

Simulation and railway timetabling norms

M. Vromans¹, R. Dekker² & L. Kroon^{1,3}

¹*School of Management, Erasmus University Rotterdam, the Netherlands*

²*School of Economics, Erasmus University Rotterdam, the Netherlands*

³*Netherlands Railways, Dept. of Logistics, Utrecht, the Netherlands*

Abstract

Speed differences between local trains and intercity or high-speed trains are one of the main causes of the high utilization rate, and therefore of the sensitivity for disturbances of the Dutch railway system. In this research a practical case is evaluated, where a situation with large speed differences is compared with a situation with small speed differences. The speed differences are decreased by redistributing the stops over the train services: fewer stops of the local services are compensated by extra stops of the intercity trains.

The simulation software SIMONE is used to evaluate the different scenarios on punctuality. The program is capable of simulating nationwide networks and communicates with the semi-automatic timetabling program DONS. SIMONE is used both for (practical) comparison of two or more timetable options, as well as for research purposes.

Keywords: railways, reliability, simulation, heterogeneity, homogeneity.

1 Introduction

In railway traffic the same infrastructure is used by cargo and passenger operators. These passenger services range from local stop trains to international high-speed connections in the Netherlands. Combined with the high utilization of the infrastructure and the high level of interconnectivity between trains, the reliability of the Dutch railway traffic has become very vulnerable. Passenger transfers, rolling stock circulations and crew schedules all play their role in the relations between trains. However, the shared use of the same infrastructure by different train services, with different speeds and different origins and destinations, is probably the main reason for the propagation of delays



throughout the network. This makes national railway systems much more vulnerable for disturbances than metro systems, since usually all metros stop at all stations and lines are independent from each other.

Reliability is, together with door-to-door travel time, probably the predominant performance measure for railway traffic. The predictability of the arrival times is a big factor in deciding to use rail or road transport, both in passenger and cargo services.

We start our paper with a short section on punctuality and reliability. In the next section we discuss heterogeneity and state our hypothesis. This is followed by the introduction of two new timetable measures. In section 4 a practical case is worked out. This includes detailed simulation for the comparison of different timetables. The results are related to the new timetable measures. Finally the results are discussed and conclusions are drawn.

2 Punctuality and reliability

Performance measures are needed in railway research to quantify the reliability of the railway traffic. Some of these measures are already in use in everyday railway practice to inform passengers, and for authorities to guard the quality of public transport.

The two measures used in this paper are punctuality and average arrival delay. The latter one speaks for itself. Punctuality is probably the most widely used reliability measure in practice (Schaafsma, [3]), both in the Netherlands and abroad. This measure calculates the percentage of trains arriving within a certain number of minutes from the scheduled arrival time.

When researching railway reliability, it is important to make a distinction between primary and secondary delays. Primary delays are initial delays caused on a train, independently from other trains. These include malfunctioning rolling stock, malfunctioning infrastructure, bad weather conditions, excessive alighting and boarding times, accidents at road-railroad crossings, and so on. Secondary delays are also referred to as knock-on delays, and are those delays on trains that are caused by earlier delays of other trains. These appear because of the shared use of the infrastructure, rolling stock connections, transfers in crew schedules, passenger transfers, and dispatching actions. In our study we consider the primary delays as given, and aim at developing timetables that both absorb primary delays fast, and cause as few secondary delays as possible.

3 Homogeneity, heterogeneity and headways

Railway traffic is considered homogeneous when all trains have similar characteristics; especially the *same average speed*, resulting from running times and stopping times. Good examples of homogeneous railway traffic are metro systems where all trains have the same running times and stop at all stations. However, for national railway networks, railway traffic cannot be fully homogeneous. Usually cargo trains and passenger trains share the same infrastructure. But probably more important, there is a large differentiation in



passenger services, ranging from short distance trains –which dwell at all stations underway– to international high speed connections –with high speeds and only stopping at a few large stations–, partly sharing the same infrastructure. When there are large differences in timetable characteristics for trains on the same track, the railway traffic is called heterogeneous.

3.1 Possibilities for homogenization

Homogenization of a railway system means that differences in running times of different trains along a railway line are decreased. There are several options for homogenization. Slowing down express trains by adding extra running time supplement is one option. Speeding up short distance services is much harder: running time supplements have to be decreased or faster rolling stock has to be purchased.

Smarter solutions include overtaking, shorter lines for the local stop services and equalizing the number of stops of the services. The latter means that some stops from the local services are shifted to the long distance services.

3.2 Hypothesis

When differentiated services share the same infrastructure over large distances with many small stations, timetabling becomes very complicated. Heterogeneity usually leads to many small headway times, which increases delay propagation in the operations and consequently hurts reliability. This leads to our hypothesis, which is worked out in this paper:

Hypothesis: the heterogeneity of the timetable has a negative influence on the punctuality and reliability of the railway system.

3.3 Heterogeneity measure

To prove this hypothesis we first develop a heterogeneity measure of the timetable. Then simulation is used to show the importance of heterogeneity of a timetable. Besides the fact that we use the measure in this paper to compare different timetables, it is also intended to be useful for the development of timetables for real world operations. In this paper we assume cyclic timetables.

Carey [4] describes several timetable measures based on headways and their distribution. In this section we develop new measures, which are also based on headways, and have some additional favorable properties.

The first measure that we use is the Sum of Shortest Headway Reciprocals, and is defined as: $SSHR = \sum_{i=1}^n (h_i^-)^{-1}$, with h_i^- the smallest scheduled headway between train i and $i+1$ on the track section, and train n is followed by train 1.

This measure takes all headways into account. The reciprocal is taken, because, given the number of trains per cycle, the sum of non-inverted headways is fixed.



Furthermore, this measure takes the surrounding tracks into account, by taking the smallest headway over the track. This incorporates the speed differences in the measure. The spread of the trains over the cycle is also taken into account.

Figure 1(a) shows a fully homogeneous situation, with the trains nicely spread over the hour. This leads to an SSHR of $\frac{1}{15} + \frac{1}{15} + \frac{1}{15} + \frac{1}{15} = 0.27$. The slightly heterogeneous situation in figure 1(b), and the very heterogeneous situations in figure 1(c) have an SSHR of $\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = 0.44$ and $\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 2$, respectively.

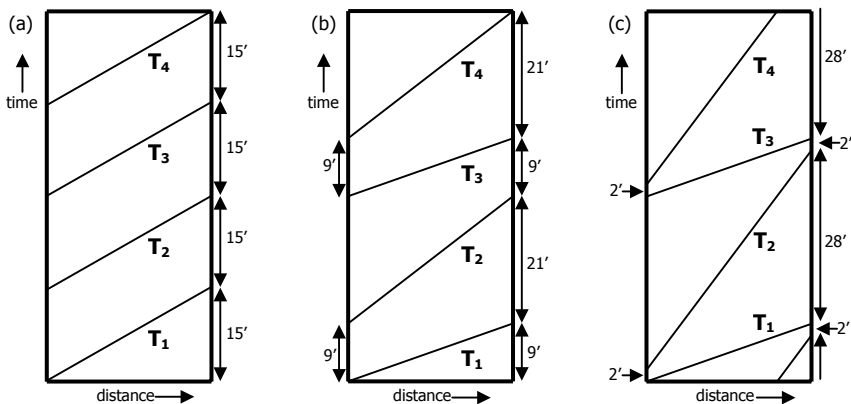


Figure 1: Three time-distance diagrams with homogeneous (a), slightly heterogeneous (b) and very heterogeneous (c) railway traffic.

Compare also the example in figure 1(a) with the situation where four homogeneous trains are not nicely spread, but have headways of 5', 25', 5' and 25': the SSHR increases from 0.27 to $\frac{1}{5} + \frac{1}{25} + \frac{1}{5} + \frac{1}{25} = 0.48$.

However, headways at arrival seem to be more important than headways at departure. The first reason is that delays at arrival are, on average, larger than at departure. Secondly, faster long distance trains can be caught behind short distance trains towards the end of a railway section. Therefore we have developed a second measure, which only depends on the arrival headways between every pair of subsequent trains, the Sum of Arrival Headway Reciprocals: $SAHR = \sum_{i=1}^n (h_i^A)^{-1}$, with h_i^A the headway at arrival between train i and $i+1$, and train n is followed by train 1.

In homogeneous situations the SAHR is equal to the SSHR, so the SAHR is 0.27 for the example in figure 1(a). In heterogeneous cases the SAHR is always less than the SSHR. The timetables represented in figures 1(b) and 1(c) have an SAHR of $\frac{1}{9} + \frac{1}{21} + \frac{1}{9} + \frac{1}{21} = 0.32$ and $\frac{1}{2} + \frac{1}{28} + \frac{1}{2} + \frac{1}{28} = 1.07$, respectively.

Unfortunately the SAHR does not take the track into account anymore. It is in fact a single location measure. Still, the arrival headways can only be evenly distributed when the timetable is not too heterogeneous.

3.4 Experiments

For this paper we have chosen the option of shifting stops from the short distance services to the long distance services to equalize the number of stops for obtaining a more homogeneous timetable.

The cyclic timetables described in this paper are developed with the timetabling tool DONS [1]. Any railway network can be defined with this software. On this infrastructure, train lines can be defined, including their intermediate stops, their rolling stock types, transfer connections and other characteristics. DONS will then provide a feasible timetable or, if this is not possible, tell which constraints following from the input make a timetable infeasible.

For the comparison of the timetables, simulation of railway traffic has been used. The simulations reported on in this paper are performed with SIMONE. This simulation model is in use both by ProRail, the Dutch infrastructure manager, and by Netherlands Railways, the predominant Dutch passenger operator. It is used both for timetable comparisons and for scientific research. A more detailed description of SIMONE is given by Middelkoop and Bouwman [2].

4 A practical case

To support our hypothesis, a practical case has been worked out. We have compared a real-life heterogeneous timetable with a more homogeneous timetable for a busy line in the Netherlands.

The case that is elaborated here consists of the lines from The Hague Central (Gvc) and Rotterdam Central (Rtd) to Utrecht Central (Ut), which merge at the non-level crossing Moordrecht Junction (Mda). This part of the network has double track everywhere, except for the section between Moordrecht Junction and Gouda Goverwelle (Gdg), which has four tracks. The distance from The Hague to Utrecht is 61 km, from Rotterdam to Utrecht 55.8 km.

4.1 The heterogeneous and homogeneous situations

With some adjustments, the 2003 rush-hour timetable has been taken for the heterogeneous situation [5]. Because cargo trains are only a relatively small portion of the railway traffic, these are skipped, resulting in a *three-train-system*: long distance trains, interregional trains and short distance trains, represented by solid (—), dashed (---), and dotted lines (·····), respectively in figure 2(a). In this figure, all lines have a cycle time of 30 minutes.

The homogeneous situation has fewer stops in the short distance services and more stops in the long distance services. The final dwell pattern of the homogeneous situation is shown in figure 2(b). There is no distinction anymore



between slower and faster services. The resulting SSHR and SAHR are given in table 1 below.

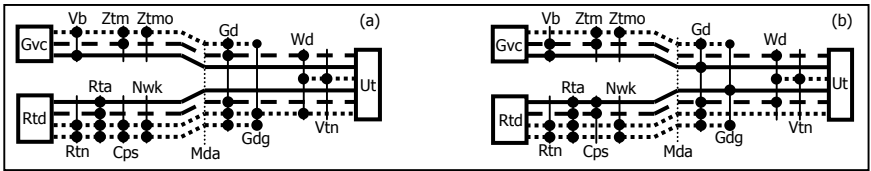


Figure 2: Train lines and dwelling patterns of the heterogeneous situation (a) and the homogeneous situation (b) of the practical case. Besides the junction Mda, all abbreviations indicate a station.

Table 1: Tables of SSHR and SAHR for the different tracks in the case.

SSHR		Hetero- geneous	Homo- geneous
From	To		
Gvc	Gd	0.76	0.67
Rtd	Gd	3.18	1.47
Gd	Ut	3.47	2.97
Ut	Gdg	5.52	3.02
Gd	Gvc	1.06	0.69
Gd	Rtd	2.83	1.19

SAHR		Hetero- geneous	Homo- geneous
From	To		
Gvc	Gd	0.64	0.62
Rtd	Gd	1.36	1.17
Gd	Ut	2.70	2.52
Ut	Gdg	2.55	1.77
Gd	Gvc	0.67	0.61
Gd	Rtd	1.74	1.07

4.2 Experimental design and simulation results

The simulation experiments consist of fifty runs of 1320 minutes, including 120 minutes of warm-up time. The primary delay distributions, all exponential, are given in table 2.

Absolute running time disturbances are independent of the running time and have the averages given in table 2. Opposed to that, the average size of relative running time disturbances is equal to a given percentage of the planned running time of the train on the track where it is disturbed.

Table 3 shows the results of the simulations. The homogeneous situation leads to a significantly better reliability than the heterogeneous situation.

The approximately 30% decrease in average arrival delay is quite substantial, but the most surprising result is the consistency of this extent for the different delay distributions. The average delay per train decreases with 28% to 35% for all experiments.

Obviously, because the primary delays are equal for the heterogeneous and the homogeneous case, the difference is made by the secondary delays. Delay propagation drops by 62% to 90%.

Another point worth noticing is the fact that a few large primary delays have a larger negative impact on the reliability than many small primary delays with the same total size of the initial delays. See for example experiments 4 and 5. One explanation is the fact that the timetable can absorb one small delay here and



another small delay somewhere else. However, it may be impossible to absorb one large delay.

Table 2: The experiments and their primary disturbances.

Experiment	DWELL TIME DISTURBANCE		ABSOLUTE RUNNING TIME DISTURBANCE		RELATIVE RUNNING TIME DISTURBANCE		TOTAL PRIMARY DISTURBANCES (in minutes per hour)
	probability for disturbance	average disturbance (in minutes)	probability for disturbance	average disturbance (in minutes)	probability for disturbance	average disturbance (in % of running time)	
1	100%	0.6					95.9
2	100%	0.8					128.0
3	10%	1	5%	1			38.7
4	20%	1	10%	1			78.2
5	50%	0.5	20%	0.5			86.2
6	75%	0.5	30%	0.5			129.2
7					30%	40%	112.8
8					60%	20%	114.0
9	20%	1			30%	20%	89.3
10	50%	0.5			30%	20%	96.9

Table 3: The results of the case. The presented average delays have standard deviations between 0.3% and 2.4% in the heterogeneous case, and between 0.3% and 3.3% in the homogeneous case.

Experiment	AVERAGE ARRIVAL DELAY (per train measurement in minutes)			3-MINUTE DISPUNCTUALITY (% of trains delayed)			TOTAL INCURRED SECONDARY DELAYS (in minutes per hour)		
	Heterogeneous	Homogeneous	Improvement (relative in %)	Heterogeneous	Homogeneous	Improvement (relative in %)	Heterogeneous	Homogeneous	Improvement (relative in %)
1	1.08	0.72	33.4	8.8	3.1	64.6	30.0	3.2	89.3
2	1.85	1.22	34.0	21.1	10.4	50.8	50.3	8.4	83.4
3	0.60	0.40	33.7	5.0	3.0	39.9	20.9	4.2	80.2
4	1.40	0.92	34.5	15.0	9.0	39.9	41.3	10.8	74.0
5	1.13	0.75	33.6	8.7	3.9	55.4	28.8	3.9	86.6
6	2.02	1.35	33.2	23.1	12.0	48.2	47.5	8.7	81.7
7	2.21	1.55	29.9	25.9	17.6	32.1	60.2	22.9	62.0
8	1.71	1.22	28.6	18.3	10.8	41.1	38.7	10.4	73.1
9	1.30	0.89	31.4	12.4	7.5	39.6	35.0	8.7	75.2
10	1.28	0.88	31.0	11.4	5.9	48.0	32.1	6.3	80.4

A final remark should be made on the dwell time disturbances. These disturbances are randomly generated at all stops, which means that local stop trains are more vulnerable for this kind of disturbance. Implicitly, average realized running time differences are increased, which means that this disturbance has an additional heterogenization effect.



4.4 SSHR, SAHR and the results

In figures 3 and 4 the SSHR and SAHR improvements are compared with the average delay reduction. For this comparison, the network is divided into six sections: Gvc-Gd, Rtd-Gd and Gd-Ut and vice versa. The average delay for each section, as given in the figures, is the average delay of all trains at the endpoint of the section minus the average delay of these trains at the start of the section.

The upper right of each line-segment in figure 3, represented by a circle, is the SSHR and average delay in the homogeneous situation. The other endpoint, represented by a square, is the result of the homogenization. Figure 4 shows the same comparison for the SAHR. The graphs represent the results of experiment 7. The figures for the other experiments are quite similar.

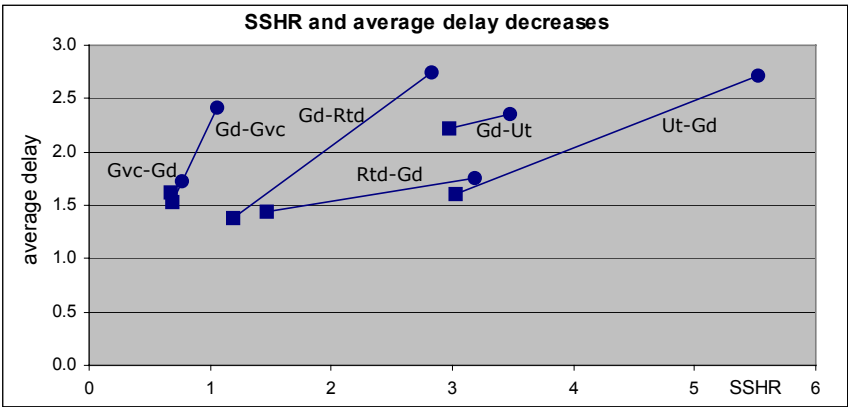


Figure 3: The relation between the decrease in SSHR and the reduction of the average delay per section.

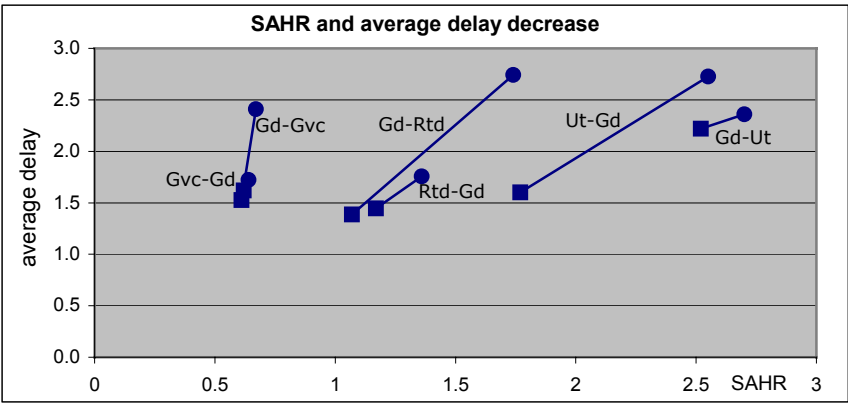


Figure 4: The relation between the decrease in SAHR and the reduction of the average delay per section.



- Figures 3 and 4 show that homogenization of the timetable leads to both a reduction in the SSHR and SAHR measures, and a reduction in average delays for all sections.
- The decrements in the SSHR are quite different for the different sections: Gvc-Gd only shows a small difference, while Ut-Gd shows a large decrease. The same is true for the SAHR.
- The relative reductions in the SSHR are not equal, or almost equal, to those of the SAHR. This means that the SSHR and SAHR are two quite distinct measures. See for example the section Rtd-Gd.
- In general, a larger decrease in the SSHR leads to a larger reduction in delays. Still, the line-segments for Rtd-Gd and Gd-Ut are rather flat. This means that the delays do not decrease as much as the SSHR might indicate. Therefore the SSHR can be used as an indication in what direction the reliability goes, but it is not an absolute measure. For the starting sections, the departure reliability from the first stations is very high. This means that small departure headways hardly have a negative influence on the reliability. The reduction of the SSHR on Rtd-Gd is mainly based on a fairer distribution of the departures; the arrivals are hardly affected by the homogenization.
- Also, a larger decrease in the SAHR leads to a larger reduction in delays. Still, sections with a relatively large delay reduction (a steep line-segment) can be observed (e.g. Gd-Gvc), as well as sections with a relatively small reduction (e.g. Gd-Ut). Due to the large reduction of the delays of the section Ut-Gd, as predicted by the SAHR, there is also a large reduction in departure delays for the section Gd-Gvc. A better starting reliability implies fewer secondary delays, which explains the large delay reduction on Gd-Gvc.

5 Discussion and conclusions

The practical case in section 4 shows that homogenization of the timetable can decrease delay propagation significantly and increase punctuality. In [6] Vromans et al. describe a theoretical case with comparable results. Simulation of the timetables has been performed using a wide range of primary delays. The results of these diverse scenarios are so similar, that we may conclude that the results are valid for any kind of disturbances of relatively small size.

The results from the detailed simulations support our hypothesis: *the heterogeneity resulting from the line plan and the timetable has a negative influence on the punctuality and reliability of the railway system.*

The SSHR and SAHR are both able to forecast reliability changes. The SSHR is applicable to tracks between stations, whereas the SAHR can be used for a station, or at least all arrivals from a certain track at that station.

Although a large decrease in SSHR or SAHR leads to a large decrease in delays in most cases, it is hard to predict the exact size of the delay reduction. Still, the relatively simple rule of thumb for timetable design to minimize the SSHR and the SAHR may improve the reliability of the offered services.

Minimizing the SSHR or the SAHR implies equalizing headways. Although a reduction in these measures indicates a reduction in delays, minimizing the sum



of SSHRs or SAHRs over the network is not necessarily optimal. This can for example be seen from the SSHR in the case, where a large reduction on one section (Rtd-Gd) has much less influence on the reliability than the small reduction of the SSHR on another section (Gd-Gvc).

Homogenization affects more characteristics of the railway traffic than just reliability. Although reliability is one of the most important features, some other characteristics should also be taken into account.

- *Travel time* for passengers can be affected significantly by homogenization. The planned travel time may increase for some passengers, but decrease for others. The number of passenger transfers and the transfer times may also change.
- *Infrastructural needs* may change due to altered train lengths, but also due to other locations for overtakings, and a changed coordination of trains at large transfer stations.
- The *rolling stock* can be homogenized as well. Furthermore, the total required number of rolling stock units can also change.
- Homogenization by one operator can lead to *additional time-slots* in the timetable, which may be assigned to other operators. Evidently this would, in the end, lead to an increased SSHR and SAHR. Therefore, network wide cooperation is necessary for a beneficial introduction of a homogenized timetable.

References

- [1] Hooghiemstra, J., An automatic timetable generator for the Dutch railway network, in *Computers in Railways IV*, Vol 2, Computational Mechanics Publications, Southampton, pp. 109-116, 1994.
- [2] Middelkoop, D., Bouwman, M., Train network simulator for support of network wide planning of infrastructure and timetables, *Computers in Railways VII*, eds. J. Allen et al., Computational Mechanics Publications/WIT Press, Southampton, pp. 267-276, 2000.
- [3] Schaafsma, A., Dynamisch railverkeersmanagement (Dynamic rail traffic management). PhD. thesis, Technical University Delft, Delft, the Netherlands, 2001. In Dutch.
- [4] Carey, M., Ex ante heuristic measures of schedule reliability, *Transportation Research Part B*, 33(7), pp. 473-494, 1999.
- [5] Nederlandse Spoorwegen, Spoorboekje, 15 December 2002 t/m 13 December 2003 (Timetable). Nederlandse Spoorwegen, the Netherlands, 2002. In Dutch.
- [6] Vromans, M.J.C.M., Dekker, R., Kroon, L.G., Reliability and heterogeneity of railway services, *ERIM Research Series 2003 090*, Rotterdam, the Netherlands, 2003.

