best distribution model for non-negative arrival delays, departure delays and the free dwell times of trains in the absence of hindrance from other trains. The distribution evaluation based on the presented parameter fine-tuning method results in rather satisfactory distribution fits to the datasets, which can be used as input of the developed and other delay propagation models.

The developed delay propagation model reflects the constraints of the signalling system and the traffic management rules on the dynamic use of track capacity. The interdependences of the arrival and departure times of different train lines are taken into account. We model the stochastic variations of track occupation times due to fluctuation of train speed in case of (no) hindrance by conditional distributions. In addition, we consider the dependences of free dwell times of trains on the corresponding arrival delays by conditional distributions. As a result, this model enables a higher accuracy of the estimation regarding the knock-on delays of trains and the impact on the punctuality of trains than existing analytical model(s). The model validation based on the local network of The Hague HS station and the scheduled arrival, departure and headway times reveals that the model estimates the arrival and departure delay distributions very well. In case of other stations with different track layout and interlocking routes, the model can be applied by decomposing the station network into a number of platform tracks and conflicting route nodes (crossings and switches) and analyzing the interdependences of different lines correspondingly. A precondition, however, is that there are suitable estimates of the distributions of the event and process times available.

The developed delay propagation model has been applied for maximizing the station capacity utilization and improving the timetable design in a case study of The Hague HS station. The mean knock-on delay of the passing trains at a route node in the station area increases exponentially with a decrease of the scheduled buffer time between the train paths. The model enables the determination of the maximal frequency of trains passing the critical track sections within the station area, while assuring a maximally acceptable knock-on delay at a certain confidence level and a desired punctuality level. Given the statistical parameters (mean and standard deviation) of the input delays of the trains at the boundaries of the station network, i.e., at the approach signals of the home signals of the station, and those of the primary delays during the running and dwell processes within the network, the running and dwell time supplements and the buffer times between scheduled train paths can be minimized by applying the developed delay propagation model, while a given maximal knock-on delay and a desired punctuality level can be achieved. This enables a very precise estimation of the required running and dwell time margins, as well as of the buffer times.

Samenvatting

Om aan de continue groei in mobiliteit en toenemende vraag naar spoorvervoer te voldoen werken de meeste railinfrastructuurmanagers niet alleen aan het uitbreiden en vernieuwen van het spoornetwerk en het verbeteren van het seinsysteem ter vergroting van de spoorcapaciteit, maar ook aan een betere benutting van de bestaande capaciteit. Het is welbekend dat de benuttingsgraad van de netwerkcapaciteit de kwaliteit van treindiensten beïnvloedt, zoals betrouwbaarheid en punctualiteit die belangrijk zijn voor zowel reizigers als vervoerders. Er bestaat een wisselwerking tussen de gewenste servicekwaliteit van reizigers, de verwachte kosten en baten van de vervoerders en efficiënt capaciteitsgebruik onder veranwoordelijkheid van infrastructuurmanagers, dienstregelingontwerpers en treindienstleiders. Om de capaciteitsbenutting te maximaliseren en het dienstregelingsontwerp te verbeteren tot een gewenste maat van betrouwbaarheid en punctualiteit is het noodzakelijk treinvertragingen en vertragingsvoortplanting te analyseren en modelleren. Omdat de knelpunten van tweesporige railnetwerken meestal op de stations liggen ligt de focus van dit proefschrift op de modellering van treinvertragingen en vertragingsvoortplanting op stations.

We hebben een analytisch kansmodel ontwikkeld gebaseerd op bloktijdtheorie dat de volgvertragingen veroorzaakt door rijwegconflicten en late overstapaansluitingen schat inclusief de impact op de punctualiteit van treinen op het station. Om een juiste formulering en realistische schatting van de vertragingsvoortplanting te krijgen is een diep inzicht in de stochastische karakteristieken van de treinverkeersafwikkeling verkregen door een gedetailleerde statistische analyse van empirische data van spoorbezettingen en vrijgaven op station Den Haag HS, dat verschillende gelijkvloerse kruisingen en samenkomende spoorlijnen bevat. Een aantal bestaande theoretische kansverdelingen voor treinactiviteiten en procestijden die gebruikt kunnen worden als invoer voor het vertragingsmodel zijn getoetst met de formele Kolmogorov-Smirnov (K-S) toets. Uiteindelijk gebruiken we het ontwikkelde vertragingsvoortplantingsmodel om de de stationscapacaciteitsbenutting te maximaliseren bij een gewenste betrouwbaarheid en punctualiteitsniveau.

De statistische analyse is niet alleen uitgevoerd voor treinvertragingen maar ook voor de voortplanting van vertragingen tussen treinparen en het schatten van de daadwerkelijke capaciteitsbenutting van de kritieke spoorsecties, i.e., de perronsporen en wissels op het emplacement. De analyseresultaten laten zien dat rijwegconflicten tussen treinseries veel volgvertragingen veroorzaken en leiden tot een punctualiteitsafname van de opvolgende treinserie. Om de kans op volgvertragingen voor een treinserie te verminderen is het niet alleen noodzakelijk om de vertragingen van de voorgaande treinserie te verlagen maar ook te vroege aankomsten van de treinen van de betreffende treinserie zelf te voorkomen. Een van de meest opzienbarende uitkomsten is dat de bloktijden van het blok op het

baanvak net voor het station voor sommige vaak gehinderde treinseries zelfs groter is dan de routeblokken op het emplacement inclusief het perronspoor. In geval van hinder voor opvolgende treinen door vertraagde voorgaande treinen op het station kan het knelpunt stroomopwaarts verschuiven.

Om een aantal veel gebruikte kansverdelingen voor treinactiviteiten en procestijden te toetsen met de K-S toets, presenteren we een nieuwe methode om de parameters van de kansverdelingen te schatten. Vervolgens vergelijken we verschillende verdelingen met de geschatte parameters via de K-S toetsresultaten. Hieruit volgt dat de verschoven lognormale verdeling – met de vroegste aankomsttijd als shift parameter – over het algemeen als het beste model van de kandidaatverdelingen beschouwd kan worden voor zowel de aankomsttijden op het perronspoor als die bij het voorsein voor het emplacement. De Weibullverdeling is over het algemeen het beste model voor de niet-negatieve aankomstvertragingen, de vertrekvertragingen en de vrije halteertijden van treinen die niet gehinderd zijn door andere treinen. De kansverdelingsevaluatie gebaseerd op de gepresenteerde parameterschattingsmethode geeft goede verdelingsschattingen van de dataverzamelingen die als invoer van de ontwikkelde – en andere – vertragingsvoortplantingsmodellen kunnen worden gebruikt.

Het ontwikkelde vertragingsvoortplantingsmodel weerspiegelt de nevenvoorwaarden van het seinsysteem en de verkeersbeheersingsregels van het dynamische gebruik van spoorcapaciteit. Ook de afhankelijkheden tussen de aankomst- en vertrektijden van verschillende treinseries worden meegenomen. We modelleren de stochastische variaties in de spoorbezettingstijden vanwege fluctuaties in snelheden in het geval van wel en geen hinder door geconditioneerde verdelingen. Daarnaast beschouwen we ook de afhankelijkheid van de vrije halteertijden op aankomstvertragingen door geconditioneerde verdelingen. Hierdoor geeft dit model een hogere nauwkeurigheid van de schattingen van volgvertragingen en het effect op punctualiteit dan bestaande analytische modellen. De modelvalidatie gebaseerd op het lokale netwerk van Den Haag HS en de geplande aankomst-, vertrek- en opvolgtijden toont aan dat het model een goede schatting geeft van de aankomst- en vertrekverdelingen. In het geval van andere stations met verschillende sporenlayouts en rijwegen kan het model worden toegepast door decompositie van het emplacement in perronsporen en conflicterende rijwegknopen (kruisingen en wissels) en een analyse van de corresponderende afhankelijkheden van de verschillende treinseries. Een voorwaarde hierbij is de beschikbaarheid van geschikte schattingen van de verdelingen van treinactiviteiten en procestijden.

Het ontwikkelde vertragingsvoortplantingsmodel is toegepast voor de maximalisatie van de stationscapaciteitsbenutting en verbetering van het dienstregelingsontwerp in een casestudie voor Den Haag HS. De gemiddelde volgvertraging van een passerende trein over een rijwegknoop op het emplacement neemt exponentieel toe met een verlaging van geplande buffertijd tussen de treinpaden. Het model kan worden gebruikt om de maximale frequentie te bepalen van het aantal treinen dat de kritieke spoorsecties passeert voor een maximaal acceptabele volgvertraging bij een gegeven betrouwbaarheidsniveau en gewenst punctualiteitsniveau. Gegeven de statistische parameters (gemiddelde en standaard deviatie) van de invoervertragingen van de treinen op de rand van het emplacement, i.e., bij de voorseinen van het station, en de primaire vertragingen van de rij- en halteertijden in het netwerk, kunnen de rijtijd- en halteertijdmarges en buffertijden tussen de geplande treinpaden worden geminimaliseerd door toepassing van het ontwikkelde

vertragingsvoorplantingsmodel, bij een gehandhaafd maximaal volgvertragingsniveau en gewenst punctualiteitsniveau. Dit geeft een zeer precieze schatting van de benodigde rij- en halteertijdmarges en buffertijden.

About the author

Jianxin Yuan was born in Shanxi Province, China, on 15 September 1966. From 1984 to 1991, he studied in Tianjin University. He from this University received his BSc (with honor) in Applied Mathematics in 1988 and his MSc (with honor) in Coastal and Offshore Engineering in 1991. Afterwards, he worked in Tianjin University and Tianjin Harbor Engineering Consultant Company from 1991 to 1998.

In October 1998, he started the MSc study at the UNESCO-IHE Institute for Water Education, Delft, sponsored by the Netherlands Fellowship Programme and the National Institute for Coastal and Marine Management (RWS RIKZ). In September 2000, he was awarded the MSc in Hydraulic Engineering after the defence of a thesis about the validation of the shallow water wave model SWAN.

Since October 2000, he has been employed by the Department of Transport and Planning, Delft University of Technology to conduct the PhD research within the TRAIL Research School. This research, under the supervision of Professor Dr.-Ing. Ingo Hansen, resulted in a number of publications in proceedings of several international conferences, e.g. Computers in Railways, the journal Transportation Research Part B, and this PhD diesertation. After his PhD defence in 2006, he will continue working at the Department of Transport and Planning as a post-doc researcher.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

Nat, C.G.J.M., van der, *A Knowledge-based Concept Exploration Model for Submarine Design*, T99/1, March 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Westrenen, F.C., van, *The Maritime Pilot at Work: Evaluation and Use of a Time-to-boundary Model of Mental Workload in Human-machine Systems*, T99/2, May 1999, TRAIL Thesis Series, Eburon, The Netherlands

Veenstra, A.W., *Quantitative Analysis of Shipping Markets*, T99/3, April 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Minderhoud, M.M., *Supported Driving: Impacts on Motorway Traffic Flow*, T99/4, July 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoogendoorn, S.P., *Multiclass Continuum Modelling of Multilane Traffic Flow*, T99/5, September 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoedemaeker, M., *Driving with Intelligent Vehicles: Driving Behaviour with Adaptive Cruise Control and the Acceptance by Individual Drivers*, T99/6, November 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Marchau, V.A.W.J., *Technology Assessment of Automated Vehicle Guidance - Prospects for Automated Driving Implementation*, T2000/1, January 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Subiono, *On Classes of Min-max-plus Systems and their Applications*, T2000/2, June 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Meer, J.R., van, *Operational Control of Internal Transport*, T2000/5, September 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Bliemer, M.C.J., *Analytical Dynamic Traffic Assignment with Interacting User-Classes: Theoretical Advances and Applications using a Variational Inequality Approach*, T2001/1, January 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Muilerman, G.J., *Time-based logistics: An analysis of the relevance, causes and impacts*, T2001/2, April 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, T2001/3, May 2001, TRAIL Thesis Series, The Netherlands

Willems, J.K.C.A.S., *Bundeling van infrastructuur, theoretische en praktische waarde van een ruimtelijk inrichtingsconcept*, T2001/4, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Binsbergen, A.J., van, J.G.S.N. Visser, *Innovation Steps towards Efficient Goods Distribution Systems for Urban Areas*, T2001/5, May 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Rosmuller, N., *Safety analysis of Transport Corridors*, T2001/6, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Schaafsma, A., *Dynamisch Railverkeersmanagement, besturingsconcept voor railverkeer op basis van het Lagenmodel Verkeer en Vervoer*, T2001/7, October 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Bockstael-Blok, W., *Chains and Networks in Multimodal Passenger Transport. Exploring a design approach*, T2001/8, December 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, T2002/1, February 2002, TRAIL Thesis Series, The Netherlands

Vis, F.A., *Planning and Control Concepts for Material Handling Systems*, T2002/2, May 2002, TRAIL Thesis Series, The Netherlands

Koppius, O.R., *Information Architecture and Electronic Market Performance*, T2002/3, May 2002, TRAIL Thesis Series, The Netherlands

Veeneman, W.W., *Mind the Gap; Bridging Theories and Practice for the Organisation of Metropolitan Public Transport*, T2002/4, June 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Van Nes, R., *Design of multimodal transport networks, a hierarchical approach*, T2002/5, September 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Pol, P.M.J., *A Renaissance of Stations, Railways and Cities, Economic Effects, Development Strategies and Organisational Issues of European High-Speed-Train Stations*, T2002/6, October 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Runhaar, H., *Freight transport: at any price? Effects of transport costs on book and newspaper supply chains in the Netherlands*, T2002/7, December 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Spek, S.C., van der, *Connectors. The Way beyond Transferring*, T2003/1, February 2003, TRAIL Thesis Series, Delft University Press, The Netherlands

Lindeijer, D.G., *Controlling Automated Traffic Agents*, T2003/2, February 2003, TRAIL Thesis Series, Eburon, The Netherlands

Riet, O.A.W.T., van de, *Policy Analysis in Multi-Actor Policy Settings. Navigating Between Negotiated Nonsense and Useless Knowledge*, T2003/3, March 2003, TRAIL Thesis Series, Eburon, The Netherlands

Reeven, P.A., van, *Competition in Scheduled Transport*, T2003/4, April 2003, TRAIL Thesis Series, Eburon, The Netherlands

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, T2003/5, June 2003, TRAIL Thesis Series, The Netherlands

Soto Y Koelemeijer, G., *On the behaviour of classes of min-max-plus systems*, T2003/6, September 2003, TRAIL Thesis Series, The Netherlands

Lindveld, Ch..D.R., *Dynamic O-D matrix estimation: a behavioural approach*, T2003/7, September 2003, TRAIL Thesis Series, Eburon, The Netherlands

Weerdt, de M.M., *Plan Merging in Multi-Agent Systems*, T2003/8, December 2003, TRAIL Thesis Series, The Netherlands

Langen, de P.W, *The Performance of Seaport Clusters*, T2004/1, January 2004, TRAIL Thesis Series, The Netherlands

Hegyi, A., *Model Predictive Control for Integrating Traffic Control Measures*, T2004/2, February 2004, TRAIL Thesis Series, The Netherlands

Lint, van, J.W.C., *Reliable Travel Time Prediction for Freeways*, T2004/3, June 2004, TRAIL Thesis Series, The Netherlands

Tabibi, M., *Design and Control of Automated Truck Traffic at Motorway Ramps*, T2004/4, July 2004, TRAIL Thesis Series, The Netherlands

Verduijn, T. M., *Dynamism in Supply Networks: Actor switching in a turbulent business environment*, T2004/5, September 2004, TRAIL Thesis Series, The Netherlands

Daamen, W., *Modelling Passenger Flows in Public Transport Facilities*, T2004/6, September 2004, TRAIL Thesis Series, The Netherlands

Zoetman, A., *Railway Design and Maintenance from a Life-Cycle Cost Perspective: A Decision-Support Approach*, T2004/7, November 2004, TRAIL Thesis Series, The Netherlands

Bos, D.M., *Changing Seats: A Behavioral Analysis of PR Use*, T2004/8, November 2004, TRAIL Thesis Series, The Netherlands

Versteegt, C., *Holonic Control For Large Scale Automated Logistic Systems*, T2004/9, December 2004, TRAIL Thesis Series, The Netherlands

Wees, K.A.P.C. van, *Juridische Aspecten van ADAS (Advanced Driver Assistance Systems)*, T2004/10, December 2004, TRAIL Thesis Series, The Netherlands

Tampere, C.M.J., *Human-Kinetic Multiclass Traffic Flow Theory and Modelling: With Application to Advanced Driver Assistance Systems in Congestion*, T2004/11, December 2004, TRAIL Thesis Series, The Netherlands

Rooij, R.M., *The Mobile City. The planning and design of the Network City from a mobility point of view*, T2005/1, February 2005, TRAIL Thesis Series, The Netherlands

Le-Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, T2005/2, April 2005, TRAIL Thesis Series, The Netherlands

Zuidgeest, M.H.P., *Sustainable Urban Transport Development: a Dynamic Optimization Approach*, T2005/3, April 2005, TRAIL Thesis Series, The Netherlands

Hoogendoorn-Lanser, S., *Modelling Travel Behavior in Multimodal Networks*, T2005/4, May 2005, TRAIL Thesis Series, The Netherlands

Dekker, S., *Port Investment - Towards an Integrated Planning of Port Capacity*, T2005/5, June 2005, TRAIL Thesis Series, The Netherlands

Koolstra, K., *Transport Infrastructure Slot Allocation*, T2005/6, June 2005, TRAIL Thesis Series, The Netherlands

Vromans, M., *Reliability of Railway Systems*, T2005/7, July 2005, TRAIL Thesis Series, The Netherlands

Oosten, W., *Ruimte voor een democratische rechtsstaat. Geschakelde sturing bij ruimtelijke investeringen*, T2005/8, September 2005, TRAIL Thesis Series, The Netherlands

Le-Duc, T., *Design and control of efficient order picking*, T2005/9, September 2005, TRAILThesis Series, The Netherlands

Goverde, R.M.P., *Punctuality of Railway Operations and Timetable Stability Analysis*, T2005/10, October 2005, TRAIL Thesis Series, The Netherlands

Kager, R., *Design and implementation of a method for synthesis of travel diary data*, T2005/11, October 2005, TRAIL Thesis Series, The Netherlands

Boer, C., *Distributed Simulation in Industry*, T2005/12, November 2005, TRAIL Research School, The Netherlands

Pielage, P., *Conceptual Design of Automated Freight Transport Systems. Methodology and Practice*, T2005/14, November 2005, TRAIL Research School, The Netherlands

Groothedde, B., *Collaborative Logistics and Transportation Networks. A Modeling Approach to Hub Network Design*, T2005/15, November 2005, TRAIL Research School, The Netherlands

Valk, J.M., *Coordination among Autonomous Planners*, T2005/16, December 2005, TRAIL Thesis Series, The Netherlands

Krogt, R.P.J. van der, *Plan Repair in Single-Agent and Multi-Agent Systems*, T2005/17, December 2005, TRAIL Thesis Series, The Netherlands

Bontekoning, Y.M., *Hub exchange operations in intermodal hub-and-spoke networks. A performance comparison of four types of rail-rail exchange facilities*, T2006/1, February 2006, TRAIL Thesis Series, The Netherlands

Lentink, R., *Algorithmic Decision Support for Shunt Planning*, T2006/2, February 2006, TRAIL Thesis Series, The Netherlands

Ngoduy, D., *Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations*, T2006/3, April 2006, TRAIL Thesis Series, The Netherlands

Vanderschuren, M.J.W.A., *Intelligent Transport Systems for South Africa. Impact assessment through microscopic simulation in the South African context*, T2006/4, August 2006, TRAIL Thesis Series, The Netherlands

Ongkittikul, S., *Innovation and Regulatory Reform in Public Transport*, T2006/5, September 2006, TRAIL Thesis Series, The Netherlands

Yuan, J., *Stochastic Modelling of Train Delays and Delay Propagation in Stations*, T2006/6, October 2006, TRAIL Thesis Series, The Netherlands