

Punctuality and Capacity in Railway Investment: A Socio-Economic Assessment for Finland

Luca Corolli ^{a, 1}, Giorgio Medeossi ^{a, 2}, Saara Haapala ^{b, 3},
Jukka-Pekka Pitkänen ^{b, 4}, Tuomo Lapp ^{b, 5}, Aki Mankki ^{b, 6}, Alex Landex ^c

^a TRENOLab s.r.l.s., via Maniaco 7/A, 34170 Gorizia, Italy, Phone: +39 0481 30031
E-mail: ¹ l.corolli@trenolab.com, ² g.medeossi@trenolab.com

^b Ramboll Finland, Espoo, Finland, E-mail: ³ saara.haapala@ramboll.fi,

⁴ jukka-pekka.pitkanen@ramboll.fi, ⁵ tuomo.lapp@ramboll.fi, ⁶ aki.mankki@ramboll.fi

^c Ramboll Denmark, Copenhagen S, Denmark, E-mail: alex.landex@ramboll.dk

Abstract

This paper presents two methods designed to provide quantitative data for analysing the socio-economic impacts of rail network improvements developed for the Finnish Transport Agency. The first is a capacity estimation method; it adapts the UIC 406 method to the characteristics of the Finnish rail network. The second method estimates delay propagation based on the key characteristics of lines; in this case distinct formulas were developed using regression for single- and double-track lines. The proposed methods were evaluated based on actual and simulated data from Finland and the UK. They provide network saturation and delay data for evaluation of capital improvements by network managers. The study results were approved and adopted by the Finnish Transport Agency.

Keywords

Capacity estimation, Delay propagation, UIC 406, Mathematical regression, Finland.

1. Introduction

The Finnish Transport Agency (FTA) requires preparation of socio-economic assessments for all major infrastructure investments. This requirement covers many types of railway projects from track rehabilitation to major network improvements. Unfortunately, there is currently no established quantitative method for assessing the capacity and traffic punctuality impacts of railway investments, and therefore they are only assessed qualitatively.

This paper presents results of research conducted for the FTA to develop quantitative methods for assessing the capacity and traffic punctuality impacts of railway investments for use in FTA's socio-economic assessments (Finnish Transport Agency, 2018). The first method assesses railway line capacity, enabling the rail network manager to determine line saturation, and thereby estimate the effect of investments on capacity. The second method evaluates delay propagation given a set of line parameters, enabling the network manager to estimate the effect of investments on train punctuality. Both methods were developed with the aim of being easy to apply by non-experts in socio-economic analyses.

This paper is organised as follows: Section 2 describes the capacity analysis method, focusing on its interpretation for Finland and results obtained by applying it to a real single-track line. Section 3 describes the delay propagation methods developed using regression for use on single- and double-track lines. Finally, Section 4 presents conclusions.

2. Capacity analysis method

The main concern in socio-economic assessments is railway network utilisation, making capacity consumption the key performance indicator. Railway capacity can be defined as the maximum throughput of a given set of trains on a specific line section or station area. Many methods have been developed to estimate railway line capacity including UIC 405 (UIC, 1996), CAPACITY (Pitkänen, 2005), and CAP1/CAP2 (Moreira et al., 2004). A basic way to calculate capacity consumption is to determine the share of time reserved for train operations during a given time period. The result is a percentage, as shown in Equation (1):

$$Cc\% = 100 \times \frac{\text{Time reserved for train operations}}{\text{Analysed time period}} \quad (1)$$

The most widely used method for estimating capacity consumption in Europe is UIC 406 (UIC, 2004). A key shortcoming of this method is that it does not clearly define many important parameters, leading to a wide room for interpretation (Lindner, 2011). As a result, multiple interpretations have been proposed including the UK's Capacity Utilisation Index (CUI) and Denmark's Train Mix (Landex, 2008).

An alternative method for capacity consumption estimation uses capacity indices. For example, heterogeneity indices have been developed based on the observation that heterogeneity has a clear negative correlation to disturbance tolerance (Vromans, 2005). Similarly, rail yard conflict indices have been developed based on railway layout, conflict probability, or minimum train headways (Pitkänen, 2005).

In addition to timetable-based calculation methods, capacity can also be estimated using microscopic simulation. Simulation is typically used when detailed information on the impact of various alternative infrastructure scenarios or fault situations is needed. An advantage of simulation models is that they can take human behaviour into account using stochastic parameters. A drawback is that they typically require users to define a complete microscopic model, which can be time consuming.

In the Finnish context, a study (Pitkänen, 2005) was aimed at calibrating the SBB's CAPACITY method for application in Finland. An important finding during model calibration was that results are always dependent on specific infrastructure, rolling stock and timetable assumptions, making it very difficult to study independent measures. In socio-economic assessments, these parameters frequently differ between alternatives, making comparison impossible.

An important requirement of the socio-economic assessments being considered in this research is that they should be tackled using macroscopic analysis. Therefore, microscopic methods (i.e., simulation) are not suitable. As a result, it was decided to develop an interpretation of the UIC 406 method based on characteristics of the Finnish rail network. The goal of developing a UIC 406 interpretation for Finland was to create a simple and accurate method for estimating capacity applicable to both single- and double-track lines.

2.1 UIC 406 interpretation for Finland

Developing an interpretation of UIC 406 for Finland started with Equation (1). Defining the equation denominator (the time period) is straightforward; defining the numerator (the time reserved for train operations) is more complicated.

Determining the time reserved for train operations depends on many parameters including features of the Finnish interlocking system and rolling stock. These parameters are listed and discussed in Table 1.

Parameter	Description	Notes
K	Capacity consumption	Measured in percentage
T	Analysed time period	Suggested value is 60 minutes
h_A	Sum of minimum headway times	Sum of the time intervals between two consecutive trains running in the same direction
t_D	Sum of driving time differences	Sum of time intervals between two consecutive trains running in the same direction with different driving times
t_O	Sum of occupation times	Sum of time intervals between two consecutive trains running in opposite directions on a single-track line
t_{EPD}	Earliest possible departure time, compared with the beginning of the time period	Time interval referring to the impact of partial trains in the beginning of the time period
t_M	Sum of supplementary time for maintenance	Time that the line section is not available for normal operations due to maintenance
t_S	Sum of station and crossing times	Amount of time needed for switch turning operations during the time period

Table 1. Parameters used to determinate time reserved for train operations. All time measurements are expressed in minutes.

Using the parameters listed in Table 1, Equation (1) can be expressed as:

$$K = \frac{h_A + t_D + t_O + t_{EPD} + t_M + t_S}{T} \quad (2)$$

The first step in calculating this equation is to define the set of trains to be analysed. Next, the data must be prepared for each of the parameters. This is described below.

Definition of the set of trains to be analysed. Capacity consumption is typically calculated for hourly time periods. Trains are assigned to time periods based on the time of departure from the first station they leave in the studied area. For double-track sections, areas can span over multiple locations (i.e., stations, halts, or junctions). Single-track sections, on the other hand, are only defined between two consecutive locations.

Calculation of minimum headway times (h_A). Minimum headway times depend on the driving speed and signalling. Block sections can vary by direction and therefore headway values must be calculated separately for each direction. Theoretically, the minimum headway time depends on the driving speeds of two consecutive trains. Defining:

- n block sections factor: $n = 1$ for single block sections, $n = 2$ for multiple block sections
- d average block section length, in km
- s weighted average speed, in km/h

Let us denote with h_i the minimum headway time for a train i , that is $h_A = \sum_i h_i$. For each train i , headway h_i is calculated as shown in Equation (3) (notice that 60 = mins/hour):

$$h_i = \frac{(n * d * 60)}{s} \quad (3)$$

Calculation of running time differences (t_D). The running time difference describes the extra time needed when a slow train is followed by a faster train. This calculation, for double-track sections, depends on the operations of consecutive trains. Let us denote with t_i the additional headway to be assigned to a train i . If a train i is followed by a slower or equally fast train, there is no additional headway: $t_i = 0$. Otherwise, t_i is calculated as the difference between the running times of the two trains over the area being analysed. The total t_D value is then calculated as the sum of all t_i values, i.e. $t_D = \sum_i t_i$.

Calculation of occupation times (t_O). In this context, the term occupation time describes the reserved period after an operation on a single-track line. It is equal to the train running time on the line section being analysed.

Calculation of earliest possible departure times (t_{EPD}). This parameter is used to describe the impact of trains that only partially operate during the analysed time period, i.e. that span over multiple time periods in the studied area. In the following, we call such trains “partial trains”. Four cases can be identified:

1. There is no partial train in the analysed time period: $t_{EPD} = 0$
2. There is only one partial train t in the scenario, departing before the beginning of the scenario and arriving at destination during the timetable period:

$$t_{EPD} = at_i + hl - rt_{i'} - bg$$

where:

- at_i arrival time of train i
 - hl headway of the last line section
 - $rt_{i'}$ running time of first train i'
 - bg beginning of the time period
3. Multiple partial trains (arriving during the considered period) are present in the scenario: only the last partial train is considered.
 4. There is at least one train running through the scenario, i.e. departing before and arriving after the scenario period: t_{EPD} is set to the length of the time period, resulting in full capacity consumption (100%).

Calculation of station and junction crossing times (t_S). Station and junction crossing times consist of the extra occupation time needed to account for turning a switch between two train operations. They are location-specific and should be provided by a signalling specialist. t_S is calculated as the sum of these values.

Calculation of supplementary time for maintenance (t_M). Timetables may or may not include planned capacity reservations for maintenance work or shuntings. If these operations are planned and known, they can be included in the analysis by simply adding their total duration, in minutes, to the t_M parameter.

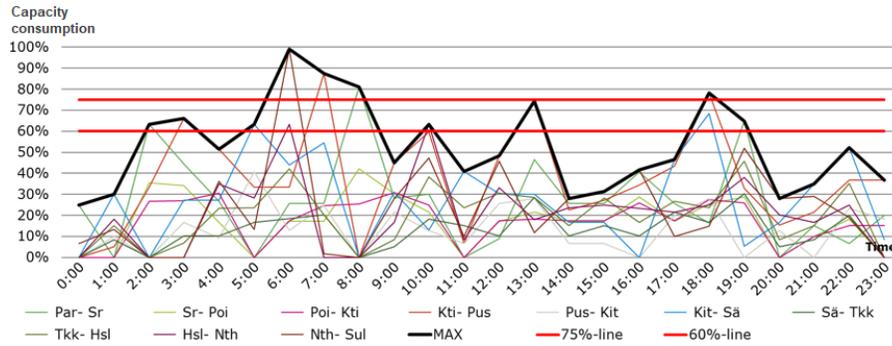


Figure 1. Capacity consumption in different single-track line sections.

2.2 Capacity consumption analysis results

Once capacity consumption values are defined for all line segments and time periods being analysed, the consumption value for the whole line can be determined. Line capacity is given by the largest value during the considered time period. For single-track sections, both directions are considered together, so the maximum value between the two directions is taken. For double-track sections, the two directions are considered separately.

When capacity consumption values are calculated over a full day, peak times typically stand out. The sharpness of these peaks gives important information on the likelihood of delays on track sections. For track sections with both passenger and freight trains, UIC has set two threshold values for congestion: 75% in peak hours, and 60% off-peak.

Figure 1 illustrates capacity consumption for different sections of a single-track line with mixed operations in Finland. Each line depicts the variation in capacity consumption over the whole day for a particular line segment. The thick black line highlights maximum values, while the two red straight lines indicate the UIC threshold values. As shown in Figure 1, during peak hours there is congestion in multiple areas, with capacity consumption remaining above 75% for three hours. This indicates a high risk of unpunctuality and little room for effective delay recovery. Similarly, the 60% off-peak threshold is exceeded three times.

3. Delay propagation method

The second method developed to better quantify socio-economic assessments of railway investments was a delay propagation method. This method calculates the relationship between capacity-related parameters and delays. Several well-accepted methods using capacity to calculate delays are already available for double-track lines (Landex, 2008). Conversely, for single-track lines, no direct relationship can be consistently identified following the theoretical evidences first identified by Potthoff (Potthoff, 1962). As a result, single- and double-track lines must be analysed separately using ad-hoc methods. These methods require large sets of data with a wide range of capacity usage. Such data can be obtained by either analysing operational data from several lines (including some heavily used lines) or using simulation (which allows testing several increasing traffic density scenarios and analysing the corresponding simulated delays). The next two sections describe the distinct methods for analysing delay on single- and double-track lines.

3.1 Single-track line delay propagation method

Based on theoretical considerations and an analysis of actual Finnish data, the delay propagation P for a group of trains on a single-track line can be defined as a function of:

- the number of crossings N_x in the timetable
- the margin t_m (it is a function of the running time: 10% for passenger trains, and 12.5% for freight trains)
- the initial delay $t_{d,i}$, with early- and late-running trains accounted for separately: $t_{d,i}^+$ and $t_{d,i}^-$. Notice that early arriving trains are considered as trains with negative initial delay, i.e. they contribute to the $t_{d,i}^-$ parameter.

Since initial delay is the delay given as input and final delay is delay given as output, in the following text we call them “input delay” and “output delay”, respectively. Both input and output delays include all delays regardless of the cause of the delay. For input delays this is not an issue since infrastructure investments can only affect delays that propagate in the track section affected by the investment. For output delays, days with heaviest delay propagation within each line need to be filtered out of the data set since they include major train or infrastructure failures, which are not related to railway investments.

As part of this research, one year’s worth of input data were aggregated by day and line. These data were supplemented by simulation data since historical data do not cover all possible parameter combinations. The simulations were run using OpenTrack software (Nash and Huerlimann, 2004) on timetables with 12 different numbers of crossings per train (each corresponding to a specific headway value) and 5 different input delay values. This showed how delays changed altering one parameter at a time. One hundred simulations were run for each combination of crossings and input delay, for a total of $12 \cdot 5 \cdot 100 = 6,000$ simulations.

The simulation results are illustrated in Figure 2 which shows the relationship between input delay (x-axis) and output delay (y-axis). The lines show the output delay variation for a given headway (in seconds), while the vertical bars show the average output delay across all headway values. As shown in Figure 2 the relationship between output and input delay appears to be slightly super-linear.

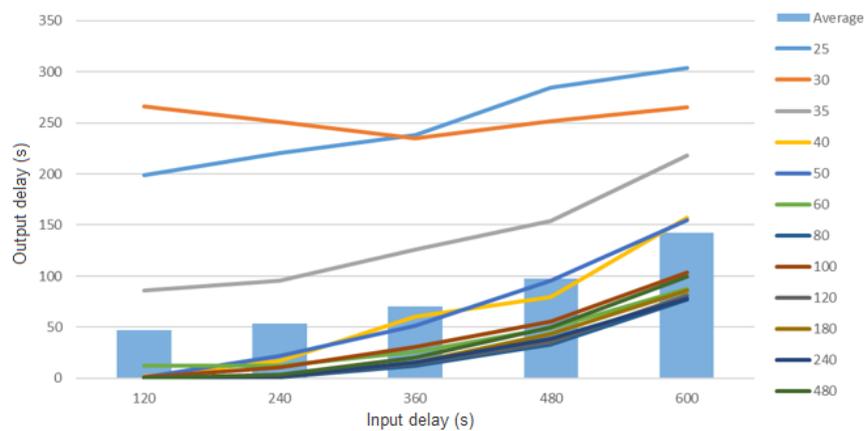


Figure 2. Simulation results analysis: output delay vs input delay

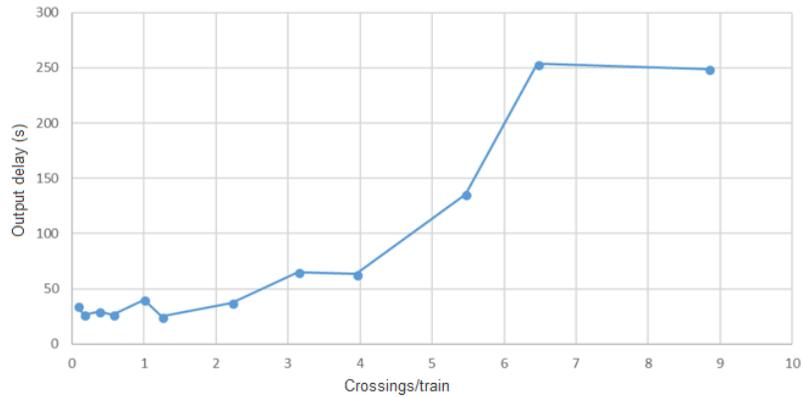


Figure 3. Simulation results analysis: output delay vs crossings.

Figure 3 illustrates the relationship between number of crossings and output delays. For small numbers of crossings there is no appreciable tendency for delays to increase. At approximately 3 crossings/train output delay starts increasing, and after 4 crossings/train it rapidly grows. The growth in output delay does not continue beyond 7 crossings/train since deadlocks in the simulation prevent trains from arriving at all.

The delay propagation model was developed to obtain a mathematical formula for estimating output delays based on input delays and crossings. In this case mathematical regression, an approach consistent with other railway delay propagation research (Marković et al., 2015) was used to develop the formula.

The first step in a regression analysis is to examine the data to determine the type of relationship. Figure 4 illustrates a quadratic trend line plotted for the relationship between input delay and output delay, while Figure 5 illustrates a quadratic trend line plotted for the relationship between crossings/train and output delay. In both cases quadratic approximations appear to be reasonable. Since quadratic equations are also easy for non-experts to apply, they were chosen for use in developing the assessment method.

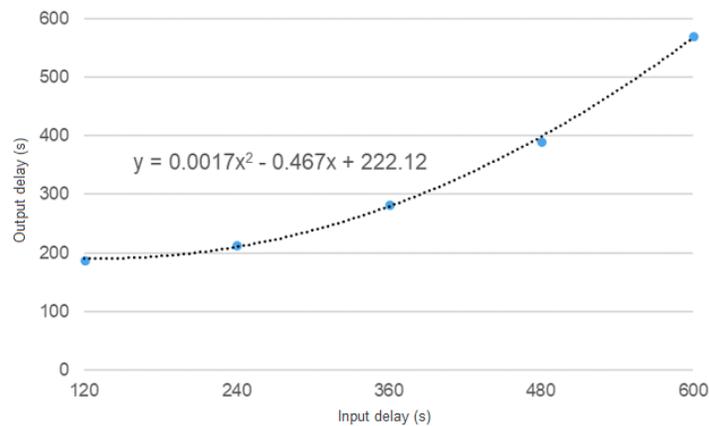


Figure 4. Output delay vs input delay: quadratic trend line.

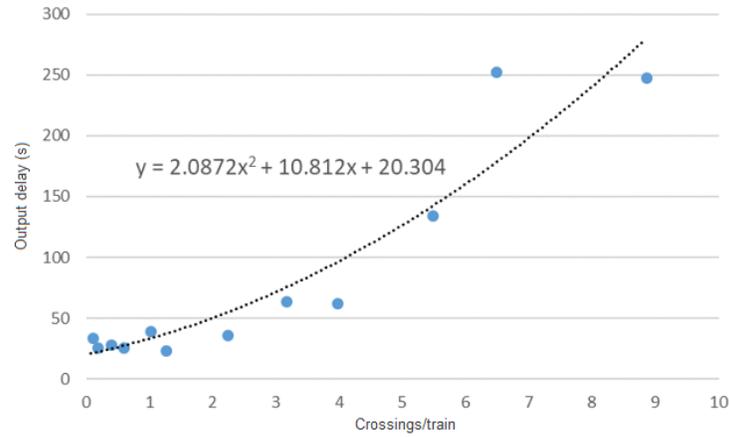


Figure 5. Output delay vs crossings per train: quadratic trend line.

The regression model for the single-track case is a combination of two quadratic formulas: one considering positive input delays and one considering the number of crossings. Denoting with $N_{x,t}$ the average number of crossings per train, and with $\beta, \gamma, \delta, \epsilon$ the regression parameters, the total expected output delay for a group of trains can be calculated with Equation (4):

$$t_{d,e}^+ = f(t_{d,i}^+, N_{x,t}) = \beta \cdot t_{d,i}^{+2} + \gamma \cdot t_{d,i}^+ + \delta \cdot N_{x,t}^2 + \epsilon \cdot N_{x,t} \quad (4)$$

The parameters were obtained by running a regression on the simulation results using the XLSTAT data analysis Excel add-on. The parameters found were: $\beta = 0.0005$, $\gamma = -0.152$, $\delta = 2.127$, $\epsilon = 10.392$. Next a goodness of fit indicator was calculated to evaluate results. Denoting with $t_{d,o}^+$ the observed (measured) positive output delays, goodness is defined in Equation (5):

$$\text{goodness} = 1 - \frac{\sum(|t_{d,e}^+ - t_{d,o}^+|)}{\sum t_{d,o}^+} \quad (5)$$

The goodness measure was calculated using the identified parameters and Finnish historical data from 12 railway lines. The goodness was equal to just 10.8%, calling for an alternative approach. Thus, a mixed approach was studied. In the mixed approach, simulated data were used to estimate crossings parameters δ and ϵ (since real data do not have a sufficient range of crossings values), and real data were used to determine the input delay parameters β and γ . The regression analysis of real input delay data resulted in a negligible value for β (so it was removed from the formula), and $\gamma = 0.918$. The goodness measure calculated with these parameters was equal to 61.0% which is reasonable. The final proposed formula for estimating total output delay for a group of trains on a single-track line is presented in Equation (6):

$$t_{d,e}^+ = 0.918 \cdot t_{d,i}^{+2} + 2.127 \cdot N_{x,t}^2 + 10.392 \cdot N_{x,t} \quad (6)$$

3.2 Double-track line delay propagation method

The key parameters used to evaluate delay propagation on a double-track line are:

- the buffer times, i.e. the additional spacing between trains provided to reduce the risk of delay propagation. It is especially important to examine cases when the buffer time is limited (so called “critical headways”). Buffer times are included using a set denoted with B , with $b \in B$ buffer thresholds. Buffer thresholds are indices to denote buffers of size s_b . Each buffer, measured in minutes, ranges from a minimum $s_{min}(b)$ to a maximum $s_{max}(b)$, thus $s_{min}(b) < s_b \leq s_{max}(b)$
- the initial delay $t_{d,i}$ (referred to as “input delay” in the following text)
- the running and stop time margins t_{mr} and t_{ms}

As for the single-track case, the formula for estimating output delays for a group of trains from a set of input parameters can be obtained using mathematical regression. First, the input data were prepared aggregating values for all parameters for each line, direction, day, and time-band. The total expected output delay for a group of trains can be calculated using Equation (7):

$$t_{d,e}^+ = \beta \cdot \sum_{b \in B} (w(b) \cdot bf_b) + \gamma \cdot t_{mr} + \delta \cdot t_{d,i}^+ + \epsilon \cdot t_{d,i}^- \quad (7)$$

Parameter bf_b is the number of buffers in a threshold b , and $w(b)$ is the weight associated to buffer b . Thus, the effect of buffer times is evaluated considering the criticality of having a small buffer time, with $w(b)$ defined to reflect this criticality: $w(b) = 2^{-s_{max}(b)}$.

Input data include 10 double-track lines with both directions separately accounted for. One-year worth of traffic data were considered, defining one train group per day/line. Buffer thresholds were subdivided into five 1-minute wide groups, from 0 up to 5 minutes. Train groups without buffers between 0 and 1 minute (the most critical ones) were not considered.

Regression performed on the input data provided the following parameter values: $\beta = 22.443$, $\gamma = -0.033$, $\delta = 1.029$, and $\epsilon = -0.001$. All parameters have a reasonable practical interpretation, and the corresponding goodness is 73.91%. Thus, they may be used in Equation (7) to create Equation (8) for estimating delay propagation on double-track lines:

$$t_{d+,e} = 22,443 \cdot \sum_{b \in B} (w(b) \cdot bf_b) - 0.033 \cdot t_{mr} + 1,029 \cdot t_{d,i}^+ - 0,001 \cdot t_{d,i}^- \quad (8)$$

Regression results were tested to evaluate the impact of timetable changes, by applying the proposed delay propagation formula to 4 scenarios from the UK’s Crossrail project. The input delay was set at zero to simplify the analysis. Results showing the estimated effect of all timetable-dependent parameters on output delay are illustrated in Table 2.

Scenario	SC0	SC1	SC2	SC3
Number of trains	40	48	22	11
0–1 min buffers	397	559	37	0
Buffer weight	210.668	291.219	36.688	0
Margin	547.5	678	0	0
Expected delay [s/train/day]	100.4	114.1	36.4	0

Table 2. Validation of the double-track line delay propagation method. Scenario N is denoted with SCN (e.g. SC0 = Scenario 0)

The base scenario (SC0) represents the current timetable. SC1 adds 8 trains to the base timetable, resulting in a large number of small buffers. SC2 and SC3 have lighter traffic levels: SC2 has about half the trains from the base scenario, and SC3 further divides the number of trains in half. This test case study shows that the proposed mathematical model is sensitive to train frequency and provides reasonable results.

4. Conclusions

This paper discusses research carried out for the Finnish Transport Agency to develop quantitative methods for evaluating the socio-economic impacts of railway investments. Two methods were developed, the first determines capacity consumption and the second determines delay propagation. These methods are designed to provide railway network managers with simple formulas for evaluating the impacts of railway line investments without performing complex simulations.

The capacity consumption method was developed by applying the characteristics of the Finnish railway (e.g., interlocking, rolling stock) to the UIC 406 capacity formula. The paper describes the development of the parameters and highlights the differences between single- and double-track line cases. The method was then applied to a Finnish line to illustrate use of capacity over the course of a day.

The delay propagation forecasting method was developed using mathematical regression with both simulated and historical traffic data. The regression results were evaluated using a goodness measure. Separate methods were developed for the single- and double-track line cases to account for the different factors triggering delay propagation.

References

- Finnish Transport Agency, 2018. "Capacity and Punctuality in Railway Investment Socio-Economic Assessment". Technical report.
- International Union of Railways (UIC), 1996. "Links between railway infrastructure capacity and the quality of operations" (UIC code 405 OR). Paris, France.
- International Union of Railways (UIC), 2004. "UIC leaflet 406", France
- Landex, A., 2008. "Methods to estimate railway capacity and passenger delays". PhD thesis, Copenhagen, Technical University of Denmark.
- Lindner, T., 2011. "Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity". *Journal of Rail Transport Planning & Management* 1, pp.49-57.
- Marković, N., Milinković, S., Tikhonov, K.S., Schonfeld, P., 2015. "Analyzing passenger train arrival delays with support vector regression". *Transportation Research Part C: Emerging Technologies*, vol. 56, pp.251-262.
- Moreira, N., Garcia, L., Catarrinho, P., 2004. "Network Capacity". *Computers in Railways IX Conference*, Dresden, Germany.
- Nash, A., Huerlimann, D., 2004. "Railroad simulation using OpenTrack"; *Computers in Railways IX*, WIT Press, Southampton, pp. 45-54.
- Pitkänen, J-P., 2005, "Radan välityskyvyn mittaamisen ja tunnuslukujen kehittäminen". Helsinki: Ratahallintokeskus, Liikennejärjestelmäosasto (in Finnish language).
- Potthoff, G., 1962. "Verkehrsströmungslehre I - Die zugfolge auf Strecken und in Bahnhöfen", Transpress, Berlin, Germany (in German language).
- Vromans, M., 2005. "Reliability of Railway Systems". Rotterdam: The Netherlands TRAIL School.