Modelling public transport disruptions and impact using smart-card data

Dong Zhao[1], Adriana-Simona Mihăiţă[1], Yuming Ou[1], Hanna Grzybowska[2]

***Abstract*— Evaluating disruptions in public transport (PT) utilisation is challenging due to often stochastic traveller behaviour and missing data information on affected services. This paper proposes a new approach for modelling PT patronage and disruption impact using integrated data-driven modelling and the Fourier transform technique.**

Firstly, using tap-on and off information of smart-card data, we estimate in-vehicle passenger numbers to integrate as well as trips passing through the incident area. Secondly, considering the PT patronage pattern as a periodic function, we employ the Fourier transform to convert it into a sum of simpler trigonometric functions to filter out the one representing common data noise successfully and generate an accurate profile for a typical day. Thirdly, we introduce an enhanced sensitivity test to improve the model's ability to identify the impact of the disruption. Finally, multiple impact measurement methods are compared to capture the disruption impact.

The findings demonstrate the effectiveness of leveraging in-vehicle count to maximise data volume and enhance impact identification. The PT patronage pattern can be effectively modelled using the Fourier transform. The utilisation of the enhanced sensitivity test can effectively filter out unnecessary trigonometric components, resulting in a refined model capable of accurately identifying the impact of disruption.

***Index Terms*— Public transport; disruption analysis; smart-card; patronage pattern modelling; Fourier transform.**

I. Introduction

A. Background and motivation

In a multi-modal transport network, a road traffic disruption always extends beyond its own network. These disruptions can affect related bus lines, train services, and, subsequently, the overall transport network. The bus network shares mobility rights with other road transport modes, and is very sensitive to road traffic disruptions. If an incident occurs on public transport (PT), passengers who are affected by the disruption may seek alternative modes of transportation to continue their journeys. This increased demand for alternative transport services can create unexpected changes in the transportation system. When disruptions occur on major transportation modes like trains, the consequences can be significant. For example, in the case of a train strike in London [1], the disruption in the train network can lead to city-wide congestion and have a widespread impact on commuters. With limited or no train services available, hundreds and thousands of passengers are affected, and the demand for alternative travel modes, such as shared bicycles, buses or subway services, increases dramatically [2], [3].

These unexpected changes in demand can strain the transport network, especially if it exceeds the network's capacity to handle such disruptions. However, it is also essential to consider scenarios where the impact is relatively small, such as a localized disruption caused by a road accident or a vehicle breakdown. These situations are more common in daily life and may not have a city-wide influence, but they could be one of the influence factors for unstable traffic [4]. This paper aims to tackle this issue by examining the daily PT patronage profile and investigating the impact of road incidents on this patronage pattern. By understanding the daily movement of PT users in the network, we gain insights for making informed decisions to enhance the PT service and ease the cooperation between private vehicles and PT.

Analyzing the patronage pattern can guide improvements in the supply of PT services, ensuring they align with the needs and preferences of passengers. However, when compared to the travel pattern analysis for private vehicles, analysis regarding PT travel patterns has received relatively less attention in the literature due to limited data availability and modelling methods. To model the PT travel behaviour, data collection can be carried out through various means, such as census or surveys [5], smart-card data [6], or utilizing the General Transit Feed Specification (GTFS) data [4]. Data-driven methods often rely on machine learning and deep learning algorithms [7], [8], [9], [10], [11]. These techniques utilize the available data to extract meaningful insights and make accurate impact predictions [12], [13], [14], and some of them will focus on also predicting how long disruptions will last in the network [15], [16]. On the other hand, simulation-based methods are commonly used in the literature to estimate PT patronage patterns. These methods involve creating simulations following assignment models that replicate real-world scenarios, allowing researchers to observe and analyze the change in behaviour and dynamics of PT systems [4], [1], [14], [17], [18], [19] so that the essential regarding the impact on the network can be captured and modelled. Some propose as well optimisation techniques under disruptions to ease impact [20].

In this research paper, we analyze the PT patronage patterns using real smart-card data and compare the normal versus incident circumstances in order to estimate the impact on PT users. As an observation, the PT patronage appears to have a daily repetitive occurrence, and we draw inspiration from signal analysis principles. Therefore, we employ a

[1] University of Technology Sydney, Ultimo, NSW 2007, Australia. Corresponding authors contact: Dong.Zhao@student.uts.edu.au

[2] Data61, CSIRO, Eveleigh, NSW 2015, Australia and School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

Manuscript 1084 submitted to 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). Received May 28, 2023.

Fourier transform function to fit the data and filter out the noise, enabling us to model the underlying pattern of PT patronage effectively. In the past, one study utilized the method to model road traffic patterns by analyzing traffic volumes derived from mobile data [21]. Another study employed the graph Fourier transform (GFT) for a similar purpose [22]. However, these studies primarily focused on modelling private vehicle traffic and not public transport patronage. By establishing a robust model capable of accurately representing traffic patterns, we take an additional step in our research by not only modelling the PT patronage pattern but also applying the model to identify the effects of incidents on patronage.

B. Paper Contributions

To summarise, the main theoretical and methodological contributions of this paper are the following:

- We introduce a novel method for dynamically modelling PT patronage patterns by incorporating the Fourier transformation to effectively reduce the influence of noise and enhance the accuracy and reliability of the model;
- We employ the frequency domain following the Fourier transform to segregate the components of PT patronage patterns, enabling us to identify and to isolate the significant elements within these patterns;
- We apply a synergistic approach that combines analytical techniques with data-driven methods to identify the impact of incidents on PT passengers;
- We perform the ability of multiple measuring metrics, such as: correlation measures (Pearson's correlation coefficient), distance measures (Chebyshev distance, Wasserstein metric, Minkowski difference and Cosine similarity) and statistical tests (change, Percentage change and Symmetric percentage change);
- We propose a new application by integrating big data resources, among which GTFS data, smart-card and incident log data when excavating the information for analysing the network vulnerability.

This paper is organised as follows. In Section II, the dynamic PT patronage pattern model considering the Fourier transform is discussed, and the details of incident impact identification metrics are also included. The application of the proposed methods to a real network is presented in Section III, and the results of the case study are demonstrated in Section IV, where detailed modelling processes are demonstrated. Finally, the research conclusion and the future directions are provided in Section V.

II. Methodology

A. Entity of PT patronage

1) Number of boarding and alighting: Given smart-card data, we can produce the number of boarding and alighting for each PT stop at each time spot or for each time interval, in order to represent the PT patronage. Each record r in the smart-card data, based on the availability for use in this research, can be expressed as:

$$r(i,u_i,j,u_j,b,d), \tag{1}$$

where each parameter in a record r represents the tap-on stop i, tap-on time u_i, tap-off stop j, tap-off time u_j, the bus number b and the date of the recording d. To obtain the number of boarding and alighting for a certain stop $i, i \in \{1 \dots I\}$ during a period of time (τ), we only need to count the number of records, denoted as $N_i^{boarding}(\tau_a), a \in \{1 \dots A\}$, where a represents the a^{th} time interval of a day, A is the total number of time interval defined for a day; similarly, the number of alighting at a specific stop during a time interval can be expressed as $N_j^{alighting}(\tau_a)), j \in \{1 \dots I\}$. The number of boarding and alighting people under an impact of disruption can be expressed as $N_i^{boarding'}(\tau_a)$ and $N_j^{alighting'}(\tau_a))$. The total number of boarding persons, counted during the time interval τ can be denoted, according to Iverson bracket notation, as:

$$N_i^{boarding}(\tau_a) = \sum_{n=1}^{N} [u_i(n) \in \tau_a], \tag{2}$$

where $u_i(n)$ represent the n^{th} tap-on time recorded in the data set; if this time belongs to τ_a is true, then $[u_i(n) \in \tau_a]$ is 1; otherwise, this record is not counted.

Based on the given definition, we observe that the patronage data relying on boarding and alighting focuses on the bus stops where passengers get on and off while ignoring the passed stops along a trip. However, when it comes to trips passing through an area affected by an incident with persons already boarded at other stations, this counting method fails to account for the actual number of passengers impacted. Therefore, to accurately determine the number of disrupted passengers specifically caused by the incident, an in-vehicle passenger count is also necessary. This method allows us to count the number of passengers affected within the impacted area and provides a more precise measure of the disruption's impact.

2) Number of in-vehicle passengers: Since we hold the information around the time and location when passengers got on/off the bus, we can define the number of in-vehicle passengers for each time interval at each PT stop. For each smart-card data record r, we are able to calculate the number of time intervals (τ) through which this trip passes from the start until the end:

$$K = \frac{u_j - u_i}{\tau}, k \in 0 \dots K, \tag{3}$$

For each τ passed by a trip, a passenger is counted at each k^th interval τ_k. K is estimated by rounding the number of τ in order to improve processing accuracy. In our study, we consider 15 minutes as the time interval, so if a record of patronage starts at 8 am and ends at 9 am, instead of counting this record by boarding time (interval 8:00-8:15) once, we count it four times, for 8 am, 8:15, 8:30 and 8:45 slot to represent that this passenger is inside the vehicle from 8 am to 9 am.

According to the above definition, each record in the smart-card data set has added another element regards to the in-vehicle time, represented by $v_{i,k}$, where k denotes the k^{th} time interval τ that this trip is passing through:

$$v_{i,k} = u_i + k\tau. \tag{4}$$

Manuscript 1084 submitted to 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). Received May 28, 2023.

Therefore, each record in the data set is updated by adding an in-vehicle time $\upsilon_{i,k}$ which equals the u_i, and $K-1$ new records are filled in the data set which corresponding to their boarding location i, boarding time u_i, in-vehicle time $\upsilon_{i,k}$, PT number b and date d, denoted as:

$$r(i,u_i,\upsilon_{i,k},b,d). \tag{5}$$

Following the definitions, the number of in-vehicle passengers becomes:

$$N_i^{in-veh}(\tau_a)=\sum_{n=1}^{N}[\upsilon_{i,k}(n)\in\tau_a]. \tag{6}$$

B. Modelling PT patronage via the Fourier Transform

By taking into account the total number of passengers onboard, it now becomes possible to identify the number of affected passengers by comparing the change between a travel pattern on the day of the incident and on a typical day. To estimate the travel pattern, we require to use the incident log data, including information regarding incident time, duration, and location. In order to obtain the travel pattern for a typical day, one can apply either the traditional approach of averaging the patronage counts during non-incident days, or the Fourier transform and filter out noises from daily travel patterns.

In this work, we propose to use the Fourier Transform to analyse time-dependent signals in the frequency domain; it is a tool for decomposing a complex and repetitive behaviour pattern by summing up the sines and cosines functions. This concept can be adapted for the PT patronage estimation because such a patronage pattern is time-varying and repetitive over a certain time, exhibited by distinct peaks during the morning and afternoon periods. The seasonality enables the patronage patterns to be predictable by using the discrete Fourier transform function. According to [23], the frequency-domain function following the discrete Fourier transform can be expressed as:

$$f(\tau)=\frac{\alpha_0}{2}+\sum_{h=1}^{H}(A_h\sin(h\omega\tau+\varphi_h)+B_h\cos(h\omega\tau+\varphi_h)). \tag{7}$$

where A_h is given to describe the amplitude of the *sine* function while B_h is the amplitude of the *cosine* function. These two parameters indicate how much sine and cosine functions should be included to estimate the function of the travel pattern. $\omega\tau$ indicates the frequency component of the trigonometrical function and φ_h is the phase of the trigonometrical function.

By decomposing the patronage pattern function into multiple sine and cosine waves, it becomes possible to convert the data from the time domain to the frequency domain and following this, capture the regularity exhibited on the frequency scale. In the frequency domain, we have the magnitude spectrum by frequency (or the power spectrum in signal analysis). Those frequencies with significant magnitudes are considered major frequencies, which indicate the dominant power contained within the signal, while frequencies with low magnitudes are treated as noise. Utilizing this information, we can filter the useful data from noise data based on magnitude. This is how we de-noise or approximate any arbitrary function by summing up a determined set of trigonometric functions.

C. Measurements of impact

There are several methods for measuring the impact of the incident according to the change of patronage with and without the incident. The options for measurement include correlation measuring, such as Pearson's correlation coefficient; the metrics related to distance, such as the Chebyshev distance, the Wasserstein metric, the Minkowski difference and Cosine similarity, as well as the statistical tests, such as the change, the Percentage change and the Symmetric percentage change.

1) Change: A common method to identify the impact of disruptive events on road networks is to simply calculate the change of patronage with and without disruptions. As described in Eq. (6) and Eq. (7), in this paper, the number of counts is calculated based on the average typical day of the week. Therefore, the change due to a disruption event for each time interval t can be expressed as:

$$I^{change}(N,N')=N-N', \tag{8}$$

where N represents a set of patronage counts for a typical day and N' for the incident day.

2) Percentage change and symmetric percentage change: To avoid the problems triggered by the value for the disrupted situation being zero, we adopt the symmetric percentage change, as well, which is given as Eq. (10).

$$I^{Pchange}(N,N')=\frac{N-N^{'}}{N+\lambda}\times 100\%, \tag{9}$$

where λ is a smoothing factor used to avoid computing problems when dividing by zero.

$$I^{Pchange}(N,N')=\frac{N-N^{'}}{\frac{N+N'}{2}}\times 100\%, \tag{10}$$

By using the symmetric percentage change, the result that approaches either 2 or -2 means that there is no similarity between the two data sets, while if the result ranges to zero, it indicates that these two data sets have high similarity.

3) Cosine similarity: The Cosine similarity is reflected by the Cosine distance, which is the dot product of the number of in-vehicle passengers affected by an incident $N^{'}$ and the number of counts for a typical day, as:

$$I^{\text{Cosine}}(N,N')=\frac{N\cdot N'}{\|N\|\|N'\|}=\frac{\sum_{m=1}^{m}N_mN'_M}{\sqrt{\sum_{m=1}^{M}N_m{}^2}\sqrt{\sum_{m=1}^{M}N'_m{}^2}}, \tag{11}$$

4) Chebyshev distance: Chebyshev distance is defined as the maximum distance along any coordinate dimension which measures the greatest discrepancy in values between the corresponding coordinates of the vectors being compared:

$$I^{\text{Chebyshev}}(N,N')=\max_m(|N_m-N_m^{'}|), \tag{12}$$

5) Wasserstein distance: The metric serves as a distance function defined between probability distributions on a given metric space. N and N' are two measures on a metric space $\mathbb{R}\times\mathbb{R}$; the Wasserstein distance between these two measures is defined as the integration of the distance between any two matched points times the amount of the mass of moving from one point to another. Thus Wasserstein distance is given by:

$$I^{\text{Wasserstein}}(N,N') = inf_{\pi \in \Gamma(N,N')} \int_{\mathbb{R}\times\mathbb{R}} |N-N'| \mathrm{d}\pi(N,N'), \quad (13)$$

where π is the joint probability measure on $\mathbb{R}\times\mathbb{R}$ with marginals N and N'.

6) Minkowski difference:

$$I^{\text{Minkowski}}(N,N') = \|N-N'\|_p == (\sum |N-N'|^p)^{1/p}, \quad (14)$$

where p is the order of the norm of the difference between N and N'. When $p=1$, the Minkowski distance is the same as the Manhattan distance, while $p=2$, such distance is the same as the Euclidean distance. The value of p is 3 in this paper based on the experiment's comparison result.

7) Pearsons correlation coefficient (PCC): PCC is a way of quantitatively measuring the linear correlation; it assesses the extent to which changes in one variable are associated with corresponding changes in another variable, both in terms of direction and magnitude.

$$I^{\text{PCC}}(N,N') = \frac{\sum(n_m-\overline{n})((n'_m-\overline{n'})}{\sqrt{\sum((n_m-\overline{n})^2 \sum(n'_m-\overline{n'})^2}}. \quad (15)$$

III. Case Study

A. Network characteristics

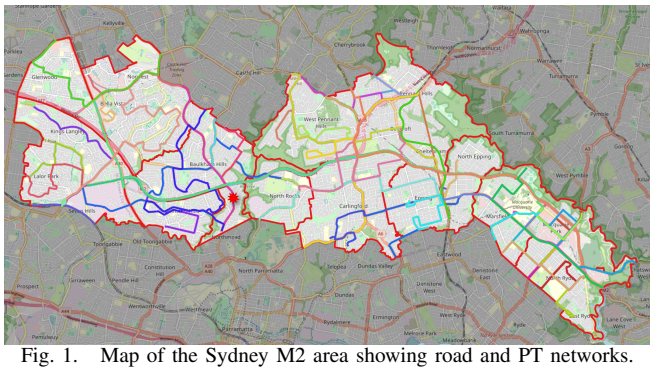

Fig. 1. Map of the Sydney M2 area showing road and PT networks.

The study focuses on the zones in the North-West region of Sydney, encompassing the M2 motorway and comprising significant residential and commercial sectors. The geographical extent of this area, depicted in Fig. 1, aligns with the boundaries defined by the Statistical Area Level 2 (SA2) [24] in digital mapping, where there are 79 bus lines consisting of 3,799 bus stops.

B. Sample of a hypothesised incident

To evaluate the feasibility of metrics on identifying the impact of patronage patterns due to the traffic disruption, we consider comparing the modelled typical profile with a hypothesised incident one and reflecting the impact by metrics as mentioned in Section II-C.

The hypothesised incident scenario is created following the details of a real incident (as introduced in Section III-C below). However, rather than directly comparing the day of the incident with a modelled typical day, we measured the difference in passenger count during the incident duration. We then incorporated this change into the patronage of the modelled typical day, effectively creating a hypothesis day of the incident. The details of the disruption are designed to mirror the scenario presented in Section III-C, with a start time of 2017-04-05 at 10:03:59 and a duration of 50 minutes. To simplify the modelling process, we approximate the time unit to 15 minutes. This makes the hypothesised incident start from 10:00:00 (time interval 40) until 10:45 (time interval 43). During the disruption, the impact manifests as an increase in patronage, with 2,397 additional passengers distributed across three 15-minute intervals (927, 790 and 680 passengers, respectively). To assess the impact, we manually adjust the count within the modelled typical day scenario for each time index interval. This allows us to observe the resulting changes reflected in the impact measuring metrics.

C. Sample of a real incident

In order to measure the impact of an incident, the sample incidents selected from the incident log data set should follow the considerations:

- such incident duration is long enough to be able to display the impacts through the PT patronage;
- such incident is away from the PT-only lane; because the impact on patronage for an isolated PT could be minor [4];
- such incidents should be located in prominent residential and commercial sectors considering the uneven distribution of patronage data. A sufficiently large dataset of patronage is necessary to ensure a discernible impact;
- such an incident possesses a substantial potential to impact PT patronage significantly.

Consequently, one sample for this case study is given as follows:

- Start time: 2017-04-05 10:03:59
- Duration: 50 min
- Type: Bus Breakdown

In this research, we have selected and tested multiple sample incidents to ensure that our findings are applicable to a wide range of scenarios. However, for the purpose of showcasing the results and the limited space, we specifically selected this particular incident in this paper. More data analysis results can be found in supplementary material [25]. All results displayed in the following sections correspond to this sample incident.

IV. Results and Discussion

A. Profiling a typical day patronage pattern using the real data

To capture the count pattern for a typical day, we explore two methods. The first method involves calculating the average of multiple non-incident days, where the day of the week matches that of the incident. This approach takes into account the observation that different days of the week exhibit distinct patronage patterns, as shown in Fig. 3. In the case of an incident occurring on Thursday, 5 April 2017, we gather the remaining non-incident Thursdays in April 2017 and compute the average count for each day. As mentioned in Section II-A, to optimize data input and streamline data processing, we have chosen to calculate the total number of in-vehicle passengers at each PT stop instead of separately considering boarding and alighting counts.

This approach effectively doubles the data size by incorporating the combined number of tap-on and tap-off events,

as depicted in the three plots in Fig. 2. This consolidation simplifies the analysis while maximizing the available data as follows: *a*) Fig. 2 demonstrates the number of in-vehicle passengers with and without an incident, where the red dash line in the figure highlights the duration of the sample incident. Comparing this figure with *b*) and *c*) in Fig. 2, which is generated by using the number of boarding and alighting, we observe that the total number of in-vehicle count on a typical day is 17,779 and that on the incident day is 18,014, whereas the sum of count for tap-on on a typical day is 9,310, and on the incident day is 9,565; additionally, the tap-off on a typical day is 11,224, and the number of tap-off on the incident day is 11,342. This comparison highlights that if we would utilize only the original tap-on data then we would have approximately 47% of trips being ignored when compared to using the in-vehicle passenger count. Similarly, when using the original tap-off data, approximately 37% of trips are overlooked.

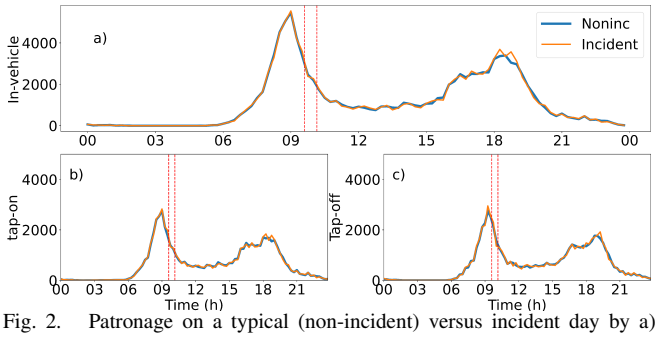

Fig. 2. Patronage on a typical (non-incident) versus incident day by a) in-vehicle passengers, b) tap-on passengers, c) tap-off passengers.

B. Modelling a typical day patronage pattern using the Fourier transform

All sub-figures in Fig. 2 present challenges in distinguishing between a typical profile and an incident profile. Despite observing numerous fluctuations and variations between the two profiles in both figures, it remains challenging to discern which changes are specifically attributed to the incident under investigation. This difficulty comes from the complex nature of the system, making it hard to ascertain the impact of the incident on the PT patronage pattern. The complexity of the system necessitates the purification of patterns by removing unnecessary noise. This requirement motivates the application of the Fourier transform, as it allows for the decomposition of complex patterns into individual simple and determined components. By examining the performance of each component separately, we can approximate the incident's impact as noise within the seasonal travel pattern.

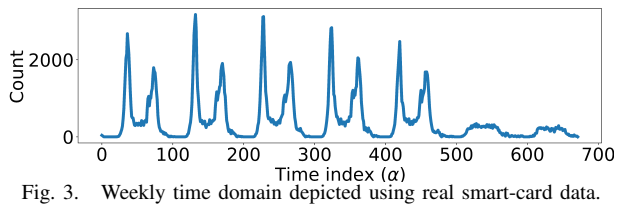

Fig. 3. Weekly time domain depicted using real smart-card data.

For the specific incident under consideration, we collect a week of data encompassing the incident day and apply the Fourier transform to convert the time domain into the frequency domain. After removing the noise components in the frequency domain, we reverse the denoised frequency domain back to the time domain to form the profile for a typical day. From this denoised time domain, we select the travel pattern for Thursday as our representative pattern. According to the weekly travel pattern depicted in Fig. 3, we only select the data for Monday to Friday. These weekdays show a similar pattern, which makes them suitable for our analytical purposes. To better visualise the tendency in the plot, we convert the date and time information into date-time-index (α) by 15 minutes.

Frequency domain: After applying the analytical process outlined in Section II-B, we convert the seasonal time domain into the frequency domain. The frequency domain representation is depicted in Fig. 4, with the left plot showcasing the overall magnitude spectrum and the right plot specifically highlighting the dominant components. In this representation, a frequency of 1/15 minutes is utilized, implying that each frequency value corresponds to the number of occurrences of a repeating event within a 15-minute interval. The properties for the top three components (highlighted by the red dots in the right sub-plot of Fig. 4) are displayed in Table I and processed by adding period information to make the data more understandable.

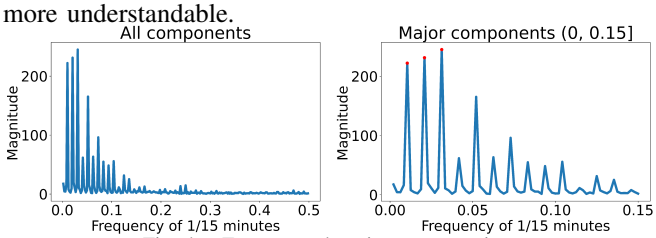

Fig. 4. Frequency domain representation.

Period and seasonality: To enhance comprehension, we convert the frequencies to periods (see Table I) by taking the reciprocal of the frequency, represented in hours and days. Each period value means the duration of time of one cycle in a repeating event. This conversion facilitates a clearer understanding of the data. The highest magnitude

TABLE I

SAMPLES OF THE FOURIER TRANSFORM OUTCOMES.

No.	FFT result	Mag	Freq ($\frac{1}{15}$ min)	Pd (h)	Pd (d)
1	53423-104789j	245	0.031	8	0.33
2	-70066+86141j	231	0.021	12	0.50
3	-98011+41997j	222	0.010	24	1.00

Notes: FFT result is the Fourier transform outcome, in complex value; Mag is magnitude; Freq is frequency per 15 minutes; Pd(h) is period by hour; Pd(d) is period by day.

(given as the first row of data shown in Table I matching the highest bar in the right figure of Fig. 4), representing the most notable seasonal patterns, exhibits a period of 8 hours (in Table I), indicating that this event repeats every 8 hours, aligning with the off-peak hours. The second highest magnitude corresponds to a period of 12 hours, reflecting the morning and afternoon peaks that repeat every half day. The event with the third highest magnitude repeats daily (repeated every 24 hours), suggesting it captures the overarching function that describes the seasonality of the patronage pattern throughout the day.

Decomposition of travel pattern: The Fourier transform, as defined in Section II-B, involves transforming a function

into a series of increasing high-frequency periodic functions. In the context of PT travel patterns, we can decompose the pattern into repetitive sub-functions. By examining the periods obtained through the Fourier transform, as shown in Table I, we can match the period information to the actual seasonality. At this stage, we can effectively filter out the noise and capture the dominant characteristics of the travel pattern. In other words, the Fourier transform allows us to isolate and analyze the significant ingredients of the PT travel pattern while disregarding irrelevant fluctuations or noise. Fig. 5 demonstrates the major frequency components in the PT patronage pattern plotted by sines and cosines. We can observe that the periodic peaks of each sub-function

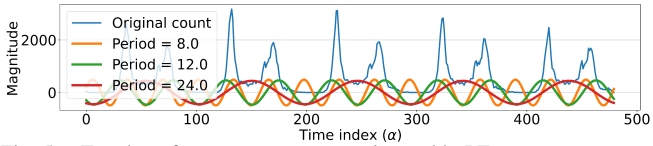

Fig. 5. Top three frequency components in weekly PT patronage pattern.

align with the count pattern depicted by the actual count data (represented by the blue wave in Fig. 5). The orange wave, with a period of 8 hours, corresponds to three off-peak periods in the time domain representation (blue wave). The green wave, with a period of 12 hours, matches two peak hour periods. Lastly, the red wave, with a period of 24 hours, captures the day and night periodicity. This alignment demonstrates how the sub-functions derived from the Fourier transform effectively capture the characteristic patterns present in the actual count data.

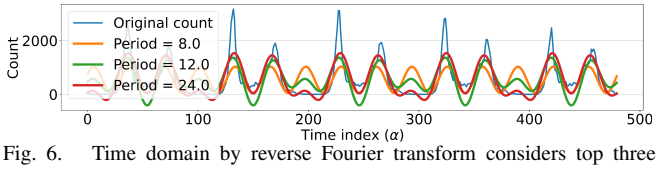

Fig. 6. Time domain by reverse Fourier transform considers top three components.

Given that the components derived from the Fourier expansion exhibit harmonic frequencies, phases, and amplitudes, we can cumulatively combine them to construct the desired approximate function, as shown in Fig. 6. In the time-domain representation, the yellow line represents the pattern considering only the top component (No.1 in Table I); incorporating the top two components results in the result plotted in green, while considering all three top components yields the red line in the time-domain representation.

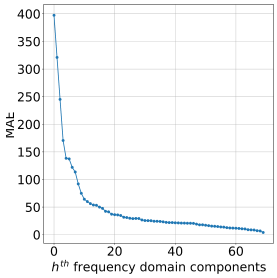

Fig. 7. Sensitivity test for modelling using various frequency components, reflected by MAE.

As we incorporate additional components and transform the frequency domain back to the time domain using the inverse Fourier transform, the modelled pattern increasingly aligns with the original pattern. This is proved by results in Fig. 7.

Filtered time domain representations: Upon transforming the time-domain representation into a frequency-domain one, we have broken down the travel pattern into several periodical functions. Our next step is to determine which periodic functions should be incorporated to construct a modelled travel pattern that is not only clean in structure but also accurately reflects the tendency of patronage pattern. Therefore, we conduct a sensitivity test to evaluate the impact of different periodic functions on the model performance. The outcome of the sensitivity test is presented in Fig. 7, which extends the findings presented in Fig. 6. This figure illustrates the modelling performance ranging from incorporating the top component to including all 70 components.

In Fig. 7, we can observe that adding more component functions during the modelling process leads to a reduction in mean absolute error (MAE), which indicates an improved model performance. Notably, the curve shows a deeper slope within the range of 0 to 10 compared to that between 20 and 30 and further. This implies that the advantages gained by including less than 10 components are insufficient while including more than 30 components appears to be unnecessary. So the ideal range should be between 10 to 20.

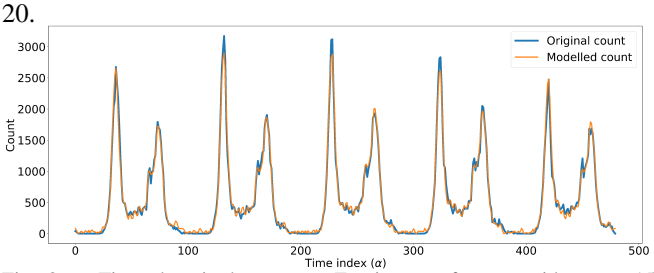

Fig. 8. Time domain by reverse Fourier transform considers top 15 components.

Consequently, at this stage, we consider the repetitive components and their meaning in the real world to determine whether the frequency/period should be included, as the explanation for Table I. And we decided to model the PT patronage pattern by using the top 15 components while excluding the remaining components. To showcase the efficacy of our modelling approach, we employ the inverse Fourier transform to generate a time-domain representation of the model, as depicted in Fig. 8.

Nevertheless, given that the primary objective of this paper is to assess the impacts of disruptions on the PT patronage pattern, we must consider if the modelled typical profile can reflect the impacts effectively. To this end, we enhance the sensitivity test by including a similarity test between the modelled typical and incident profiles; the detailed explanation can be found in Section IV-C.2.

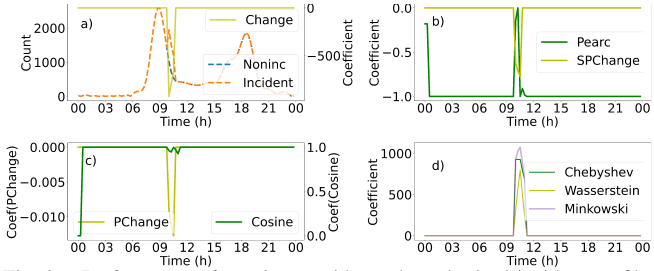

Fig. 9. Performance of metrics considers a hypothesised incident profile.

C. Incident impact measurement

1) Measure the impact of the hypothesised incident: Through the application of a hypothesized incident, we are

able to control the level of noise (variables), enabling us to assess the performance of the metric directly.

The presented Fig. 9 illustrates the outcomes of all metrics (as shown in Section II-C) employed to evaluate the resemblance between the profile of the modelled typical day (using 15^{th} significant periodic components) and the incident day. In this figure, we can observe that all metrics perform well when identifying the change in patronage.

2) Measure impact using real incidents: However, when measuring the change using data with noises (measuring the impact using real incident and modelled typical profile), the count change Fig. 10-a) and symmetric percentage change Fig. 10-b) follow the flow of the peak and off-peak hours; the percentage change, Cosine similarity Fig. 10-c) and Pearson's correlation Fig. 10-b) prove to be less effective as they closely align with the large percentage change. In contrast, distance measurements in Fig. 10-d) exhibit better performance. Out of the Chebyshev distance, Wasserstein distance, and Minkowski difference metrics, the Chebyshev distance yields a more straightforward shape that effectively illustrates the observed trend. Consequently, we solely present the results utilizing the Chebyshev distance in the subsequent visual representations.

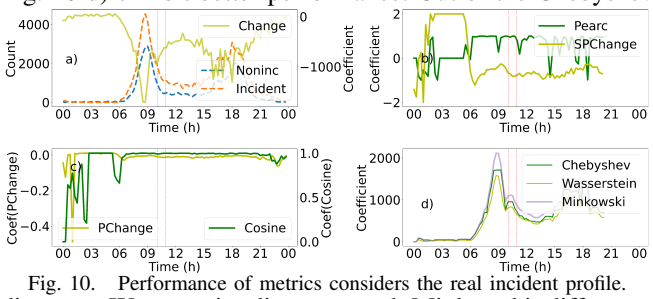

Fig. 10. Performance of metrics considers the real incident profile.

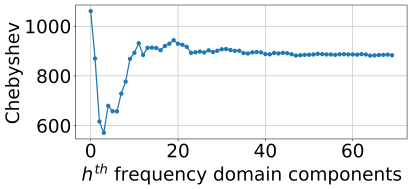

Fig. 11. Enhanced sensitivity test for modelling using various frequency components, reflected by Chebyshev distance.

As mentioned in the last paragraph of Section IV-B, it is crucial to assess the ability of the modelled profile to capture the impact accurately. This approach enables us to determine the optimal combination of components that can effectively represent the PT patronage pattern for a typical day so that we are able to encapsulate the influence of a real-life incident effectively.

The result is displayed in Fig. 11 following metrics of Chebyshev distance. It is apparent that incorporating the top three components yields the most satisfactory outcome. However, upon visualising the measurement results over time, as depicted in Fig. 12-a), we notice substantial disparities during peak hours in the morning and afternoon. Thus, solely accounting for the similarity between an ordinary day and an incident day might not be adequate to best encapsulate the disruption's effect on the PT patronage pattern; this implies that the model generated using the top three components may not necessarily be the most optimal choice for impact identification purposes.

Fig. 12 displays a series of snapshots that show the performance of identifying the disruption impacts. Each snapshot indicates the performance of adding a different number of frequency domain components when modelling; the model performance is reflected by the Chebyshev distance (green line). For example, Fig. 12-a) shows the result when adding the top three components, where the similarity of a typical (blue line) and incident (orange line) day is evaluated by the Chebyshev distance. The red dash lines highlighted the incident duration. However, it is difficult to discern any noticeable differences during the incident in this sub-figure. Instead, the peak measurements are predominantly evident during the morning and afternoon peak hours.

When we include a greater number of components in the modelling of the typical day, for instance, as shown in Fig. 12-c), where the top 11 components are considered, we can observe a distinct sub-peak during the incident duration (as indicated by the two red dashed lines). However, it's important to note that the peak measurements persist during the morning and afternoon peak hours, which can be attributed to the significant flow and substantial fluctuations that typically occur during these times.

The observed performance trend in Fig. 12 aligns with the findings presented in Fig. 11. When incorporating the top three components, the model primarily captures the characteristics of the morning and afternoon peaks. However, by including the top 11 components, the model is able to capture not only the impact during peak hours but also the concurrent noise resulting from the incident. Furthermore, as more components are added, the model's performance remains relatively stable.

V. Conclusion

The proposed method in this paper aims to dynamically model the PT patronage patterns and identify the impacts of road incidents on PT users. The proposed method applies the Fourier transform to decompose complex patterns into distinct waves; this allows the dominant components of the patronage pattern to be capturable and used as the reference (typical) profile for traffic analysis. One specific application showcased in this paper is impact identification. The presence of peak hours makes it challenging to capture the current incident impacts accurately. However, through an enhanced sensitivity test that considers the performance of impact identification, we can improve the modelling ability of the typical day. This improvement enables us to capture the current impact effectively. Multiple sample incidents are tested using this method, and the results obtained are robust. However, due to word limitations, only the results of one sample incident are presented in this paper. More data analysis results can be found in supplementary material [25].

The proposed modelling method allows us to identify the optimal typical profile. As for future directions, further exploration can focus on decomposing the patterns and effectively identifying the noise generated by disruptions, such as recurring congestion or incidents. This would allow for quantifying the impacts of these disruptions by using wave functions, for example. Additionally, the model has

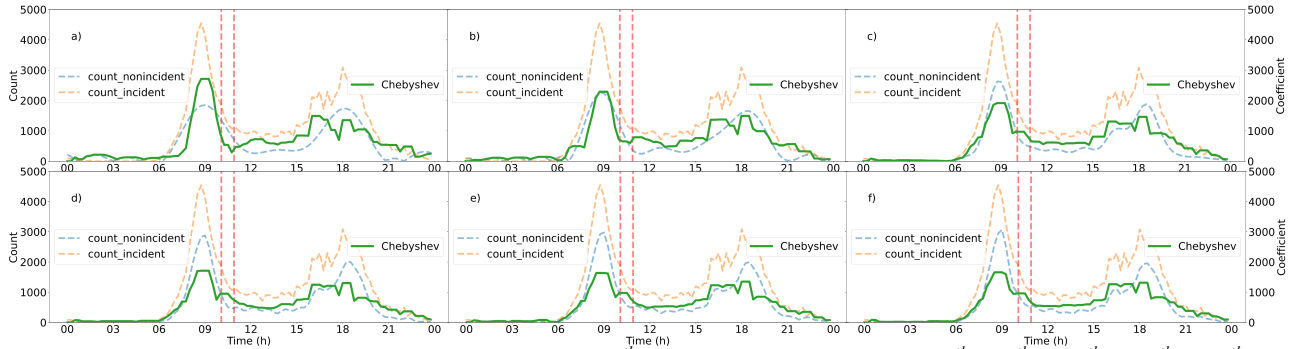

Fig. 12. Performance of real incident impact detection by including the n^{th} frequency domain components: a) 3^{th}, b) 6^{th}, c) 11^{th}, d) 15^{th}, e) 18^{th}, f) 30^{th}.

the potential to be expanded to incorporate spatial analysis. By introducing an additional spatial dimension, it becomes possible to capture the evolution of impacts based on location. This extension would provide valuable insights into the spatial dynamics of disruptions and their effects on PT patronage patterns.

ACKNOWLEDGMENTS

This work has been done as part of the ARC Linkage Project LP180100114.

REFERENCES

[1] M. Saberi, M. Ghamami, Y. Gu, M. H. S. Shojaei, and E. Fishman, "Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London," *Journal of Transport Geography*, vol. 66, pp. 154–166, jan 2018.

[2] Y. Ou, A.-S. Mihăiţă, and F. Chen, "14 - big data processing and analysis on the impact of covid-19 on public transport delay," in *Data Science for COVID-19*, U. Kose, D. Gupta, V. H. C. de Albuquerque, and A. Khanna, Eds. Academic Press, 2022, pp. 257–278. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323907699000104

[3] Y. Ou, A.-S. Mihaita, and F. Chen, "Dynamic train demand estimation and passenger assignment," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.

[4] D. Zhao, "Bus Bunching Modelling and Control: A Passenger-oriented Approach," Master Dissertation, University of Sydney, jul 2019. [Online]. Available: https://ses.library.usyd.edu.au/handle/2123/23899

[5] M. . Rahimi, Z. . Ghandeharioun, A. . Kouvelas, F. Corman, M. Rahimi, Z. Ghandeharioun, and A. Kouvelas, "Multi-modal management actions for public transport disruptions: an agent-based simulation," in *ETH*. ETH, 2021, pp. 1–7. [Online]. Available: https://doi.org/10.3929/ethz-b-000501419

[6] B. Assemi, A. Alsger, M. Moghaddam, M. Hickman, and M. Mesbah, *Improving alighting stop inference accuracy in the trip chaining method using neural networks*. Springer, 2020, no. 1.

[7] K. Saleh, A. Grigorev, and A.-S. Mihaita, "Traffic accident risk forecasting using contextual vision transformers," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 2086–2092.

[8] A. Grigorev, A.-S. Mihăiţă, K. Saleh, and M. Piccardi, "Traffic incident duration prediction via a deep learning framework for text description encoding," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 1770–1777.

[9] A. Grigorev, A.-S. Mihaita, S. Lee, and F. Chen, "Incident duration prediction using a bi-level machine learning framework with outlier removal and intra–extra joint optimisation," *Transportation Research Part C: Emerging Technologies*, vol. 141, p. 103721, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X22001589

[10] T. Wen, A.-S. Mihăiţă, H. Nguyen, C. Cai, and F. Chen, "Integrated incident decision-support using traffic simulation and data-driven models," *Transportation Research Record*, vol. 2672, no. 42, pp. 247–256, 2018. [Online]. Available: https://doi.org/10.1177/0361198118782270

[11] A. Grigorev, A. Mihaita, S. K, and M. Picardi, "Traffic incident duration prediction via a deep learning framework for text description encoding," in *Proc. of the IEEE Intelligent Transport Systems Conference 2022, Macao, China*, 2022.

[12] T. Wen, A.-S. Mihăiţă, H. Nguyen, C. Cai, and F. Chen, "Integrated Incident Decision-Support using Traffic Simulation and Data-Driven Models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 42, pp. 247–256, dec 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0361198118782270

[13] Y. Ou, A. S. Mihaita, and F. Chen, "Dynamic Train Demand Estimation and Passenger Assignment," *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020*, sep 2020.

[14] S. Shafiei, A.-S. Mihăiţă, H. Nguyen, and C. Cai, "Integrating data-driven and simulation models to predict traffic state affected by road incidents," *https://doi.org/10.1080/19427867.2021.1916284*, 2021. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/19427867.2021.1916284

[15] A. Mihaita, Z. Liu, C. Cai, and M. Rizoiu, "Arterial incident duration prediction using a bi-level framework of extreme gradient-tree boosting," *Proc. of ITS World Congress (ITSWC 2019), Singapore*, Oct. 2019.

[16] S. Shafiei, A. Mihaita, and C. Cai, "Demand estimation and prediction for short-term traffic forecasting in existence of non-recurrent incidents," *Proc. of ITS World Congress (ITSWC 2019), Singapore*, Oct. 2019.

[17] D. Zhao, A.-S. Mihaita, Y. Ou, S. Shafiei, H. Grzybowska, K. Qin, G. Tan, M. Li, and H. Dia, "Traffic disruption modelling with mode shift in multi-modal networks," *IEEE International Conference on Intelligent Transportation*, pp. 2428–2435, nov 2022.

[18] S. Shafiei, A. Mihaita, B. Nguyen, Hoang, C. D. B, and C. Cai, "Short-term traffic prediction under non-recurrent incident conditions integrating data-driven models and traffic simulation," in *Transportation Research Board (TRB) 99th Annual Meeting, Washington D.C.*, 2020.

[19] T. Wen, A. S. Mihăiţă, H. Nguyen, and C. Cai, "Integrated incident decision support using traffic simulation and data-driven models," *Transportation Research Board - 96th Annual Meeting, Washington, D.C.*, Oct. 2018.

[20] T. Mao, A. Mihaita, and C. Cai, "Traffic signal control optimisation under severe incident conditions using genetic algorithm," *Proc. of ITS World Congress (ITSWC 2019), Singapore*, Oct. 2019.

[21] Y. Abera and D. Hailemariam, "Spatio-temporal Mobile Data Traffic Modeling Using Fourier Transform Techniques," *9th International Conference on Information and Communication Technology Convergence: ICT Convergence Powered by Smart Intelligence, ICTC 2018*, pp. 20–24, 2018.

[22] N. B. Chindanur and P. Sure, "Low-dimensional models for traffic data processing using graph fourier transform," *Computing in Science and Engineering*, vol. 20, no. 2, pp. 24–37, 2018.

[23] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: machine learning, dynamical systems, and control*. Cambridge University Press, 2021.

[24] Australian Bureau of Statistics, "Australian Bureau of Statistics," 2021. [Online]. Available: https://www.abs.gov.au/

[25] D. Zhao, A.-S. Mihăiţă, Y. Ou, S. Shafiei, and H. Grzybowska, "Appendix: Modelling public transport disruptions and impact using smart-card data," 2023. [Online]. Available: https://www.researchgate.net/publication/371124886_APPENDIX_Modelling_public_transport_disruptions_and_impact_using_smart-card_data