Relationships between capacity, speed heterogeneity, and robustness against delays in railway networks

Lars Wittrup Jensen

Abstract

The optimal utilisation of railway capacity is essential to obtain high service quality and an effective railway system. In this context the amount of capacity that can be consumed without causing excessive delays (the practical capacity) is a key parameter in the planning process. Previous research has mainly dealt with this subject by examining lines and line sections and not more complex networks. This paper therefore investigates the relationship between practical capacity and maximum capacity in networks as well as the relationship between speed heterogeneity, infrastructure occupation, and robustness against delays in networks. Based on this, an initial suggestion for additional time rates required to obtain the practical capacity is given for a network in Southwestern Denmark.

Keywords Railways · Capacity analysis · UIC406 · Robustness · Delays · Heterogeneous operation

1 Introduction

The available infrastructure capacity in a railway system is not easily derived as the actual number of trains that can be handled depends on several operational parameters as well as the characteristics of the infrastructure and rolling stock.

The UIC (2004) (International Union of Railways) defines the operational parameters that affect capacity as the number of trains, stability (or robustness against delays), average speed, and speed (running time) heterogeneity. These four parameters together define how capacity is utilised given the infrastructure, rolling stock, and the probability of disturbances and disruptions. Different timetables may utilise capacity in different ways according to these parameters.

The relationship between the four different operational parameters are well-known in practice and literature (described in Section 1.1). However, existing research has mainly dealt with one or more of these relationships on lines and line sections. Examples of such research can be found in Abril et al (2008); Landex (2008); Harrod (2009). In networks, trains on different branch lines may have an impact on each other via a third, common, branch or a junction. Defined in this paper as network dependencies. These dependencies have an impact on network-wide capacity results. In the context of heterogeneity and stability/robustness, there is little or no
literature regarding the relation with capacity on a network-wide scale. This paper is therefore concerned with the answer to the following questions related to capacity, heterogeneity, and robustness against delays (stability) in networks to provide better insight on the subject for both practitioners and researchers:

1. What is the relationship between maximum capacity and practical capacity in networks – or formally infrastructure occupation and capacity consumption?
2. What is the relationship between speed (running time) heterogeneity, infrastructure occupation, and robustness against delays (in networks)?

The first question deals with the relationship between practical capacity (the maximum capacity obtainable while still obtaining a robust service) and maximum capacity (theoretical capacity). In this context, the focus in the paper is on a network in Southwestern Denmark and to compare the results from this case with recommendations for line sections by the UIC (2004, 2013). As shown by Jensen et al (2017) capacity analyses of networks (and routes) yields higher infrastructure occupation values than analyses of line sections (the latter being the standard in capacity analyses). This comes from the higher amount of dependencies in the network (and longer sections with a speed heterogeneity). This also means that additional buffer times are implicitly added to a compressed network timetable than to a compressed line section timetable. Thus, this paper seeks to study how using a network-wide analysis approach, rather than a line section approach, impacts the necessary additional time rate to be added to the infrastructure occupation to conclude if the practical capacity of a network has been fully utilised.

The second question deals with the buffer times that are made inherently available due to speed heterogeneity between different trains as described above. In heterogeneous operation, heterogeneity leads to higher robustness compared to homogeneous operation. This may allow for higher values of infrastructure occupation. As described by (UIC, 2004), this is an known aspect for lines, although not studied in great detail. Analysing networks rather than lines means that more dependencies are taken into account thus providing even more "hidden" buffer time. For this second question, this study thus seeks to find the relation between heterogeneity, infrastructure occupation, and robustness in networks and how this influence recommendations for infrastructure occupation, i.e. the relation between practical and maximum capacity.

The remainder of this section describes how the four parameters, the number of trains, robustness (stability), average speed, and heterogeneity, affect capacity on a line section. Section 2 describes the methodology and cases used, while Sections 3.1 and 3.2 present and discuss the results obtained. Conclusions and a summary are given in Section 4.

1.1 The four operational parameters

Fig. 1 depicts the utilisation of infrastructure capacity by two different types of operation given the four operational parameters defined by (UIC, 2004).

The average speed of trains affects the minimum headway time as higher speeds reduce the running time through a signalling block. However, the braking distance increases proportionally to the square of the speed. Therefore, the approach time to a signalling block increases at increasing speeds which results in longer minimum headways. As calculated by Abril et al (2008), the relation between average speed and minimum headway time can generally be described as the minimum headway times increase with increasing speeds, except at lower speeds where an increase in speed decreases minimum headway time. In Fig. 1 the low average speed in metro-like operation means that minimum headway times are reduced compared to higher speeds (mixed-train operation).
A robust and stable system can be obtained by adding running and dwell time supplements and buffer times. Adding supplements results in an increase in the minimum headway times as time supplements decrease the average speed and buffer times have to be inserted between trains. The increased minimum headway times mean that less trains can be operated if heterogeneity and average speed of train services are left unchanged. As described and proven by Landex (2008) there is a non-linear relationship between capacity and total delay for different amounts of input delay. Large buffer times and time supplements consume a high amount of capacity with little effect on achieved robustness in daily operation. However, if there are small or no supplements and buffer times, very large delays will inevitably arise. Buffer times and time supplements should therefore be optimised to obtain the required robustness while optimising capacity utilisation and shorten travel times.

A high speed heterogeneity leads to longer line headways between trains of different speeds. This is illustrated in Fig. 2 where (A) shows a cyclic sequence of six trains with trains of the same type bundled. This leads to a low cycle time (the time the train sequence occupies the infrastructure) compared to the most heterogeneous case (C) where no bundling of train types is done. In (C) train types are thus operated alternately. This is very common as it allows different train services to operate in equal intervals. (B) illustrates a scenario with a heterogeneity between (A) and (C). In Fig. 1 the metro-like operation has a low heterogeneity to allow for a higher frequency (number of trains). The mixed traffic operation has a high heterogeneity due to fast and slow trains that serve both short and long-distance passengers as well as freight.

Lastly, the number of trains (service frequency) is naturally a significant factor in the utilisation of infrastructure capacity. In the context of Fig. 1, running more trains means that heterogeneity, minimum headway times, robustness/stability or a combination hereof have to be reduced.

Besides utilising the capacity in different ways to obtain high heterogeneity, stability (robustness), average speed, or frequency, the infrastructure capacity may also be increased. This can be done by upgrading the signalling system and/or optimising the block sections to reduce critical block occupation time and thereby the minimum line headway times. The infrastructure capacity may of course also be increased by constructing more tracks and upgrading junctions to avoid conflicting routes, though this is a costly measure.
Fig. 2 Three different orders (sequences) of the same set of six trains divided on three slow and three fast trains. In A) the two train types are bundled (as homogeneously as possible), and thus this sequence consumes less capacity than C) where the sequence is completely heterogeneous.

2 Methodology

2.1 Model

The framework and model developed in Jensen et al (2017) is used as the tool to study the relationships between different parameters of capacity in the paper. The model samples different sequences (orderings) of trains (see Fig. 2) to obtain a span of capacity. The model is able to analyse complete networks rather than line sections. Robustness against delays is assessed using a stochastic simulation model that makes it possible to investigate the consecutive delays that trains in a train sequence suffer given a sampled input delay. Consecutive delays is also labelled as knock-on or secondary delays by other authors. The model by Jensen et al (2017) thus provides the possibility to assess the relationship between heterogeneity, infrastructure occupation, and consecutive (knock-on) delays in networks.

In the model of Jensen et al (2017) constraints can be used to exclude train sequences that are undesirable from a service perspective, e.g. due to bundling of similar trains. For the theoretical analysis in Section 3.2, no train sequences have been excluded to make the results less biased for the regression analysis (i.e. the decisions of which sequences to exclude is a subjective decision). For the analysis in Section 3.1, the same exclusions of undesirable sequences have been used as in the practical case of Jensen et al (2017). The output provided by the model is analysed visually using plots complemented by statistical analysis and measures to draw the conclusions.

2.2 Definition of measures

To investigate the impact of heterogeneity, an indicator of heterogeneity is needed. Jensen (2015) proposed a heterogeneity indicator for headway heterogeneity in cyclic timetables. Based on this indicator, the following heterogeneity indicator for running time or (average) speed heterogeneity in cyclic operation is proposed:
\[ RTH = 1 - \left( \sum_{i=1}^{n-1} \min \left( \frac{t_{i+1}}{t_i}, \frac{t_{n}}{t_1} \right) + \min \left( \frac{t_{1}}{t_n}, \frac{t_{n+1}}{t_1} \right) \right) \cdot \frac{1}{n} \]  

(1)

The heterogeneity indicator in Eq. 1 can only be used on line sections. That is, on sections where the number of trains and the order of trains does not change. For a network consisting of a set of connected edges aggregation therefore must be done to obtain a single heterogeneity value for the network. Three simple aggregation methods in the form of a simple average, a weighted average, or a maximum as expressed in Eq. 2, 3, and 4 are therefore tested in Section 3.2.1 for their ability to capture speed (running time) heterogeneity in a network.

\[ RTH^{Avg} = \frac{1}{|E|} \cdot \sum_{e \in E} RTH_e \]  

(2)

\[ RTH^{WAVg} = \frac{1}{|E|} \cdot \sum_{e \in E} n_e/n \cdot RTH_e \]  

(3)

\[ RTH^{Max} = \max_{e \in E} RTH_e \]  

(4)

2.3 Case networks

Two networks are used for the study conducted in this paper. The first network is a simple synthetic network (or corridor) as depicted in Fig. 3. As the figure shows, the network consists of four edges shaped like a Y. Five train types traverse the network with varying speed as represented by the running time in Fig. 3. Two trains of each type is added as an input to the model summing up to 10 trains total. The minimum headway time is set to 120 seconds. This simple network thus represents a corridor with converging routes and heterogeneous train operation, and is used for the analysis between heterogeneity and infrastructure occupation in Section 3.2.1.

The second network is the network of Southwestern Denmark from Jensen et al (2017) where one or more of four infrastructure scenarios are assessed. The network consists of 161 km of double track lines traversed by 18 trains in a period of one hour. The network is depicted schematically in Fig. 4 with tracks used in normal operation. The four infrastructure scenarios consist of the base scenario (the current infrastructure) and three upgrade scenarios. In the base scenario, the junctions TL and SNO are at-grade junctions causing track conflicts in opposing directions. In the upgrade scenarios, the junctions TL, SNO or TL and SNO are upgraded to out-of-grade junction resulting in less track conflicts.

Table 1 provides an overview of the train types that should traverse the network, the number of trains per type, distribution parameters for the stochastic simulation, and the route used. The train mix forms a heterogeneous operation. E.g. between OD (Odense) and SNO (Snoghøj) the slowest train type (RE-A) is 17 minutes slower than the express train type that uses 31 minutes between OD and SNO. For the case a minimum (block) headway time of 150 seconds is used.

Perturbed scenarios for the stochastic simulation are constructed by sampling initial delays from the Weibull distribution at entrance stations OD, FA, VM and ES and at each stop in the network (Jensen et al, 2017). Initial delays are truncated at 10 minutes as this corresponds to the smaller delays occurring in normal operation. A running time supplement of 9% is used for all trains in the stochastic simulation.
Fig. 3 Synthetic network consisting of four edges in a Y-shape. The network is traversed by five different train types with route and running time (minutes) per edge depicted.

Fig. 4 Case network in Southwestern Denmark.

Table 1 Train types in network with route used, number of trains per hour and Weibull distribution parameters for input delay at the entrance to the network.

<table>
<thead>
<tr>
<th>Train</th>
<th>Route</th>
<th>#/h</th>
<th>Shape</th>
<th>Scale</th>
<th>Running time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Express train</td>
<td>1: OD(4) → FA(2)</td>
<td>1</td>
<td>0.9</td>
<td>70</td>
<td>31 minutes</td>
</tr>
<tr>
<td></td>
<td>2: FA(6) → OD(2)</td>
<td>1</td>
<td>0.9</td>
<td>80</td>
<td>31 minutes</td>
</tr>
<tr>
<td>IC-A</td>
<td>1: OD(4) → FA(2)</td>
<td>1</td>
<td>0.9</td>
<td>60</td>
<td>33 minutes</td>
</tr>
<tr>
<td></td>
<td>2: FA(6) → OD(2)</td>
<td>1</td>
<td>1</td>
<td>65</td>
<td>33 minutes</td>
</tr>
<tr>
<td>IC-B1</td>
<td>3: OD(3) → ES(1)</td>
<td>1</td>
<td>0.9</td>
<td>60</td>
<td>67 minutes</td>
</tr>
<tr>
<td></td>
<td>4: ES(3) → OD(1)</td>
<td>1</td>
<td>1.1</td>
<td>55</td>
<td>67 minutes</td>
</tr>
<tr>
<td>IC-B2</td>
<td>31: KD(2) → VM(2)</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>50 minutes</td>
</tr>
<tr>
<td></td>
<td>42: VM(1) → KD(1)</td>
<td>1</td>
<td>0.9</td>
<td>60</td>
<td>50 minutes</td>
</tr>
<tr>
<td>RE-A</td>
<td>5: OD(4) → FA(3)</td>
<td>1</td>
<td>1.1</td>
<td>45</td>
<td>48 minutes</td>
</tr>
<tr>
<td></td>
<td>6: FA(5) → OD(1)</td>
<td>1</td>
<td>1.1</td>
<td>45</td>
<td>48 minutes</td>
</tr>
<tr>
<td>RE-B</td>
<td>7: FA(4) → ES(2)</td>
<td>2</td>
<td>1</td>
<td>70</td>
<td>64 minutes</td>
</tr>
<tr>
<td></td>
<td>8: ES(3) → FA(1)</td>
<td>2</td>
<td>1.1</td>
<td>45</td>
<td>64 minutes</td>
</tr>
<tr>
<td>Freight</td>
<td>9: OD(3) → VM(2)</td>
<td>2</td>
<td>1.4</td>
<td>200</td>
<td>72 minutes</td>
</tr>
<tr>
<td></td>
<td>10: VM(1) → OD(1)</td>
<td>2</td>
<td>1.4</td>
<td>200</td>
<td>72 minutes</td>
</tr>
</tbody>
</table>

3 Results

3.1 Practical and maximum capacity – the need for additional time rates

The maximum capacity of a railway system cannot be utilised as this inevitably leads to delay propagation due to a lack of critical buffer times between trains. To account for this UIC (2004, 2013) recommends that the infrastructure occupation rate should not exceed 60-85% depending
on type of operation. These maximum infrastructure rates may be converted into additional time rates according to the formula:

$$\text{Additional time rate} = \frac{1}{\text{max. infrastructure occupation rate}} - 1 \quad (5)$$

UIC (2004, 2013)'s recommendations for line sections as additional time rates are given in Table 2.

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Peak hour</th>
<th>Non-peak hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedicated (sub)urban passenger</td>
<td>18%</td>
<td>43%</td>
</tr>
<tr>
<td>Mixed traffic</td>
<td>33%</td>
<td>67%</td>
</tr>
<tr>
<td>Dedicated high-speed</td>
<td>33%</td>
<td>67%</td>
</tr>
</tbody>
</table>

Table 2: Additional time rates as suggested by UIC (2013) for line sections.

Fig. 5 shows box plots for additional time rates for the four infrastructure scenarios for the Southwestern Denmark network. The results have been derived by calculating the relative difference between infrastructure occupation and capacity consumption for each train sequence assessed by the model of Jensen et al (2017). As described by Jensen et al (2017), the difference
between infrastructure occupation and capacity consumption is the addition of critical buffer times. The relative difference between infrastructure occupation and capacity consumption of a train sequence is thus the additional time rates that have to be added to the occupation time of a train sequence to account for robustness. This also defined in UIC (2013).

The results in Fig. 5 are illustrated as box plots as the additional time rates may differ according to the delay propagation generated by a sequence, and thus the critical buffer times needed, depending on the input delay sampled (and the sequence itself). The maximum value thus represents the additional time rate necessary to account for the maximum delay propagation across all train sequences, the 75th percentile for the necessary additional time rate to account for the delay propagation of 75% of train sequences, and so on for the median, 25th percentile, and minimum.

Fig. 5 shows three things. Firstly, it shows that the maximum values of additional time rates are much higher than the 75th percentile. This is not unexpected as it is caused by a few extreme cases. Therefore, it is also recommended not to use the maximum additional time rate in operations planning as it will lead to an underutilisation of capacity. In this regard the 75th percentile serves as a better recommendation for the additional time rate.

Secondly, the figure shows that the 75th percentile is significantly lower than the UIC recommendations for both peak hour and daily (non-peak) operation. This result strengthens the hypothesis that "hidden" buffer times in a network lead to a lower need of additional buffer times as described earlier.

Lastly, it can be observed from the figure that the 75th percentile increase from the base scenario over the upgrade scenarios of Sno and Tl, respectively, to the upgrade of both Sno and Tl (from 16.2% to 21.2%). This is most likely caused by the fact that dependencies between trains are reduced (or removed) in the junctions. Thus, some "hidden" buffer times are removed resulting in a higher need of additional time rates (to add the necessary buffer times), again strengthening the hypothesis mentioned earlier on "hidden" buffer times.

Given the results, a preliminary recommendation for an additional time rate for a network similar to the network of Southwestern Denmark could be the average of the 75th percentiles. That is 18.3%, which should be compared to the UIC recommendations for mixed traffic in the peak hour (33%).

3.2 Heterogeneity, infrastructure occupation, and robustness

A heterogeneous train operation induces implicitly added buffer times in the beginning or ending of line sections which yield additional buffer times and thus a reduced risk of delay propagation. On the other hand, a high heterogeneity leads to higher infrastructure occupation as depicted in Fig. 2.

In the following two sections, the relationship between heterogeneity, infrastructure occupation, and robustness is investigated for networks, not just for line sections. In Section 3.2.1, the correlation between heterogeneity and infrastructure occupation is studied. Furthermore, this section also studies the different aggregation methods for the heterogeneity indicators listed in Section 2. Lastly, the relationship between heterogeneity and robustness is investigated in Section 3.2.2 using the stochastic model output in the form of consecutive delays and the best heterogeneity indicator found in the analysis conducted in Section 3.2.1.

3.2.1 Heterogeneity and infrastructure occupation in networks

As described in Section 1.1 there is a clear dependency between heterogeneity and infrastructure occupation on a line. To investigate the relationship between infrastructure occupation and speed
(running time) heterogeneity in a network, a regression analysis has been conducted on the simple (Y-shaped) network described in Section 2.3.

Specifically, the relation between infrastructure occupation and heterogeneity has been investigated as linear, (natural) logarithmic, reciprocal, and squared. For all three heterogeneity methods Eq. 2, 3, and 4, this study finds that it is possible to obtain a good fit with both a linear, logarithmic, reciprocal, and squared regression model for the simple network. By examining residual plots, it is found that the linear and logarithmic model perform slightly better than the reciprocal model which in turn perform a little better than the squared model. Between the linear and logarithmic model, the difference is insignificant. As the linear model is the simplest, this model is suggested as a reflection of the relationship between infrastructure occupation and heterogeneity in a simple network.

Fitted linear models for the three heterogeneity models result in a root mean square error (MSE) of 10.65, 10.62, and 12.02 for the average, weighted average, and the maximum aggregation methods. The $R^2$-values are, in the same order, 0.76, 0.76, and 0.69 for the three aggregation methods. Thus, the weighted average, Eq. 3, has the lowest MSE, and together with the average aggregation method the highest $R^2$ indicating that these heterogeneity indicators reflect the relationship between infrastructure occupation and heterogeneity better than the maximum aggregation. The difference between the average and the weighted average is small, and may be explained by the fact that the network is quite small and simple. Thus, the extra information contained in the weighted average is not necessarily utilised. As the weighted indicator does contain extra information, this is used in the further analysis in this section.

For the network in Southwestern Denmark, a regression analysis has been conducted similarly to the one for the simple network above. In Fig. 7 the correlation between infrastructure occupation and heterogeneity index using the weighted average is shown in the upper right and lower left part of the figure. The figure clearly shows that there is a non-linear relationship between
Fig. 7 (Bottom left and top right) Correlation between infrastructure occupation and heterogeneity (index) for the network of Southwestern Denmark (Pearson correlation coefficient is 0.71). (Top left) Histogram showing the distribution of heterogeneity index values. (Bottom left) Histogram showing the distribution of infrastructure occupation values. Heterogeneity indices aggregated across edges using weighted average (eq. 3).

A regression analysis of the output for the Southwestern Denmark case base scenario shows that the best regression model is the reciprocal model followed by the natural logarithmic model. The linear and squared model does not provide a good fit. The fitted reciprocal model is shown in Fig. 8. As the figure shows, the reciprocal model does not model the non-linear relationship between infrastructure occupation and speed heterogeneity perfectly. The regression model may therefore be further improved by more advanced transformations (e.g. power transformations or splines). To maintain simplicity, the reciprocal relationship is accepted as a general model.

Lastly, it should be noted that assessing output data from the upgrade of both junctions instead of the base scenario leads to a more linear correlation between infrastructure occupation and heterogeneity. Again this is probably caused by the reduced amount of dependencies in the network. The increased linearity from the base scenario to the upgrade of both junctions is expressed by an increase in the Pearson correlation coefficient from 0.71 to 0.80 (where 1
is a perfect positive linear correlation, 0 is no correlation, and -1 is a perfect negative linear correlation).

3.2.2 Heterogeneity and robustness (consecutive delays)

Using the weighted average heterogeneity index described in Eq. 3, the model of Jensen et al (2017), and the network of Southwestern Denmark, the relationship between heterogeneity, infrastructure occupation and consecutive delays have been calculated for a large sample of random sequences for the four infrastructure scenarios provided in the case.

The results of these calculations are summarised in Fig. 9 and 10, where Fig. 9 shows the results for the base network and Fig. 10 the results for the upgrade scenario of both junctions. The results depicted in the two figures represent the two extremes, why results for the individual upgrade scenarios of each of the junctions (TL and SNO) are not shown.

In the two figures, one point denotes the results (infrastructure occupation, heterogeneity index, and sum of consecutive delays) for a single sequence based on a single sample of input delays. That is one iteration out of the total amount of iterations completed. Thus, the same sequence is present in the figures multiple times based on different samples of input delay. This leads to the vertical banding seen in both figures (especially in the right part of Fig. 9) as the only parameter changing is the consecutive delay with infrastructure occupation and heterogeneity index remaining constant (not affected by different samples of input delay).

From Fig. 9 two things can be observed. One is the large cluster to the left which seems to indicate that sequences with low heterogeneity leads to higher delay propagation than sequences with higher heterogeneity. The other is the cluster to the right which shows a tendency towards higher infrastructure occupation leading to higher consecutive delays. The upwards tendency in the cluster to the right in Fig. 9 is not equally apparent in Fig. 10, which shows a more
Fig. 9 The relationship between infrastructure occupation (values increasing from left to right on the x-axis), heterogeneity index (as calculated using eq. 3), and consecutive delays for the base scenario in the Southwestern Denmark case. One point in the graph denotes one sequence based on a single sample of input delays (one iteration).

Fig. 10 The relationship between infrastructure occupation (values increasing from left to right on the x-axis), heterogeneity index (as calculated using eq. 3), and consecutive delays for the upgrade scenario of both junctions in the Southwestern Denmark case. One point in the graph denotes one sequence based on a single sample of input delays (one iteration).

linear relationship. The median consecutive delay in the base scenario is more than double of the median consecutive delay in the upgrade scenario. This leads to suggest that the upwards tendency in the right cluster in Fig. 9 is amplified due to additional dependencies in the network present in the base scenario, but not present in the upgrade scenario due to the upgrade of both at-grade junctions.

In Fig. 11 the same data is represented as histograms (note that the consecutive delay is on a natural logarithmic scale). The figure shows the same tendency as Fig. 9 and Fig. 10, however more aggregated. From Fig. 11 it can be observed that high heterogeneity (moving from left to right) leads to lower maximum delay propagation (top point on box plots), but not necessarily
lower average delay propagation (denoted by diamonds). The tendency of higher heterogeneity leading to lower delay propagation is stronger in the upgrade case. This is ascribed to the reduced amount of dependencies in the upgrade case of TL and SNO, as described earlier.

The output on infrastructure occupation, heterogeneity, and consecutive delays has not made it possible to obtain a good regression model based on these parameters. However, correlation statistics for the output data show a negative Pearson correlation coefficient of -0.15 for the upgrade of both junctions and -0.09 for the base scenario. With a reciprocal transformation (similar to the one done in Section 3.2.1) of the heterogeneity index the Pearson coefficients increase slightly to 0.11 for the base case and 0.16 for the upgrade of both junctions.

Acknowledging that there is some correlation between speed (running time) heterogeneity and robustness in networks, a link may drawn to the analysis of additional time rates in Section 3.1 which can now be extended by the heterogeneity of train sequences.

An analysis of the results of such a relation indicates the same tendencies as seen earlier in this paper. Thus, it is concluded that there is clearly a trend where higher heterogeneity leads to reduced maximum required additional time rates. That is the practical capacity is higher with a high heterogeneity if robustness against maximum delay propagation has to be handled. This does not imply that a train sequence with higher heterogeneity is always better than one with lower heterogeneity in networks, as this also depends on the dependencies in the network and the train order. However, if the only information available is the heterogeneity index, it can be concluded that high heterogeneity leads to lower maximum delay propagation. Average delay propagation, contrary to maximum delay propagation, decrease with higher heterogeneity (negative correlation) to some point, after which the correlation turns positive as illustrated by the diamond symbols (the average) in Fig. 11. This is mainly ascribed to network dependencies as the shift from negative to positive correlation occurs at higher heterogeneity index for the upgrade scenario.

4 Summary and conclusions

In the introduction, the following questions were raised related to the relationship between robustness and capacity in networks:
1. What is the relationship between maximum capacity and practical capacity in networks – or formally infrastructure occupation and capacity consumption?

2. What is the relationship between speed (running time) heterogeneity, infrastructure occupation, and robustness against delays (in networks)?

In the context of the first question raised, regarding maximum capacity versus practical capacity, our results show that the additional time rates, which can be used to obtain the practical capacity (capacity consumption) from infrastructure occupation values, are lower in networks than the recommendations given for line sections by UIC (2004, 2013). This is in line with expectations and is caused by the additional amount of buffer times available in the compressed network timetable.

Complementing this, it was observed that the infrastructure scenario with most dependencies (due to more track conflicts) in the Southwestern Denmark case needs a lower additional time rate compared to the scenario with less dependencies (due to an upgrade of junctions). This is a consequence of less buffer times due to less dependencies (track conflicts). Specifically, it was found that an additional time rate of approximately 18.3% is sufficient for the case network of Southwestern Denmark. The corresponding recommendation from UIC (2004, 2013) (line sections) is 33% for mixed traffic operation in the peak hour.

Regarding the relationship between heterogeneity, infrastructure occupation, and robustness in networks (the second question raised), the relationship between heterogeneity and infrastructure occupation was first investigated. Given a proposed heterogeneity indicator for speed (running time) heterogeneity, it was found that for networks the relationship between heterogeneity and infrastructure occupation is best represented by a reciprocal function. For simpler network it was found that a linear function also represents the relationship well. The correlation found between infrastructure occupation and heterogeneity is significant enough to be described by a (reciprocal or linear) model obtained using regression.

Using the developed heterogeneity index, the relationship between heterogeneity, infrastructure occupation, and robustness (consecutive delays) was investigated. It was found that train sequences with low heterogeneity have a risk of higher maximum delay propagation (consecutive delays) than train sequences with higher heterogeneity. The study also shows how network dependencies, in the form of track conflicts, may lead to higher consecutive delays for a large proportion of the possible train sequences.

Acknowledgements The research described in this paper is part of the research project RobustRailS funded by the Danish Council for Strategic Research (now part of Innovation Fund Denmark). The author would like to thank the council for making it possible to conduct this research.

References

