

# A conditional Bayesian delay propagation model for large-scale railway traffic networks

Yixuan Zhang<sup>1</sup>, Ruimin Li<sup>2</sup>, Ting Guo<sup>1</sup>, Zhidong Li<sup>1</sup>, Yang Wang<sup>1</sup>, Fang Chen<sup>1</sup>

<sup>1</sup>Faculty of Engineering & Information Technology, University of Technology, Sydney, NSW 2007

<sup>2</sup>Sydney Trains, Transport for New South Wales, NSW 2008

ruimin.li2@transport.nsw.gov.au

## Abstract

Reliability is one of the critical success factors for both passenger and freight rail service delivery. One major factor that significantly impacts reliability performance is delays spanning over spatial and temporal dimensions. One way to increase reliability is to avoid systematic delay propagation through better timetable design to reduce the interdependencies between trains caused by route conflicts and train connections. In this paper, we aim to predict the propagation of delays on a railway network by developing a conditional Bayesian delay propagation model. In the model, the propagation satisfies the Markov property that determination of delay propagation for the future of the process is based solely on its present state, and that the history does not have an influence on the future. For the cases of delay caused by cross line conflicts and train connection, throughput estimation is considered in the model. The proposed model benefits from scalable computing time and complexity advantages over the Markov property. Implementation of actual operational data shows the feasibility and accuracy of the proposed model when compared to traditional probability models. The proposed model can be used for timetable evaluation and operations management decision support.

## 1 Introduction

Delay and its propagation is one of the major factors that seriously influence the railway systems performance. In scheduled railway traffic networks, a single delayed train may cause a domino effect causing secondary delays over the entire network, which is the main concern to planners and train controllers (Goverde, 2010). Despite the application of buffer times and systematic delays (which is highly related with timetable design), train delays are inevitable due to many stochastic factors (for instance, increasing passenger demand, equipment failures, passenger behaviour, asset failures, driver behaviour and other factors like weather conditions). In a complicated network system, delays are usually not solely contained to one train, station or line. They have a propagation effect that can spread throughout the network, which bring secondary delays and cause a series of traffic problems such as disordered operations of the railway network that results in reduced service quality, decreased punctuality and reliability. In general, delays can result in negative economic and social impacts that affect daily life. Therefore, it is important to analyse

delay factors, develop delay management and prediction strategies to maintain the reliability of the railway network system.

In real-world train operations, delay prediction relies heavily on the experience and intuition of a local train controller rather than a network-wide computational instrument (Martin, 2016). Given the complex structure of a railway network and interdependent train operations between a large set of origins and destinations, a local train controller's estimation of delays and the subsequent decisions are strongly dependent on the state of traffic on the network and generally limited to a local geographical area or line. Whereas in reality, not only the trains on same service line should be considered for secondary delays, but also the route conflict of crossing or merging lines, crew-relief, and train connections can all cause secondary delays. With the rapid growth of demand and the large and dense network structures in big cities, the domain knowledge and expertise of local train planners/controllers must be supported by an advanced computational tool that enables them to respond to an emergency in real time with optimal or approximate optimal solutions.

In this paper, we aim to predict the propagation of delays on Sydney Trains railway network by developing a conditional Bayesian delay propagation Model. Creation of such a delay propagation model for a complex network like Sydney Trains has been hindered by two fundamental limitations. Firstly, it is a challenge to discover the delay patterns at the station level by collecting and incorporating the mass of train operational data from a big network. Secondly, there has been a lack of models capable of simultaneously taking the complex railway network structure and the delay dependencies among multiple trips into consideration for propagation/prediction tasks. To overcome these limitations, a conditional Bayesian model is proposed in this study. A Bayesian-based methodology is a representational tool meant to capture complex structures and "organize one's knowledge about a particular situation into a coherent whole" (Darwiche, 2009). At the same time, it allows for the incorporation of massive historical data in identifying the contingencies between multiple events and updating the state of different variables given real-time data.

The proposed conditional Bayesian model in this paper assumes that the propagation satisfies the Markov property, where if can determine delay propagation for the future of the process based solely on its present state, and the history do not have an influence on the future, despite one might know the process's full history. The introduction of the Markov assumption helps us to easily discover the delay dependency and reduce the model complexity. For the problem of complicated network structure, we divided the context of delay propagation into four types based on the following scenarios: The impact of delays on: (1) the current trip which caused the primary delays (self-propagation); (2) the following trips along the same service line (backward propagation); (3) the trips from cross service lines that interact with intersection stations of current trip (cross-line propagation); and (4) the trips which have train connection<sup>1</sup> of current trip (train-connection propagation). The four scenarios contain all the dependencies of different trips based on the railway

---

<sup>1</sup> Train connection: A train will start a new trip after reaching the destination of one trip. We call there is a train connection between these two trips as they share the same train. The preceding trip's delay will always cause the following trip's secondary delay at the origin.

network and operation timetable, which further reduce the complexity of delay propagation. Therefore, our proposed model benefits from scalable computing time and simplified complexity over the Markov property. Implementation on Sydney Trains operation data has demonstrated the feasibility and accuracy of the proposed model comparing to traditional probability models. The proposed model can also be used for timetable evaluation.

The rest of the paper is organised as follows: In Section 2, we summarise some related work about delay propagation analysis. In Section 3, we first describe some background knowledge and problem setting about delay propagation. Then, we discuss the details of our proposed model. In Section 4, we present a case study results of the method introduced in Section 3. Finally, we conclude our work in Section 5.

## 2 Related works

Several empirical methods have been studied in machine learning literature to model and analyse delay propagation. (Chen and Harker, 1990) improved empirical estimation method by taking the uncertainty in the actual train departure time explicitly into consideration. (Carey and Kwiecinski, 1994) developed a simple stochastic method to measure delay propagation by using stochastic simulation between trains and derive the relationships between scheduled headways and delay propagation. Although many methods have been studied and applied to the analysis of delay propagation, such methods have never been used successfully over a complex network. And some research has been conducted to solve this issue. (Florio and Mussone, 1998) proposed a solution which can be carried out for a real-world case with medium complexity of railway scheme.

However, runtime delay and dwell delay is not taken into consideration in the model based on real operation data. (Goverde, 2001) introduced the use of regression analysis to predict departure delays from arrival delays and capture the variation of run-time and dwell time in the model. This method is easy to be implemented but cannot handle complex scenarios very well. (Yuan et al., 2002) proposed a stochastic model to forecast the distribution of departure delays based on the distribution of late arrival delays and dwell delays using Monte Carlo sampling method.

Nevertheless, this method has its limitation with lack of dependence analysis of dwell delays on late arrival delays and scheduled dwell times. (Goverde, 2010) presented an advanced model which can be applied to the whole network to calculate the delay propagation of initial delays within a timetable period. The author also designed a graph algorithm to compute the propagation of train delays by storing the propagated delays in the bucket. But, this algorithm was designed based on a linear system without taking the distribution of run-time delay and dwell time delay into account. Recent work by (Cerreto et al., 2018) developed a clustering method to identify the delay patterns which can be used for big datasets generated by the railway signal system. However, the author used a hard assignment method (K-means), which cannot be interpreted in a probabilistic way and cannot be used to analyse and measure the delay impacts with conditions. (Wu et.al, 2019) further trailed deep learning methods such as LSTM in predicting primary delays, but the work is still preliminary and has not been tested on big data set.

### 3 Methodology

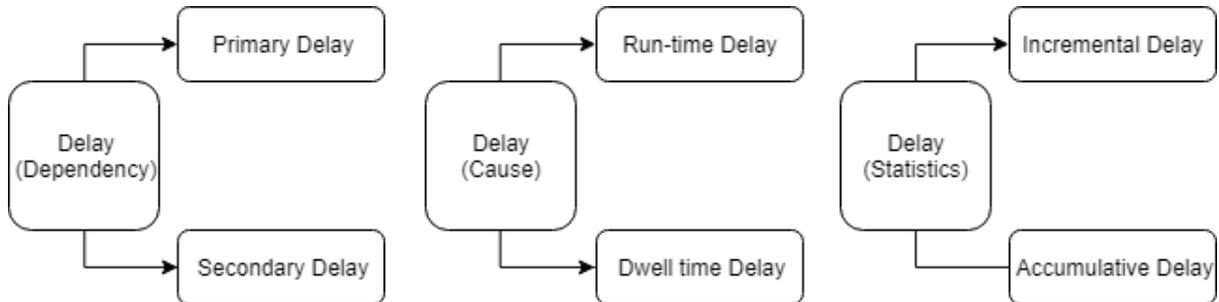
The optimization of capacity utilization and timetable design requires the prediction of the reliability and punctuality level of train operations, which is determined by the train delays and delay propagation. The propagation of train delays is more likely to occur during the arrival/departure of trains at stations because the crossing or merging of lines and platform tracks are in most cases bottlenecks in highly used railway networks. In the following section, we will introduce the problem formulation of delay propagation and the related background, then a conditional Bayesian model will be proposed for delay propagation prediction. The model can be used for estimating future impact when a delay happened, not only for a single train but also the following trains, crossing-line trains and connected trains.

#### 3.1 Problem formulation and background

##### 3.1.1 Train delay

Delay is defined as the variation time between actual time and scheduled time. In the railway networks, the delay can be further decomposed into different types according to different definitions as shown in Figure 1. In this paper we will consider both primary delay and secondary delay. We would divide them based on delay causes, then estimate the distribution of incremental delay on station level by using historical data, and finally give the accumulative delay and influences to the whole railway network.

Figure 1. Delay types under different standards<sup>2</sup>.



Run-time and dwell time delay are considered as variation time with respect to run-time and dwell time of a given train, respectively. The incremental run-time from Station  $i$  to Station  $j$  is defined as:

$$R_{ij} = t_j^a - t_i \tag{1}$$

Where  $t_j^a$  is the arrival time at Station  $j$  and  $t_i$  is the departure time at Station  $i$ . Then the incremental run-time delay from Station  $i$  to Station  $j$  is computed by  $Actual(t_j^a) - Scheduled(t_i)$ . Similarly, the incremental dwell time from Station  $i$  to Station  $j$  is:

---

<sup>2</sup> Based on the dependency relationship, delay can be divided into primary delay and secondary delay. Here, we defined the earliest self-caused delay as primary delay, and all delays caused by this particular delay are called secondary delay. Based on delay causes, delay can be divided into run-time delay and dwell time delay, which are defined as the variation time with respect to run-time and dwell time respectively. Delay can also divide into incremental delay and accumulative delay according to whether considering the stacking effect of delay of the same trains.

$$D_j = t_j - t_j^a \tag{2}$$

Therefore, the incremental dwell time delay can be computed by  $Actual(t_j) - Scheduled(t_j) = Actual(t_j - t_j^a) - Scheduled(t_j - t_j^a)$ .

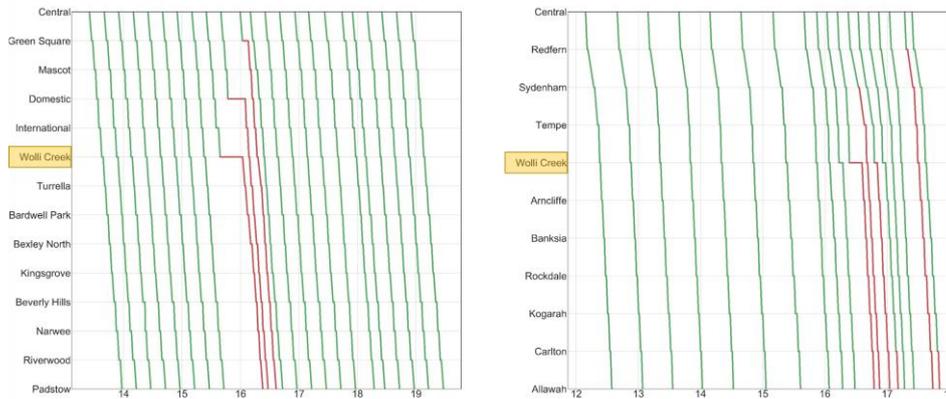
### 3.1.2 Delay propagation

As discussed in the Introduction section, the delay can have a domino effect and one delayed single trip (primary delay) can cause secondary delays to following trains and crossing-line trains, this phenomenon is defined as delay propagation as shown in Figure 2. Specifically, 4 typical scenarios have been specified to capture all the delay propagation possibilities as shown in Figure 3.

- **Self-propagation:** If a train  $T_1$  has a delay at Station 3, the delay will propagate and influence  $T_1$  itself at the following stops.
- **Cross-line propagation:** If a train  $T_1$  has a delay at Station 3, the delay may propagate and influence trains that are from cross lines arriving at Station 3 during the time period that  $T_1$  parked at Station 3 unscheduled.
- **ackward propagation:** If a train  $T_1$  has a delay at Station 3, the delay may propagate and influence the following trains that would arrive at Station 3 during the time period that  $T_1$  parked at Station 3 unscheduled.
- **Train connection propagation:** As trains always need to run round trips or connected trips each day, there is always train connection effects<sup>3</sup>: a train arrives late at the destination will cause a start delay for its next trip.

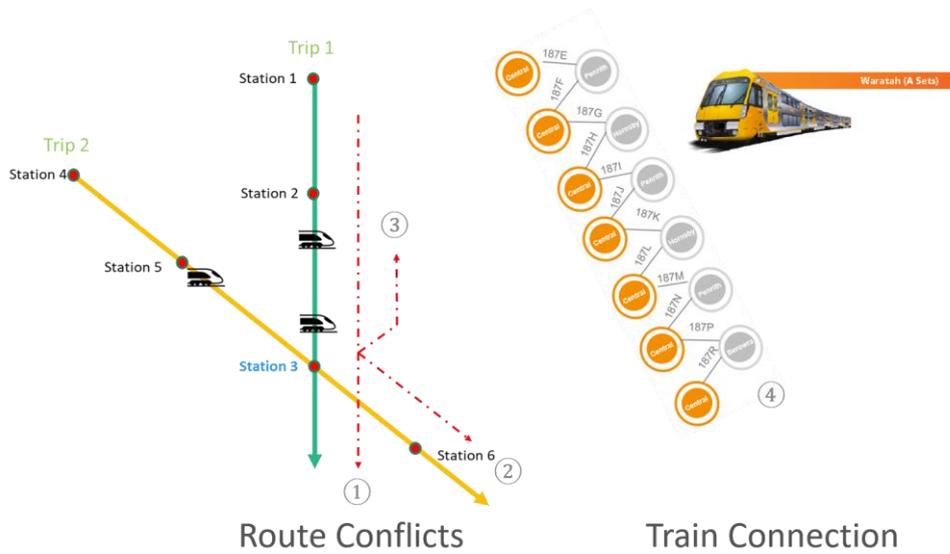
Cross-line and backward effects of a train’s delay are also termed as *route conflict effects*.

**Figure 2. Delay propagation and impacts.**



<sup>3</sup> Beside train connection effect, crew caused delay also contributes to the start delay for the following trips. Crew caused delays are the delays that caused by the late arrival of drivers or guards. Based on our research, the late arrival of crew are always caused by the delay of preceding trains the crew stay aboard.

Figure 3. Four scenarios of delay propagation.



### 3.2 Preliminary

Delay propagation could cause a huge impact on daily railway network operations. Reliable prediction of delay propagation would support train planners/controllers' operation decision making. A conditional Bayesian model incorporates with Markov property is developed to predict and estimate delay propagation, which can capture the complex real-world scenarios by including both run-time delay and dwell delay into the model. To better understand the model, we first give a brief introduction of the background knowledge, and then the detailed model design will be proposed.

#### 3.2.1 Bayesian inference

A Bayesian inference framework is adopted in the proposed model. Bayesian inference has been studied in many empirical machines learning literature by applying Bayes' Theorem  $P(A | B) = P(B | A)P(A)/P(B)$  to model any random variables. The denominator in Bayes' Theorem is also called evidence or normalising constant, so we normally rewrite the formula into  $P(A | B) \propto P(B | A)P(A)$ , where  $P(A | B)$  is the posterior distribution and is proportional to the likelihood  $P(B | A)$  times prior  $P(A)$ . Usually, the Bayesian inference framework is constructed in such a way that the prior is formed first and the posterior distribution is then computed using Bayes' Theorem. However, when conjugate priors are not applicable, it is necessary to compute the intractable integrals, which are not possible to compute the posterior distribution analytically.

#### 3.2.2 Markov property

The Markov property is described as the independence of the future from the past, given the present. To be more formally, if we define a one-parameter process  $X$  is a Markov process with respect to a filtration  $\{\mathcal{F}\}_t$  when  $X_t$  is adapted to the filtration, then for any  $s > t$ ,  $X_s$  is independent of  $\mathcal{F}_t$  given  $X_t$ , which can be expressed as  $X_s \perp \mathcal{F}_t | X_t$ . The Markov property holds in train delay propagation cases because the self-propagation caused delays in any station are influenced only by the delay status of the immediate preceding station. More specifically, if a train at Station 1 has a primary delay and propagates the delay to Station 2, and Station 2 further

propagates delay to Station 3, the delay on Station 3 will only be considered being affected by the delay in Station 2. Formally,  $P(\tau_{i+1}^{delay} | \tau_i^{delay}, \dots, \tau_1^{delay}) = P(\tau_{i+1}^{delay} | \tau_i^{delay})$ .

### 3.3 Conditional Bayesian delay propagation

Herein we will propose our conditional Bayesian model for different delay propagation scenarios as introduced in Section 3.1.2.

#### 3.3.1 Delay self-propagation

Delay self-propagation within a single trip is illustrated as Scenario ① in Figure 3. As shown in the figure, the delay of a trip at one station can be propagated to the next station. Therefore, it is obvious that the accumulative delay (secondary delay) at the next station is highly dependent on the delay at the preceding stations by following the Markov property. We use a delay example, which occurs at two consecutive stations, to explain how to model the delay propagation from one station to another from Bayesian theory perspective within a single trip and how it can be easily extended to the whole trip.

If we define  $R_{12}$  as the normal runtime<sup>4</sup> from Station 1 to Station 2, and  $D_2$  as the normal dwell time at Station 2. We assume both  $R_{12}$  and  $D_2$  follow normal distributions and can be learnt by using historical data, expressed as  $R_{12} \sim \mathcal{N}(\mu_{12}^R, \sigma_{12}^R)$  and  $D_2 \sim \mathcal{N}(\mu_2^D, \sigma_2^D)$ .

If there is a primary/secondary delay occurred at Station 1 (departure delay of Station 1:  $\Delta t_1$ ), it will affect the driver's behaviour (e.g. speed up to make up the delay in the following trip) and the passenger flow pattern (e.g. more passengers on the platforms) with arrival delay ( $\Delta t_2^a$ ) of Station 2. Therefore, based on Markov property, the runtime and dwell time between Station 1 and 2 will follow:

$$R'_{12} = R_{12} + (\Delta R_{12} | \Delta t_1),$$

$$D'_2 = D_2 + (\Delta D_2 | \Delta t_2^a), \text{ where } \Delta t_2^a = \Delta t_1 + (\Delta R_{12} | \Delta t_1)$$

In which,  $(\Delta R_{12} | \Delta t_1)$  and  $(\Delta D_2 | \Delta t_2^a)$  are the incremental delays caused by  $\Delta t_1$  for runtime and dwell time, respectively.

#### Incremental runtime delay:

When there is a delay occurred at a Station 1 ( $\Delta t_1$ ), the driver tends to accelerate the train by turning on more power in order to make up the delay in the following trip, hence influencing the actual runtime to the next station (Station 2). For different values of  $\Delta t_1$ , the incremental runtime delay could be different. If  $\Delta t_1$  is small, the driver can easily catch up with no delay at Station 2, while when  $\Delta t_1$  is very large, there is no way for the driver to overtake the delay at Station 2 even if the train will be running at full speed. Therefore, in our proposed model, we will divide  $\Delta t_1$  into several bins: (1)  $\Delta T^1: 0 < \Delta t_1 \leq 3mins$  ; (2)  $\Delta T^2: 3mins < \Delta t_1 \leq 5mins$  ; (3)  $\Delta T^3: 5mins < \Delta t_1 \leq 10mins$ ; and (4)  $\Delta T^4: \Delta t_1 > 10mins$ . For different bins, we will learn different normal distribution to fit them respectively, i.e.  $(\Delta R_{12} | \Delta t_1) = (\Delta R_{12} | \Delta T^i) \sim \mathcal{N}(\mu_{12}^{\Delta T^i}, \sigma_{12}^{\Delta T^i})$ .

<sup>4</sup> Normal runtime means the runtime with no delay happened at the departure/arrival stations.

Therefore, the incremental runtime delay by giving  $\Delta t_1$  can be re-written as:

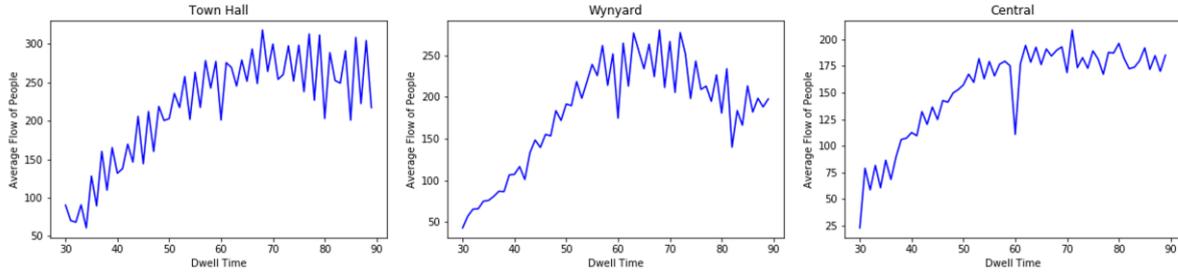
$$R'_{12} \sim \mathcal{N}(\mu_{12}^R, \sigma_{12}^R) + \mathcal{N}(\mu_{12}^{\Delta T^i}, \sigma_{12}^{\Delta T^i}) = \mathcal{N}\left(\mu_{12}^R + \mu_{12}^{\Delta T^i}, \sqrt{(\sigma_{12}^R)^2 + (\sigma_{12}^{\Delta T^i})^2}\right)$$

### Incremental dwell time delay:

When there is an arrival delay happened at a Station 1 (arrival delay of Station 2:  $\Delta t_2^a$ ), the number of passengers will increase and therefore influence the actual dwell time when the delayed train arrives at Station 2. Based on the historical train running data of Sydney Trains, we found that there was a good linear relationship between the actual dwell time and the increasing passengers to some extent. From Figure 4, we can see that the strong linear relationship exists when the dwell time is between 30 seconds and 60 seconds (30 seconds is the scheduled dwell time for most stations in Sydney Trains Timetable). This fact inspired us to develop and train a piecewise function to fit the relation:

$$(\Delta D_2 | \Delta t_2^a) = \begin{cases} \alpha_2(\Delta t_2^a) + \beta_2, & \Delta t_2^a < 60s, \\ c_2, & \Delta t_2^a \geq 60s. \end{cases}$$

**Figure 4. The relationship between dwell time and average passenger flow on three stations. (Based on our experiments, we found that most of other stations have similar patterns.)**



Therefore, the incremental runtime delay by giving  $\Delta t_1$  can be re-written as:

$$D'_2 \sim \mathcal{N}(\mu_2^D, \sigma_2^D) + (\Delta D_2 | \Delta t_2^a) = \begin{cases} \mathcal{N}(\mu_2^D + \alpha_2(\Delta t_2^a) + \beta_2, \sigma_2^D), & \Delta t_2^a < 60s \\ \mathcal{N}(\mu_2^D + c_2, \sigma_2^D), & \Delta t_2^a \geq 60s. \end{cases}$$

### Accumulative departure delay:

In the process of delay propagation in a single trip, we only consider the accumulative delay at each station rather than the incremental runtime/dwell time. On the other hand,  $\Delta t_1$  always follows a normal distribution propagated from preceding stations rather than fixed values when  $\Delta t_1$  is a secondary delay, i.e.  $\Delta t_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ . Therefore,  $(\Delta R_{12} | \Delta t_1)$  and  $(\Delta D_2 | \Delta t_2^a)$  are the delay components that we should consider. They can be expressed under Bayesian theory as:

$$P(\Delta R_{12} | \Delta t_1) = \sum_{b=\Delta T^1}^4 \mathcal{N}(\Delta R_{12} | \mu_{12}(b), \sigma_{12}(b)) P(b | \Delta t_1), \text{ where } b = \{\Delta T^i, i = 1, 2, 3, 4\}$$

$$\text{where } a_i < \Delta T^i \leq b_i \text{ and } P(\Delta T^i | \Delta t_1) = \int_{a_i}^{b_i} \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right) dx$$

And,

$$P(\Delta t_2^a) = P(\Delta t_1 + (\Delta R_{12} | \Delta t_1)) = \sum_{b=\Delta T^1}^4 \mathcal{N}(\mu_1 + \mu_{12}(b), \sqrt{(\sigma_1)^2 + (\sigma_{12}(b))^2}) P(b | \Delta t_1)$$

where  $b = \{\Delta T^i, i = 1, 2, 3, 4\}$

$$(\Delta D_2 | \Delta t_2^a) \sim \mathcal{N}(\alpha_2(\Delta t_2^a) + \beta_2, \sigma_2^D) P(\Delta t_2^a < 60s) + \mathcal{N}(c_2, \sigma_2^D) P(\Delta t_2^a \geq 60s),$$

Where  $P(\Delta t_2^a < 60s) = \int_0^{60} f(x) dx$ ,  $P(\Delta t_2^a \geq 60s) = \int_{60}^{+\infty} f(x) dx$  and we have  $f(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)$ .

Finally, with the given departure delay of Station 1:  $\Delta t_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ , the departure delay of Station 2  $\Delta t_2$  is:

$$\Delta t_2 = \Delta t_1 + (\Delta R_{12} | \Delta t_1) + (\Delta D_2 | (\Delta t_1 + (\Delta R_{12} | \Delta t_1)))$$

Any delay propagation between two consecutive stations for a single trip can be calculated by using the above equation.

### 3.3.2 Cross-line propagation, backward propagation and train connection propagation

The other delay scenarios have a common delay propagation structure: delay can be propagated from unscheduled arrival time from a preceding train (cross-line train, preceding train or itself with round-trips). Because of this common delay propagation structure, cross-line, backward and train connection propagation can share the same model. Note that Train  $j$  and Train  $k$  (from a different origin to Station 2) may have conflict area at Station 2 and in the scheduled timetable, Train  $j$  arrives at Station 2 prior to Train  $k$ . To estimate the delay propagation from Train  $j$  to Train  $k$ , we have:

- Train  $j$ 
  - Scheduled Departure time at Station 2:  $t_2^j$
  - Delay propagation from Station 1a to Station 2:
 
$$\Delta t_2^j = \Delta t_1^j + (\Delta R_{12}^j | \Delta t_1^j) + (\Delta D_2^j | \Delta t_2^{a,j})$$
  - Estimated Departure time at Station 2:  $t_2^j + \Delta t_2^j$
- Train  $k$ 
  - Scheduled Arrival time at Station 2:  $t_2^{a,k}$
  - Delay propagation from Station 1b to Station 2:
 
$$\Delta t_2^{a,k} = \Delta t_1^k + (\Delta R_{12}^k | \Delta t_1^k)$$
  - Estimated Departure time at Station 2:  $t_2^{a,k} + \Delta t_2^{a,k}$

We then have  $Z = |t_2^{a,k} + \Delta t_2^{a,k} - t_2^j - \Delta t_2^j|$ , which is the absolute value of variation time between estimated Departure time at Station 2 of Train  $k$  and estimated Departure time at Station 2 of Train  $j$ . We also define  $S$  as the minimum interval time between Arrival time at Station 2 of Train  $k$  and Departure time at Station 2 of Train  $j$  to guarantee Train  $k$  can run smoothly through Station 2 not being affected by Train  $j$ . This value can be computed from historical statistics.

Next, we can define delay propagation from Train  $j$  to Train  $k$  as:

$$\begin{aligned} (\delta t | \Delta t_2^{a,k}) &\sim P(Z - S < 0) \\ &= P(|t_2^{a,k} + \Delta t_2^{a,k} - t_2^j - \Delta t_2^j| - S < 0) \end{aligned}$$

Now, we can incorporate estimated delay propagation from Train  $j$  to Train  $k$  into departure delay estimation equation and get estimated accumulative departure delay at Station 2 of Train  $j$  as the equation below (we eliminate notation  $j$  to keep the formula clear):

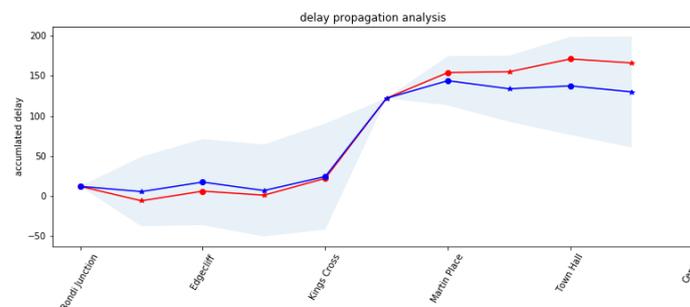
$$\Delta t_2 = \Delta t_1 + (\Delta R_{12} | \Delta t_1) + (\Delta D_2 | \Delta t_2^a) + (\delta t | \Delta t_2^{a,k})$$

## 4 Case study

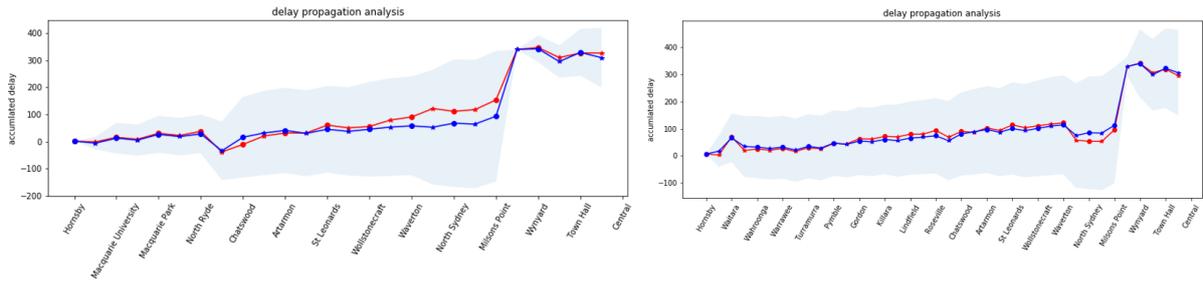
In this section, we predict the delay propagations by using the proposed conditional Bayesian model for different scenarios introduced in Section 3.1.2 and compare them with the observed values. When a primary delay at the given station is specified, the predicted means and confidence intervals of secondary delays for the impacted following, cross-line, and connected trips are calculated.

In Figure 5, we show the predicted delay propagation pattern (the blue line is the mean and the blue band is the confidence interval) and the actual running records of the trip (the red line). The dots represent the predicted/actual dwell time at stations, and the stars indicate the predicted/actual runtime between two consecutive stations. The model is applied from the beginning of the trip (by taking the starting time as the input of the proposed model and predicting the dwell/running time for the following stations). It can be seen that when there is no delay, the proposed model can be used for normal running/dwell time prediction (the accumulated delay can be considered as the noise to the scheduled running/dwell time), the means of the predicted values are similar to the actual ones. When we specified a delay happened between Kings Cross and Martin Place (a runtime delay), we just updated the input of the model and re-run it for the following stations. The predicted delay propagation pattern matches the actual one well, and the difference between actual and predicted mean at the trip destination (Central Station) is less than 30 seconds, which demonstrates the reliability of the proposed model. Similar performance can be found for other delay scenarios. Figure 6 and 7 show the predicted delay propagation on the cross-line trip/ following trip. Figure 6 shows that the cross line trains would be affected by the primary delay (left figure) occurred at Station Milsons Point and the predicted delays of the cross-line trip (right figure) were close to the actual ones, with difference within 10 seconds (the difference between actual values and predicted means). Figure 8 shows the start delays (right figure) caused by the train-connection delay (left figure). The delay of the preceding trip was propagated to the following trip as the two trips were using the same train.

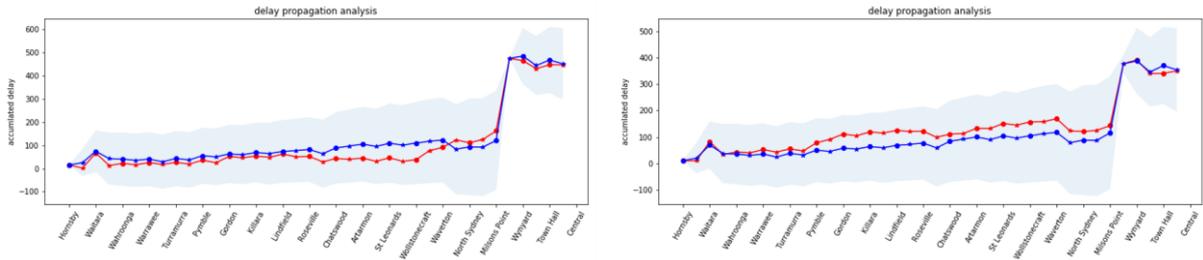
**Figure 5. Delay self-propagation.**



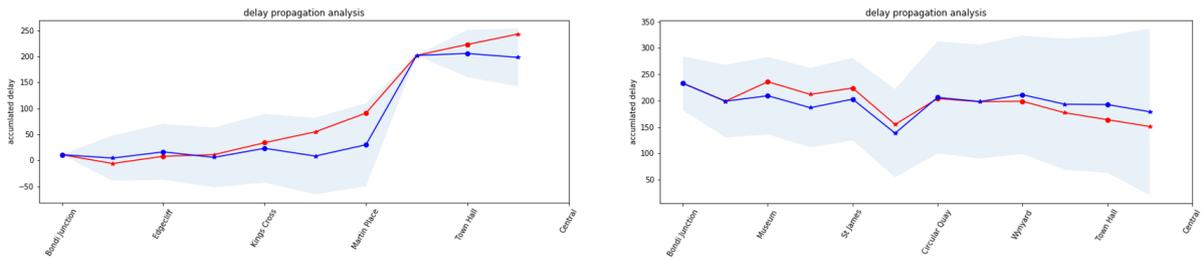
**Figure 6. Delay cross-line propagation.**



**Figure 7. Delay backward propagation.**



**Figure 8. Delay train-connection propagation.**



## 5 Conclusion:

In this paper, we proposed a conditional Bayesian delay propagation model. The model takes Markov property into consideration and handles complex realistic application scenarios. Based on the case study results, it can be seen that the proposed model can not only provide reliable predicted secondary delays for the impacted following, cross-line, and connected trips when a primary delay is captured. The model can also give the confidence intervals of the predicted values for railway operation managers and site controllers to better understand the potential influence of the delay and handle it properly.

## 6 References:

- BOWMAN, L. A. & TURNQUIST, M. A. 1981. Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General*, 15, 465-471.
- CAREY, M. & KWIECINSKI, A. 1994. *Stochastic approximation to the effects of headways on knock-on delays of trains.*
- CERRETO, F., NIELSEN, B. F., NIELSEN, O. & S. HARROD, S. 2018. *Application of Data Clustering to Railway Delay Pattern Recognition.*
- CHEN, B. & HARKER, P. 1990. *Two Moments Estimation of the Delay on Single-Track Rail Lines with Scheduled Traffic.*

- CHIGUSA, K. 2018. *A framework for identification of systematic service deterioration in urban rail systems*. Massachusetts Institute of Technology.
- FLORIO, L. & MUSSONE, L. 1998. An Analytical Model For The Simultaneous Calculation Of Capacity Of Lines, Junctions And Station Tracks. *WIT Transactions on The Built Environment*, 37, 10.
- FRUMIN, M. S. 2010. *Automatic data for applied railway management: passenger demand, service quality measurement, and tactical planning on the London overground network*. Massachusetts Institute of Technology.
- FURTH, P. G. & MULLER, T. H. 2006. Service reliability and hidden waiting time: Insights from automatic vehicle location data. *Transportation Research Record*, 1955, 79-87.
- GOVERDE, R. M. P. 2001. Delay Propagation and Process Management at Railway Stations.
- HANSEN, I. A. Improving railway punctuality by automatic piloting. ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585), 2001. IEEE, 792-797.
- LUKE J., MARTIN, W., 2016. Predictive Reasoning and Machine Learning for the Enhancement of Reliability in Railway Systems. International Conference on Reliability, Safety and Security of Railway Systems: 178-188.
- PARBO, J., NIELSEN, O. A. & PRATO, C. G. 2016. *Passenger perspectives in railway timetabling: a literature review*. *Transport Reviews*, 36, 500-526.
- ROB, M., GOVERDE, P. 2010. *A delay propagation algorithm for large-scale railway traffic networks*. *Transportation Research Part C: Emerging Technologies* 18.3: 269-287.
- Sydney Trains. 2018. Sydney Trains 2017-2018 Annual Report.
- VROMANS, M. 2005. *Reliability of railway systems*.
- WILSON, N. H., NELSON, D., PALMERE, A., GRAYSON, T. H. & CEDERQUIST, C. 1992. Service-quality monitoring for high-frequency transit lines. *Transportation Research Record*.
- WU, J., CAI, C., ZHOU, L., DONG, F. & SHEN, J., 2019. Towards a General Prediction System for the Primary Delay in Urban Railways. *IEEE Intelligent Transportation Systems Conference*
- YUAN, J., GOVERDE, R. & HANSEN, I. A. 2002. *Propagation of Train Delays in stations"*.