Australasian Transport Research Forum 2013 Proceedings
2 - 4 October 2013, Brisbane, Australia
Publication website: http://www.patrec.org/atrf.aspx

Rail simulation and the analysis of capacity metrics

James Gray

Melbourne, Victoria

Email for correspondence: jamespaulpetergray@gmail.com

Abstract

Population growth is changing the form of Australian cities. The limits of metropolitan rail systems are being stretched, such that infrastructure and operational changes are required to provide additional capacity. Combined with an increasing desire in the modern world to ensure that decisions are better informed by evidence, there is a need to develop objective methods to assess and quantify railway capacity.

It is generally accepted that the capacity of a railway represents the number of trains that can reliably operate in a given section of track in a given time period. Reliable operation however is a subjective measure, with capacity a balancing act between travel time, punctuality and service frequency. This paper investigates how operational modelling can balance these factors and help assess rail capacity. The use of simulation modelling is attractive because it can easily manage the large number of variables that impact operational performance.

A parameter approach to capacity assessment is proposed and validated on the Melbourne metropolitan network. A number of metrics are compared and contrasted for their ability to inform the level of sectional rail capacity. Both an on-time performance threshold and the average arrival delay are found to match with observed capacity benchmarks. The build-up of both input and inherent delay may also be attractive measures in certain circumstances, while there is a level of subjectivity in the selection of the capacity thresholds with these measures, a precise threshold is less important when the desire is only to seek a relative assessment between different timetable and/or infrastructure options.

1. Introduction

Changes to the structure of the Melbourne metropolitan rail network, alongside a mix of demographic, economic and environmental factors have led to an increase in patronage on the Victorian railways since 1980. Following a recent surge (from the early 2000s), boardings on the Melbourne network have risen to a level in excess of the previous historical record.

This growth has stretched the limits of the rail system, such that infrastructure and operational changes are now required to provide additional capacity. Eddington (2008, pg. 77) notes that in Melbourne the patronage rises initially led to a significant decline in network performance (crowding, punctuality and passenger satisfaction). There is a need to determine and assess railway capacity in a robust manner to better understand how changes to operational methods, delays and parameters impact capacity.

As a major transport mode, poor levels of railway performance have potential to impact a wide variety of stakeholders through financial penalties for private operators, political implications for the government and economic costs for businesses and the community.

Currie (2010) highlights how rail overcrowding is becoming an endemic problem around the world and notes the difficulties associated with obtaining project funding and the timescales required to develop and construct infrastructure solutions. It is vital that proposed capacity solutions are properly developed to ensure they:

- Represent the most effective and efficient option available
- Can be justified to funding bodies in a robust manner
- Are consistent with longer term strategies

It is generally accepted that the capacity of a railway represents the number of trains that can reliably operate in a given section of track in a given time period. However, railway capacity remains an elusive property because reliable operation is a subjective measure and difficult to uniquely define. In practice capacity is a balancing act between:

- Travel time
- Punctuality
- Service frequency

To be able to assist decision making modelling needs to provide an objective measure of capacity that enables reliable operation to be quantified. The railway industry currently uses both traditional (operational experience, analytic desktop assessment) and simulation (parameter, detailed simulation) approaches to inform decisions of rail operations.

The complexity involved in detailed simulation assessment means that long timeframes are often required to develop robust timetables, particularly if different timetable variants are required for different infrastructure options. The examples presented in this paper are based around parameter simulations and demonstrate that while simpler to timetable this approach is still a robust and objective method for calculating practical rail capacity.

This paper investigates measures to quantify the operational capacity of passenger railways. It outlines a robust and transferable methodology that can demonstrate the capacity benefits that can be achieved through operational and/or infrastructure changes to a railway.

Beginning with a review of philosophies for decision making, we outline the major factors that can impact operational capacity and previous research into rail capacity before demonstrating the use and capability of simulation to help analyse rail operations.

2. Philosophy of decision making

The ability to make appropriate decisions is important in a range of environments. Common techniques used to support decision making include analysis, debate, emotion, experience, intuition and popular opinion. While the thought processes behind personal decisions can be subjective and secretive, business and political decisions usually need to be associated with a degree of confidence and transparency in how they have been reached.

Garvin and Roberto (2001) describe two approaches for decision making:

- Inquiry is a collaborative problem solving exercise where discussion is used to test and evaluate ideas, minority views are cultivated and considered as potential alternatives and the outcome results in collective ownership of the decision.
- Advocacy is a contest where discussion is used as a means to persuade and lobby people, minority views are discouraged or dismissed and the outcome results in winners and losers.

In a review of public transport performance, the Victorian Auditor-General supports the *"inquiry"* approach and highlights the importance of the wider industry adopting a *"well-structured, evidence-based approach to understanding future threats to performance and developing ways of addressing them"* (VAGO, 2012, pg. 23).

Argyrous (2009) outlines the merits of both evidence and ideological based decision making in the public sector:

- Evidence based decision making involves undertaking a systematic assessment of all potential options, it requires a detailed level of analysis of multiple options meaning it is in general, a time consuming process.

- Ideological based decision making represents a more high level approach, where decisions can be made in a more timely manner. Ideological based decisions can be made quicker because they rely on the application of common beliefs and attitudes that are already held at the outset of the analysis. These beliefs and attitudes can often be formed from practical experience in the industry.

3. Modelling and simulation

The use of modelling to inform train planning and project development has increased in popularity over the last 20 years. It has attained particular prominence over the last decade following the introduction of several integrated simulation packages capable of undertaking timetable development, simulation and analysis of rail operations (see for example, Abril et al, 2008; Hansen and Pachl 2008).

Modelling is useful because rail operations are complex, it aims to *"capture some important aspects of a real situation in a way that permits easy comprehension and manipulation to aid decision"* (Jessop, 1990, pg.13). While conventional analysis techniques can struggle to handle the large number of variables that are often involved in operational investigations, simulation can provide an objective approach for quantifying operational benefits.

Modelling generally takes one of three forms (outlined in table 1). While desktop and spreadsheet models are suitable for conflict free and delay free situations, numerical simulations are more easily able to consider a broader range of variables and operational issues (multiple interacting lines, delays and signal conflicts).

Table 1: Three common forms of modelling assessment

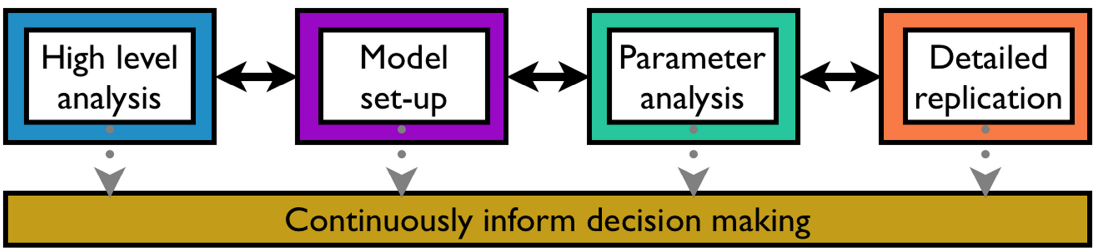
Analytic/ desktop models	This class includes basic high-level, analytic methods and/or basic calculations that are suited to being solved by hand or with the assistance of a spreadsheet. Desktop models can be developed within a relatively short timeframe without requiring specialist software training and/or programming skills.
Parameter simulation	Parameter models help to develop a timely and intuitive understanding of operational performance by considering the key operational factors in isolation, they seek to understand *"why an impact is occurring?"* rather than precisely calculate *"the magnitude of the impact"*.
Detailed simulation	Detailed replication investigations are comprehensive studies that consider the full range of parameters in the development of the operational model. Typical considerations in the railway context include infrastructure, signalling, rolling stock, crew rostering, safe-working rules, timetables and train-train connections.

An integrated analysis should use a range of different approaches on an iterative basis:

- High level analysis and experience to identify relevant issues
- Use of a valid and calibrated model based on relevant data and insight
- Break operations down to see how individual factors (parameters) affect performance
- Create a full timetable to test for operational performance

An integrated approach also allows for the outputs from the high-level analysis and parameter simulations to be used to optimise the inputs of detailed simulations (figure 1).

Figure 1: Diagram outlining how different approaches for the analysis of rail operations can inform each other as part of an integrated decision making process.



Simulation is about more than quantifying a figure for use in the economic analysis of a business case. Combined with a level of practical operational knowledge, simulation is a valuable tool to assist the development of concept service plans and the identification of project infrastructure requirements.

4. Operational capacity

A high level overview of issues associated with rail capacity is given by Connor (2010); a more technical analysis is found in the transport and quality of service manual (TCQSM, 2003). The key influences on operational capacity and railway performance are explored in table 2.

Table 2: Key influences on operational capacity

Signalling system	Railway signalling exists to ensure that a safe separation is maintained between trains operating on a rail network. The minimum separation that can be provided is known as the signalling headway and has a major impact on capacity because it constrains the number of trains that can be scheduled in a given time period.
Train delays	A delay represents *"a deviation in time from a scheduled event"* (Yuan, 2008). Typical delays in passenger railway operations include dwell delay (passenger movements), incident delays (environmental, infrastructure and/or rolling stock faults) and train-train interactions (between consecutive trains and/or at junctions).
Performance allowance	Rail timetable planners generally schedule trains to travel from A to B with more than the technical minimum running time to enable an acceptable level of operational reliability to be achieved. This excess time (known as performance allowance) is generally between three to ten percent of the sectional run time and is applied to mitigate the impact of operational variations associated with weather conditions, driver behaviours and/or train loads (UIC, 2000; Pachl, 2004, p. 178; Weeda and Wiggenraad, 2006).
Buffer time	Buffer time is present in a timetable when the spacing between trains is greater than the minimum headway. The amount of buffer time in a schedule relates to the level of utilisation of the system, with lower buffer times facilitating higher levels of utilisation. It is good practice to include both performance allowance and buffer time when designing a timetable. Performance allowance allows recovery from small delays, while buffer time reduces the transfer of delay to subsequent trains (Pachl, 2004, p. 180).
Pathing time	Pathing time is generally applied in timetables on the approach to junctions to synchronise the timetabled presentation of branchline services at junction merges and provides for reliable operation of the railway as a network.
Stopping pattern	Capacity is maximised when all trains in the timetable are scheduled with the same stopping pattern (running time). Vromans et al (2004) have developed a metric to compare the level of heterogeneity present in railway timetables. Corresponding simulations demonstrate that homogeneous timetables represent a significantly more reliable operation than heterogeneous timetables (containing a greater mix of stopping patterns).
Passenger loading	The recent surge of patronage in Melbourne initially led to higher levels of crowding and extended dwell times. Dwell times are railway and station specific as they depend on the number of boardings, alightings and train loads. For accurate analysis it is best to consider statistical delay profiles based on data obtained from the actual network being investigated (Yuan et al 2006; Yuan 2008).
Loading diversity	Loading diversity occurs when passengers do not load evenly onto trains. While crowding forces people to even out loads to some degree (rather than endure an uncomfortable journey or risk being unable to get on the train) entirely consistent loads are rarely seen in practice (TCRP, 2003, pg. 5-5). A high level of diversity reduces capacity because it results in (a) unused passenger capacity on some carriages and (b) extended dwell times due to congestion on the most heavily loaded carriage.

5. Rail capacity assessment

The theoretical capacity of a railway is a clearly defined quantity that represents the maximum number of trains that can be scheduled in a perfect world scenario. However, because railway delays are (somewhat) inevitable, reliable operations can not be maintained on a regular basis when operating at the limit of theoretical capacity.

Knowledge of the level of practical capacity is vital to making effective and efficient use of resources across a rail network, practical capacity is more difficult to define because it is influenced by a range of factors including the service plan and the levels of timetable robustness and primary delay.

This section undertakes a review of research and methods that exist in the literature in regards to the analysis of rail capacity.

5.1 Analytic/desktop models

The analytical method assesses line capacity based on the minimum signalling headway for the given section. Figure 2 indicates how headway can be used to calculate the theoretical capacity of trains across a one hour period. Practical capacity can then be estimated by introducing a factor to enable the system to be robust to the typical level of delay it is likely to experience on a day to day basis.

Figure 2: Analytic calculations for theoretical and practical capacity.

Theoretical Capacity (tph) = 3600/Headway (seconds)

Where Headway is the *"minimum time interval between successive trains running at line speed on clear signal aspects"* (Wayth, 2008).

Practical Capacity (tph) = Capacity threshold X 3600/Headway (seconds)

Where the *"capacity threshold"* is typically in the order of 75% (SRA, 2003).

The analytic method is simple to use, though it has limited ability to (a) easily determine capacity for mixed stopping patterns or (b) predict the level of system performance. The UIC 406 compression method (UIC, 2004) provides a method to develop a *"theoretical headway"* equivalent applicable in mixed traffic situations and assesses the ability of a timetable to recover from delay by quantifying the amount of buffer time present in the timetable.

The compression method takes the scheduled timetable and closes in pathways (without modifying running times, dwell times or train order) to the minimum achievable spacing based on a precise calculation of signal blocking times. The process effectively removes buffer time from the schedule to calculate the level of capacity that is utilised by the given operation. The UIC code recommends that during peak periods[1] the level of capacity utilisation should not exceed 75-85%.

5.2 Parameter simulation

Dicembre and Ricci (2011) provide an example of a parameter study into the impacts that signal block length has on railway capacity. Parameter studies aim to understand how the system will operate prior to a detailed timetable having been developed. This approach can reduce the level of risk associated with developing a new timetable and also provide an ability to confirm ideas and inform expectations at an early stage of project development. Parameter studies can help to inform decisions that need to be made quickly.

A key to parameter studies is to use a simple approach for the modelling, so that the performance impact of making one change at a time can be investigated. The changes can be to either the service plan, input parameters, infrastructure options or model assumptions. Parameter studies are similar in intent to the response surface methodology outlined by Lindfeldt (2010), though because they are simpler, parameter studies are likely to be more easily applied to practical railway problems.

[1] Lower thresholds (60-70%) apply in mixed traffic situations and across extended periods of time.

5.3 Detailed simulation

Detailed simulation outputs can identify and inform a wide range of operational issues including (a) capacity and punctuality performance estimates, (b) infrastructure bottlenecks and requirements and (c) timetable bottlenecks and requirements. Nugent (2007) provides a summary of the simulation approach, highlighting the importance of dynamic simulation because of the strong link between on-time running and capacity.

The quality and consistency of the timetable development phase dictates the accuracy of the investigation. The timetable construction phase can be time consuming and tricky to optimise, particularly when investigations seek to understand the impact of small changes in operational conditions. A quality timetable is required to ensure that optimal use is made of infrastructure. Consistency in operational parameters and the timetable approach are required to ensure that a fair comparison can be made between different options.

A disadvantage of the detailed simulation approach is that it can be hard to test the sensitivity of a detailed timetable to different assumptions (e.g. including an additional train in the peak hour). Essentially the options available to do this are (a) to rewrite the entire timetable or (b) make localised changes to the timetable. Care needs to be taken with both these approaches, rewriting an entire timetable is a lengthy process, while localised changes are at risk of scheduling a sub-optimal pathway, with a poorer level of performance.

Simulation of concept service plans can be used to validate that a given service plan can operate at an acceptable level of performance. A disadvantage is the timeframe required to develop (a) dynamic models capable of precisely predicting performance and (b) detailed service plans capable of closely replicating reality (particularly if multiple options need to be assessed).

6. A simulation approach for capacity assessment

Of the approaches outlined in the previous section, parameter simulation is best suited to assess capacity because it easily allows for localised timetable changes to be investigated on a consistent basis[2].

6.1 Concepts for defining capacity

This paper compares and contrasts the suitability of a number of different metrics to assess the upper capacity limit (table 3).

Table 3: Different concepts for defining capacity

Punctuality threshold	There is a commonly held view that the capacity threshold is reached when the addition of further trains would lead to an unacceptable level of operational punctuality performance. Here the benefit of providing additional transport capacity is outweighed by the cost associated with a reduced level of punctuality (SRA, 2003, pg. 34).
Delay saturation	An alternative metric is described by practical capacity being the threshold of the growth of knock-on delay when a timetable becomes saturated with trains (Dingler et al 2009). Delay is an attractive measure because it is easy to obtain, analyse and understand.
Localised timetable stability	Goverde (2008) defines the local stability of a timetable to be achieved when output delays do not exceed the input delays to the track section. The radial nature of the Melbourne network limits the practicality of this measure because passenger delays and operational complexities are generally concentrated in the central area, nevertheless there is potential to consider the use a threshold level of delay growth for radial networks.

[2] Detailed simulation would necessitate a longer timeframe to assess capacity due to the workload associated with developing multiple detailed timetables.

6.2 Methodology for capacity assessment

The parameter assessments undertaken in this paper follow the basic approach of Radtke and Bendfeldt (2001). While simulation packages are unable to define a single numerical value that represents the upper limit of practical capacity, they can demonstrate the capability of a railway to recover from delays. By iteratively increasing the number of scheduled trains, the upper capacity limit can be determined as the point where delay can no longer be absorbed in the system.

The assessment uses the infrastructure and delay sets from a calibrated City Loop and Inner City (CLIC) model (Gray and Daly, 2007). Input delay data was based on a comprehensive dwell and station departure time survey undertaken in October 2006 (section 6.3 outlines this as a period when the system was at capacity). The CLIC model takes account of the departure delay from the cordon station (North Melbourne for the *"northern group"*), dwell delay, terminal platform occupancy and potential delays from other groups.

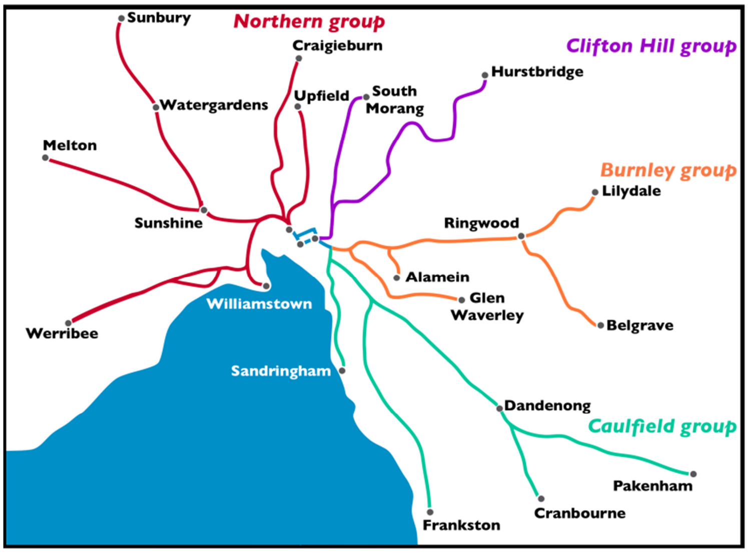
The parameter technique schedules 45 equally spaced trains in each simulation set, to model train frequencies ranging from one to 45 tph[3]. Dynamic simulations used 400 timetables to ensure convergence. Train runtimes were based on the relevant working timetable (i.e. the relevant amounts of pathing and performance allowance were considered).

6.3 Determining a capacity benchmark

The final component required to assess capacity is a set of operational parameters suitable for use as a capacity benchmark. The benchmark needs to correspond to a period where the railway was operating at capacity.

The Melbourne electric rail network essentially comprises a hub and spoke network, with the spokes divided into four geographical groups (figure 3). In the central area, each group is allocated one of the four *"city loop"* lines, with trains that are unable to be accommodated in the city loop running direct to the central terminus (Flinders Street station).

Figure 3: The Melbourne electric train network



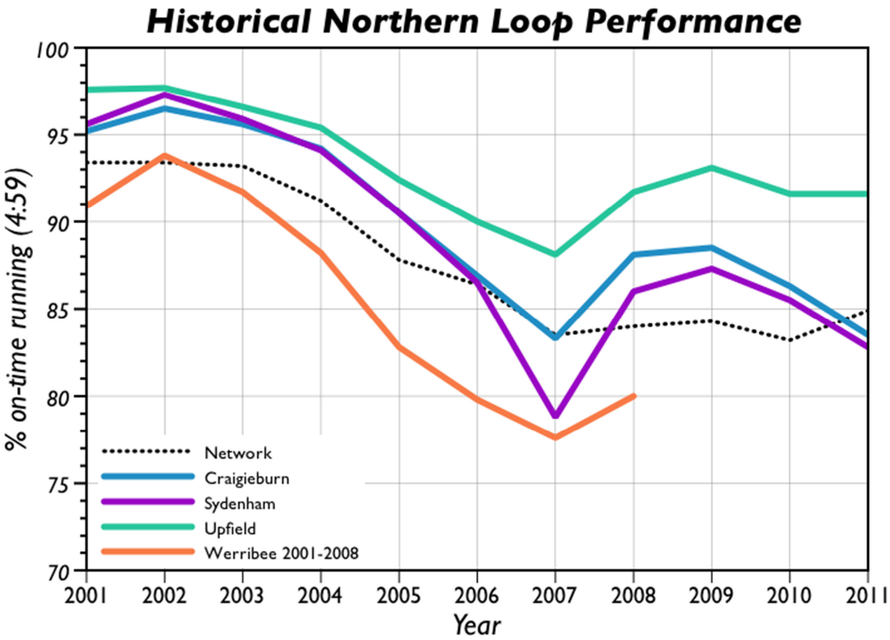
The highest train throughput operating on each group occurs in the *"city loop"* sections, where trains from multiple lines merge into one set of tracks. The *"northern group"* of lines, serves three of the four major growth corridors in Melbourne and as a result capacity is a particular constraint on this group.

[3] Note however, that for clarity figures 5 to 8 only display the results up to the 30 tph simulations.

Patronage on the Melbourne metropolitan rail network increased steadily following the construction of the *"city loop"* in the early 1980s. A further surge in patronage[4] since 2004 initially led to overcrowding and a corresponding decline in punctuality across the network; these are signs of a system struggling to cope.

Figure 4 demonstrates how the performance of the *"northern loop"* lines declined from 2002 to 2007. Before operational changes made in association with a new timetable in September 2007 (for the extension of electrification to Craigieburn) led to improved performance.

Figure 4: Historical on-time performance of lines operating in the *"northern loop"*.



Further improvements followed the diversion of peak period Werribee line services from the *"northern loop"* in November 2008. Clearly, the late 2006 *"northern loop"* operation was capacity constrained and is suitable for use as a capacity benchmark.

7. Results

A series of metrics were then evaluated to assess their ability to quantify the practical capacity limit:

- On-time running threshold
- Growth of knock-on delay
- Local timetable stability

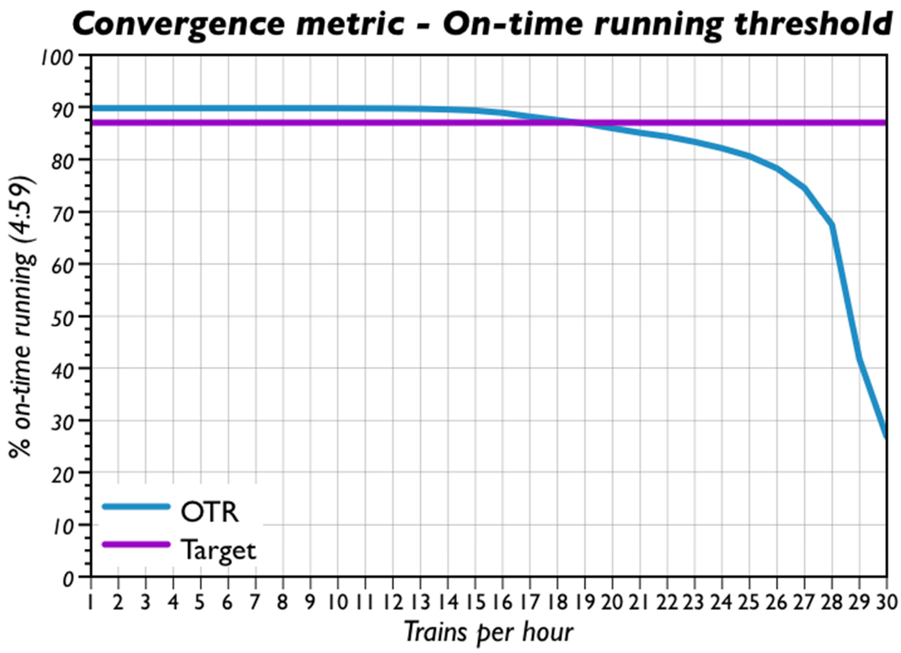
On-time running is a common measure for railway performance in Australia, the growth of knock-on delay considers the transfer of delay from a late train onto subsequent trains, while local stability compares the difference in delay entering and exiting a track section.

7.1 On-time performance threshold

The first metric considered is the five minute on-time running measure (figure 5). This is a simple measure and is also the current definition for public compensation in Victoria.

[4] The patronage surge has been attributed to a combination of population growth and a mode shift from road to rail. The mode shift has been linked to a range of environmental and economic factors including CBD job growth, road congestion, petrol and car parking costs (Gaymer, 2010).

Figure 5: On-time running performance to assess capacity of the *"northern loop"* with assessment parameters equivalent to the operation in October 2006.



Inherent delay represents the impact of primary delays on system performance. At low tph, the trains are so far apart, that there are no knock-on interactions between trains to reduce the level of on-time running and the level of delay reflects the levels of presentation delay, dwell delay and running time margin in the service plan.

The inherent level of performance (90% at five minutes on-time running) is maintained up until 15 tph (4 minutes headway).

Eventually as service level increases, simulation performance falls below the on-time running threshold at 19 tph (3 minutes 10 seconds). This finding is consistent with the actual operation in 2006 which scheduled a maximum of 19 tph in the peak hour.

It should be further noted that performance begins to decline dramatically by 25 tph (2 minutes 24 seconds headway) and significantly again by 28 tph (2 minutes 9 seconds headway). This is not surprising given the minimum signalling headway (2 minutes 16 seconds), the method indicates extremely poor performance results when trains are scheduled at close to the theoretical headway.

This study confirmed the ability to use an on-time running threshold to assess capacity and demonstrating a match to actual operations.

7.2 Growth of knock-on delay

An alternative concept, sees capacity defined as the point at which the knock-on delay between trains rises to an unacceptable level. There are however various options through which knock-on delay can be measured, two of which are considered in this paper:

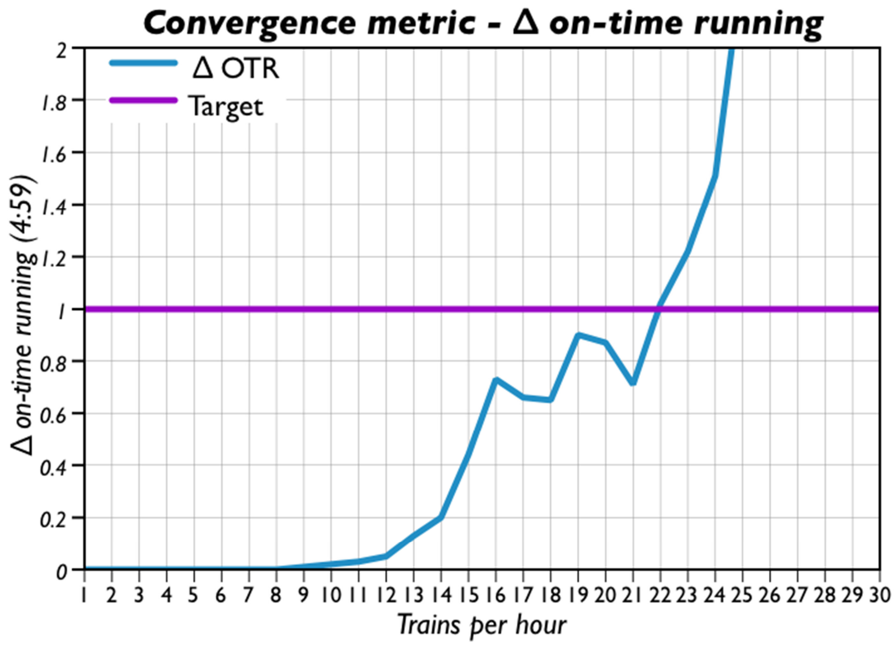
- Gradient of on-time running curve (ΔOTR)
- Average arrival delay

7.2.1 Gradient of on-time running

The second metric, the gradient of the on-time running curve (i.e. the gradient of figure 5) is a measure of the amount of knock-on delay. In effect, this measure represents the difference between consecutive on-time running curves at a defined level of lateness (figure 6).

The assumption being that knock-on delay is unacceptable when the level of performance (at the five minute on-time running measure) decreases by more that one percent as each train is added (figure 6).

Figure 6: Δ On-time running performance to assess capacity of the *"northern loop"* with assessment parameters equivalent to the operation in October 2006.



Compared to the on-time running threshold, this measure indicates a higher level of sectional capacity is achievable from the system (22 tph). Though as the curve is uneven between 15 and 24 tph there are issues in using this measure to provide a definitive assessment of capacity.

Investigations indicate that the simulations have converged so the unevenness is a combination of the software only being able to schedule trains to the second (particularly given the small margins between consecutive simulation sets) and difficulties associated with obtaining the gradient of the on-time running curve from a limited number of data points. Accordingly, the ΔOTR measure only represents an approximation to the gradient of the on-time running curve. Furthermore, the measure is difficult to obtain from the software which reduces the practicality of using the measure for the assessment of rail track capacity.

7.2.2 Average arrival delay

The third metric considered is the average level of arrival delay at Flinders Street (figure 7). The first thing to note is that compared to the five minute on-time running metric, deviation from the inherent delay level occurs much sooner i.e. 8 tph for the arrival delay and 15 tph for the on-time running measure.

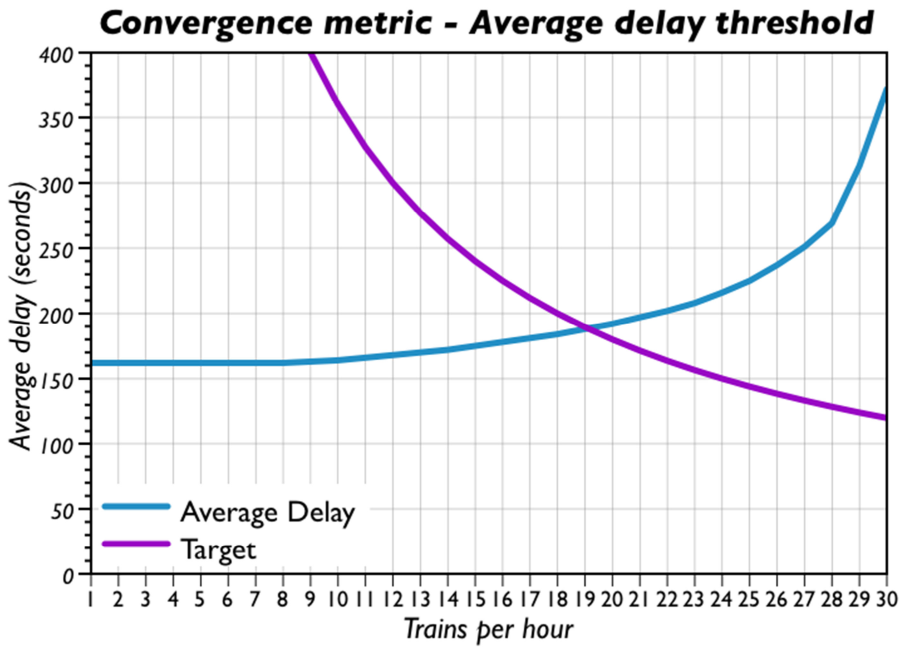
This difference reflects the lack of sensitivity in the five minute on-time running measure and also that on-time running is a rounded measure (the measure does not change until a train is five minutes late), whereas the average arrival delay is more sensitive to incremental changes in lateness.

The second point of note is that the shape of the curve is opposite to that of the on-time running curve, the average arrival delay increases along with the train frequency. With this metric, capacity would be said to be reached when the given threshold is exceeded.

The difficulty in using the average delay is that it is not obvious how to set an objective threshold. To achieve an acceptable comparison, the delay threshold needs to be set above the level of inherent delay and below the accepted on-time running measure for the network. We note that a three minute average delay threshold would estimate a capacity of 16 tph, while a three minute 30 second threshold would estimate capacity at 23 tph.

Such an estimate is highly sensitive, particularly given the delay metric is also responsive to the level of presentation delay, which is generally beyond the control of the section being assessed.

Figure 7: Average arrival delay performance used to assess capacity of the *"northern loop"* with assessment parameters equivalent to the operation in October 2006.



The recommended measure is that the level of delay does not exceed the average scheduled spacing between trains (as at this point trains are usually running in the next scheduled pathway). Similar, to the on-time running measure (figure 5), the average arrival delay indicates a capacity of 19 tph for the 2006 operation.

The on-time running and average arrival delay metrics both incorporate the level of presentation delay and sectional dwell delay in their computations. They represent sectional capacities, though they can are also impacted by the level of delay incurred outside the section. These measures would indicate a lower level of capacity if services entered the section with a higher level of lateness.

7.3 Localised timetable stability

Goverde (2008, pg.119) notes that a section of track can be considered locally stable if the *"sum of output delays is smaller than the sum of input delays"*. The correlation for sectional capacity is that the system is within its capacity if the output delays from the section are smaller than the input delays to the section.

By itself however, this is not an appropriate definition for the central areas of the Melbourne network where (a) the nature of boardings and alightings means that dwell delay is somewhat inevitable and (b) the five minute threshold for on-time running implies that it is acceptable to have an increase in the level of delay across some track sections.

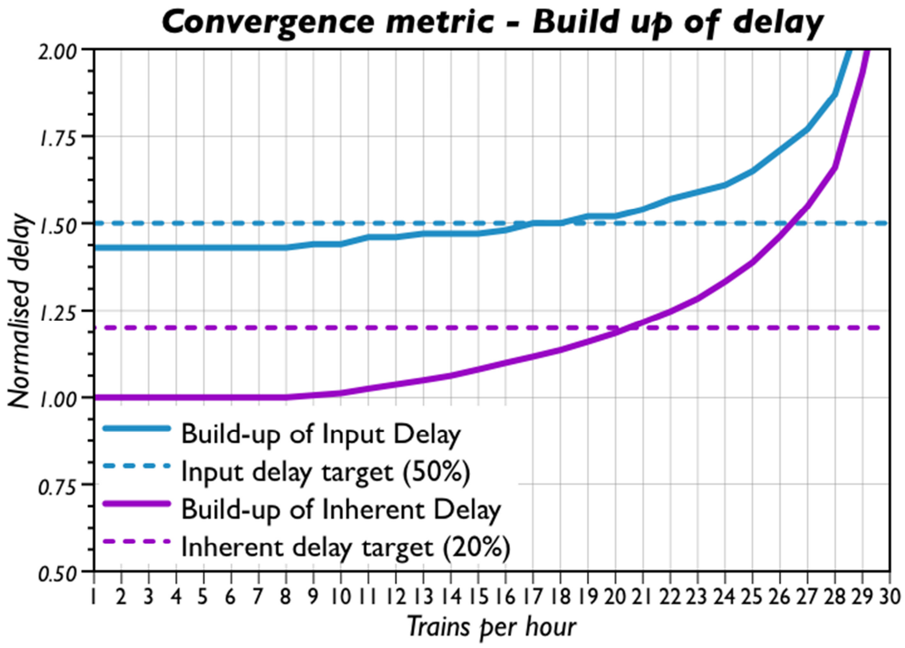
The principle can be applied by accepting a small increase in delay (i.e. allowing output delay to be higher than the level of input delay) across the section. This provides a more objective means of assessing sectional capacity, applicable to a centrally dominated system such as

Melbourne, by nominating a proportion by which the level of delay is able to increase over the section. Sectional stability is investigated in this section by considering the build up of both input and inherent delay.

7.3.1 Build-up of input delay

The normalised level of delay for both the input and inherent delay measures are demonstrated in figure 8, with a 50% threshold being deemed appropriate for the input delay. Using the input delay measure, capacity of the system is estimated to be reached at 18 tph.

Figure 7: Average arrival delay performance used to assess capacity of the *"northern loop"* with assessment parameters equivalent to the operation in October 2006.



The drawback of normalising delay to the input delay is that the result remains sensitive to the level of lateness as trains present to the cordon, it measures the relative proportion of delay inside and outside the section of interest not the overall delay to the passenger.

7.3.2 Build-up of inherent delay

Subsequently, inherent delay is a more attractive metric of sectional capacity because it measures the actual build-up of knock-on delay in the system. The inherent delay metric computes the increase in delay compared to the level of delay in the system assuming that trains do not interact. The build-up of inherent delay is calculated by normalising the arrival delay at the terminal station for each simulation, by the arrival delay in the one train per hour simulation.

Assuming that a 20% threshold is deemed an appropriate threshold for the increase in inherent delay, capacity of the system is estimated to be reached at 20 tph.

Similarly, to the other methods considered in this section, the growth of inherent delay, appears capable of assessing capacity, though the need to define an arbitrary threshold to quantify capacity, limits the ability to use it to investigate an absolute measure of capacity.

Furthermore, we note that significant divergence occurs above 28 tph in all three of the delay measures considered in this section. This corresponds to the theoretical capacity of the signalling system being exceeded.

8. Discussion

This paper has developed and compared a number of different metrics to assess the capacity of a railway using a parameter simulation approach. An on-time running threshold has been demonstrated to provide a suitable metric for capacity. An advantage of using this in the Victorian context is that it is aligned with existing performance measures so is likely to be easily understood and accepted across the industry.

On-time running is a good indicator of capacity on a network like Melbourne with a hub and spoke configuration, because it considers the impact of both (a) the level of delay in the section being investigated and (b) the presentation performance entering the section. The measure is reflective of the public experience of train performance.

Further investigations considered measuring capacity by tracking the growth of knock-on delay (i.e. ΔOTR, growth of arrival delay) and local timetable stability (i.e. build up of input and inherent delay).

While ΔOTR is in theory an attractive measure for tracking the growth of knock-on delay, it needs to be processed separately (outside of the software) and resulted in an irregular curve (effectively the metric attempts to determine the gradient of a curve from a limited number of discreet points). Further development would be required before it could be recommended for wider use.

The growth of knock-on delay can be measured by tracking station arrival delays. This measure incorporates both the level of delay presenting to and dwell delay in the section and when compared to the scheduled train spacing provides a clear and transparent indication of capacity.

Dingler et al (2009) favour the use of delay to measure capacity, though note the difficulty in defining a generic delay threshold. This is primarily because each *"railroad's customers or commodity groups all may have different acceptable levels of service and corresponding tolerance of delay."*

Because different systems will have different levels of input and inherent delay, it is impossible to define a generic threshold for capacity. There is potential to normalise these measures and define an acceptable percentage increase in delay.

A weakness in the delay methods is the general lack of objectivity in determining an appropriate threshold. If an arbitrary threshold is defined, the metric is still likely to be suitable for assessing the relative difference between options even if it is unable to define the absolute level of capacity.

The risk in using the normalised measures for absolute capacity, is that they could significantly under (or over) estimate capacity because they only consider behaviour inside the section under investigation.

The best metric, will depend somewhat on the ability to obtain accurate data to enable performance to be monitored. In this regard, a metric that is aligned with how a given railway actually measures performance will generally be suitable for performance assessment.

This paper focuses on the quantification of sectional capacity. The level of capacity achievable in a given track section may be reduced when considering the section as part of a complex, interacting network. Factors such as line lengths, single line sections, at-grade conflicts and other timetable constraints reduce the level of capacity that can be achieved on a network basis. Future work should explore the sectional and network capacity relationship.

9. Conclusions

Population growth is changing the form of new world cities; increasing levels of patronage are driving a need to better understand rail capacity, with a range of tools available to support operational analysis. While practical operational knowledge needs to remain a cornerstone of capacity analysis, there are opportunities to use modern simulation software to better inform decision making by developing a broader knowledge base to quantify project impacts.

A method is presented to evaluate capacity by varying the number of trains per hour to assess the impact to operational performance. As capacity is largely a subjective measure, it is important to develop an objective basis by which capacity improvement projects can be analysed, compared and justified.

The ability to measure railway capacity has been considered for three definitions of capacity:

- On-time running threshold
- Growth of knock-on delay
- Localised timetable stability

On-time running was shown to be a good measure to determine capacity as it captures information about presentation performance and dwell delay in the section. Although they are associated with a greater level of subjectivity in the selection of the threshold, measures based on the average level of delay and timetable stability can also be informative in certain circumstances.

Acknowledgements

This work is based on research conducted for a Masters thesis in Rail Operations Management at CQ University. The author would like to thank Mark Daly for reviewing a draft of this paper.

The views and opinions expressed in this paper are the personal views of the author and do not necessarily reflect the views of the author's employer or the Victorian government.

References

Abril, M., F. Barber, L. Ingolotti, M. A. Salido, P. Tormos, and A. Lova (2008). An assessment of railway capacity. Transportation Research Part E 44, 774–806.

Argyrous, G. (2009). Evidence or idealogically based decision making. UNSW Press.

Connor, P. (2010, October). Making the most of existing capacity: Good operating is the key. Modern Railways, 61–65.

Currie, G. (2010). Quick and effective solution to rail overcrowding. Journal of the Transportation Research Board 2146, 35–42.

Dicembre, A. and S. Ricci (2011, December). Railway traffic on high density urban corridors: Capacity, signalling and timetable. Journal of Rail Transport Planning & Management 1(2), 59–68.

Dingler, M. H., Y.-C. Lai, and C. Barkan (2009). Impact of train type heterogeneity on single-track railway capacity. Journal of the Transportation Research Board (2117), 41–49.

Eddington, R. (2008). Investing in transport: East West link needs assessment. Technical report, A study by Sir Rod Eddington.

Garvin, D. A. and M. A. Roberto (2001, September).What you don't know about making decisions. Harvard Business Review, 108–116.

Gaymer, S. (2010). Quantifying the impact of attitudes on shift towards sustainable transport modes. ATRF.

Goverde, R. (2008). Timetable stability analysis, Chapter 7, pp. 118–134. Railway Timetable & Traffic: Analysis - Modelling - Simulation. Eurail Press.

Gray, J. and M. Daly (2007, November). CLIC 2006 model calibration - Methodology.Technical report, Public Transport Division, Department of Infrastructure, Victoria.

Hansen, I. A. and J. Pachl (Eds.) (2008). Railway Timetable & Traffic: Analysis - Modelling - Simulation. Eurail Press.

Jessop, A. (1990). Decision and forecast models: with transport applications. Chichester, England: Ellis Horwood limited.

Lindfeldt, O. (2010). Railway operation analysis - Evaluation of quality, infrastructure and timetable on single and double-track lines with analytical models and simulation. Ph. D. thesis, Division of Traffic and Logistics, Department of Transport and Economics, Stockholm.

Nugent, M. (2007). Using simulation techniques to quantify the capacity benefits of signalling upgrades. In IRSE technical convention Melbourne.

Pachl, J. (2004). Railway operation and control. VTD Rail Publishing.

Radtke, A. and J.-P. Bendfeldt (2001). Handling of railway operation problems with RailSys. In Proceedings of the 5th World Congress on Rail Research, Cologne, Germany.

SRA (2003). Capacity utilisation policy: Network utilisation strategy. Technical report, Strategic Rail Authority.

TCQSM (2003). Transit Capacity and Quality of Service Manual (2nd ed.).Transit Cooperative Research Program, Transportation Research Board.

UIC (2000). UIC 451-1: Timetable recovery margins to guarantee timekeeping. Technical report, International Union of Railways.

UIC (2004). UIC code 406: Capacity. Technical report, International Union of Railways.

VAGO (2012). Public transport performance. Technical report, Victorian Auditor-General's Office.

Vromans, M., R. Dekker, and L. Kroon (2004). Simulation and railway timetabling norms. In Computers in Railways IX, pp. 685–694.

Weeda, V. A. and P. B. L.Wiggenraad (2006). Joint design standard for running times, dwell times and headway times. In Computers in Railways X, pp. 611–620.

Wayth, R. (2008, August). Signalling design principles on the Victorian rail network. Technical report, Sinclair Knight Mertz.

Yuan, J. (2008). Statistical Analysis of train delays, Chapter 10, pp. 170–181. Railway Timetable & Traffic: Analysis - Modelling - Simulation. Eurail Press.

Yuan, J., R. M. P. Goverde, and I. A. Hansen (2006). Evaluating stochastic train process time distribution models on the basis of empirical detection data. In Computers in Railways X, pp. 631–640.