# A Closed Form Railway Line Delay Propagation Model 

Harrod, Steven; Cerreto, Fabrizio; Nielsen, Otto Anker

Published in:
Transportation Research. Part C: Emerging Technologies

Link to article, DOI:
10.1016/j.trc.2019.02.022

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Harrod, S., Cerreto, F., \& Nielsen, O. A. (2019). A Closed Form Railway Line Delay Propagation Model. Transportation Research. Part C: Emerging Technologies, 102, 189-209.
https://doi.org/10.1016/j.trc.2019.02.022

[^0]
# A Closed Form Railway Line Delay Propagation 

## Model

Fabrizio Cerreto ${ }^{*}$, Steven Harrod ${ }^{\text {a }}$, Otto Anker Nielsen ${ }^{\text {b }}$.
Department of Management Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

* Corresponding author: facer@dt.dk
${ }^{\text {a }}$ stehar@dtu.dk
${ }^{\text {b }}$ oani@dtu.dk


#### Abstract

Delays of railway services can be measured by the aggregate delay over a time horizon due to an event that delay a given train. Timetables for railway service may dampen aggregate delay by adding either supplement time or buffer time to the minimum driving time. The evaluation of these variables is often performed by time consuming numerical analysis with simulation tools and hence with some degree of stochasticity of the outcome. This paper proposes instead an analytical closed form formulation of aggregate delay with a polynomial form. The function returns the aggregate delay of a railway line resulting from an initial, primary, delay. This can be used to get theoretical insights into railway delays and as part of larger railway scheduling systems, where simulation models would require too much calculation time. Analysis of the function recommend a balance between the two remedial measures, supplement and buffer. Further, the effect of different threshold values in delay measurement is depicted, giving information valuable in the design of service contract. Numerical analysis of an example railway line shows that the polynomial function provides guidance and insight even when theoretical assumptions are violated.


KEYWORDS: Rail transportation, Train delays, Timetable robustness, Timetable design, Delay propagation

## 1. Introduction

Operational stability and robustness are important for railway transport. Not only are the passengers or users of the service sensitive to these measures of quality (Parbo et al., 2016), but railways are usually integrated systems or networks, where failures at one location of the system often affect other locations and services, sometimes quite catastrophically. A railway network planner is thus faced with many decisions about what quality of service to provide and what resources to allocate to deliver this service. Much of the literature demonstrates that there are frequently multiple feasible alternatives to allocate resources, and each alternative has a unique performance profile with characteristic statistics, especially in punctuality and robustness (Caimi et al., 2009). The analysis of these alternatives frequently requires laborious and inconclusive modeling with simulation software, which is time consuming in both model programming and analysis run-time (Parbo et al., 2014).

Cumulative, aggregate, or total delay, is a common performance measure used in several fields, from operation monitoring, to timetable planning and optimization. Goverde and Meng (2011) developed a data analysis tool to build conflict trees based on track occupation data. The conflict trees depict the realized delay propagation across consecutive trains on a railway infrastructure, so that it is possible to identify primary delays and the overall disturbance generated in form of secondary delays. Goverde and Meng assess the severity of individual disturbances measuring the total delay that they generated. Following a comparison of methods and data used to assess performance quality by 22 metro operators worldwide, Barron et al. (2013) describe measures of total effect of disruptions as the best representation of service quality, as they provide a better understanding of how incidents affect operation and customers. In particular, total vehicle hours of delay reflect the operator's interest in the vulnerability to network disruptions. Alternative forms of aggregate delay are currently used in transport industry for service quality assessment and to rule contracts between railway undertakers and infrastructure managers. Transport for London uses the Lost Customer Hours as a performance measure in metro operation. The measure consists in the total delay given to passengers, meant as estimated travel time extension due to incidents (TfL Investment Programme Management Office, 2008). In Europe, the

Performance Regime produces cash flows between railway undertakers and infrastructure managers as incentive to improve service quality. In Italy, every minute of train delay is valued $2 €$ (Rete Ferroviaria Italiana, 2015), and the aggregate line delay can be reconducted to the total cash flow generated by an individual primary delay.

Cumulative delay is also used as a metric in the planning phase to evaluate timetables before real operation. It is often compared to given initial delays to evaluate how the timetable is stable against disturbances, and how large are the consequences of given disturbances assuming different timetables. Salido et al. (2012) compare timetables using the cumulative delay resulting from simulation, and define a timetable more robust than another, if, for a given disturbance, the cumulative delay generated is smaller. Cumulative delay is also found in optimization timetabling problems, as objective function to minimize (Toletti, 2016; Törnquist, 2007). Total vehicle delay appears in delay management problems: Ginkel and Schobel (2007) optimize the operational departure time for a connecting service at a transfer station, given that the primary service is delayed. A bicriteria model is set up to minimize the number of passenger missed connections and the total delay recorded by vehicles. Harker and Hong (1994) use cumulative line delay to evaluate dispatching choices on a railway network in a Nash noncooperative game. The model is set up to seek the network optimal dispatching strategy, given that the single divisions of the network act to minimize the aggregate deviation of operation from the schedule of trains running on its area of control. The model is eventually used in the pricing of train slots, according to the value of time attributed to individual trains.

This paper contributes to the literature with an analytic closed form formulation of aggregate railway line delay propagation in response to a primary delay. This function may supplement or replace the application of simulation models for exploration of alternatives when appraising different timetable alternatives. The formulation is closed form under a set of timetable-structure assumptions that this paper later shows, using microsimulation, to be robust to deviation from these assumptions. The advantage of such a formulation is the rapidity and avoidance of simulation as compared to the oftenused micro simulation models in the industry. It facilitates a quick evaluation of the expected cumulative
delay, and makes it possible to evaluate structural design factors, such as running time supplement, headway buffers, and to design service contract performance measures. Lastly, cumulative delay calculated in a closed form can efficiently be implemented in optimization models for timetabling reducing, in this way, the time lost in simulation to estimate delays.

The formulation is derived from a finite series of deviations from the service plan (secondary delays) caused by a singular initial disruption (primary delay). The primary delay is propagated to following trains and recovered on the individual trains' downstream paths. The results may be used to establish bounds of expected performance for simulation models, and possibly reduce the use of simulation models in some applications. The scientific novelty of this model is the explicit and simultaneous inclusion of headway buffers and running time supplements to reduce the individual train delay in the propagation process, whereas previously proposed methods considered only one type of timetable slack at a time, or used queuing theory to evaluate the interferences between vehicles (Hasegawa et al., 1981; Huisman, 2002; Landex, 2008; Mattsson, 2007; Pyrgiotis, 2012; Salido et al., 2012). The results are also relevant for industrial managerial purposes, giving the possibility to investigate the effect of timetable design parameters in advance, such as contract delay threshold, running time supplement and headway buffer.

### 1.1. Structure of the Paper

A literature review is presented in Section 2. Section 3 defines the timetable structure and relevant measures for railway performance. Section 4 explains the finite series delay structure applied in this paper, followed by the solution of the polynomial total line delay. In Section 5, the formulation is validated and tested for robustness with a microsimulation of a suburban railway line north of Copenhagen. The extended model to include primary delays at unspecified station and train is presented in Section 6. The novelty of such a model and inferences on the results are discussed in Section 7, followed by concluding remarks in section.

## 2. Literature Review

Prior relevant literature defines suitable measures of railway performance and methods of estimating that performance. Section 3 presents literature concerning suitable performance measures for railways, while performance estimation methods are reviewed here. Performance estimation may be categorized as statistical or analytical, and the analytical literature is reviewed first.

Hasegawa et al. (1981) applies a hydrodynamic analogy to model railway traffic. The study models the delay propagation as a shock wave in a compressible fluid and finds the total delay as a cubic function of the primary delays by means of propagative velocity. Mattsson (2007) applies a different analogy to standard road traffic-flow to calculate the capacity utilization of a double-tracked railway line, as the percentage of time used by a train sequence. The method is similar to previous approaches (Gibson et al., 2002; UIC, 2004), and the capacity used by a timetable has been used later as a measure of it reliability (Haith et al., 2014). Meester and Muns (2007) apply phase-type distributions to derive the net delay distribution on connected railway network segments given the distribution of primary delays on each segment. The derivation asserts that recursive calculation of the solution may be attained with just three operations: sum, nonnegative excess beyond a bound, and maximum. The paper states that a phase-type distribution, a distribution of the absorption time of a continuous time Markov chain, can be contained in the three operations in closed form. However, the method depends on the assertion of independence of the primary delays. The method is demonstrated for a sample network of 24 directional line segments with seven transfer points.

Goverde $(2010,2007)$ presents an efficient delay propagation algorithm where timetables are modeled as timed event graphs (using max-plus algebra) and initial delays are known. The algorithm is very fast and in a few seconds can calculate the delay propagation over a large network consisting of many interdependent services, such as the Dutch national railway timetable. Goverde uses this method to compute performance indicators, including delay propagation statistics such us total secondary delay, and settling time. However, the model offers no functional relationship, and results must be calculated for each scenario separately. Most closely related to this paper, Landex (2008) proposes a delay
propagation model computing the transfer of delay between trains through the scheduled buffer times. This model is used to study the relation between capacity consumption and the development of the disruptions, but does not consider the recovery of train delays according to the timetable allowance. A similar method is proposed by Huisman and Boucherie (2001). Huisman and Boucherie model the delay propagation in absence of a timetable, assuming that all trains run at their maximum possible speed. The first train is assumed to run within its minimum running time, whereas the following trains increase their running time in terms of the delay of the previous train, reduced by the headway buffer. Using queuing theory, expected interference between consecutive trains are estimated, but the influence of primary delay is only considered implicitly as distribution of free-flow running times. Relevant analytical approaches for delay propagation span also other means of transportation. Pyrgiotis (2012) describes a mixed algorithmic and analytical model to propagate delays in airport networks, following the single aircrafts rosters. The analytical section of this model uses queueing theory to assign delays to flights due to congestion at airports, based on the airport capacity and the time-dependent traffic demand. The delay assigned to a flight is then propagated to the downstream airports in the aircraft's roster, with an approach similar to the papers described above. The aircraft delay is reduced in every trip segment by means of the scheduled slack, and it is increased by congestion, using the queuing theory.

Various statistical studies of railway performance exist applying either probability theory or simulation. Carey and Kwiecinski (1994) propose a stochastic approximation of the secondary delays of trains and validate the approximation with simulation studies of railway lines with common infrastructure of two and three aspect signals. The study finds that when trains have exponential delays between stations, the expected train time between stations is directly dependent on the headway between trains plus a constant. Carey (1999) defines many statistical measures of service plan configuration and many probability distributions of train performance, and offers many propositions on their sensitivity to changes in conditions. Gibson et al.(2002) use regression analysis on data recorded in the British railway network to define an exponential functional relation between the line capacity utilization and
the expected reactionary delays, by means of line section and route-specific constants. They furthermore identified several factors influencing the expected reactionary delays: geography, time of operation, and speed heterogeneity. Olsson and Haugland (2004) present extensive statistical analysis of recorded data in Norway and identify some influencing factors on railway punctuality, and some frequent patterns in the development of delay along the train paths. Strong correlation is found, for specific train categories, between the delay at the origin of the trip and at the destination. Olsson and Haugland identify critical values of initial delay over which the probability to encounter major delays along the way increases considerably. This pattern is triggered by the prioritization of the trains and the consequent dispatching choices of the operators. Goverde and Meng (2011) present a tool based on track circuit occupation data to identify realized conflicts and delay propagation from realized operation. The tool is able to identify a primary delay and the disturbance generated to each individual train. It is so possible to identify what are the critical blocking sections and signals that generate most conflict during operation using standard statistics and sorting after adapted performance metrics. Unfortunately, such a detailed information is very seldom available, either because it is not logged, or because operators keep it confidential.

Simulation-based delay propagation models exist also at mesoscopic level. Zhu and Schnieder (2000) propose a model to assess railway timetable performance using Colored Petri Nets, where lined between stations are divided in actual blocking section that can host one train at a time. As opposite, the station capacity is only modeled as number of tracks and track length. Given primary delay distribution, the authors evaluate the timetable performance measuring the Delay degree, which is the sum of the train delays recorded at specific station. However, the complexity of Colored Petri Net models increases exponentially with the number of stations, and the authors point out the imperative need to reduce the model complexity. Salido et al. (2012) proposes analytical measures of robustness and validates them using simulation. The model uses aggregate delay as performance statistic, and considers the headway buffer between trains, whereas running time supplement is not considered. Salido uses advanced simulation models that reschedule the traffic after disruption, and eventually suggests that analytical models of this type are a potential supplement or replacement for simulation
modeling. Furthermore, Salido observes a linear relationship between primary and aggregate delay when the railway system is lightly utilized, whereas the relationship is nonlinear for heavily utilized railway systems. Cerreto (2015) introduces a method to reduce computation time in simulation models shrinking the number of simulation runs required with a heuristic process called the skimming method. Instead of simulating all combinations of trains and delays, a composite profile of train delay is estimated from an initial simulation analysis, and this composite delay function is used to calculate the aggregate system delay. Lastly, Jensen et al. (2017) estimates the capacity consumption of a timetable and relates it to robustness against delays using a framework based on stochastic simulation.

## 3. Timetable Structure and Measures of Performance

A variety of measures are used in practice and in the literature to characterize the railway traffic density, to measure the risk of disruption, to measure the magnitude of disruption, and to measure the recovery from disruption back to planned services. Goverde and Hansen (2013) describes the train performance measures and summarizes timetable design methods typical of the German and Netherlands railway systems, and provides a broad summary of definitions of performance measures. These measures are generally applicable to most railway systems, including Denmark's.

In all cases, this paper and much of the cited literature presume operation according to a deterministic, feasible, timetable. A feasible timetable is one in which all train paths are planned in advance, and it is known, at a minimum, that all trains may operate without conflict with each other and on the known track infrastructure, if there are no errors, failures, or disruptions during the timetable operation. Of course, disruptions and errors in planning do occur, and so timetables are evaluated according to their robustness and stability. These terms are closely related and sometimes used interchangeably. Robustness is most closely associated with the ability of the planned train operations to absorb natural variation and delays due to weather, operator error, passenger actions (holding doors open, failure to buy a ticket, etc.), and smaller mechanical malfunctions. Stability refers to the ability of the planned train operations to return to their original schedule after a disruption has occurred. Lastly
timetable resilience refers to the adaptability of a timetable, the flexibility of the timetable structure to allow dynamic, tactical changes in operating plan in response a disruption.

Two fundamental timetable design measures are the supplements and the buffers (Goverde and Hansen, 2013). A supplement is additional time beyond the minimum operating time between timing points that allows a train to experience disruptions and yet still attain scheduled arrivals (Figure 3-1). This measure is specific (and potentially unique) to each train between a given pair of timing points, and directly supports timetable robustness.


Figure 3-1: Definition of timetable supplement for single train path.

A buffer is additional time between trains so that disruptions and delays of the leading train are less likely to cause interference with the following train (Figure 3-2). The buffer is a component of the headway (the total time between passing trains), but not the same as the headway. The headway equals the minimum safe separation time between trains plus the buffer. The capacity or number of trains on the railway line is strictly determined by the headway, but clearly the buffer is a decision variable, that other things being equal, determines the tradeoff between capacity and stability.


Figure 3-2: Group of train paths with separating buffer time.

Delays to trains are classified as primary or secondary. Primary delays are events happening to or "owned" by a specific train, such as a driver mistake, a passenger incident, unusual crowds, etc. Secondary delays are delays experienced as the result of conflict or interference with another train that has deviated from its plan (Figure 3-3). For example, if train 1 experiences an excessive station stop time due to a handicapped passenger boarding or unusual crowds, and following train 2 is delayed in entering the station, and following train 3 is delayed behind train 2 , then train 1 is experiencing a primary delay, and trains 2 and 3 are experiencing secondary delays.

The settling time is the time it takes from a primary delay until the trains have returned to their planned schedules. Salido et al. (2008) define a settling time performance measure called ( $t, k$ ) robust. The measure states that if a timetable suffers a disruption of $t$ time units, and returns to stability or the original plan in $k$ time units, then it is $(t, k)$ robust. An unstable timetable, from the earlier definition of stability, would never return to its planned schedules, and would thus have an infinite or undefined settling time. The settling time may serve as a performance measure for the relative stability of a timetable.


Figure 3-3: Calculation of secondary delay as a function of primary delay.

In the analysis that follows, total cumulative delay of the timetable system is the performance measure of interest. That is, the total deviation from the timetable of every train at every station (measuring point) over the time horizon of the analysis period. Thus, a train that is five minutes late at three sequential stops would register fifteen minutes of system delay. Törnquist (2007) finds that the total cumulative delay is strongly correlated with other performance measures. If the value of service is measured as passenger-minutes of delay (number of passengers times minutes of delay for each passenger), then under conditions of uniform traffic demand, the total cumulative train delay is a reasonable estimate of the loss of utility of the train service. Of course, if the passenger traffic is not uniformly distributed, then the following analysis must be modified, or may not be applicable.

## 4. A Finite Series Model of Delay in Two Dimensions

This section proposes a model that is a closed form function that calculates the total cumulative deviation (delay, there is no earliness allowed) from the service timetable at all measurement points, as a function of timetable supplement, timetable buffer, and a single initial delay to one train. This paper suggests that cumulative deviation is appropriate for estimating the utility loss to passengers in the system under relaxed assumptions of uniformity of traffic. In the following section this model will be shown to be a reasonable guide even when the initial conditions of uniformity are violated.

This model has a two-dimensional analysis horizon or domain. Many of the prior cited papers define the analysis horizon in terms of the length of line or the number of train path segments. This model specifically includes the secondary delays to trains, and thus the second dimension of the analysis horizon is the number of trains included in the cumulative delay statistic. This model will consider trains on a single line with a single direction of movement (e.g. on a double-track railway), which is a common operating plan in Europe and urban North America. The time horizon of the model then begins with the train and location of the primary delay, and ends with the return of the last train to schedule within the allowed service parameter (delay threshold).

Table 1 presents the terms and labels for sets and parameters in the model. Without loss of generality, the timetable measurement points are called "stations", even though they can just as well be any geographic location where the train must adhere to the timetable. Subscripts $i$ and $s$ specify the train and station that the parameters refer to, respectively. When parameters are later used without a subscript, they are constant and identical for all trains in the formulation, so that $a$, for example, is the running time supplement constant and identical for every train between any pair of consecutive stations. $s^{*}{ }_{i}$ is a component of the settling time, and refers to the last station after a disruption at which train $i$ deviates from its planned timetable, assuming that the timetable is stable. $i^{*}$ is a companion component of the settling time, and refers to the last train at a given station after a disruption which deviates from its planned timetable. $\delta$ is the delay threshold, which means that delays below a specified magnitude will be ignored in the calculation of utility loss. It is common to contractually define a train as "late" in Denmark and many other countries according to a fixed threshold of delay (Hofman et al., 2006).

| S | The ordered set of stations of the analysis, $\{1,2, \ldots,\|\mathrm{~S}\|\}$ |
| :--- | :--- |
| $R$ | The ordered set of trains in the analysis, $\{1,2, \ldots, \mathrm{R} \mid\}$. Lower numbered trains <br> precede higher numbered trains. |
| $p$ | Primary delay |
| $d_{i, s}$ | Individual delay of train $i$ at station $s$ |
| $t_{i, s}$ | Minimum running time of train $i$ between stations $s-l$ and $s$ |
| $h_{i, s}$ | Minimum time separation (headway) between trains $i-1$ and $i$ at station $s$ |
| $\delta$ | Delay threshold |
| $a_{i, s}$ | Running time supplement of train $i$ between stations $s-l$ and $s$ |
| $b_{i, s}$ | Buffer time at station $s$ between trains $i-l$ and $i$ |
| $s^{*}{ }_{i}$ | Last delayed station for train $i$ |
| $i^{*}{ }_{s}$ | Last train delayed at station $s$ |
| $\varphi_{p}$ | Timetable settling time for delay $p$ |

### 4.1. Derivation of Cumulative Delay

The cumulative delay, $\Gamma$, represents the unweighted utility loss experienced by the railway service due to a disruption. It is the sum of all individual delays as measurement points in the timetable over the analysis horizon, and is presented in Equation (1).

$$
\begin{equation*}
\Gamma=\sum_{\substack{i \in R \mid d_{i, s} \geq \delta \\ s \in S}} d_{i, s} \tag{1}
\end{equation*}
$$

$p$ defines the primary delay, the delay of the first train at the first station, thus $d_{l, l}=p$. This delay will propagate to, or impose secondary delays on, the following trains through the expression of individual delay given in (2)
$d_{i, s}=\max \left\{\left(d_{i, s-1}-a_{i, s}\right),\left(d_{i-1, s}-b_{i, s}\right), 0\right\} \forall i>1, s>1$
$d_{1, s}=\max \left\{\left(d_{i, s-1}-a_{i, s}\right), 0\right\} \forall s>1$
$d_{i, 1}=\max \left\{\left(d_{i-1, s}-b_{i, s}\right), 0\right\} \forall i>1$
$d_{1,1}=p$

What Equation (2) represents is that every train delay incidence is defined either by delay originating with the train or by secondary delay imposed by another train. A train is protected by or may recover from its own delay by the timetable supplement $a_{i, s}$. A train is likewise shielded from the obstruction of a preceding train by the buffer $b_{i, s}$. A train then experiences a delay if either or both of these limits are exceeded, and the larger of the two values determines the functional train delay. The first train in the series of the analysis horizon $(i=1, s>1)$ is only affected by its own delay, and trains at the location of the primary delay $(s=1, i>1)$ are only affected by preceding trains.

### 4.2. Relaxed Formulation for Homogeneous Trains and Line Segments

Consider now a railway line composed of equitemporal (not necessarily equidistant) line segments and identical train compositions dynamic performance, such that $a, b, h$, and $t$ are constant values throughout the formulation. These are not wholly unrealistic assumptions for suburban railways or metros operating periodic timetables. Equation (2) becomes Equation (3)

$$
\begin{gather*}
d_{i, s}=p-(s-1) a-(i-1) b \forall i \geq 1, s \geq 1 \mid p \geq(s-1) a+(i-1) b \\
d_{i, s}=0 \text { elsewhere } \tag{3}
\end{gather*}
$$

The conditions linking $p, i, s, b$, and $a$ define a two-dimensional region where individual train delay, $d_{i, s}$, is positive. Outside this region, trains have returned to their original planned timetable, or recovered. This region, where trains are recovering or settling back into their planned timetable, is defined the recovery region. The recovery region in turn defines the settling time of the initial delay, $p$.

Consider that a positive value of $\delta$ further defines a recovery region only where $d_{i, s} \geq \delta$, and Equations (4) and (5) yield solutions of Equation (3) for the boundary values of the recovery region in dimensions of number of stations and number of trains. The extreme points of the recovery region are defined at $s^{*}{ }_{1}$ and $i^{*}{ }_{1}$. Since both $i$ and $s$ must be integers, these solutions are returned as floor functions. When the primary delay is both less than the running time supplement plus the delay threshold, and less than the buffer time plus the delay threshold, the recovery region is undefined because no active recovery occurs.

$$
\begin{align*}
& \left.s_{i}^{*}=\left\lfloor\frac{p+b-\delta}{a}-i \frac{b}{a}\right\rfloor+1 \right\rvert\, p \geq a+\delta  \tag{4}\\
& \left.i_{s}^{*}=\left\lfloor\frac{p+a-\delta}{b}-s \frac{a}{b}\right\rfloor+1 \right\rvert\, p \geq b+\delta \tag{5}
\end{align*}
$$

Figure 4-1 depicts the boundaries of the recovery region. The diagonal boundary has an approximate slope of $a / b$. The settling time will be the greater of the times necessary to traverse the recovery region, either along the station axis $(s)$ or the train axis $(i)$, as shown in Equation (6).

| $\varphi_{p}=\max \left\{(t+a) s_{1}^{*},(h+b) i_{1}^{*}\right\}$ | $(6)$ |
| :--- | :--- |



Figure 4-1: Recovery region bounds

### 4.3. Symmetric Systems

Consider the system where $c=a=b$, which will be called "symmetric", because not only are the control values of timetable supplement and buffer equal, but the ratio $a / b$ implies the region is symmetric with respect to the number of trains and stations. Then the recovery region is defined only by the primary delay and a single parameter, $c$, as shown in Equations (8) and (7). Checking the values of $S^{*}{ }_{1}$ and $i^{*}{ }_{1}$, it can be seen that indeed the region is symmetric with an equal number of trains and stations.

| $\left.s_{i}^{*}=\left\lfloor\frac{p-\delta}{c}\right\rfloor-i+2 \right\rvert\, p \geq c+\delta$ | (7) |
| :---: | :--- |
| $\left.i_{s}^{*}=\left\lfloor\frac{p-\delta}{c}\right\rfloor-s+2 \right\rvert\, p \geq c+\delta$ |  |

### 4.4. Calculation of Cumulative Delay

The resulting summations for cumulative delay, $\Gamma$, are shown for the general case in Equation (9), and for the symmetric case in Equation (10). The summation in Equation (9) operates first in the
dimension of the stations for individual trains, returning the cumulative delay recorded on one train's whole itinerary, and then it sums the cumulative delay across the individual train itineraries.

$$
\begin{gather*}
\Gamma=\sum_{\substack{i \in\left\{1,2, \ldots,\left\lfloor\frac{p+a-\delta}{b}-s \\
s \in\left\{1,2, \ldots\left[\frac{p-\delta}{a}\right]+1\right\}\right.\right.}} p+a+b-s a-i b  \tag{9}\\
\Gamma_{\mathrm{S}}=\sum_{\substack{i \in\left\{1,2, \ldots,\left\lfloor\left\lfloor\frac{p-\delta}{c}\right]-s+2\right\} \\
s \in\left\{1,2, \ldots\left[\frac{p-\delta}{c}\right]+1\right\}\right.}} p-c(s+i-2) \tag{10}
\end{gather*}
$$

The floor functions in these summations prevent them from resolving into manageable functions. If the floor functions are relaxed (removed), the summations resolve into the following polynomials: the general case in Equation (11), and the symmetric case in Equation (12).

$$
\begin{array}{r}
\Gamma=\frac{p^{3}}{6 a b}+\frac{3(a+b) p^{2}}{12 a b}+\frac{\left(a^{2}+3 a b+6 b \delta-6 \delta^{2}\right) p}{12 a b} \\
+\frac{-a^{2} \delta+9 a b \delta-3 a \delta^{2}-9 b \delta^{2}+4 \delta^{3}}{12 a b} \\
\Gamma_{s}=\frac{p^{3}}{6 c^{2}}+\frac{p^{2}}{2 c}+\frac{\left(2 c^{2}+3 c \delta-3 \delta^{2}\right) p}{6 c^{2}}+\frac{4 c^{2} \delta-6 c \delta^{2}+2 \delta^{3}}{6 c^{2}} \tag{12}
\end{array}
$$

Naturally there is a question of how much error is introduced by relaxing the floor functions.
Figure 4-2 shows that the difference, with and without the floor function, is very small for a delay threshold of 3 (minutes) and supplements and buffers of one or two (minutes).


Figure 4-2: Plot of symmetric cumulative delay for $\delta=3$ and $c=1$ (top curve) and $\delta=3$ and $c=2$ (bottom curve).

Summations with and without floor function overlaid.

### 4.5. Inferences from the Polynomial Function

Figure 4-3 presents the contour of a symmetric system and shows that while timetable "slack", in the form of supplements and buffers, certainly is valuable in damping the damage of primary delays, its incremental value quickly declines. The figure suggests that supplements and buffers can be applied excessively, wasting resources without accomplishing proportional reductions in system delay.


Differential calculus of aggregate delay against timetable slack expresses the marginal reduction of aggregate delay given by increments of timetable slack. The symmetric case is represented

Figure 4-3: Contour of symmetric case for $\delta=3$
by equation (13) and depicted in Figure 4-4.

$$
\begin{equation*}
\frac{\partial \Gamma_{s}}{\partial \mathrm{c}}=-\frac{p^{3}}{3 c^{3}}-\frac{p^{2}}{2 c^{2}}+p\left(\frac{4 c+3 \delta}{6 c^{2}}-\frac{2 c^{2}+3 c \delta-3 \delta^{2}}{3 c^{3}}\right)+\frac{8 c \delta-6 \delta^{2}}{6 c^{2}}-\frac{4 c^{2} \delta-6 c \delta^{2}+2 \delta^{3}}{3 c^{3}} \tag{13}
\end{equation*}
$$

Being (12) a polynomial form also in $c$, its first derivative keeps characteristic of continuity. The figure shows that excessive timetable slack does not contribute to reduce the aggregate line delay, reaching a plateau of near-zero marginal decrement. The plateau corresponds to timetable slack large enough to prevent any delay propagation to following trains or downstream stations. In such a system, the only delay registered is the primary delay assigned to the first train at the first station, which is unavoidable by means of timetable slack.


Figure 4-4: Contour of marginal reduction of aggregate delay against timetable slack and primary delay, with $\delta=3$.

Further evidence that there is an optimal value for timetable supplement and buffer is offered in Figure 4-5. This contour plots Equation (6) with a running time between stations of $t=5$, a minimum headway of $h=5$, and a delay threshold of $\delta=3$. Note that both timetable supplement $(a)$ and buffer $(b)$ must be present to control settling time, and that excessive values of either actually worsen the settling time. This figure also suggests the optimal ratio $a / b=1$, visible by the median wrinkle of the settling time envelop.


Figure 4-5: Contour of settling time with $\delta=5, h=5, \delta=3$ and $p=5$.

## 5. CASE STUDY

In this section a contemporary suburban railway in Denmark is simulated and comparisons are made between the measured and theoretical system delay. The simulation is performed in OpenTrack (Nash and Huerlimann, 2004). The subject line is the Hillerød suburban railway on its northern segment from Hellerup to Hillerød (29 km.). On this segment there are eleven stations inclusive of the terminal, Hillerød, and the junction Hellerup. Hellerup is not the end of the line. All trains continue through Hellerup, through Copenhagen, and on to destinations much further south of Copenhagen.

On this line there are two services with distinct stopping patterns, and this is where the case study deviates significantly from the theoretical model. Service "A" is an express to Hillerød, which does not stop at the $2^{\text {nd }}, 3^{\text {rd }}, 6^{\text {th }}$, and $7^{\text {th }}$ stations from Hellerup. Service " $E$ " is a local service that terminates at the $8^{\text {th }}$ station (Holte, 11 km .). Figure 5-1 shows the simulation model track block structure and the train


Figure 5-1: Simulation model of Hillerød Suburban railway in OpenTrack showing track blocks and speed profiles of "A" service (blue) and "E" service (purple)

## Hellerup - Hillerød



Figure 5-2: Graphical timetable (stringline) for Hillerød Suburban railway. Schedule in black, simulated operation colored.
performance speed profiles for both services. Figure 5-2 shows the graphical timetable or stringline diagram for the two services between Hellerup and Hillerød. It should be further noted that the heterogeneous service pattern of Figure 5-2 deviates from the homogeneous service pattern in the theoretical model of Figure 3-3.

### 5.1. Experimental Design of Simulation

Two simulation analyses are presented: primary delays experienced by the A service and primary delays experienced by the E service. In each case, primary delays are simulated at the Hellerup station from a uniform distribution of $[0,600]$ seconds, and 100 replications are sampled. Only northbound traffic to Hillerød is studied. Cumulative delay is measured across both services, A and E, on the line.

The experimental sample deviates significantly from the theoretical ideal. Unlike the theoretical model of Section 4, the stations in this sample are not uniformly distributed. As previously described, the train sample consists of a mix of two different service patterns, stopping at different stations. Further,
the supplement and buffer times are not uniformly distributed. Table 2 presents the actual timetable supplements present in the simulated Hillerød Suburban railway. Table 3 presents the buffers between train paths.

Table 2: Timetable supplements on Hillerød Suburban railway as simulated, seconds

| Service | $2^{\text {nd }}$ | $3^{\text {rd }}$ | $4^{\text {th }}$ | $5^{\text {th }}$ | $6^{\text {th }}$ | $7^{\text {th }}$ | $8^{\text {th }}$ | Stations from Hellerup |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  | $9^{\text {th }}$ | $10^{\text {th }}$ | $11^{\text {th }}$ |
| $A$ |  |  |  | 82 |  |  | 40 | 67 | 45 | 62 |
| E | 21 | 20 | 83 | 10 | 64 | 9 | 22 |  |  |  |

Table 3: Buffers between pairs of trains (first train following second train), seconds

| Pairing | $2^{\text {nd }}$ | $3^{\text {rd }}$ | $4^{\text {th }}$ | $5^{\text {th }}$ | $6^{\text {th }}$ | $7^{\text {th }}$ | $8^{\text {th }}$ |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| A follow $E$ | 83 | 236 | $9^{\text {th }}$ | $10^{\text {th }}$ | $11^{\text {th }}$ |  |  |
| E follow $A$ | 442 | 168 | 367 |  |  |  |  |

Further, it should be noted that the timetable supplement, $a$, and headway buffer, b , are severely asymmetric. They are far from the recommended level of $a=\mathrm{b}$ of Section 4.5. The average buffer for all six measures of services A and E is 240.5 seconds. The average supplement at each of ten destination stations is 43,75 seconds. The delay threshold, $\delta$, is zero, and all delays of any magnitude are included in the cumulative delay. These extremes lead to induced error in the polynomial approximation of Equation (9). In this case, the polynomial is continuously undervaluing the cumulative delay as seen in Figure 5-3.


Figure 5-3: Difference from polynomial function, Equation (11), to exact summation, Equation (9) for parameters in this section (values in seconds).

### 5.2. Results of the Simulation

The simulation results from OpenTrack are summarized in Figure 5-4. It is hypothesized that the delays experienced by service E result in significantly lower cumulative system delays than those experienced by service A because of the difference in the headway buffer outside of Hellerup. Primary delays on service A greater than 83 seconds propagate to the following trains, whereas service E can receive primary delays up to 442 seconds without triggering delay propagation. Note that the polynomial approximation is lower than the simulated result from both services up until primary delays of magnitude 450 seconds and greater, where the polynomial occupies the interval between the $A$ and E service results. At primary delays of 300 seconds, the polynomial underestimates the cumulative delay caused by primary delays on the A service by $39 \%$.

To test the proposal that asymmetry between $a$ and $b$ is a factor in the polynomial fit to experimental results, another cycle of simulations is performed with a 30 second increase of the supplement throughout the timetable, so that $a$ is now 73.75. The buffer remains unchanged. The correlation between the polynomial and the E service result is now very strong, as shown in Figure 5-5, and the higher A service cumulative delay is likely because of the close headway between A and E services leaving Hellerup.


Figure 5-4: Comparison of results from simulations of delays on $A$ and $E$ lines with estimated delay from polynomial function.


Figure 5-5: Fit of polynomial to simulation results with $a=73.75$ and $b=240.5$

## 6. Model Extension: a Universal Polynomial Form For Primary Delays

## at UnSpecified Stations

The polynomial model presented in section 4 rests on the selection of different equations in cases of full, partial, or absent recovery before the last station of the line and before the last train scheduled. In real operation, incidents occur at different locations on a railway line, and trains can
experience primary delays at any station. In this section, a universal equation is derived as an expansion of the previous cases. The equation is valid in any of the mentioned cases, and can be used to analyze the effects of primary delays at different locations on a railway line, or selecting specific areas of interest. The delay recovery domain is split in regions, and the polynomial form is integrated with logical functions to include or exclude specific sections from the delay summation domain.

### 6.1. Study region and Delay recovery region

Delay propagation studies can be limited to limited sections of the railway lines. For example, the most congested section of railway lines within a node could result of greater interest than marginal lines. In other cases, the lines can be divided in different homogeneous study regions, suburban railway networks can be split in sections according to the scheduled traffic volume and average headway between trains. The study region is the domain of interest in the two dimensions of the model, stations and trains, and is defined by the number of stations $S$ and the number of trains $R$. The recovery region, defined in section 4, is the set of trains $i$ and stations $s$ where the individual delay exceeds a given threshold $\delta$. The recovery region shapes as a triangle, which vertices are the first train at the first station, where primary delay is generated $(1,1)$, the last delayed train at first station $\left(i_{1}^{*}, 1\right)$, and the last delayed station for the first train $\left(i_{1}^{*}, 1\right)$. The cumulative delay is the summation of the individual delays of trains at stations within the study region, where the individual delay is greater than or equal to the delay threshold, so the summation domain extends to the area resulting from the overlap of study region and recovery region. The study region and the recovery region overlap depends on the values of primary delay $p$, running time supplement $a$, headway buffer $b$, delay threshold $\delta$, the number of stations $S$ and the number of trains $R$ in the study region. Keeping the definition of cumulative delay as the overall effect of a primary delay over the area of interest, namely the study region, the equations proposed in section 4 are modified so that the cumulative effect is calculated assuming the existence of enough trains and station to fully recover the primary delay. The result is defined unbounded cumulative delay, the recovery region can extend limitless. To reduce the cumulative delay to only include the study
region, individual sub-recovery areas ae identified, and removed from the unbounded cumulative delay if the system meats specific requirements, described below.


Figure 6-1: Study region and recovery region overlap

Figure 6-1 depicts a study region entirely included in the recovery region. This case shows all the possible delay sub-recovery areas to include in the general formulation for the cumulative delay. The individual areas and the specific equations are explained in the following sections. The general formulation of the cumulative delay over individual areas keeps the same form of (9), as a summation of individual delays, but the summation domains are different for each area.

### 6.2. Unbounded cumulative delay and sub-recovery regions

If no restriction is imposed to the number of trains and stations, the delay will always be completely recovered. The formulation of unbounded cumulative delay corresponds to equation (9), where the summation domain extends from the point of primary delay to the last delayed train and station.

When the study region terminates before the Last Delayed Station, namely $s_{1}^{*}>S$, the unbounded cumulative delay must be reduced by an amount corresponding to the cumulative delay in the exceeding area, called here Line progress sub-recovery. The summation domain extends in this case from the first station outside the study region $S+1$ to the last delayed station for the first train $s_{1}^{*}$.

$$
\begin{align*}
& \Gamma_{e}=\sum_{\substack{i \in\left\{1, \ldots, i_{s}^{*}\right\} \\
s \in\left\{S+1, \ldots, s_{1}^{*}\right\}}} p+a+b-s a-i b  \tag{14}\\
& =\frac{p^{3}}{6 a b}+\frac{p^{2}(3 a+3 b-6 a S)}{12 a b}+\frac{p\left(a^{2}+3 a b-6 a^{2} S-6 a b S+6 a^{2} S^{2}+6 b \delta-6 \delta^{2}\right)}{12 a b} \\
& +\frac{-a^{3} S-3 a^{2} b S+3 a^{3} S^{2}+3 a^{2} b S^{2}-2 a^{3} S^{3}-a^{2} \delta+9 a b \delta-6 a b S \delta-3 a \delta^{2}-9 b \delta^{2}+6 a S \delta^{2}+4 \delta^{3}}{12 a b}
\end{align*}
$$

Similarly to the line progress sub-recovery, when the study region terminates before the Last Delayed Train, namely $i_{1}^{*}>R$, the unbounded cumulative delay must be reduced by an amount corresponding to the cumulative delay in the exceeding area, called here Train fleet sub-recovery.

In the general case, the mentioned exceeding areas can overlap. This happens when the delay cannot be recovered by any of the trains in the study region, before the last station. In these cases, the overlapping exceed should be reintroduced to prevent the subtraction twice. The last delayed station for the first train outside the study region is a new type of boundary for the summation domain. The formulation is derived from (4) and it is defined as $s_{o}^{*}$

$$
\begin{equation*}
\left.s_{o}^{*}=s_{R+1}^{*}=\left\lfloor\frac{p+b-\delta}{a}-(R+1) \frac{b}{a}\right\rfloor+1 \right\rvert\, p \geq a+\delta \tag{15}
\end{equation*}
$$

The summation domain extends from the first station outside the study region to $s_{o}^{*}$.

### 6.3. Universal formulation

The existence of the individual sub-recovery areas mentioned in this section depends on the relation between system parameters, $R, S, a, b, \delta$, and the primary delay $p$. The universal formulation proposed in this section includes logical controls on the specific delay sub-recovery to select only the
regions that are active. The logical controls include the formulation from individual areas only if their specific dimension is positive.

The line progress sub-recovery region is controlled by $l_{e}$, the number of excluded stations, defined as the difference between the last station in the study region and the last delayed station for the first train. Similarly, line progress sub-recovery region is controlled by $l_{h}$, the number of excluded train paths, defined as the difference between the last train in the study region and the last delayed train at the first station. The sub-recovery overlap is controlled by $l_{o}$, the overlap length between the last delayed station for the first train outside the study region and the last station in the study region.

| $l_{e}=s_{1}^{*}-S=1-S+\frac{p-\delta}{a}$ | $(16)$ |
| :---: | :--- |
| $l_{h}=i_{1}^{*}-R=1-R+\frac{p-\delta}{a b}$ | $(17)$ |
| $l_{o}=s_{o}^{*}-S=1-S+\frac{b+p-b(1+R)-\delta}{a}$ | $(18)$ |

Equation (19) is a closed form function that returns the cumulative delay on a railway line as a function of the primary delay $d(p)$, in any condition of recovery, given the system variables $a, b, \delta, R$, $S$.

$$
\begin{equation*}
\Gamma=\Gamma_{p} * \frac{\max \left(s_{1}^{*}, 0\right)}{s_{1}^{*}}-\Gamma_{e} * \frac{\max \left(l_{e}, 0\right)}{l_{e}}-\Gamma_{h} * \frac{\max \left(l_{h}, 0\right)}{l_{h}}+\Gamma_{o} * \frac{\max \left(l_{o}, 0\right)}{l_{o}} \tag{19}
\end{equation*}
$$

The logical control on the unbounded cumulative delay is introduced in this formulation as a replacement of the constraint $p \geq(s-1) a+(i-1) b$ in (3), to simplify the formulation.

## 7. MODEL DISCUSSION

The polynomial form proposed in section 4 is third degree. It is a cubic function of the primary delay, if the measurement horizon extends fully over the length of line and number of trains necessary for the disturbance to be absorbed by the timetable running time supplements and the headway buffers (the recovery region). This agrees with the earlier findings of Hasegawa et al.(1981). This paper differs from Hasegawa in that it explicitly models the discrete summation of delays, considering three parameters: supplement, buffer, and threshold for measurable delay. This results in a polynomial function of primary delay instead of the purely cubic function of delay in Hasegawa. If the measurement horizon is restricted to less than the full recovery region, the polynomial reduces to second degree and over very short horizons it is linear.

The polynomial is an approximation of the discrete summation, and is robust over a wide range of parameters. Investigation of the contour of the polynomial finds that, in the examples considered, running time supplement and headway buffer should be equal values. Further, excessive values of running time supplement and headway buffer may result in poor timetable stability. When supplement and buffer are identical, the cumulative delay function simplifies further to a form that is easy to work with.

The model extension presented in section 6 describes the more generic case where primary delays occur on unspecified trains at generic stations, and allows a more extended analysis of line delays, and of the influence of systems parameters such as the delay threshold. It is possible, in this way, to inference the effect of strategic decisions for performance assessment in transportation contracts. The delay threshold allows for some flexibility in the planning phase, under the assumption that small delays are not perceived by the passengers. The calibration of the delay threshold in service contracts between service providers and transport authorities has influence on daily operation, and different strategies in the delay threshold dimensioning lead to different dispatching strategies to pursue the measured punctuality. Punctuality penalties are a relevant share of operations budget of transportation companies, especially in cases where a performance regime is applied. For example, the

European Performance Regime (2013) draws the guidelines for performance management in the European countries, and every single minute of delay of a train can cost to the service contractor up to $2 €$ (Rete Ferroviaria Italiana, 2015). The relation between punctuality measurement methods, delay thresholds, and distribution of running time supplement in train paths across several countries in Europe is described by Schittenhelm (2011). Suburban and regional railway services in Europe admit thresholds between 3 and 5 minutes, whereas long distance services are allowed to reach from 5 to 15 minutes of delay before penalties are applied. The closed form introduced in (19) allows to infer the effect of different values of the delay threshold in a given timetable, and to evaluate the most appropriate value of delay threshold $\delta$.

Figure 7-1 presents the cumulative delay function and shows the effect of different delay threshold strategies. Intuitively, no cumulative delay is recorded for primary delays smaller than the delay threshold. The figure shows also the range of effectiveness of the delay threshold. The cumulative delay measure is dumped in situations of full or partial recovery.


Figure 7-1: Cumulative delay as a function of primary delay on a railway line with different values of delay threshold. $S=11$ stations, $R=5$ trains, $a=1 \mathrm{~min}, b=1 \mathrm{~min}$

The delay threshold becomes ineffective as soon as all the trains are delayed enough to be included in the summation, which corresponds to meeting the condition

| $d_{R, S}(p)<\delta$ | (20) |
| :---: | :--- |
| $p<\delta+a(S-1)+b(R-1)$ |  |

Based on the formulation provided, service contractors and transport authorities can evaluate a convenient value for delay threshold that meats the measured delay distributions and reduces operational costs improving the measured punctuality.

The novelty of this analytical approach is the ability to account explicitly for the running time supplement and the headway buffers at the same time. Previous approaches (Landex, 2008; Salido et al., 2012) only considered delay propagation to the following trains, leaving out the spatial dimension. Similarly to the model presented in this paper, Pyrgiotis (2012) proposed an airport network model to propagate delays over the air traffic and over the single aircraft path. Railway traffic is, though, much more constrained than air traffic, and in most railway lines the train sequences do not change between stations. Pyrgiotis applied queuing theory to model the interferences between aircrafts, whereas the model proposed here considers explicitly the headway buffer between trains. The two models are alike in the explicit formulation of the delay reduction given by the running time supplement along the train path or aircraft roster.

Furthermore, the polynomial formulation proposed in this paper provides insight in the relation between primary and secondary delays. This is not possible using other queuing models proposed in the past (Huisman and Boucherie, 2001) because the primary delay are accounted for implicitly in the variability of free running times. Huisman and Boucherie focus on the secondary delays induced by the speed differences in the timetable, more than the relation between primary and secondary delays. Other past models for delay propagation do not provide a functional relationship (Goverde, 2010, 2007), or use complex methods, such as Colored Petri Net, which can effectively only represent few stations (Zhu and Schnieder, 2000).

## 8. CONCLUSION

This paper contributes to the literature with an analytic closed formed function that returns cumulative railway line delay as a function of a single initial primary delay. The function can predict the output of a simulation model, with high confidence, thus being a fast analytic alternative to resource consuming simulation models. The polynomial function may thus be used for an initial screening of possible timetables, leaving simulation to later parameter fine-tuning.

The polynomial function model is shown to be robust to violations of the basic assumptions, and the form holds valid with heterogeneous running time supplements and headway buffers. The universal formulation introduced in this paper is non-specific to stations and trains, which allows to analyze the effects of primary delays occurring at any location on the railway line, and at any time of operation. Thanks to the closed formulation, it is possible to quickly evaluate the effect of different strategic choices on contract performances between operators and transport authorities. Operation design tools such as the delay threshold, running time supplement, and headway buffer, can be designed accurately investigating the expected cumulative delay with analytical approach. Differential calculus of the polynomial form shows that a limited amount of timetable slack is effective, whereas larger slack does not contribute in performance improvement and results in extending scheduled running times and delay recovery times.

Further development of this function should consider the summation and interaction of multiple primary delays at different locations. In addition, the application of the model to heterogeneous timetables should be deepened. The closed-form model could be inverted to calculate average timetable supplement and buffer time. This means that given a desired punctuality and stability of service, the necessary timetable supplement may be estimated from this function.

Acknowledgement : This work was funded by a Dean Grant from the Technical University of Denmark (DTU) and by the Danish Innovation Fund through the IPTOP project (Integrated Public Transport Optimisation and Planning).

## References

Barron, A., Melo, P., Cohen, J., Anderson, R., 2013. Passenger-Focused Management Approach to Measurement of Train Delay Impacts, in: Transportation Research Board, 92nd Annual Meeting. Transportation Research Board of the National Academies, pp. 46-53. doi:10.3141/2351-06

Caimi, G., Burkolter, D., Herrmann, T., Chudak, F., Laumanns, M., 2009. Design of a railway scheduling model for dense services, in: Networks and Spatial Economics. Springer US, pp. 2546. doi:10.1007/s1 1067-008-9091-6

Carey, M., 1999. Ex ante heuristic measures of schedule reliability. Transp. Res. Part B Methodol. 33, 473-494. doi:10.1016/S0191-2615(99)00002-8

Carey, M., Kwieciński, A., 1994. Stochastic approximation to the effects of headways on knock-on delays of trains. Transp. Res. Part B 28, 251-267. doi:10.1016/0191-2615(94)90001-9

Cerreto, F., 2015. Micro-simulation based analysis of railway lines robustness, in: 6th International Conference on Railway Operations Modelling and Analysis. International Association of Railway Operations Research, Tokyo, pp. 164-1-164-13.

European Performance Regime project, 2013. Handbook for the European Performance Regime (EPR) Guidelines for actual and potential users. Vienna.

Gibson, S., Cooper, G., Ball, B., 2002. Developments in transport policy: The evolution of capacity charges on the UK rail network. J. Transp. Econ. Policy 36, 341-354.

Ginkel, A., Schobel, A., 2007. To Wait or Not to Wait? The Bicriteria Delay Management Problem in Public Transportation. Transp. Sci. 41, 527-538. doi:10.1287/trsc.1070.0212

Goverde, R., 2010. A delay propagation algorithm for large-scale railway traffic networks. Transp. Res. Part C Emerg. Technol. 18, 269-287. doi:10.1016/j.trc.2010.01.002

Goverde, R., 2007. Railway timetable stability analysis using max-plus system theory. Transp. Res. Part B Methodol. 41, 179-201. doi:10.1016/j.trb.2006.02.003

Goverde, R., Hansen, I.A., 2013. Performance indicators for railway timetables, in: 2013 IEEE International Conference on Intelligent Rail Transportation Proceedings. IEEE, pp. 301-306. doi:10.1109/ICIRT.2013.6696312

Goverde, R., Meng, L., 2011. Advanced monitoring and management information of railway operations. J. Rail Transp. Plan. Manag. 1, 69-79. doi:10.1016/j.jrtpm.2012.05.001

Haith, J., Johnson, D., Nash, C., 2014. The case for space: the measurement of capacity utilisation, its relationship with reactionary delay and the calculation of the capacity charge for the British rail network. Transp. Plan. Technol. 37, 20-37. doi:10.1080/03081060.2013.844906

Harker, P.T., Hong, S., 1994. Pricing of track time in railroad operations: An internal market approach. Transp. Res. Part B 28, 197-212. doi:10.1016/0191-2615(94)90007-8

Hasegawa, Y., Konya, H., Shinohara, S., 1981. Macro-Model on Propagation-Disappearance Process of Train Delays. Railw. Tech. Res. Institute, Q. Reports 22, 78-82.

Hofman, M., Madsen, L., Groth, J.J., Clausen, J., Larsen, J., 2006. Robustness and Recovery in Train Scheduling - a simulation study from DSB S-tog a / s. 6th Work. Algorithmic Methods Model. Optim. Railw.

Huisman, T., 2002. Forecasting delays on railway sections. Adv. Transp. 13, 779-786.

Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. Transp. Res. Part B Methodol. 35, 271-292. doi:10.1016/S0191-2615(99)00051-X

Jensen, L.W., Landex, A., Nielsen, O.A., Kroon, L.G., Schmidt, M., 2017. Strategic assessment of capacity consumption in railway networks: Framework and model. Transp. Res. Part C Emerg. Technol. 74, 126-149. doi:10.1016/j.trc.2016.10.013

Landex, A., 2008. Methods to estimate railway capacity and passenger delays. Technical University of Denmark (DTU).

Mattsson, L.-G., 2007. Railway Capacity and Train Delay Relationships. Crit. Infrastruct. Adv. Spat. Sci. doi:10.1007/978-3-540-68056-7_7

Meester, L.E., Muns, S., 2007. Stochastic delay propagation in railway networks and phase-type distributions. Transp. Res. Part B Methodol. 41, 218-230. doi:10.1016/j.trb.2006.02.007

Nash, A., Huerlimann, D., 2004. Railroad simulation using OpenTrack. Comput. Railw. IX 45-54.

Olsson, N.O.E., Haugland, H., 2004. Influencing factors on train punctuality-results from some Norwegian studies. Transp. Policy 11, 387-397. doi:10.1016/j.tranpol.2004.07.001

Parbo, J., Nielsen, O.A., Prato, C.G., 2016. Passenger Perspectives in Railway Timetabling: A Literature Review. Transp. Rev. 36, 500-526. doi:10.1080/01441647.2015.1113574

Parbo, J., Nielsen, O.A., Prato, C.G., 2014. User perspectives in public transport timetable optimisation. Transp. Res. Part C Emerg. Technol. 48, 269-284. doi:10.1016/j.trc.2014.09.005

Pyrgiotis, N., 2012. A Stochastic and Dynamic Model of Delay Propagation Within an Airport Network For Policy Analysis. Massachusetts Institute of Technology.

Rete Ferroviaria Italiana, 2015. Prospetto Informativo della Rete 2014.

Salido, M.A., Barber, F., Ingolotti, L., 2012. Robustness for a single railway line: Analytical and simulation methods. Expert Syst. Appl. 39, 13305-13327. doi:10.1016/j.eswa.2012.05.071

Salido, M.A., Barber, F., Ingolotti, L., 2008. Robustness in railway transportation scheduling, in: Proceedings of the World Congress on Intelligent Control and Automation (WCICA). IEEE, Chongqing, China, pp. 2833-2837. doi:10.1109/WCICA.2008.4594481

Schittenhelm, B.H., 2011. Planning With Timetable Supplements in Railway Timetables, in: Annual

Transport Conference at Aalborg University. trafikdage, Aalborg, DK.

TfL Investment Programme Management Office, 2008. Business Case Development Manual.

Toletti, A., 2016. Modelling customer inconvenience in train reschedul- ing. 16th Swiss Transp. Res. Conf. (STRC 2016).

Törnquist, J., 2007. Railway traffic disturbance management-An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon. Transp. Res. Part A Policy Pract. 41, 249-266. doi:10.1016/j.tra.2006.05.003

UIC, 2004. Leaflet 406 - Capacity.

Zhu, P., Schnieder, E., 2000. Modelling and Performance Evaluation of Railway Traffic Under Stochastic Disturbances. IFAC Proc. Vol. 33, 303-312. doi:10.1016/S1474-6670(17)38163-6


[^0]:    General rights
    Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

    - Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
    - You may not further distribute the material or use it for any profit-making activity or commercial gain
    - You may freely distribute the URL identifying the publication in the public portal

    If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

