

# Data aggregation for detailed analysis of train delays

T. Richter

*Operations Monitoring and Analysis, Rail Net Denmark, Denmark*

## Abstract

The classical approach to analysis of train delays is either based on analysis of delays over a certain threshold along with their cause or of all delays at arrivals to and departures from stations, irrespective of cause. On Danish railway lines with dense traffic, these classical approaches have not been able to yield sufficient information on the dynamics of delays to allow for in depth understanding of the behaviour of delays (Section 1).

On large parts of the Danish railway network, data is available describing the occupation of block sections and important track circuits (Section 2). Based on this data a method is developed to construct an enriched train run history on track circuit level. This train run history include information on delay, change in delay, headway between trains as well as the most likely signal aspect received at each block section and which train that caused it (Section 3).

This enriched train run history allows for analysis of network performance on track circuit level on lines between junctions. Simple examples of such analysis are then demonstrated (Section 4). Such analysis can help identify i.e. infrastructure bottlenecks, timetabling inefficiencies and initial trains that cause queuing. This again makes it possible to focus the punctuality improving efforts to achieve a higher punctuality (Section 5).

*Keywords: train run history, delays, queuing, resolution of data, single line.*

## 1 Introduction

The network of the Danish Railway Infrastructure provider, Rail Net Denmark (BDK), is one of the most densely used in Europe following Switzerland and The Netherlands [1]. Capacity consumption is near the limit or at the limit on parts of the network [2]. Such intense usage of a railway network makes it prone to



escalating delays. This also raises the risk that minor timetabling faults or local temporary infrastructure shortcomings have larger consequences. Thus detailed and advanced analysis of train delays are important in order to maintain a high punctuality. Such approaches have already been demonstrated in [3, 4].

Traditionally all analysis of train delays within the Danish railway sector have been based on the notion of affected train, which are trains delayed more than 4 minutes and 59 seconds or cancelled (2 minutes 29 seconds on the Copenhagen Suburban network). All affected trains are attached to Delay Reports describing the cause and circumstances of the delays and coded in order to allow for further analysis [4]. This is the method recommended by UIC [5] and within this definition; a train is either affected or punctual.

The Delay Reports along with their affected trains and affected arrivals form the basis of the primary Key Performance Indicators (KPI) for punctuality within the Danish railway sector [3]. However during the last years, performance reports based on analyses of all delays have been introduced within BDK [4]. This is part of a general shift in mindset in the sector towards a focus on delays smaller than the affected train threshold. These analyses and reports do none the less still all focus on arrivals to and departures from stations and halts and not on what happens on the open line between these measuring points. The network of BDK has only approximately 330 such measuring points of which 166 automatically collect data from the Centralised Traffic Control (CTC). This is compared to a total network length of 1,992 km [6].

On networks, which are not running close to their capacity limit, train run data on a station by station resolution provides sufficient information on network performance since queue phenomenon do only occur at a limited scale. However, if networks are densely used queuing phenomena will occur more often and train performance between stations and junctions thus becomes of relevance in order to understand the dynamics of delays (escalation and absorption of delays).

Fortunately, on the most important lines of BDK, data is available from the digital CTC describing the occupation of important track circuits and blocks. Aggregation and analysis of this data allows a description of the dynamics.

## 2 Data describing train runs between stations

The network of BDK is controlled by several different types of CTC system where the two most moderns ones, the DCTC and XCTC systems, collects and log data on occupation of important track circuits and block sections and makes this easily available for further analysis. These systems cover approximately 80% of the network in terms of train passages at measuring points.

This data is currently only used for specific track circuits to log train arrivals and departures from stations and halts but data is technically available for all block sections and major track circuits at stations. The raw data consists of train number, station identifier, track circuit identifier, delay calculated by the CTC, and a timestamp for the reception of the telegram (Table 1). When the track circuit is occupied for the first time, the CTC records this and a new log entry is

created every minute the track remains occupied. This allows a registration of when trains have come to a halt or passed the track circuit very slowly.

Table 1: CTC telegram data.

TRAIN_NO	STAT	T_CIRCUIT	DELAY	TIMESTAMP
1011	CPH	2164	0,33	23-02-2012 05:30:05
1011	CPH	214	0,33	23-02-2012 05:30:45
1011	CPH	214	0	23-02-2012 05:32:03
1011	CPH	214	-1	23-02-2012 05:33:03
1011	CPH	209A	-1,17	23-02-2012 05:33:17
1011	CPH	209A	-1,17	23-02-2012 05:34:03
1011	CPH	209A	-1,5	23-02-2012 05:35:03
1011	CPH	209A	-2	23-02-2012 05:35:36
1011	CPH	205	-2	23-02-2012 05:35:36
1011	CPH	201	-2	23-02-2012 05:35:54
1011	KLK	2102	-2	23-02-2012 05:36:29

The entire data logging is based on the location on train numbers, which is the unique identifier of a train and which links it to its schedule, both technically and from a safety point of view. Since shunting by definition are train movements without a train number, these movements are not included in this data.

### 3 Aggregation of train run data on track circuit level

In order to extract information permitting an analysis of train behaviour at track circuit level, an aggregation of the CTC data is carried out. The aggregated data describes the properties of the train passing the track circuit, the properties of the previous train passing the track circuit and the properties of the present train when passing the previous track circuit as well as the signal aspect that the train received (Table 2). Such a record is created every time a train passes a track circuit.

Table 2: Aggregated train run data at track circuit level.

<b>Train run descriptors</b>
Train number and first run date (unique ID)
Station and track circuit number (unique ID)
<b>Performance data</b>
Timetable deviation at first telegram from track circuit
Time of first telegram
Time of last telegram
Deviation at previous track circuit (from first telegram)
Time of last telegram from last track circuit
Previous train number on this track circuit
First deviation of previous train on this track circuit
Time of last telegram of previous train on this track circuit
Signal aspect received
Train causing this signal aspect if restrictive
Train causing a tree of trains with restrictive signal aspect

The aggregation is done using a 1,500 line PL-SQL script running on the Oracle 11g Datawarehouse server of the Punctuality Reporting System. The script consists of a number of steps or functions:

#### Function `f_toghist_atns`

A master table is generated on basis of the raw track circuit occupation data (i.e.).

##### Function `f_toghist_atns_insert_values_1`

Added to master table is information about the previous track circuit, delay as well as first and last occupation of this for the present train.

##### Function `f_toghist_atns_insert_values_2`

Then information about the previous train on the same track circuit is added to the master table: timetable deviation as well as first and last occupation.

##### Function `f_toghist_signal`

Since the signal aspect that the trains receive at each track circuit is not easily available from the CTC log, the most likely signal aspect is deduced from the location of the train ahead using the following algorithm:

1. The track circuits that the train will pass during the next 15 minutes are identified through a lookup in the master table.
2. The latest train on each track circuit ahead is identified (most recent occupation before passage of train being examined).
3. Of those times-train combinations of previous occupation of track circuits ahead, the newest is identified. Thus the location of nearest occupied track circuit ahead is identified and the most likely signal can be deduced.

The function initially returns a value indicating how many major track circuits ahead the next train is located. This is then converted into a signal aspect (Table 3). Stored in the table is also information on which train caused the restrictive signal aspect and from what track circuit. It should be noted, that restrictive signal aspect will not always lead to reduced speed. This depends on speed and braking performance for the train. The optimal signal aspect on the parts of the network equipped with train control (ATC) is 3 free block sections ahead.

Table 3: Signal aspects.

Location of next train	Signal
Next block	Stop
2 blocks ahead	Restrictive ("1 green")
3 blocks ahead	Restrictive ("2 green")
4 blocks ahead	Clear ("3 green")
5 blocks ahead	Clear ("3 green")
...	Clear ("3 green")
Passage of previous train > 10 min	No train in vicinity

In the general cases, this approach to deduce the most likely signal aspect is unproblematic. However, some special cases exist which have to be treated

differently since the correct signal cannot in all cases be deduced. If no train ahead is found on the path, which the train will cover for the next 10 minutes, it is assumed, that the train received an optimal signal aspect.

Also, when the deduced signal aspect is based on the occupation of the final track circuit in the train run history, this is indicated with an additional “(S)” after the signal aspect. This can either be due to the train reaching its final destination and is shunted away afterwards or since it reaches the “edge” of data, the most prominent of which is when the trains leave the Danish network at Peberholm (PHM) towards Sweden. The consequence of this is that the nearest train ahead may be at a location further ahead than indicated by the deduced signal aspect. 8.6% of the deduced signal aspect is subject to this condition.

Function `f_toghist_init_1` and `f_toghist_init_2`

Using the knowledge of what signal aspect a train received, what train that caused it and if the signal aspect was restrictive, a tree of trains that caused restrictive signal aspect for other trains is constructed. The head consists of the train with the initial delay i.e. the root cause of the delay.

### 3.1 Data properties and quality issues

The most common flaws in data are telegrams, which are lost due to communication drops, telegrams which are received late and which thus get a “wrong” timestamp and telegrams, with a wrong calculation of delay due to lost previous timestamps. Finally, the trains sometimes lose their train number i.e. due to false occupation of track circuits and thus run without a train number until the signalman corrects the problem. The train number queue used when trains enter the CTC system may also lead to incorrect log entries in case the track circuit is occupied by a different train than originally scheduled and the signalman has not updated the number queue.

On a general level, these flaws are of a low magnitude. However, algorithms have been developed to deal with some of these flaws, either estimating the correct values or removing obviously incorrect values.

## 4 Application/demonstration

One of the busiest lines on the BDK network is Øresundsbanen, which runs from Sweden to Copenhagen Central Station (KH) serving Copenhagen Airport station (CPH) as well as the halts Tårnby (TÅT) and Ørestanden (ØRE) ( and ). During day hours on a weekday, the line serves up to approximately 13 trains an hour in each direction being a mix of local trains, long distance trains and freight trains. This leads to a planned lead time of 4 minutes – and some of the train systems stop at the halts on the line. The Capacity Consumption on the line is 74.5% in direction Copenhagen Central [7], which is above the recommended Capacity Consumption on lines with mixed traffic of 60% [8].

The classical approaches to analysis of delays on this line have not yielded sufficient information on its ability to absorb delays and the circumstances when delays escalate. Thus it is obvious to use the enriched train run history as basis



for a number of simple aggregations in order to demonstrate the potential with this data.

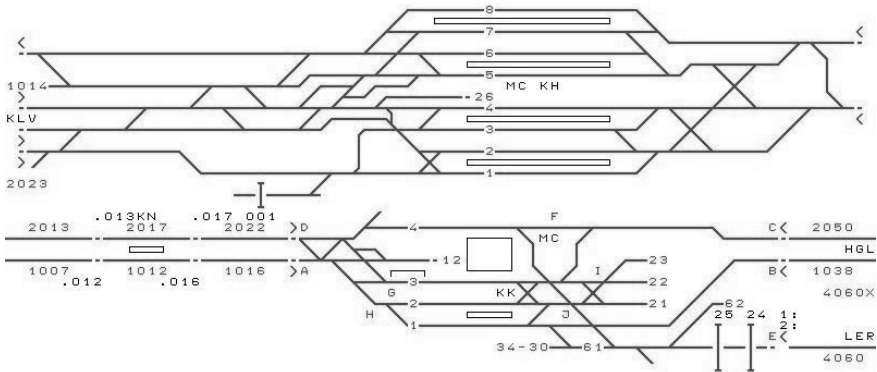


Figure 1: CTC diagram of "Øresundsbanen" - part 1.

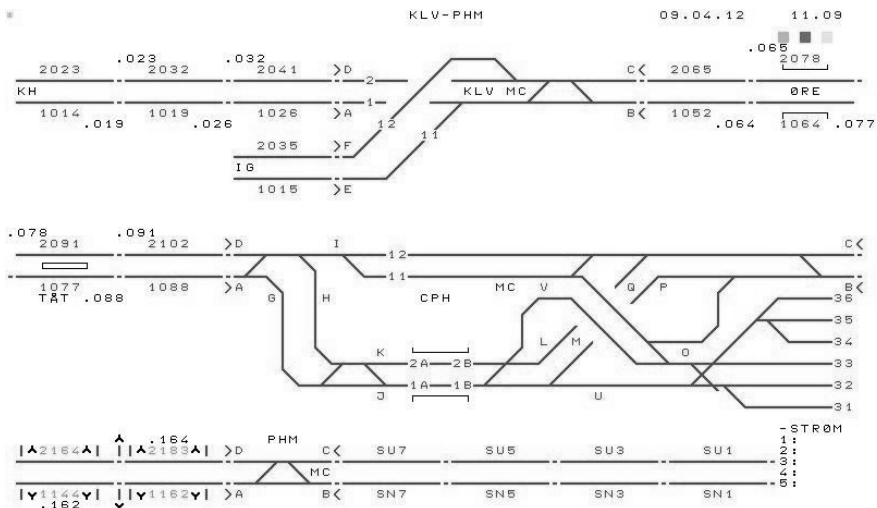


Figure 2: CTC diagram of "Øresundsbanen" - part 2.

Possible relevant aggregations are comparisons of:

- Delays at track circuits. (cf. Figure 3).
- Lead time between trains. (cf. Figure 4).
- Time loss between consecutive trains. (cf. Figure 5).
- Occupancy time of track circuits. (cf. Figure 6).
- Track circuits where trains receive restrictive signal aspect. (cf. Figure 7).
- Track circuits with initial cause for restrictive signal aspect. (cf. Figure 8).

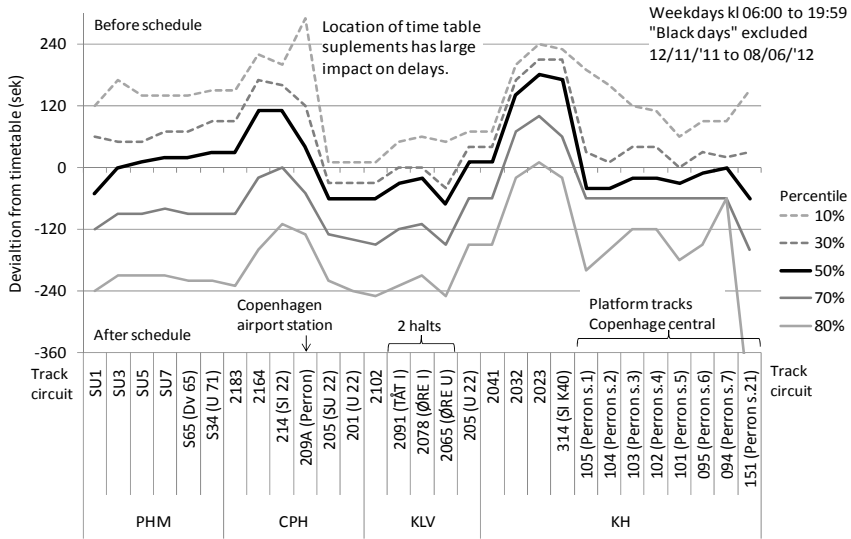


Figure 3: Delays at track circuits.

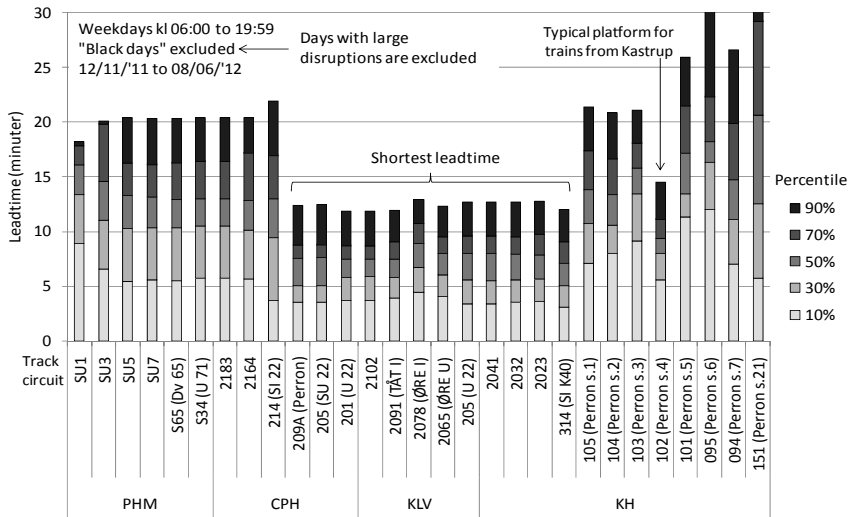


Figure 4: Lead time between trains.

The distribution of delays at track circuits indicates where the trains have the largest delays on the line – this might be due to infrastructure bottlenecks, timetabling issues or CTC deficiencies in calculating the correct delay.

Lead time between trains can be used as a proxy for capacity consumption.

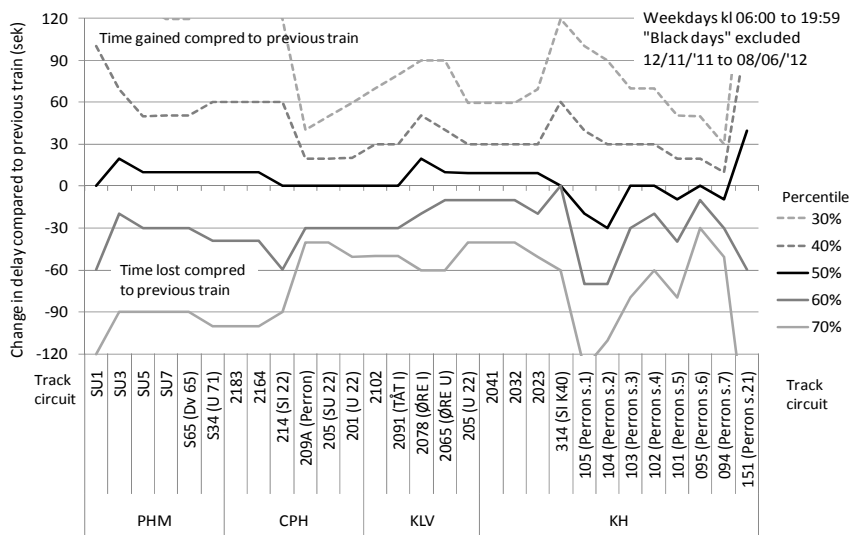


Figure 5: Time loss between consecutive trains.

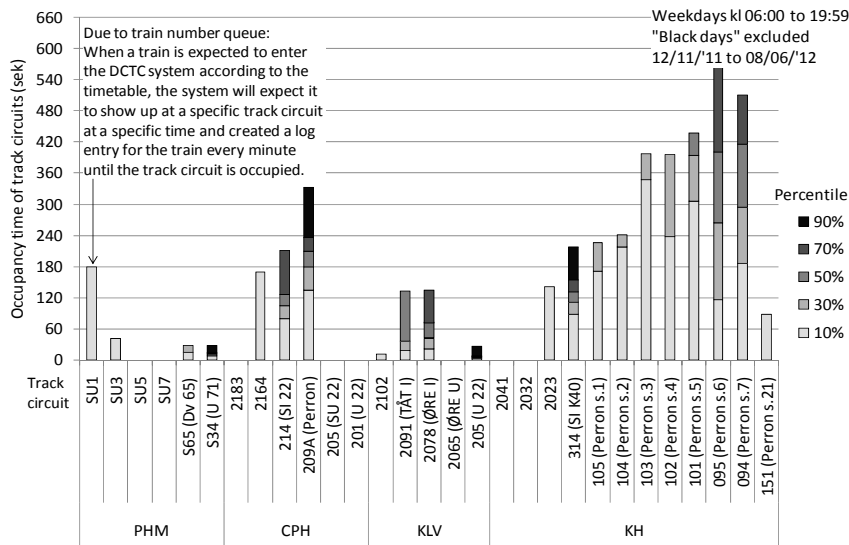


Figure 6: Occupancy time of track circuits.

The difference in delay between consecutive trains is clearly smaller where headway between trains is shorter. The halt at “Ørestaden” station (ØRE) is easily visible in this data.



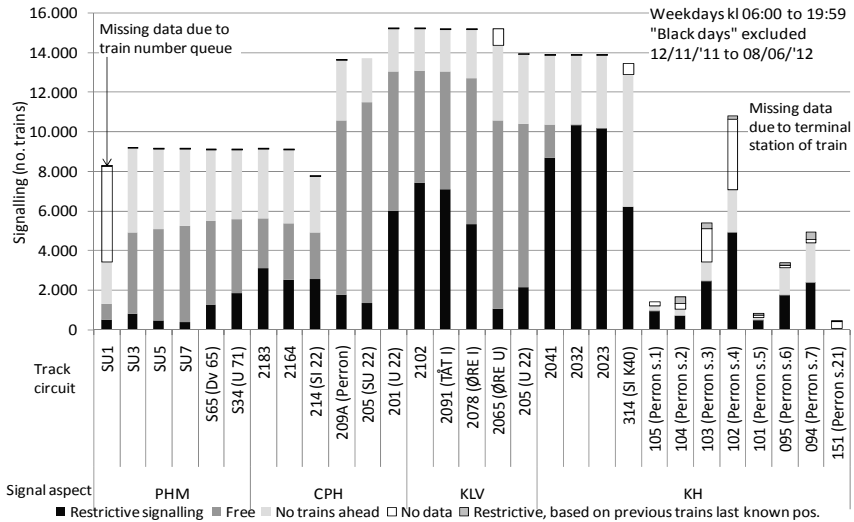


Figure 7: Track circuits where trains receive restrictive signal aspect.

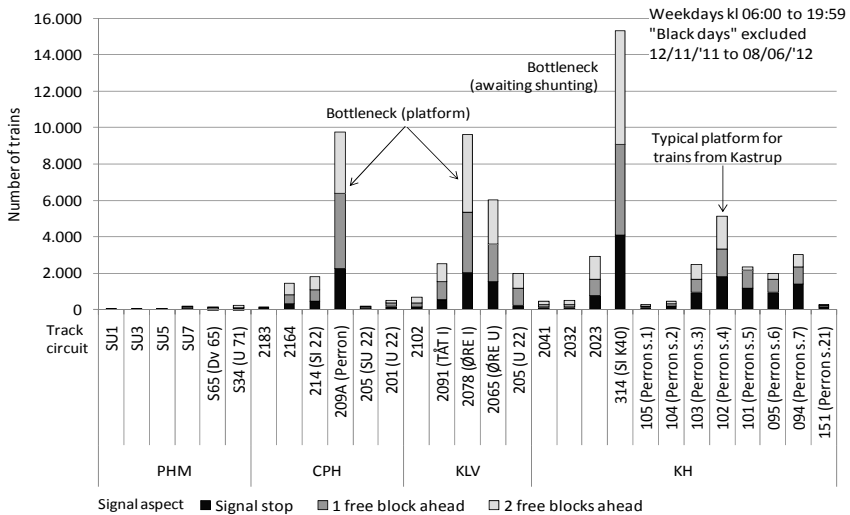


Figure 8: Track circuits with initial cause for restrictive signal aspect.

Occupancy time of track circuits indicates where trains have come to a halt. This can either be due to stop at platforms or due to queuing before platforms / stations. This can thus be used as an indicator of bottlenecks.

Another way to indicate bottlenecks is where trains have received restrictive signal aspect. This visibly happens before the halts TÅT and ØRE and approach to Copenhagen Central station (KH).

The root cause of the restrictive signal aspect is also of relevance (head of the tree with restrictive signal aspect). Entrance to Copenhagen Central station and the platforms at Kastrup station and Ørestaden are such track circuits.

Simple observations based on the enriched train run history at track circuit level indicates are that trains typically have long occupation times on track circuit 209A at Copenhagen Airport station (CPH), as well as block section 2091 and 2078 between Copenhagen Airport station (CPH) and the junction Kalvebod (KLV). This is expected since platforms are located at these track circuits. The track circuits before the platform at Copenhagen Airport station also have a long occupation time which is not surprising due to the layout of the station (one platform track in each direction). Long occupation time before Copenhagen Central station (KH) is not surprising either due to shunting. A future logical step is to expand this approach with analysis of differences between train systems and where these train systems typically receive restrictive signal aspect.

Additional information on performance can also be gained through Top 5 lists of worst performing trains and track circuits due to different criteria.

Table 4: Trains causing most restrictive signalling.

<b>Train</b>	<b>Restrictive signal/day</b>
1372	12,3
1026	12,2
1028	12,1
542	11,9
1040	11,7

Table 5: Track circuits causing the most restrictive signal aspect.

<b>Direction</b>	<b>Track circuit</b>	<b>Restrictive signal/day</b>
→ SE	CPH: 107B (Platform)	243,3
→ SE	PHM: SN1	175,9
→ KH	KH: 314 (SI K40)	119,6
→ KH	CPH: 209A (Platform)	76,1
→ KH	KLV: 2078 (ØRE I)	7

Table 6: Trains most often subject to restrictive signal.

<b>Train</b>	<b>Restrictive signal/day</b>
1062	12
1462	12
1453	11,9
487	11,7
21429	11,7

Table 7: Track circuits most often being subject to restrictive signal aspect.

Direction	Track circuit	Restrictive signal/day
→ SE	CPH: 1077 (TÅT I)	81,6
→ KH	KH: 2032	8
→ KH	KH: 202	79,6
→ SE	CPH: 1088 (TÅT U)	69,6
→ KH	KH: 2041	67,7

Table 8: Train / track circuits most often causing restrictive signal aspect.

Dir.	Train	Track circuit	Restric. sign. /day caused
→ SE	1342	CPH: 107B	8,1
→ SE	1390	CPH: 107B	7,8
→ SE	1354	CPH: 107B	7,6
→ SE	1384	CPH: 107B	7,6
→ SE	1336	CPH: 107B	7

Table 9: Train / track circuits most often subject to restrictive signal aspect.

Dir	Train	Track circuit	Restric. sig./day
→ KH	1059	KH: 2032	1
→ SE	18	CPH: 1077	1
→ SE	40	CPH: 1077	1
→ SE	136	CPH: 1088	1
→ KH	1029	KH: 2023	1

From these Top 5 lists it can i.e. be concluded that the 13xx system and the track circuit 107B at Copenhagen Airport station is the combination of train/track circuit which most often causes restrictive signalling for other trains (Table 8). It can also be concluded that train 1059 receives restricted signal aspect at the entry to Copenhagen Central station every day (Table 9). An in-depth analysis of these observations is a logical future step. It should be noted, that the restrictive signal aspect does not necessarily mean that the train has come to a complete halt.

These simple Top 5 lists only give simple examples of possible additional usages of the enriched train run history. Additional analysis can focus on differences as a function of time of day or hourly train path. Other relevant approaches are in-depth investigation queuing events, i.e. if similar delay patterns always trigger similar escalations of delays. Such investigations shall try to pinpoint specific triggers of escalations of delays.

These very simple figures and tables have thus demonstrated that the enriched train run history can provide new information on train run performance which again can lead to an improved operation after corrective actions are taken.

## 5 Discussion/conclusion

Investigations of conflicting train paths and robustness of timetables with respect to conflicting trains path have been investigated in a number of papers i.e. [9]. However, aggregations of data allowing for analysis of dynamics of delays and queuing effect on single lines between stations have received less focus.

The method described in this paper focuses on providing data to allow an examination of the dynamics of trains operation on a single line using high density delay information (delays at track circuit level). Furthermore, a number of simple usages of this have demonstrated the potential of the enriched train run history. Focus is on indicators of queuing on a single line which can be used as background for adjusting timetables, pinpoint infrastructure bottlenecks and procedural inefficiencies.

With data available and an approach to analysis of this data ready, the in-depth analysis work can begin. The interpretation of indicators of queuing effects is to be done in cooperation with other sections within BDK such as the timetabling sections and the Traffic Control Centres: Without an organisational implementation of methods such as the demonstrated one, the methods have little practical relevance. However, this method and the analysis suggested fits well into the existing Plan – Do – Check – Act loop for train run performance monitoring within BDK.

## References

- [1] Landex, A. *Methods to estimate railway capacity and passenger delays*, Ph.D. thesis, DTU Transport 2008.
- [2] Rail Net Denmark. Network Statement 2013.
- [3] Schittenhelm, B., Richter, T., *Railway Timetabling Based on Systematic Follow-up on Realized Railway Operations*, Annual Transport Conference at Aalborg University 2009.
- [4] Richter, T., *Systematic analyses of train run deviations from the timetable*, COMPRAIL XII, 2010.
- [5] UIC, *Assessment of the performance of the network related to rail traffic operation for the purpose of quality analysis – delay coding and delay cause attribution process (UIC leaflet 450-2)*, International Union of Railways (UIC), Paris, France, 2009.
- [6] Homepage of Rail Net Denmark, [www.bane.dk](http://www.bane.dk), 1 June 2012.
- [7] Traffic Operations Planning, Rail Net Denmark.
- [8] UIC, *Capacity (UIC leaflet 406)*, International Union of Railways (UIC), Paris, France, 2004.
- [9] Cule, B., Goethals, B., Tassenoy, S., Verboven S. *Mining Train Delays*, 10th International Symposium on Intelligent Data Analysis (IDA 2011).

