Estimating the robustness of train plans for single-line corridors with crossing loops

A. R. Albrecht[1] J. de Jong[2] P. G. Howlett[3]
P. J. Pudney[4]

(Received 31 January 2010; revised 6 October 2010)

Abstract

The movement of trains is often planned in advance to determine where and when trains will cross each other, and to determine arrival times for trains at their final destinations. However, random variations to departure times and travel times mean that crosses do not always occur at the planned locations and times, and excessive delays lead to trains arriving late at their destinations. We define the *robustness* of a train plan to be the expected reliability of that plan when the trains are subjected to typical departure variations and variations in travel speeds, and when standard procedures are used to recover from these delays. Robustness can be estimated by simulating the operation of the network for many different delays, but can we do a more direct calculation? We calculate the distribution of arrival times for a train with random departure and travel times that might be delayed by a second train.

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/2528 gives this article, © Austral. Mathematical Soc. 2010. Published October 26, 2010. ISSN 1446-8735.

Contents

1 Introduction C769

2 Modelling the problem C771

3 Modelling delays C772

4 Recovery procedure C775

5 Monte Carlo estimation of arrival time distributions C776

6 A more efficient calculation? C777

7 Summary C780

References C782

1 Introduction

Australian interstate rail corridors are mainly single-track, with occasional 'crossing loops' that allow one train to pull off the main line so another train can pass. The movements of trains on the rail corridors are planned in advance to determine where and when trains will pass, and to determine reasonable arrival times for the trains at their final destinations. In practice, however, random variations to train departure times and train travel times mean that crosses do not always occur at the planned locations and times. Excessive delays can lead to trains arriving late at their destinations.

One of the key requirements of rail customers is reliability—trains should arrive at their destinations on time. The reliability of the rail system is typically expressed as the proportion of trains that arrive no later than their planned arrival time.

Reliability depends on planned departure times, planned section running times, planned arrival times, actual departure times, actual section running times, and how train controllers handle variations to planned train paths.

We define the *robustness* of a train plan to be the expected reliability of that plan when the trains are subjected to typical departure variations and variations in travel speeds, and when standard procedures are used to recover from these delays. We are interested in developing a fast method for estimating the robustness of train plans. This method could be used to assess manually constructed train plans or as part of the objective function in automated train scheduling tools.

To estimate robustness, we need to know planned train departure times, planned section running times, planned arrival times, the expected distribution of departure times, the expected distribution of section running times, and how train controllers will handle variations to train paths.

There is a moderate body of literature relating to the calculation of train delays. Murali et al. [2] overview analytical and simulation models, of varying complexity, to estimate delay in rail networks.

The ARRIVAL project[1], funded by the European Commission, is developing robust optimisation methods to improve the reliability of rail transport. Cicerone et al. [1] gave a good overview. The train timetabling problem is generally set in the context of European passenger rail networks, and assumes that interactions between trains occurs at specified locations (typically stations).

We, like many, use simulation to predict the movement of trains along a single line corridor. In our case we simulate typical Australian long haul corridors where there are significant delays in train departure times and where crossing locations are not important. We use these simulations to predict the distribution of arrival times for trains. We then begin to formulate a simple method, suitable for Australian long haul rail networks, for calculating the distribution of arrival times.

[1] http://arrival.cti.gr

2 Modelling the problem

We restrict our attention to a single-line corridor with n intermediate crossing loops. Distance along the corridor is $x \in [0, X]$. Train journeys start and finish at $x_0 = 0$ and at $x_{n+1} = X$. Trains can pass each other only at crossing loop locations $x_0 < x_1 < x_2 < \cdots < x_n < x_{n+1}$. Loops are typically placed so that running times are approximately equal between each pair of adjacent loops. Table 1 shows that the mean section running time on the Sydney to Brisbane corridor is 12–13 minutes, with a standard deviation of two minutes.

The set of trains is $T = T_+ \cup T_-$ where T_+ is the set of trains travelling from $x = 0$ to $x = X$, and T_- is the set of trains travelling from $x = X$ to $x = 0$. The information specified for each train i is

- d_i, the planned departure time of train i from the origin;
- t_{ij}, the planned travel time for train i between x_j and x_{j+1};
- a_i, the planned arrival time of train i at the destination;
- Δ_i, the direction indicator: $\Delta_i = 1$ if train i travels from 0 to X, and $\Delta_i = -1$ if train i travels from X to 0;
- p_i, the priority of train i, a higher value indicates a higher priority.

At any instant there can be at most one train on each interval (x_j, x_{j+1}) and at most two trains at each crossing location $x_1, x_2, \ldots, x_n$. The corridor has unlimited capacity at x_0 and x_{n+1}.

We do not allow a fast train to overtake a slower train travelling in the same direction.

Figure 1 is an example train graph that shows the movement of trains over a two day period. The horizontal axis indicates time. The vertical axis indicates location. The horizontal lines correspond to crossing loops. Notice that trains cross only at crossing loops, and that one of the trains involved in each cross is delayed.

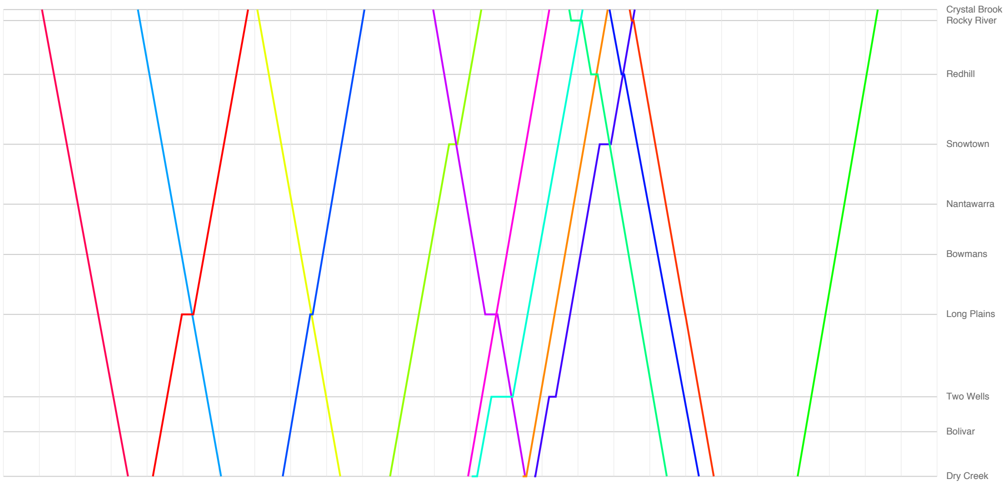

Figure 1: Example train graph.

For a given set of random departure and travel time perturbations, and a given recovery procedure for deciding where trains should cross, we wish to calculate the expected proportion of trains that will arrive no later than the planned arrival times.

3 Modelling delays

To calculate robustness, we need to model the delays that occur to each train on entering the network, and on each section of its journey. The train planning process assumes a nominal running time for each train on each journey section. However, the running times of a particular train could vary significantly from the assumed running times since there are significant variations in train length, train mass and the types and number of locomotives used to haul trains. Nominal section running times should be selected so that most trains require

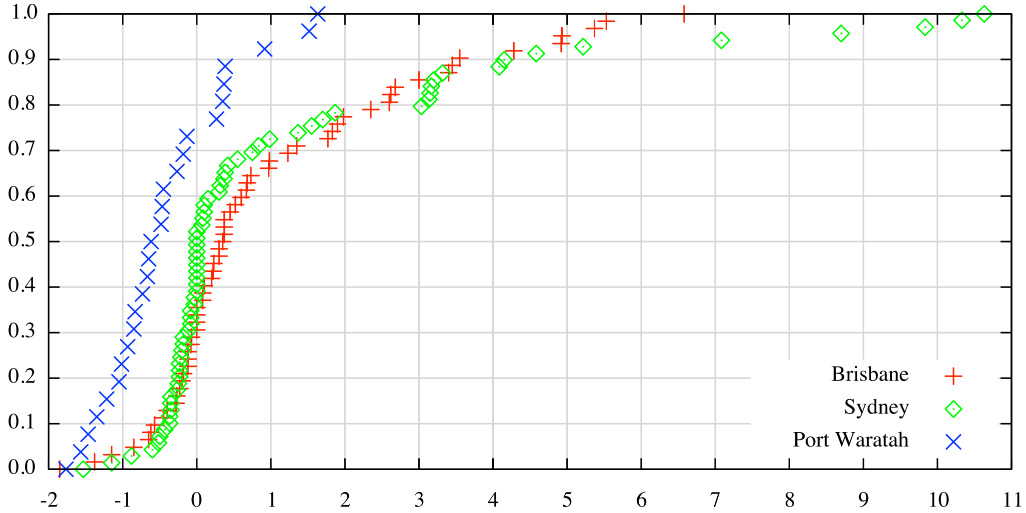

Figure 2: Cumulative distributions of departure delays for three different freight corridors.

less time, but some trains may take longer than the nominal time. Temporary speed restrictions introduce further delays.

Figure 2 shows the cumulative distribution of departure delays for trains departing Port Waratah (blue), Sydney (green) and Brisbane (red) on the main freight corridor between Sydney and Brisbane.

Figure 3 shows measured running times relative to the planned running times for one class of train travelling from Sydney to Brisbane. The vertical axis represents earliness; the horizontal axis indicates distance (increasing across the graph).

Table 1 shows some statistics for two classes of trains travelling in each of the two directions between Sydney and Brisbane. We use data such as this to construct distributions of actual departure times and running times on each section.

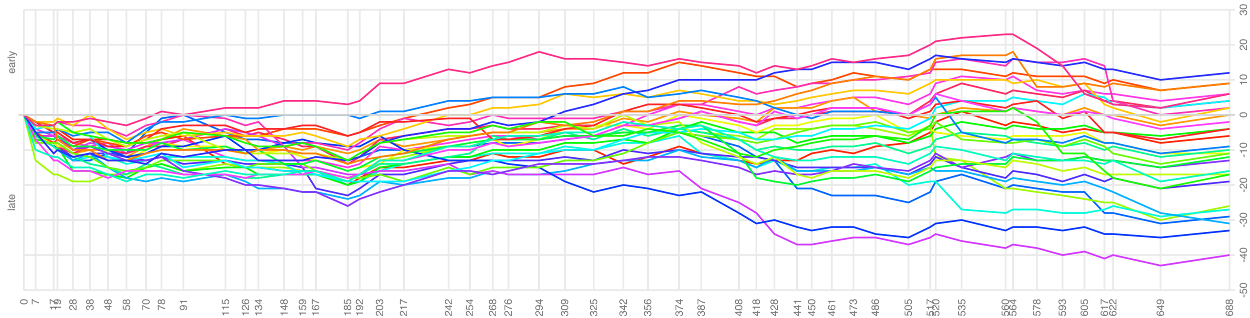

Figure 3: Earliness of trains relative to the planned section running times. The horizontal axis is distance, 0–688 km; the vertical axis shows time, from 50 minutes late to 30 minutes early.

Table 1: Statistics of section running times, all in minutes.

	SB1	BS1	SB2	BS2
mean section running time	12.4	12.4	13.0	12.8
standard deviation	2.0	2.0	2.1	2.5
mean midpoint cumulative time	300.2	350.1	327.6	387.1
standard deviation	7.5	13.3	11.4	13.5
mean finish cumulative time	642.4	645.8	703.5	702.1
standard deviation	14.0	16.8	18.1	22.2

4 Recovery procedure

When trains are not running according to the plan, the train controller must re-calculate where each cross should occur. We assume that crossing decisions are based on 'natural crossing locations'; that is, locations where the opposing trains would meet if they did not have to pass at crossing loops. The crossing location for a pair of trains is calculated as follows.

- If the two trains have the same priority, the cross should occur at the loop closest to the natural crossing location. The portion of train graph on the left of Figure 4 shows a cross at the upper crossing loop, which is closest to the natural crossing location. This scheme minimises the local delay.

- If the two trains have different priority, the cross should occur at the crossing loop immediately prior to the natural crossing location on the path of the low priority train. The portion of train graph on the right of Figure 4 shows a cross where the down (red) train has priority.

This is similar to the recovery scheme used by train controllers at the Australian Rail Track Corporation.

Given a random departure time for each train, random section running times for each train, the scheduled arrival time for each train, and the locations of

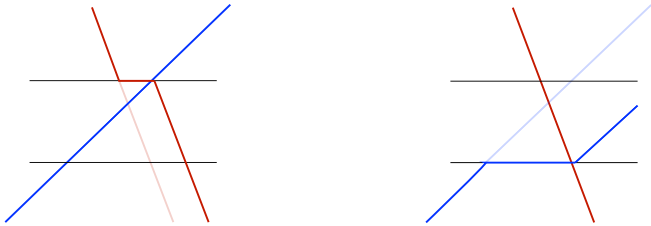

Figure 4: Crosses for trains with equal priority (left) and where the down (red) train has priority (right).

crossing loops, we use the recovery rules to simulate the movement of trains across the corridor, and hence the lateness of each train.

5 Monte Carlo estimation of arrival time distributions

The robustness of the train plan can be estimated by simulating the operation of the network for many different perturbation instances, and measuring the proportion of late trains.

Figure 5 shows the distribution of arrival times for the trains depicted in Figure 1 when departure times are selected from a normal distribution with a standard deviation of 10 minutes, and section running times are based on an average running speed of 80 km/h with a standard deviation of 2 km/h. We ran 1000 simulations, then set the required arrival time for each train to the 95th percentile value of the observed arrival times so that each train service will have a robustness of 95%. The distributions of arrival times shown in Figure 5 have been shifted so that they show arrival time relative to the 95th percentile arrival time.

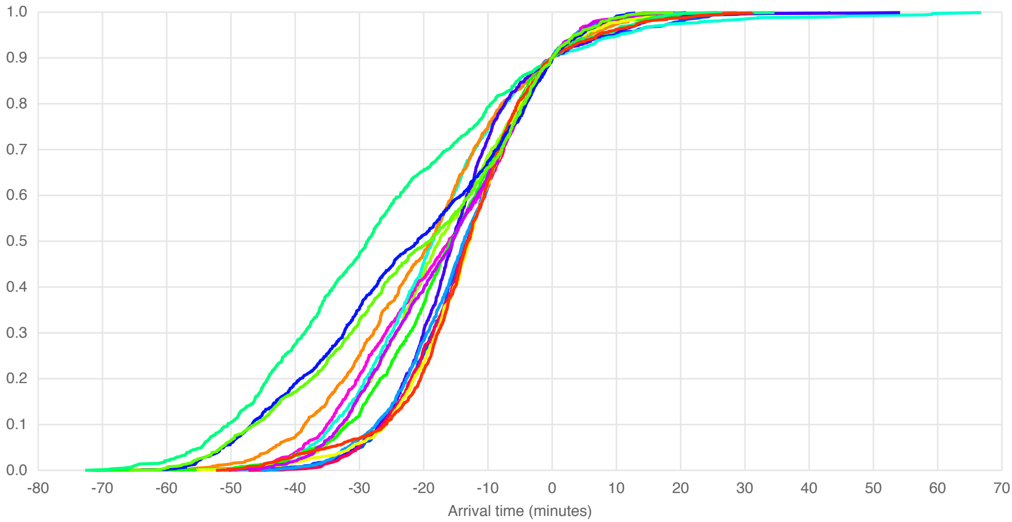

Figure 5: Cumulative distributions of arrival times from a Monte Carlo simulation.

6 A more efficient calculation?

Can we estimate robustness without simulating, many times, the detailed movement of trains on the network? Can we predict the distribution of arrival times for each train based on the number of crosses it will have and the expected delay at each cross?

First we show how to estimate the delay to each of the trains involved at a cross without knowing details such as the natural crossing location and the locations of crossing loops.

Trains with equal priority cross at the loop closest to the natural crossing location. Let d be the distance between the natural crossing location and the closest crossing loop. Suppose the mean distance between loops is δ. As the natural crossing location varies between one loop and the next, the distance d varies from 0 to $\delta/2$ and back to 0 again. The expected value for d is $\delta/4$.

Let the train speeds be v_1 and v_2. If train 1 is the first to arrive at the crossing loop, then the delay to train 1 is $d/v_1 + d/v_2$ and the delay to train 2 is zero. Alternatively, if train 2 is first to arrive, then the delay to train 2 is $d/v_1 + d/v_2$ and the delay to train 2 is zero. Since each train is equally likely to be the first to arrive at the crossing loop, the expected delay to each train is

$$d_1 = \frac{\delta}{8}\left(\frac{1}{v_1} + \frac{1}{v_2}\right).$$

The total delay is $2d_1$. Now consider the case where train 1 has priority over train 2. In this case, the crossing location is selected so that train 2 arrives at the cross first and waits for train 1, which has zero delay. The expected distance d to the crossing loop is $\delta/2$, and the delay to train 2 is

$$d_2 = \frac{\delta}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right).$$

The total delay in this case is d_2, since only train 2 is delayed.

The total delay when one train has priority is is double the total delay when the trains have the same priority.

We now look at how a single cross may affect the arrival time of a train. Consider the incoming (blue) train in Figure 6. The left diagram shows the incoming train arriving at its destination at time a before the outgoing (red) train departs at time d. If the incoming train does not cross the outgoing train, calculating the distribution of arrival times is straightforward. However, some perturbations applied to the incoming train and to the outgoing train will cause the two trains to cross, introducing further delay to the incoming train. The incoming train now arrives at time $a > u$, where u is the unimpeded arrival time. This is shown in the right diagram.

Let U be the unimpeded arrival time of the incoming train, D be the departure time of the outgoing train and A be the actual arrival time of the incoming train. These are random variables. For an instance (u, d) of (U, D), the

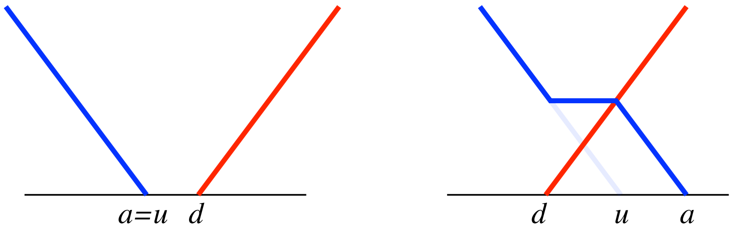

Figure 6: Unimpeded and actual arrival times.

actual arrival time

$$a = \begin{cases} u, & u \leqslant d, \\ u + \gamma, & d < u, \end{cases}$$

where γ is the delay to the incoming train.

The density of unimpeded arrival times U and departure times D are depicted in the left portion of Figure 7. The fuzzy blob indicates the density of the joint distribution. For instances (u, d) with $u > d$, the incoming train will be delayed.

There are two ways that the incoming train may arrive at time t: either the unimpeded arrival time is t; or the unimpeded arrival time is $t - \gamma$ and the train is delayed by duration γ.

The probability that the incoming train arrives before time t is

$$\Pr[A < t] = \int_{-\infty}^{t-\gamma} \int_{-\infty}^{\infty} f_{U,D}(u, \delta)\, d\delta\, du + \int_{t-\gamma}^{t} \int_{u}^{\infty} f_{U,D}(u, \delta)\, d\delta\, du\,,$$

where $f_{U,D}$ is the density of the joint probability distribution for (U, D). This calculation is depicted in the right side of Figure 7. The probability density function for A is

$$f_A(t) = F_D(t - \gamma) f_U(t - \gamma) + (1 - F_D(t))\, f_U(t),$$

where F_D is the cumulative distribution function for D.

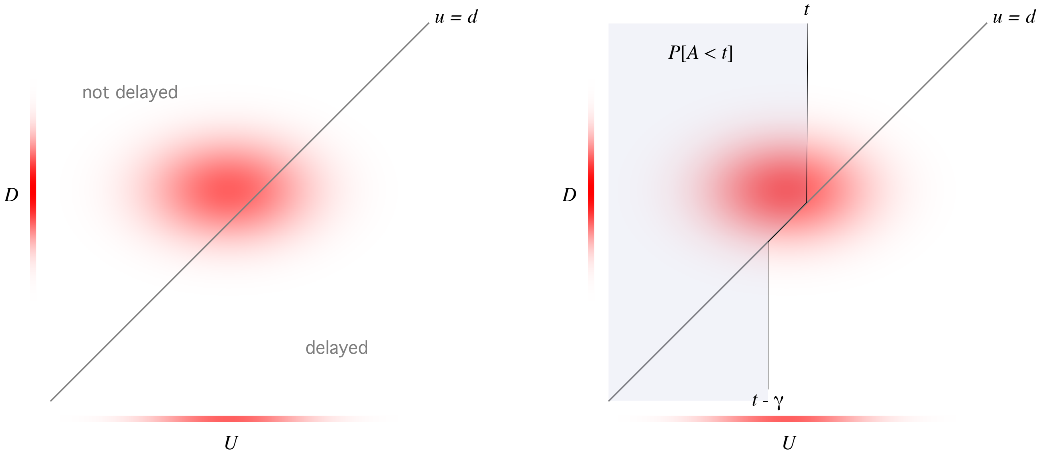

Figure 7: The joint distribution of unimpeded arrival times and departure times.

Figure 8 shows the probability density function for A when U is normal with mean 300 seconds and standard deviation 80 seconds, and D is normal with mean 350 seconds and standard deviation 50 seconds. When the incoming train is delayed, the delay is fixed at 300 seconds.

So far we have only considered the impact of a single cross. Further work is required to generalise this method to account for multiple crosses and to develop fast numerical procedures for estimating the distribution of arrival times of each train.

7 Summary

Given distributions of train departure times and section running times, and a specified procedure from recovering from delays, simulation and Monte Carlo methods predict the proportion of trains that will arrive on time, or alternatively to calculate target arrival times for which a given proportion of

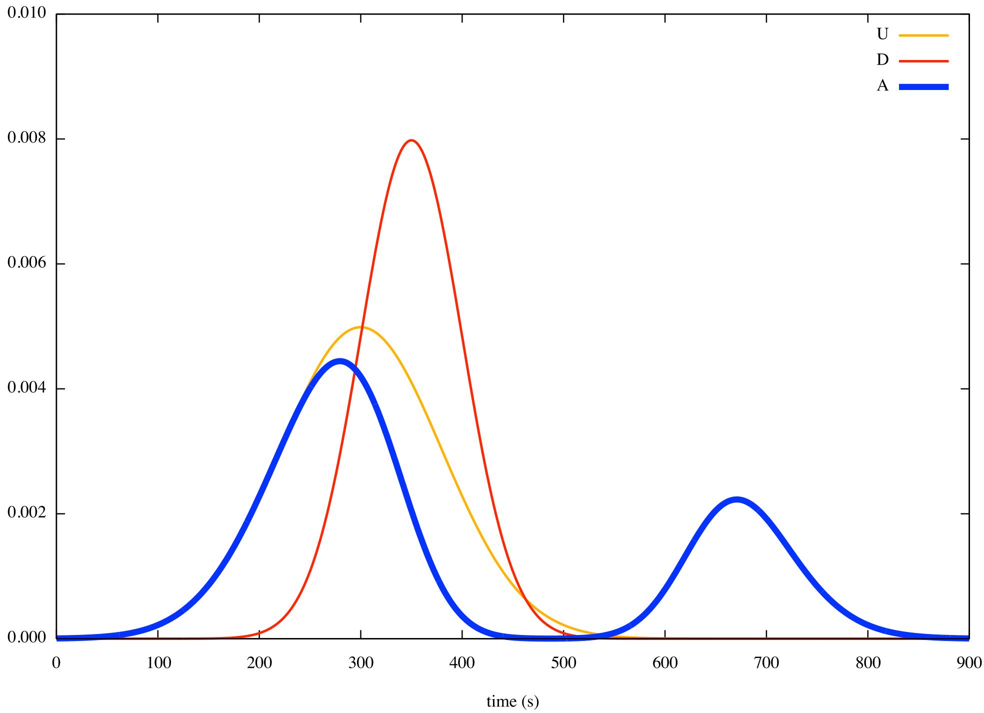

Figure 8: Probability density functions for U, D and A.

trains will arrive on time. We have also shown how we can estimate crossing delays and the distribution of arrival times for a pair of opposing trains. One interesting result is that having trains with different priorities doubles the expected total delay at a cross.

Further work is required to model the distributions of departure times and section running times, and to generalise the calculation of arrival distributions to networks with many trains and many potential crosses. The resulting calculations can be used to assess the impact on robustness of different timetables and different recovery procedures.

Acknowledgements We are grateful to the CRC for Rail Innovation (established and supported under the Australian Government's Cooperative Research Centres program) for the funding of this research which was conducted as part of Project No. R2.103, 'Train Planning Assessment Tools', and to the Australian Rail Track Corporation for providing data and advice.

References

[1] Cicerone, S., Stefano, G. D., Schachtebeck, M. and Schöbel, A. Multi-stage recovery robustness for optimisation problems: a new concept for planning under disturbances, 2009, Technical Report 0226, ARRIVAL. C770

[2] Murali, P., Dessouky, M., Ordóñez, F. and Palmer, K. A delay estimation technique for single and double-track railroads, *Transportation Research Part E*, **46**, 2010, 483–495, 2010. doi:10.1016/j.tre.2009.04.016 C770

Author addresses

1. **A. R. Albrecht**, School of Mathematics and Statistics, University of South Australia, Mawson Lakes, Australia.
2. **J. de Jong**, Department of Applied Mathematics, University of Twente, The Netherlands.
3. **P. G. Howlett**, School of Mathematics and Statistics, University of South Australia, Mawson Lakes, Australia.
4. **P. J. Pudney**, School of Mathematics and Statistics, University of South Australia, Mawson Lakes, Australia.
   mailto:peter.pudney@unisa.edu.au