Journal of Rail Transport Planning & Management 23 (2022) 100339

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

Understanding causes of unpunctual trains: Delay contribution and critical disturbances

Martin Joborn[a,b,*], Zohreh Ranjbar[a]

[a] Mobility and Systems, RISE Research Institutes of Sweden, Stora Gatan 36, SE-72212, Västerås, Sweden
[b] Communications and Transport Systems, ITN, Linköping University, Norrköping, Sweden

A R T I C L E I N F O

Keywords:
Railway punctuality
Critical disturbances
Delay contribution

A B S T R A C T

In this paper we define new concepts and metrics for improved understanding of causes to unpunctual trains. The metrics are denoted delay contribution and critical disturbance. Delay contribution can be interpreted as how much a specific disturbance contributes to the delay of a train and the critical disturbances can be interpreted as the disturbances that made the train become unpunctual. The metrics are applied in a test case with trains in southern Sweden. The results show that the metrics can provide a complementary view regarding causes to unpunctuality compared to standard methods and are able to pinpoint disturbances that made trains become unpunctual and separate them from disturbances that have less impact on the punctuality. The methods are useable in the continuous work to improve railway performance e.g., by prioritizing maintenance work that give best impact on punctuality.

1. Introduction

The perceived quality of railway transportation is important to make railways attractive, both for passenger and freight transportation. One quality aspect that is in focus in many countries is the punctuality of the trains, especially for passenger transportation. One of the overall targets of the EU research program Shift2Rail (Shift2Rail, 2015) is to reduce the unpunctuality by 50 percent, and in many railway systems, great efforts are made to improve the punctuality. The punctuality is also often in focus of the political debate, and there is a pressure to improve punctuality and make the railway more attractive so that the railway can increase its share of transport, as it is considered as an environmentally friendly means of transport. Since the punctuality is a widespread KPI for the railway performance, very often published and commented in the media, increasing the punctuality is very important both for the quality itself but also for the general perception of the quality.

Railway authorities and railway undertakings collect both punctuality data and data regarding disturbances. The overall target for the railway punctuality of passenger trains in Sweden is 95%. For several years, there have been efforts across the railway sector to improve the punctuality from about 90% to the target, but it has been very hard to achieve measurable and stable improvements. In 2019, the punctuality was 91% (Trafikanalys, 2020). Therefore, improved understanding of the problems regarding punctuality and its causes is very important. Also, it is crucial to be able to prioritize the measures that give the best effect.

In Sweden, a train is considered as punctual if the delay at the end station of the train is 5 min and 59 s or less. (In the official statistics, a train that is cancelled less than 24 h before its planned departure is considered as unpunctual. However, cancelled trains is

* Corresponding author. Mobility and Systems, RISE Research Institutes of Sweden, Stora Gatan 36, SE-72212, Västerås, Sweden.
E-mail addresses: martin.joborn@ri.se (M. Joborn), zohreh.ranjbar@ri.se (Z. Ranjbar).

https://doi.org/10.1016/j.jrtpm.2022.100339
Received 13 January 2022; Received in revised form 15 June 2022; Accepted 23 July 2022
Available online 12 August 2022

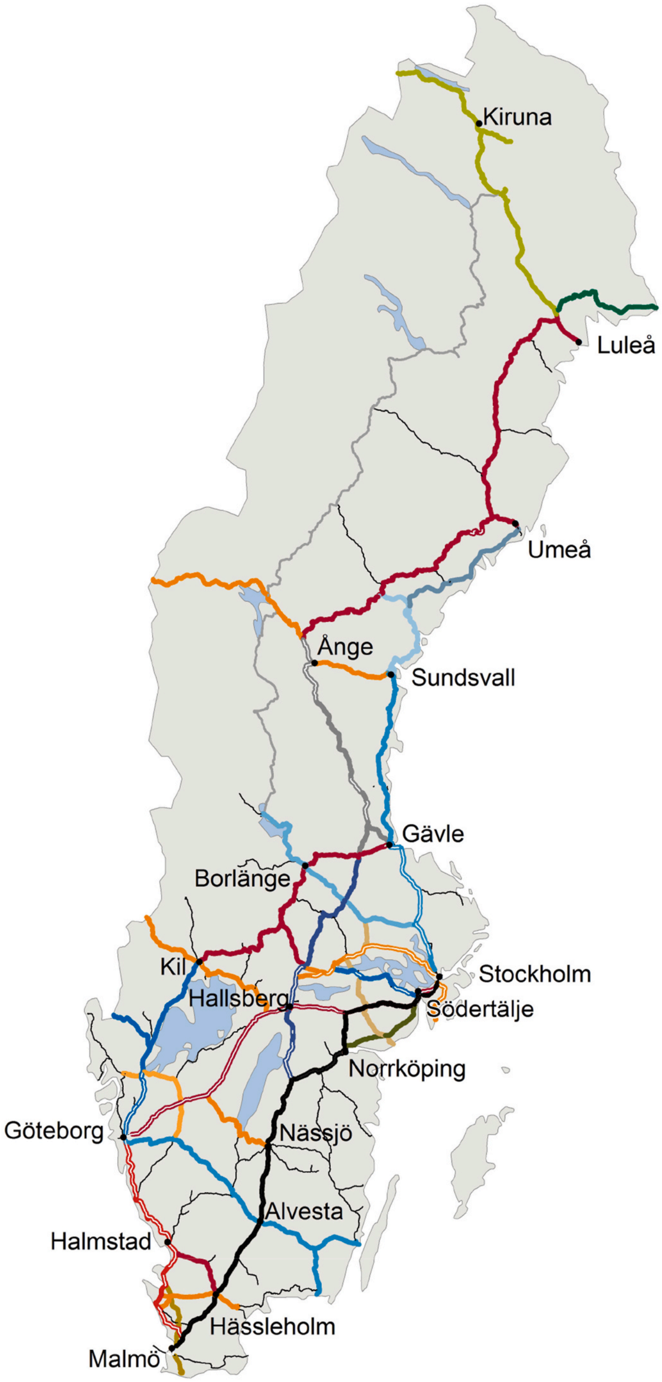

Fig. 1. Railway map over Sweden. Solid black line is the considered railway line Stockholm-Malmö.

another topic than the scope of this paper and are therefore not further discussed.) Other railway systems have other definitions of unpunctual trains, but it is common that punctuality is connected to the delay at the terminus. The punctuality threshold, i.e., how much the train can be delayed without being denoted punctual, varies in different railway systems.

Since the punctuality of a train is a binary metric (yes or no), reducing delays in general is not the same as improving the punctuality. To improve the punctuality, the important thing is to make more trains arrive earlier than the punctuality threshold. This means that a reduction of the delay has different impact – reducing the delay from 20 min to 15 has no impact on punctuality but reducing the delay from 7 min to 4 has an impact. Hence, there is also a difference in the disturbances' impacts on the punctuality, some disturbances have a direct impact while other disturbances do not. Thus, to improve the punctuality, it is important to identify

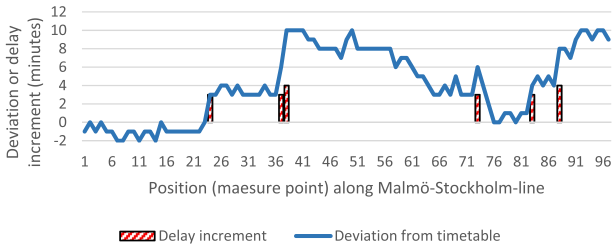

Fig. 2. Time deviation from timetable along the trip for train 542 from Malmö to Stockholm on 2020-09-29. X-axis corresponds to measure point numbers along the ride (1–97) and y-axis is deviation from timetable (in minutes). Striped (red) bars represent the delay increment of each registered disturbance, each of these equal to or larger than 3 min. Solid (blue) line represents the current (accumulated) deviation from the timetable. See also Table 1.

Table 1
Summary of the registered disturbances for train 542 2020-09-29 with final delay at terminus of 9 min. The total delay increment of the registered disturbances is 20 min. Positions correspond to measure point number along the trip. Columns with delay attribution codes are examples of the registered reasons to disturbances: Level 1 is the overall classification (where Secondary means that disturbance is consequence of delays on other trains) and level 2 is a refinement of level 1.

Registered disturbance ID	Position	Delay increment	Delay attribution code level 1	Delay attribution code level 2
754	24 (Hästveda)	3	Dispatching	Staff
598	37 (Moheda)	3	Infrastructure	Signal
598	38 (Lidnäs)	4	Infrastructure	Signal
673	73 (Åby)	3	Infrastructure	Maintenance
897	83 (Stjärnhov)	3	Dispatching	Prioritization
914	88 (Järna)	4	Secondary	Other train

the problems and take the measures that really have an impact on the punctuality. This paper aims at contributing to methods aimed at improving the understanding of how disturbances impact the punctuality.

In the remainder of this section, there is a problem description, focusing on the gathered data related to punctuality and the shortcomings of this, followed by a literature overview and the contribution of this paper. In section 2, we define new metrics and concepts aimed at improving the understanding how disturbances contribute to unpunctual trains and give an example how to calculate the metrics. In section 3, there is a case study in which the new metrics are used, and we show how these complement the traditional metrics. Finally, the paper is concluded, and the usage of results are discussed in section 4.

1.1. Problem description

In short, the problem is that, with the normally used methods, it is hard to evaluate if a disturbance has an impact on the punctuality or not since the gathered data does not include a link or correlation between disturbances and unpunctuality.

To describe the problem in more detail we study the Swedish situation looking at an example: Train 542 departing 2020-09-29 from Malmö to Stockholm. The train trip is 616 km and takes 4 h and 27 min (according to timetable). The arrival in Stockholm was 9 min after the schedule, thus the train is unpunctual. Along the train's trip from the start station to its end station, it passes 97 time measuring points. Each passage of a measure point is registered and the deviation between planned passing time according to the timetable and the actual passing time is calculated. Fig. 1 includes a map of the railway line and Fig. 2 illustrates the deviation from the timetable along the train ride. If a train is on or after its timetable and the deviation increases 3 min or more from one measure point to the following, a deviation remark is made. This is denoted as a registered disturbance, and the increase of the delay from the previous measure point to the current is denoted as the delay increment (all time registrations are in minutes, truncated towards zero). Disturbances with delay increment less than 3 min are not given a remark and also not considered as a registered disturbance. Each "problem" that cause a disturbance is given a unique ID. If one "problem" is the root cause of several disturbances, they are all assigned to the same ID. The train traffic controller at the traffic control centre at the infrastructure manager is responsible for (manually) assigning the ID and also for categorizing the cause of the delay by assigning a delay attribution code to the registered disturbance. The delay attribution codes are organized in three levels: On the top level (level 1), the attribution codes correspond to infrastructure, incidents and external factors, railway operators, dispatching and secondary causes, level 2 is a refinement of level 1, and level 3 is a refinement of level 2. In some cases, it can be difficult to assign the correct ID and delay attribution code to a registered disturbance. Still, the quality of the assignments is, by the Infrastructure Manager, considered as good enough and the data used for quality analysis in many different ways, and in this paper, we consider the data as correct. Note that data and statistics regarding delay increment does not include deviations smaller than 3 min. Table 1 summarize the registered disturbances and their delay increment in the example train, also indicated in Fig. 2. For trains running ahead of schedule – which is common among freight trains - there are no registered disturbances.

Punctuality is an indirect metrics in the sense that you cannot improve it at the place it is measured – at the end station. The improvements must be made along the train ride before the arrival at the terminus. Therefore, the analyses based on delay increment are the most important base for the sector's work to improve the punctuality. The yearly total delay increment is denoted as yearly disturbance time and is interpreted as a quality indicator of the overall railway system, and it is assumed that the punctuality will increase if the yearly disturbance time is decreased. The statistics regarding delay increment are broken down in many ways to increase the understanding and to prioritize improvements in the system. However, disturbance time and delay increment are indirect metrics with respect to punctuality and it is not certain that an improvement in the former metrics will result in an improvement in the punctuality, and it is not certain that the measures that have big impact on delay increment are efficient with respect to punctuality improvements.

Many trains travel for several hours and several disturbances may occur along the way. The example train show that during the trip the train may be behind the timetable on a part of the trip, then recover time and reduce the delay and get an increased delay again. Both a punctual and an unpunctual train may have had several disturbances. In that case, the connection between each disturbance and the final delay at the end station is very vague as the total delay increment is most often not the same as the final delay as a consequence of recovered time and small (unregistered) disturbances. Thereby, the delay increment registrations have a rather indirect connection to the punctuality. Further, the disturbances with the largest delay increment may not have the largest negative impact on the punctuality, since also small disturbances can cause trains to become unpunctual. These facts mean that using the delay increment as metric for punctuality improvements can give bad indications regarding what problems that are the most important to resolve.

In Fig. 2 it can be noted that a passenger train can be a few minutes ahead of its timetable. As long as a train is ahead of its timetable, no delay registrations are made even if they loose time compared to the timetable. For freight trains, that are commonly much ahead of their timetable, the methods presented in this paper should be adjusted to make them suitable.

Referring to the train 542 summarized Table 1, the train arrived its terminus 9 min after the timetable; thus, it is unpunctual. The total delay increment is 20 min. Hence, the correlation between the disturbances and the punctuality is very unclear just analysing the delay increments. Just studying the delay increment of the disturbances, you could conclude that disturbance 598 with a total delay increment of 7 min would be the most important but looking over the whole train ride we know that some hour after 598 occurred, the train had recovered all the delay and was on time; hence you can conclude that 598 did not contribute to the final delay and unpunctuality at all. Also, the delay increment does not capture the small disturbances or the recovered time. Such weaknesses in the analyses based on delay increment and that they can give erroneous indication of what is important in the work to improve punctuality is the main motivation for the work in this paper.

1.2. Literature review

As noted above, the punctuality threshold varies between different countries (Grechi and Maggi, 2018; Infrabel, 2018). In Germany, punctuality is measured at each stop and with a limit of 6 or 16 min for passenger and freight trains, respectively. In the Netherlands, the punctuality threshold is 3 min. In UK, starting from April 2020, punctuality is measured with 1 min resolution with the goal to improve performance and reduce delays (Network Rail, 2020). In Japan, a train is considered delayed if it is more than 1 min behind schedule. In Norway, local trains and long-distance trains are on time if they reach the final destination within 3 min and 59 s or 5 min and 59 s delay, respectively. In Australia, the limit is 5 min for both region and long services.

The punctuality of the railway system is very dependent on the construction of the timetable, see e.g. Harrod (2012) for an overview of the research connected to this area. There are many studies regarding construction of timetables to become more robust to disturbances (Andersson et al., 2015; Cerreto et al., 2016; Solinen et al., 2017).

There is also plenty of research which has been conducted to investigate and predict train delays. Kecman and Goverde (2014) present a method for dynamic estimation of process times for each train, based on actual delays and train interactions. Different modelling techniques has been used to understand which factors that influence train delays. For example, Marković et al. (2015) applied machine learning (ML) models to capture the relation between passenger train arrival delays and different characteristics of a railway system and Wen et al. (2020) used a deep learning model to predict train delays based on the interactions and delay propagation in a group of trains.

Delay propagation from one train to other trains are studied by different methods like simulations, empirical data, or theoretical approaches. Simulations have been used extensively to study delay propagation. Goverde (2010) and Lindfeldt (2015) studied the relationship between key factors, such as capacity utilization and the impact on random primary delays. Delay propagation algorithms have been proposed by Sørensen et al. (2017) and Huang et al. (2019). Also, machine learning models are applied by Wang and Zhang (2019) to investigate delay propagation and how the factors like the weather and the arrival times of two consecutive trains at the same station influence train delays. In Cerreto et al. (2016), found that the factors that primarily affect the train delay are the infrastructure layout and the performance and reliability of train vehicles.

Multiple regression models are used by Palmqvist et al. (2017) to quantify the influence of timetable, weather, and infrastructure-related variables on punctuality. Analytical models are used in Bergström and Krüger (2013) to identify the distribution of passenger train delay and to understand possible measures to improve reliability. The dwell time delays based on passenger loads has been studied in Palmqvist et al. (2020). Kristoffersson and Pyddoke (2019) studied relation between passenger punctuality with train punctuality.

Økland and Olsson (2021) used regression models on registered delay hours of disturbances to investigate influencing factors on punctuality in Norway. In Wen and Yang (2020), some delay indicators are investigated, and the impact of delays are visualized.

The work we have found that is most closely related to our paper, as described above, can be divided into three different areas:

delay prediction, delay propagation, and finding factors that influence delays. Compared to the work regarding delay prediction and delay propagation, our paper is aimed at post analysis of the actual train performance rather than creating models for train system performance. Compared to work regarding factors influencing the delays, our paper digs into more detail of each train's performance rather than utilizing aggregated data. By utilizing the details of each train's performance data, we create a link between individual disturbances, their contribution to the (total) delay, and the train punctuality, and we are not aware of other work that creates this kind of link. Further, this link also makes it possible to highlight the importance of the accumulation of very small deviations. The detailed analysis makes it possible to calculate new metrics which are related to how each specific disturbance affects the (total) delay and the punctuality for each specific train. As far as we are aware, it is also innovative to be able to explicitly separate important and less important disturbances from each other and to increase the understanding how each disturbance affects the punctuality.

1.3. Paper contribution

The contribution of this paper is a novel method to analyse the correlation between the disturbances and the delay during the train ride. The method includes two new metrics for measuring the "severeness" of a disturbance from a punctuality perspective: delay contribution and critical disturbance. Delay contribution is a metric to estimate how much each disturbance has contributed to the delay of the train, and critical disturbances point out if a disturbance makes the train "fall over threshold" for punctuality. The correlation between disturbances and delay is also useful to illustrate how delay from disturbances spread in the railway network. By utilizing the metrics, the aim is that railway companies can be "sharper" in the selection of measures to improve the punctuality.

The proposed method in the paper is designed for performance regimes where the punctuality is based on arrival to the terminating station. As mentioned above, several railway systems measure punctuality at each station. However, the method can easily be extended to be relevant also for railway systems that measure punctuality at all stations, which is discussed in the conclusion of the paper.

2. Delay contribution, critical disturbances and minor delays

As pointed out in the problem description above, it is not straightforward to tell how much each disturbance contributes to the final delay of a train. Further, it is not always easy to say if some disturbances are – from a punctuality perspective – more important than other disturbances. In this section, we define and describe new concepts and metrics aimed at overcoming this. In the next section, the new concepts are used in a case study, in order to verify that they can be used to increase the understanding of unpunctuality and its causes.

A basis in the defined concepts is to include consideration to recovered time and also to small (unregistered) disturbances. The *active disturbance* is the latest disturbance that the train has not fully recovered from. Small (unregistered) delays and recovered time is assigned to the active disturbance.

A train can be delayed without having a registered disturbance. The *minor delay* of a train is the net sum of (small) unregistered deviations (delays and recoveries) for which there is no other disturbance these can be referred to. If there is no other active (registered) disturbance, then the minor delay is considered as active. The minor delay can be interpreted as the share of the delay that does not depend on any registered disturbance.

The *delay contribution* of a disturbance can be interpreted as the share of the delay that depends on this disturbance. The delay contribution can either be larger, equal or smaller than the delay increment that corresponds to the disturbance. In contrast to the delay increment, the delay contribution also includes additional small delays and recoveries that occur after the disturbance. Studying the train ride from the start to its terminus, the calculation of delay contribution is based on the following rules:

a) If a train gets delayed by a registered disturbance, the delay contribution of this disturbance is initiated with the delay increment of the disturbance at this position. If the train already has suffered from this disturbance (i.e., the disturbance-ID is reused), then the delay increment is added to its delay contribution. This disturbance becomes the active disturbance.
b) If a train that is delayed gets another registered disturbance than the active disturbance, then the new disturbance becomes the active one and the delay contribution of previous (now non-active) disturbance stays at the level is has (until it might become active again). Note that the "new" disturbance may refer to an earlier registered disturbance that continues to be the active disturbance or becomes the active one again. It can also happen that the new disturbance has less delay contribution than the previously active one so that the previous disturbance is considered to have larger impact on the total deviation than the new disturbance has.
c) If a train reduces its delay, the recovered time is assigned to the active disturbance and the delay contribution of this disturbance is decreased by the recovered time. Vice versa, if the train gets a small additional (unregistered) delay, it is assigned to the active disturbance and the delay contribution is increased by the lost time.
d) A disturbance is *"dead"* when the train has recovered so much time that the delay is at the same level as it was when the disturbance occurred. When the disturbance dies, the delay contribution of the disturbance is zero. The disturbance that was active before the now dead disturbance now becomes active again.
e) If there is no (other) active disturbance (i.e., no non-dead disturbance), then the minor delay is considered to be the active disturbance.

Finally, assume that a train arrives terminus t minutes after scheduled time, that the punctuality threshold is q minutes and the delay contribution of disturbance i is d minutes, then i is a *critical disturbance* if $t - d < q$. A critical disturbance can be interpreted as a disturbance that caused the train to be unpunctual; without the delay contribution from the critical disturbance, the train would have

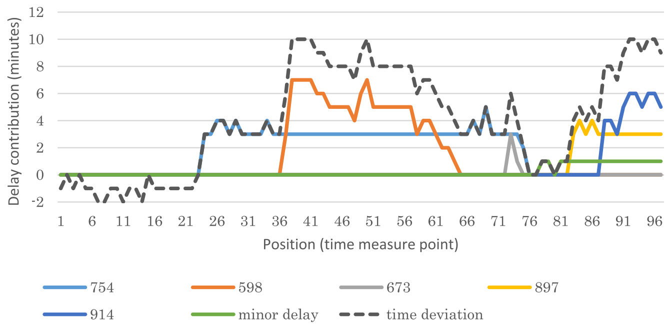

Fig. 3. Delay contribution from the different disturbances along the ride of train 542 2020-09-29.

Table 2
Delay contribution train 542 2020-09-29 (see also Table 1). Final delay at terminus is 9 minutes

Registered disturbance	Delay increment	Delay contribution at terminus	Critical disturbance
754	3	0	No
598	7	0	No
673	3	0	No
897	3	3	No
914	4	5	Yes
Minor delay	–	1	No
Total	**20**	**9**	

been punctual.

Pseudocode for the calculation of delay contribution is presented in Appendix A. As mentioned above, the data from Trafikverket is truncated to minutes and there is a 3-min threshold for disturbance registration; however the proposed method can be applied also with not-truncated time registrations and with other thresholds for both disturbance registration and punctuality.

The calculation of the metrics and concepts is illustrated using the ride of train 542 from Malmö to Stockholm 2020-09-29, and shown in Fig. 3 (see also Fig. 2 and Table 1). On the x-axis in Fig. 3 there are the different positions (time measure points) along the railway line and the y-axis corresponds to the delay contribution from each disturbance. The solid lines (of different colours) are the delay contributions from each respective disturbance along the ride. The dashed line is the delay of the train along the ride, which is also the same as the sum of the delay contributions at each point. From start up to position 23 the train is on-time (or before schedule) and all delay contributions are zero. At position 24 (see also Table 1) there is a disturbance (754) with delay increment of 3 min, thus disturbance 754 becomes the active disturbance and the delay contribution of 754 at position 24 grows to 3. At positions 26, 29, and 34 the delay increases by 1 min, and at positions 28, 30, and 35 the train recovers 1 min; the delay contribution of 754 takes the same increases and reductions as the deviation differentials. At position 37 there is a new disturbance (598) of 3 min: 598 become the active disturbance and the delay contribution of it is 3 at position 37. At position 38, there is another disturbance (4 min), but this has the same root cause as the previous and therefore this disturbance is also registered as 598. Thus, the delay contribution of 598 becomes 7 min at position 38 (while the delay contribution of 754 stays at 3 min – the same level as it had when 598 became the active disturbance.

Continuing the ride, the train recovers several minutes and at position 65 the delay is again 3 min which means that the train has recovered all the delay from 598, the delay contribution from 598 is zero and 598 becomes dead, and 754 becomes the active disturbance again.

At position 73 there is a new disturbance (673) of 3 min, thus, 673 becomes the active disturbance and delay contribution from 673 is 3. The train recovers this, and at position 75 the train has recovered from this disturbance, the disturbance dies and 754 once more becomes the active disturbance.

The train continues to recover time and at position 75 the train has recovered all the delay, 754 dies, the delay contribution from 754 is zero. The minor delay becomes the active disturbance since there is no other non-dead registered disturbance. At position 78 the train loses 1 min which is assigned to the minor delay, 1 min is recovered at position 80 and lost again at position 81.

At position 83, disturbance 897 occurs, the delay increment is 3 min, and disturbance 897 becomes the active, while the delay contribution from the minor delay stays at 1 min. At position 88 there is another disturbance (914) of 4 min. Thus, disturbance 897 becomes inactive and its delay contribution stay at 3 min.

Finally, there are some more small deviations that are assigned to 914 before the train arrives at the final destination Stockholm. The delay at the terminus is 9 min. In Table 2 the delay contributions from all disturbances are summarized. Note that the total delay contribution is the same as the delay at the terminus, while the total delay increment of the disturbances is much more.

In this way, the delay contribution can be used to divide the final delay into components, i.e., how much each disturbance contributed to the final delay. In the example train, 3 min of the final delay was caused by 897, 5 min by 914 and 1 min was caused by minor delay, while the other disturbances did not contribute to the final delay. Thus, we create a connection between the disturbances

Table 3
Overview of dataset, disturbances, and minor delays.

	All trains	Trains with 6–10 min final delay
Number of trains	4683	641
Unpunctual trains	1649	641
Unpunctual trains with no registered disturbance	258	249
Trains with one or more critical disturbance	1176	632
Trains with minor delay > 0	1629	352
Trains with critical minor delay	258	248
Trains with minor delay – alive and non-critical	1371	124
Registered disturbances	6206	1273
Disturbances – "dead"	2579	437
Disturbances – critical	1294	712
Disturbances – not dead and not critical	2333	124

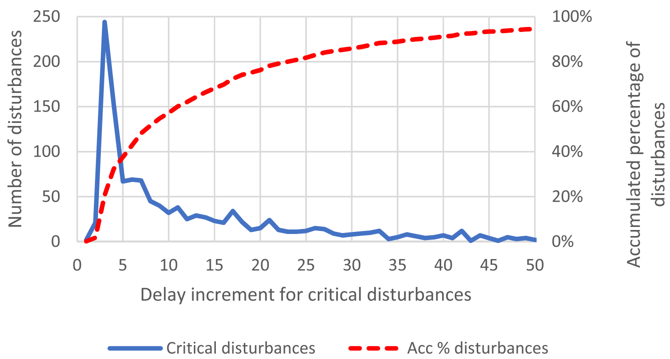

Fig. 4. Size of critical disturbances.

and the delay at the final station and a connection to the punctuality. Using the definition of critical disturbance, disturbance 914 is a critical disturbance while no other of the disturbances is critical.

An interpretation of this is that the unpunctuality was caused by disturbance 914, since if we remove the delay contribution from 914, the train would be punctual. Reducing the final delay with any one of the other delay contributions and the train would still be unpunctual, thus the other disturbances are not critical. Of course, it is impossible to really know it the punctuality of the train depends on a certain disturbance, since we cannot know what would happen if this disturbance did not happen. Nevertheless, these concepts identify that some disturbances are more important and other are less important for the final delay and the punctuality. In the next section we show that interesting conclusions can be made using these metrics and concepts.

3. Case study

In this section we describe the results of a case study. First, the case is described, followed by a characterization of the concepts defined in the previous section and finally the concepts are used to analyse the performance of the trains.

3.1. The case: high-speed trains Malmö-Stockholm

The railway line between Malmö and Stockholm is one of the most important lines in Sweden for both passenger and freight traffic. Many types of trains co-exist on this double track line: high-speed, long-distance and regional passenger trains, commuter trains, and long-distance and local freight trains. Thus, the traffic on the line is very heterogeneous with different stopping patterns and speeds. In a punctuality perspective, it is the most problematic line in Sweden, and in particular, the punctuality of the high-speed trains between Malmö-Stockholm is far below the target levels. The focus of the case study is the northbound high-speed trains from Malmö to Stockholm.

Data regarding train operations is provided from Trafikverket's (Swedish Transport Administration) database for follow-up of train operations (LUPP) for the year 2017. The punctuality for the studied trains is about 65%.

Of special interest are the trains that are "almost punctual" since small improvements in the delay pattern could make such trains become punctual. In this study we consider trains with a delay between 6 and 10 min as "almost punctual trains", and 12% of the trains belong to this group (which corresponds to 20% of the unpunctual trains). Since these trains are of special interest, we focus on such trains in some of the results shown in the report. The developed methods are also particularity suited for analysing "almost punctual trains". The dataset is summarized in Table 3.

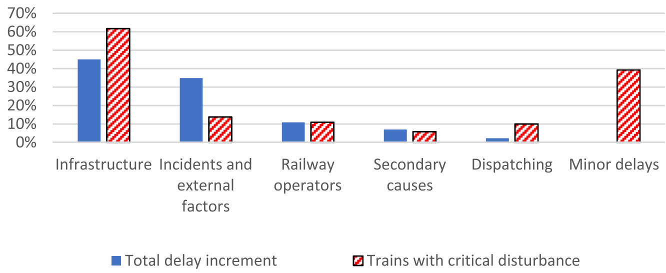

Fig. 5. Comparison between two ways of analysing the delay attribution codes (level 1) to punctuality problems.

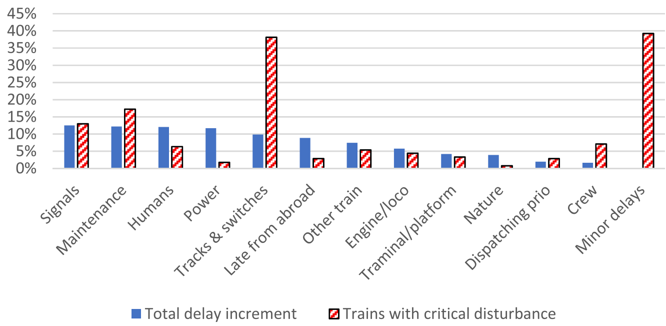

Fig. 6. Refined comparison between two ways of studying the delay attribution codes (level 2) to punctuality problems.

3.2. Characterization of critical disturbances

Table 3 summarizes the overall results of the dataset. Trains with a final delay of 6–10 min are in a separate column as these trains are of special interest as explained above. One objective with the defined metrics is to separate less and more important disturbances from each other in a punctuality perspective. The results show that most unpunctual trains have at least one critical disturbance and still a good share of the disturbances are not defined as critical – thus the critical-concept divide the disturbances and minor delays into two groups – in the next section we will illustrate that the separation is indeed valuable.

Fig. 4 illustrates the size of the critical disturbances; x-axis corresponds to the delay increment of the registered disturbances. The solid line represents the number of disturbances of different size (left y-axis) while the dashed line represents the accumulated percentage of disturbances (right y-axis). The figure shows that most of the critical disturbances are small: 38% are 5 min or less and 52% are 8 min or less.

3.3. Analyses using critical disturbances

In this section we use the developed metrics and concepts to show that they provide a complementary lens through which to study punctuality problems. We compare two ways of studying the unpunctuality: the "traditional" metrics and the new alternative view. The traditional way uses the **total delay increment** of all trains as metric, but the alternative uses **number of trains with critical disturbance** as metric. By focusing on the number of trains (instead of time) we reduce the impact of disturbances that cause large delays with respect to time and instead favour disturbances that impacts many trains.

Fig. 5 illustrates the delay attribution code (level 1) of disturbances. Solid (blue) bars represent the percentage of total disturbance time that depends on various causes (the "traditional" analysis). The striped (red) bars correspond to the proportion of trains that have a critical disturbance linked to the various causes. (Note that the striped bars add up to more than 100 percent because one train can have several critical events.) Main differences between the two analyses are:

- Minor delays are critical disturbances for 39% of the unpunctual trains. Minor delays are not at all captured in analyses based on registered disturbances and delay increment.
- Infrastructure problems are the cause of 62% of unpunctuality, but account for 45% of the volume of delay increment.
- Accidents and incidents account for 35% of the volume of delay increment but corresponds only for 14% of the unpunctual trains. This is a consequence of that this kind of disturbances often results in large delays, not proportional to the number of trains that becomes unpunctual.

Fig. 6 represents a similar comparison as in Fig. 5, but for delay attribution codes level 2. Again, solid (blue) bars correspond to the percentage of total delay increment while striped (red) bars correspond to the proportion of trains with critical disturbances. There is a

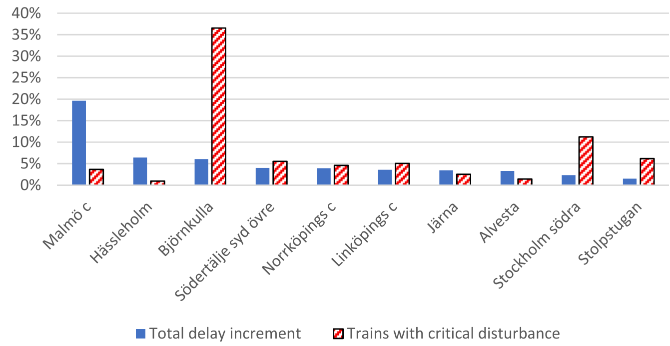

Fig. 7. Comparison between share of total delay increment versus critical disturbances at different locations.

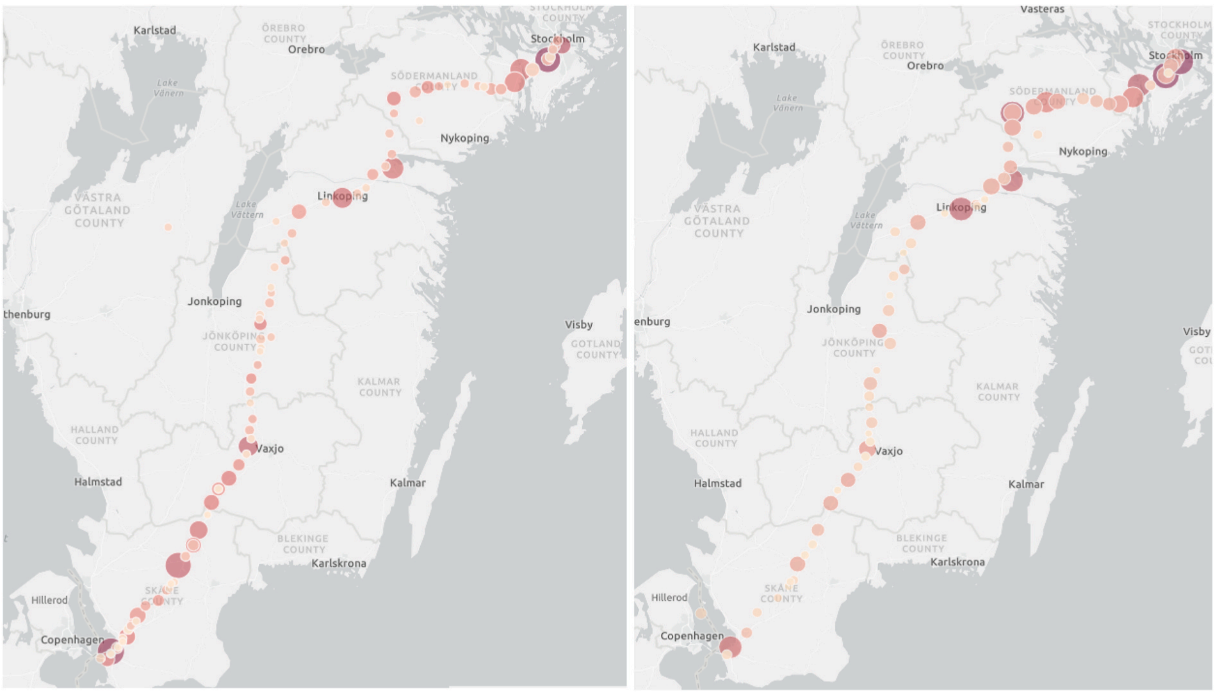

Fig. 8. Left: Map showing total delay increment of the registered disturbances at different positions. Right: Map showing number of critical disturbances at different positions and the total delay contribution of these.

clear difference between the two ways of analysing the problems along the line, in particular the level-2 code Tracks and switches stands out with a significantly higher share of critical disturbances causing unpunctual trains than delay increment.

Fig. 7 illustrates which positions that contribute to delay increment and critical disturbances, respectively (sorted according to total delay increment). Again, solid (blue) bars represent total delay increment while the striped (red) bars correspond to the number of trains with critical disturbance. The result shows large differences in conclusions. Considering delay increment, Malmö is most important, but considering critical disturbances, Björnkulla is much more important; as many as 36% of the unpunctual trains have a critical disturbance in Björnkulla, while the total delay increment in Björnkulla does not stand out at all.

When digging deeper into the punctuality data (not illustrated here) it turns out that during 2017 there was a speed reduction close to Björnkulla, (registered as a problem with Tracks and switches) which resulted in a small disturbance of 3–4 min on very many trains. Björnkulla is situated about 12 min train ride before the final station Stockholm. Many trains lost a few minutes in Björnkulla. However, when looking at the total delay increment, Björnkulla is not particularly interesting, but studying critical disturbance it shows that very many trains became unpunctual because of the small amount of lost time in Björnkulla. Thus, the conclusion is that critical disturbance is – at least in this case - capable of pinpointing important contributors to the problems of unpunctual trains that was not as clearly seen when using the "traditional" delay increment as indicator. (The Infrastructure Manager Trafikverket was also not aware that the "small" problem in Björnkulla hade this big impact on the punctuality. If they had this information, they could have prioritized the maintenance work in Björnkulla higher to reduce the negative impact of this problem.)

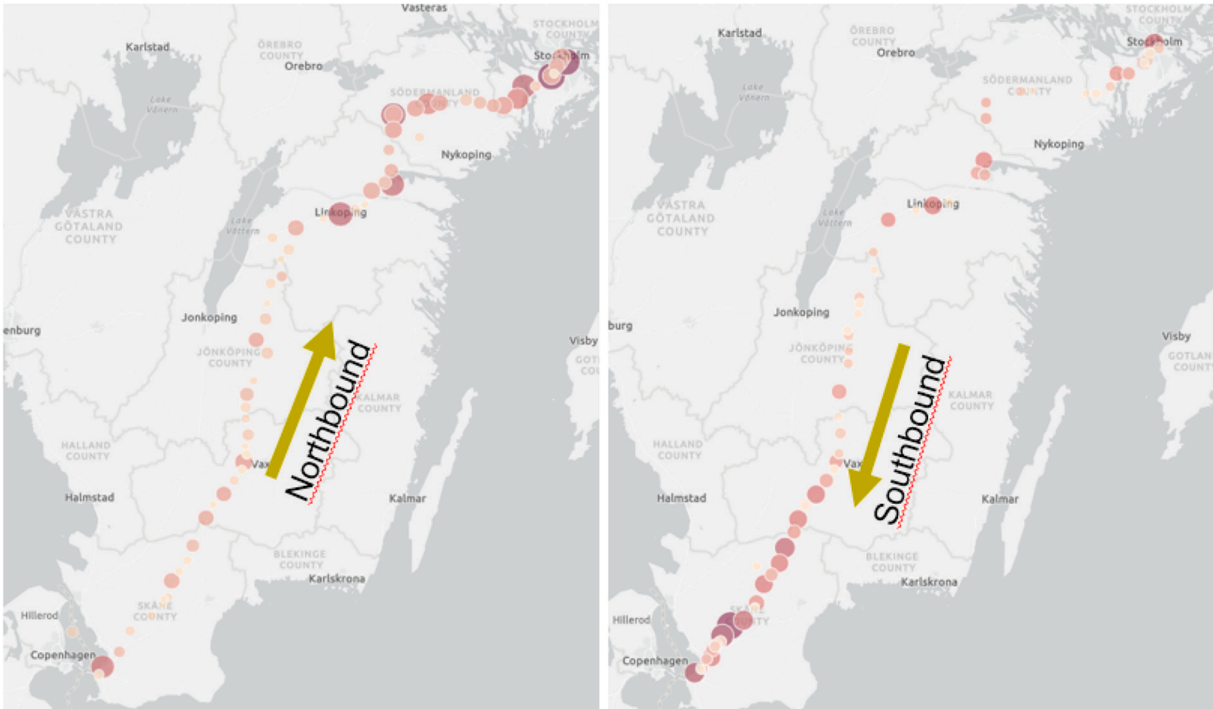

Fig. 9. Positions for critical disturbances. Left: Northbound train. Right: Southbound trains.

The later in a trip a disturbance occurs, the shorter timespan the train has available to recover from the disturbance before the train arrives at the terminus. Therefore, it is expected that critical disturbances accumulate towards the end of the journey since small disturbances early in a journey have greater chance to recover than a small disturbance towards the end of the journey, as illustrated in Fig. 8. The left map in the figure indicates the positions of the delay increments for northbound trains, while the right map indicates positions of critical disturbances. Ring size corresponds to the total time (delay increment in left picture and delay contribution in right picture) and colour corresponds to the number of registrations (the darker the colour, the more registrations). The figure shows that the positions for delay increment are spread rather evenly over the entire Malmö-Stockholm section, while critical disturbances are accumulated towards the end of the trip.

Fig. 9 illustrates that there are more critical disturbances towards the end of the journeys, the left map indicates positions of critical disturbances for northbound trains, while the right map is for southbound trains. Size of dots represent total delay contribution; colour represent number of trains (darker is more trains). A deeper analysis (not illustrated) shows that for 75% all trains, the last disturbance can be classified as a critical disturbance. (For the “almost punctual trains”, the number is even higher: for 85% of the trains with a delay of 6–10 min, the last disturbance is a critical disturbance.)

3.4. Visualizing disturbance spread in railway network using delay contribution

Delay contribution is also useful for visualizing the spread of a disturbance in the railway network. While the delay increment from a disturbance is registered at very few locations during the trip, the delay contribution “lives” with the train along the trip.

In Table 1 and in Fig. 3 we showed the impact from disturbance 598 on train 542 with respect to delay increment and delay contribution, respectively. In fact, several trains were affected by this disturbance. The actual position of the signal problem 598 was in Moheda, but since disturbances are registered by their root cause, there are registrations from disturbance 598 at several positions. Fig. 10 illustrates the spread of the registered disturbance 598 in the 14 trains that were affected by 598. The left map illustrates the positions of the delay increment registrations, while the right map illustrates the spread of delay contribution caused by 598 – one dot for each position at which the delay contribution from 598 is greater than zero. Each colour represents one train. The illustration of delay contribution gives a much better indication of how the delays caused by 598 spread in the railway network.

4. Conclusion and discussion

In this paper we have defined new concepts aimed at supporting root cause analysis for train delays and how different disturbances contribute to the unpunctuality. Two new metrics were defined. The delay contribution can be interpreted as a division of the delay onto different disturbances, which can be used to point out important disturbances from less important. The critical disturbances can be interpreted as pointing out disturbances that made the train “fall over the punctuality threshold” and become unpunctual. In a case

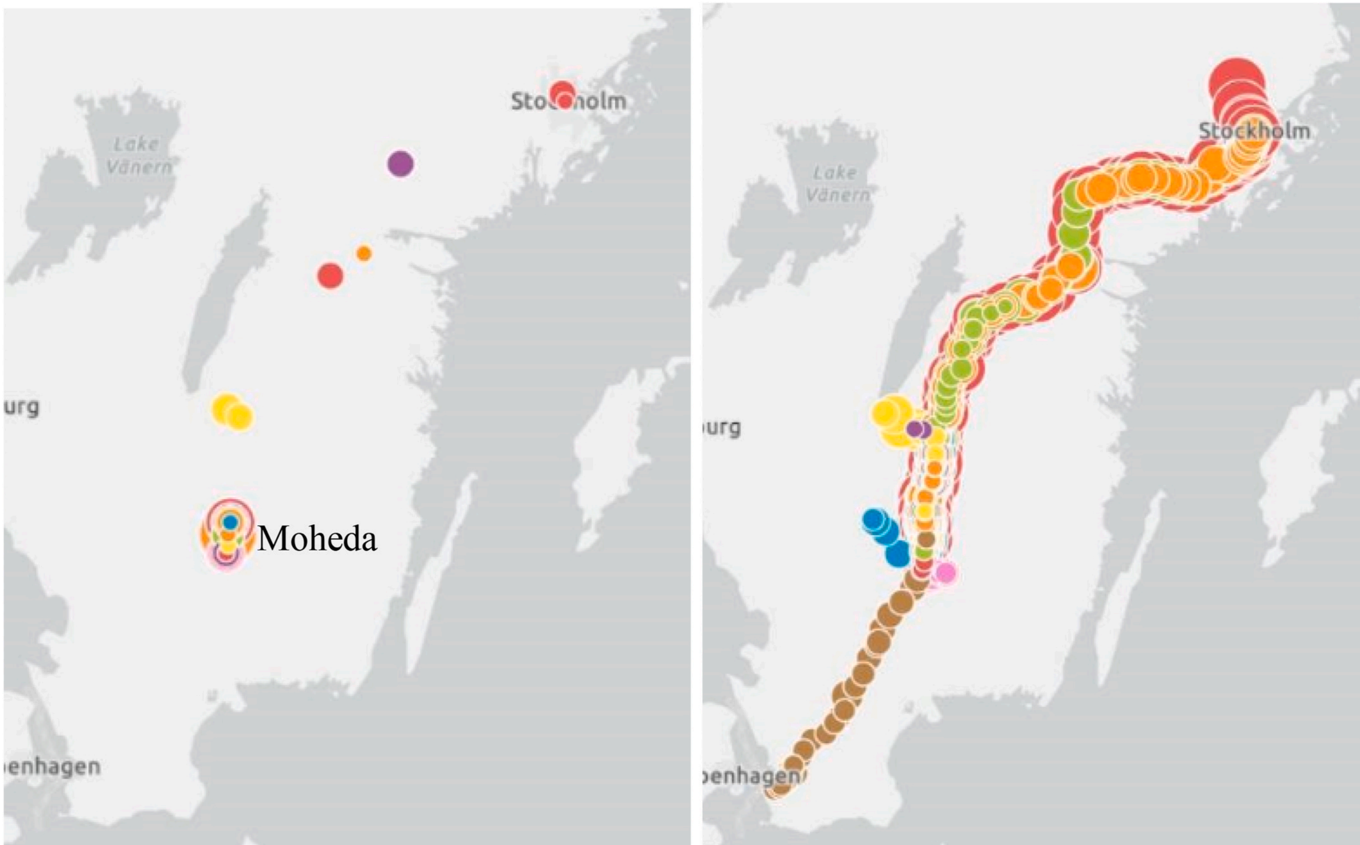

Fig. 10. Disturbance registrations (left) and delay contribution (right) in the railway network caused by disturbance 598.

study we showed how the concepts can be used and that they are able to provide an alternative view of how disturbances impact unpunctuality compared to "traditional" methods. The new concepts and metrics provide a direct connection between the disturbances and the unpunctuality.

The new concepts can also be used for analysing consequences of smaller disturbances and for visualization of disturbance spread in the railway network.

One conclusion of the work is that the usage of this kind of new metrics and concepts make the connection between disturbances and delay more explicit and thereby gives a complementary view of reasons to unpunctual trains. The explicit connection can be used to evaluate which disturbances that contribute much to a delay and which that contribute less. Further, the results also highlight the role of minor disruptions, i.e., disruptions that are so small that they are very often neglected when analysing causes of punctuality problems. Moreover, the results from the case study indicate that disturbances late in the trips have larger impact on punctuality than earlier disturbances, which is also natural since there is less travel time for the train to recover from late disturbances before arriving final station.

The Infrastructure Manager in Sweden (Trafikverket) has showed great interest in the results. A continued study has started with the aim to confirm that this kind of analysis is scalable and can be made for the whole railway traffic in Sweden and also to investigate if and how these new concepts can be used in their operations. The most direct usage of the concepts is of course in the analyses that are made regarding unpunctuality and its causes. The results and the kind of analysis performed here can also be useful in timetable planning for increasing punctuality and in the operational traffic control. It could also be useable in maintenance planning, in particular for prioritizing maintenance work that could give large impact on the punctuality.

The new concepts can be defined in alternative ways. For example, there are alternative ways how to assign recovered time to disturbances – in the used definition the recovered time is always assigned to the latest disturbance. Assigning recovered time in some other way would make the last disturbances on a trip contribute even more to the final delay, and thereby even strengthen some of the conclusions that were made from the analyses, e.g., the severeness of late disturbances on the ride would be even larger.

The method described in this paper is adapted for railway systems where the performance regime is based on a discrete threshold at the terminating station. In many railway systems, the punctuality is measured at every station. The generalization of the proposed methods to be applicable for the latter type of punctuality measure is straightforward: the delay contribution principle presented calculates the delay contribution at each measure point (including intermediate stations) and the critical disturbance can be calculated for each station instead of just the final. The details of this will be presented in future research.

Furthermore, in future research, the concepts of the paper will be further enhanced and adapted for other types of punctuality measures and traffic types. Also, the concepts could be adapted to passenger punctuality in difference to the train punctuality used here. The concepts introduced in this paper could also be applied in performance analysis systems, like train performance simulation or delay prediction, in order to further enrich the results from such analyses.

Acknowledgment

This research has been performed with support from Trafikverket via research program Capacity in the Railway System (www.kajt.org) in close collaboration to the punctuality improvement program denoted TTT. The authors would like to thank the editor and the three anonymous reviewers for their constructive comments.

Appendix A

Below, a pseudocode for the calculation of the delay contribution is presented.

Input parameters
N: Number of measure points along train ride.
p: Position (measure point) along train ride, $p = 1, 2, 3, \ldots, N$.
I: Set of all registered disturbance IDs assigned to the train along the complete ride.
Ip: Registered disturbance ID at position p. If no registered disturbance at p, then $Ip = \varnothing$.
tp: Deviation from timetable at position p.
Calculated data
Dpi: Delay contribution of disturbance i at position p.
Mp: Accumulated minor delay at position p.
S: Stack (i.e., "last in, first out memory") of disturbance IDs.
j: Active disturbance ID.
τp: Positive deviation from timetable at position p.
δp: Reminder of change in deviation from position p-1 to p (not assigned to any disturbance).
δpj: Change in deviation from position p-1 to p assigned to disturbance j.
Pseudocode
Initialize: $p := 1$, $S = \varnothing$, $Mp = 0$, $Dpi = 0 \ \forall \ i \in I$
$Tp := \max \{\mathrm{tp}, 0\}$
$S := Ip$
if $\mathrm{S} = \varnothing$: $Mp = \tau p$

(continued on next page)

(*continued*)

```
else
  j = Ip
  Dpj: = τp
endif
p: = p + 1
repeat
  if Ip ∉ S: push Ip to S
  elseif Ip ∈ S: lift Ip to top of S
  endif
  τp: = max {tp, 0}
  δp: = τp - τp-1
  Mp: = Mp-1
  Dpi: = Dp-1,i ∀ i ∈ I
  repeat
    j: = top element of S
    if S = ∅: Mp: = max{Mp-1 + δp, 0}, δp: = 0
    elseif δp > 0: Dpj: = Dp-1,j + δp, δp: = 0
    else
      δpj: = min{-δp, Dp-1,j }
          Dpj: = Dp-1,j + δpj
      δp: = δp - δpj
      if Dpj = 0: pop j from S
    endif
     until δp = 0
  p: = p + 1
until p > N
```

References

Andersson, E.V., Peterson, A., Krasemann, J.T., 2015. Reduced railway traffic delays using a MILP approach to increase Robustness in Critical Points. J. Rail Transport Plann. Manag. 5 (3), 110–127.

Bergström, A., Krüger, N., 2013. Modeling passenger train delay distributions: evidence and implications. In: Centre for Transport Studies Stockholm, Swedish National Road & Transport Research Institute (VTI), KTH Royal Institute of Technology, S-WoPEc, Scandinavian Working Papers in Economics.

Cerreto, F., Nielsen, O.A., Harrod, S., Nielsen, B.F., 2016. Causal analysis of railway running delays. In: 11th World Congress on Railway Research (WCRR 2016). Milan, Italy.

Grechi, D., Maggi, E., 2018. The Importance of Punctuality in Rail Transport Service: an Empirical Investigation on the Delay Determinants", vol. 70. European Transport/Trasporti Europei.

Goverde, R.M., 2010. A delay propagation algorithm for large-scale railway traffic networks. Transport. Res. C Emerg. Technol. 18 (3), 269–287.

Harrod, S., 2012. A tutorial on fundamental model structures for railway timetable optimization. Surv. Oper. Res. Manag. Sci. 17, 85–96.

Huang, P., Wen, C., Li, Z., 2019. Mining train delay propagation pattern from train operation records in a high-speed system. In: RailNorrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), Norrköping, Sweden, June 17th–20th, 2019, No. 069. Linköping University Electronic Press, pp. 439–451.

Infrabel, 2018. Punctuality in 2017 European Benchmark.

Kecman, P., Goverde, R.M., 2014. Online data-driven adaptive prediction of train event times. IEEE Trans. Intell. Transport. Syst. 16 (1), 465–474.

Kristoffersson, I., Pyddoke, R., 2019. A traveller perspective on railway punctuality: passenger loads and punctuality for regional trains in Sweden. In: RailNorrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), Norrköping, Sweden, June 17th–20th, 2019, No. 069. Linköping University Electronic Press, pp. 565–578.

Lindfeldt, A., 2015. Railway capacity analysis: methods for simulation and evaluation of timetables, delays and infrastructure. In: Doctoral Dissertation. KTH Royal Institute of Technology.

Marković, N., Milinković, S., Tikhonov, K.S., Schonfeld, P., 2015. Analyzing passenger train arrival delays with support vector regression. Transport. Res. C Emerg. Technol. 56, 251–262.

Network Rail, 2020. https://www.networkrail.co.uk/who-we-are/how-we-work/performance/railway-performance/.

Økland, A., Olsson, N.O., 2021. Punctuality development and delay explanation factors on Norwegian railways in the period 2005–2014. Publ. Transport 13, 127–161.

Palmqvist, C.W., Tomii, N., Ochiai, Y., 2020. Explaining dwell time delays with passenger counts for some commuter trains in Stockholm and Tokyo. J. Rail Transport Plann. Manag. 14, 100189.

Palmqvist, C.W., Olsson, N.O., Winslott-Hiselius, L., 2017. Some influencing factors for passenger train punctuality in Sweden. Int. J. Prognostics Health Manag. 8.

Shift2Rail, 2015. Shift2Rail Strategic Master Plan. https://ec.europa.eu/transport/sites/transport/files/modes/rail/doc/2015-03-31-decisionn4-2015-adoption-s2r-masterplan.pdf.

Solinen, Emma, Nicholson, Gemma, Peterson, Anders, 2017. A microscopic evaluation of railway timetable robustness and critical points. J. Rail Trans. Plann. Manag. 7 (4), 207–223. https://doi.org/10.1016/j.jrtpm.2017.08.005.

Sørensen, A.Ø., Landmark, A.D., Olsson, N.O., Seim, A.A., 2017. Method of analysis for delay propagation in a single-track network. J. Rail Transport Plann. Manag. 7 (1–2), 77–97.

Trafikanalys, 2020. Punktlighet På Järnväg 2019 Kvalitetsdeklaration. https://www.trafa.se/globalassets/statistik/bantrafik/punktlighet-pa-jarnvag/2020/kvalitetsdeklaration-punktlighet-pa-jarnvag-2020-kvartal-3.pdf.

Wang, P., Zhang, Q.P., 2019. Train delay analysis and prediction based on big data fusion. Transport. Saf. Environ. 1 (1), 79–88.

Wen, C., Mou, W., Huang, P., Li, Z., 2020. A predictive model of train delays on a railway line. J. Forecast. 39 (3), 470–488.

Wen, C., Yang, X., 2020. Visualizing train delays using tableau and the framework of a delay impact visualization system. J. Big Data Analyt. Transport. 2, 75–91.