Peak Spreading Forecast in Urban Rail Transit Demand

Ruimin Li[1], Lee Rushton[1], Matthew Jones[1]

[1] Bureau of Statistics and Analytics, Transport for New South Wales, NSW 2008

Ruimin.li@transport.nsw.gov.au

Peak spreading in the public transport system is a behavioural response to crowded trains and increased peak fares. Some passengers may shift their travel departure times to slightly before or after the peak period. Peak spreading is usually defined as a decrease in the proportion of 24-hour rail patronage during the peak hour. Understanding whether this proportion will remain constant or will change in response to these factors is critical to railway planning and timetable design.

This study collected five years of Magnetic Stripe Ticket (MST) data from 2011 to 2016 and Opal card data from its introduction to the Sydney network in 2013. The paper analyses time series Opal and MST data to identify trends in the ratio between peak and offpeak rail patronage. The analysis identified change in the proportion of 24-hour rail patronage during the AM Peak and, to a lesser extent, the PM Peak. Case studies of peak spreading patterns for two busy stations in the residential areas of Auburn and Chatswood suggest that departure time choice may be associated with socio-economic factors. The research also develops a dynamic regression model to predict peak spreading in the medium term. The results from this study could assist rail operators in policy development and planning new infrastructure.

1. Introduction

Peak spreading in the public transport system is a behavioural response to crowded trains and increased peak fares. Many commuters wishing to avoid the added delay and peak fares can choose the time they make the trip, given the trend to flexible work schedules and homeworking. Analysis of the temporal resolution of patronage on public transport services is relevant to the delivery of improved transport infrastructure and services for customers.

The importance of studying peak spreading has been recognized for some time, especially in road traffic areas (Palm and Meulen, 2014, Barnes et al., 2012, Ben-Elia and Ettema, 2010, Nelson et al., 2010). It is well accepted that the increase of road congestion is usually accompanied by the spreading of demand to peak shoulders. Most of this research was focused on the investigation of the impact of peak avoidance, including congestion pricing, monetary rewards and new information services to reduce peak trips.

Currie (2011) introduced a program in the context of public transport that aimed to encourage peak rail passengers to shift to pre-peak trains by using a free fare ticket that was valid for CBD station exit before 7am in Melbourne, Australia. It was reported that after the free fare program was rolled out in 2008, train loads increased by 41% for trains arriving in the city before 7am. Liu and Charles (2013) reviewed the effects of differential fare policy on peak spreading in urban rail transit context. They carried out a comprehensive and systematic literature review of empirical evidence from national and regional data collected in the United Kingdom, the Netherlands and selected cities in Australia. A key finding was that multiple factors affect passengers' willingness to shift transit mode or time. In the research reported by Daniels and Mulley (2013), a university in Sydney was selected to study the 'paradox' of public transport spreading. Their paper suggests that messages about travel demand should be communicated effectively to students and staff, who might be able to make better decisions in their choice of transport mode.

With more diverse data becoming available in the past decade, a deeper understanding of peak spreading and its forecasting has become possible. The traditional four-step travel demand model and Logit modelling approaches have been widely used for modelling departure time choice (Transportation Research Center, The University of Florida, 2007, Fox et al., 2015, Tirachi, Hensher and Rose 2013). Regression based time series analysis has also been reported in the literature to predict peak spreading (Ivan and Allaire, 2001, Taylor et al., 2012). Recent studies increasingly sought novel rather than traditional methods with the help of machine learning tools. For example, Hofleitner et al. (2012) used a hybrid modelling framework to estimate and predict traffic conditions based on streaming GPS data; Chiang et al. (2011) compared neural networks and Autoregressive Integrated Moving average model (ARIMA); and Oztaysi et al. (2015) applied support vector machines.

In this research we analyzed passenger peak spreading behavior using data collected from MST and Opal cards between 2011 and 2016, and 2013 and 2016 respectively. We identified the changes in peak and off-peak travel behavior and the impact of the introduction of the Opal card in the Sydney rail transit system. We selected two stations with different socio-economic profiles, namely, Auburn and Chatswood as case studies to investigate the potential causes of different peak spreading patterns. Results of the analysis were applied to peak spreading predictions, with peak spreading indicated as the change in offpeak proportions. The prediction model was developed based on a dynamic regression model by incorporating ARIMA and independent series. We estimated the model with a training dataset and tested it on a validation dataset. Error measures were calculated as validation metrics. The results demonstrate that the proposed model is a reliable tool for peak spreading prediction.

The paper is organized as follows. Section 2 introduces the datasets used in the study; Sections 3 and 4 explain the nature of available data, and present two case studies and a comprehensive investigation into peak spreading behavior; Section 5 introduces a peak-spreading prediction model based on a dynamic regression model; and Section 6 concludes the paper.

2. Data collection

The NSW government launched Opal, a smartcard ticketing system, in June 2013. The Opal card is valid on all public transport in the Sydney metropolitan area. The new ticketing system was progressively rolled out from 2013. The take-up rate reached 95% by mid-2016. Over two million Opal transactions are collected on an average weekday. The dataset contains a wide range of public transport-related data, including journeys, sales and transactions. The implementation of the Opal card system provides a rich dataset from which to derive insights that can help inform improved delivery of infrastructure and services.

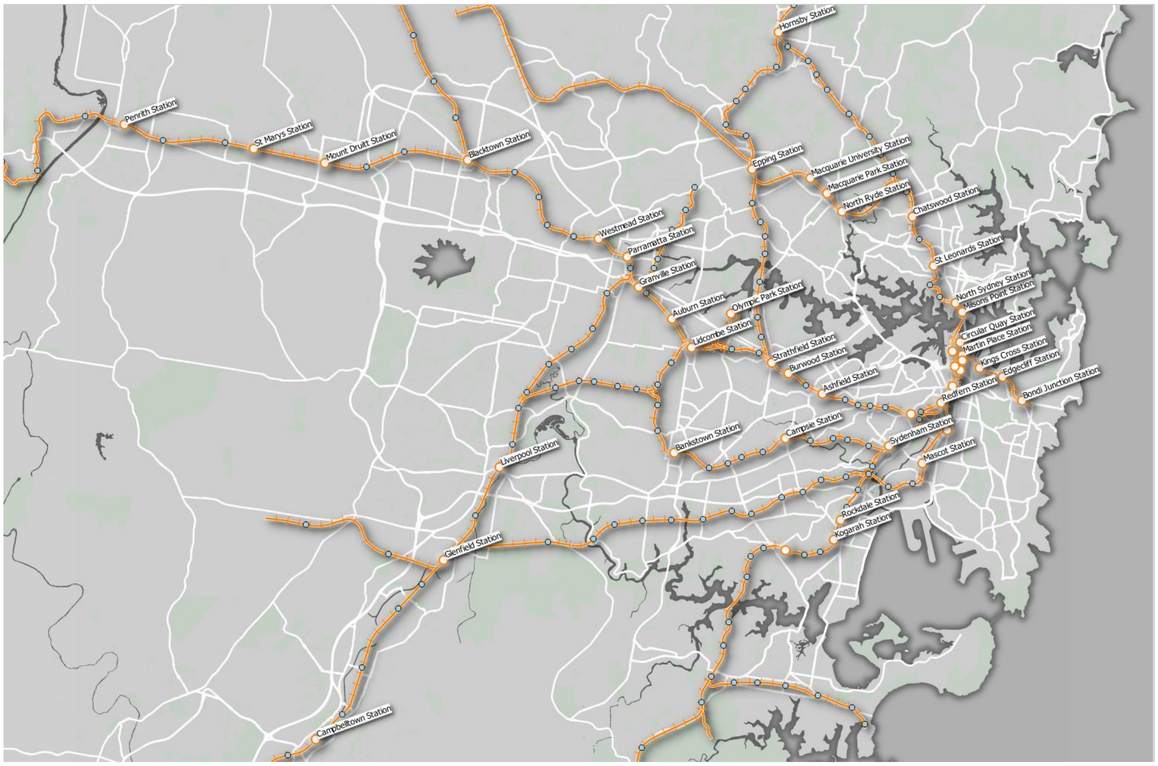
The following data were used in the analysis for all stations with barriers in the Sydney Trains network:

- MST entries and exits from March 2011 to February 2016
- Opal card entries and exits from August 2013 to February 2016.

Only stations with barriers were analyzed in this study for 24-hour reporting of patronage (48 out of more than 300 stations). Figure 1 shows the map of Sydney Trains gated stations, which represent the busiest stations across the Sydney Trains network. Opal card and MST entries and exits were aggregated into 15-minute time intervals. Median values of the proportions of time bands for each weekday in each calendar year were calculated using the following time band definitions for the AM and PM Peak:

- 4-hour AM Peak (6-10am)
- Interpeak (10am -3pm)
- 4-hour PM Peak (3-7pm)
- 11-hour Night (7pm-6am)

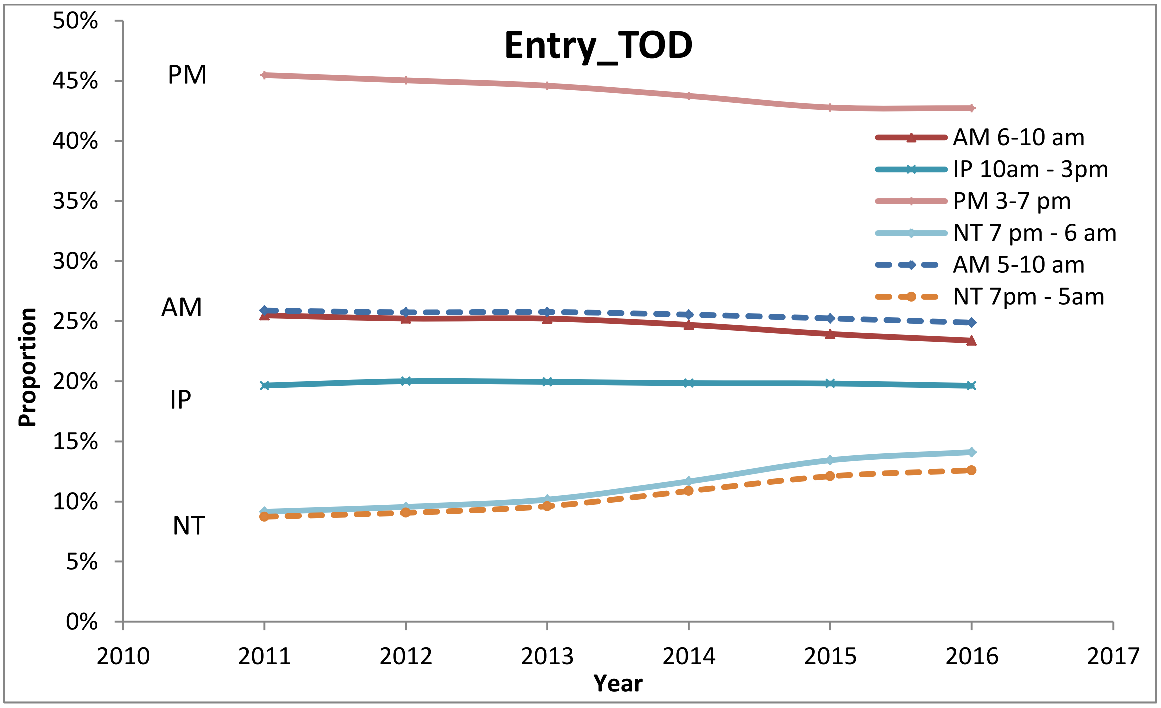
Figure 1: Sydney Trains gated stations

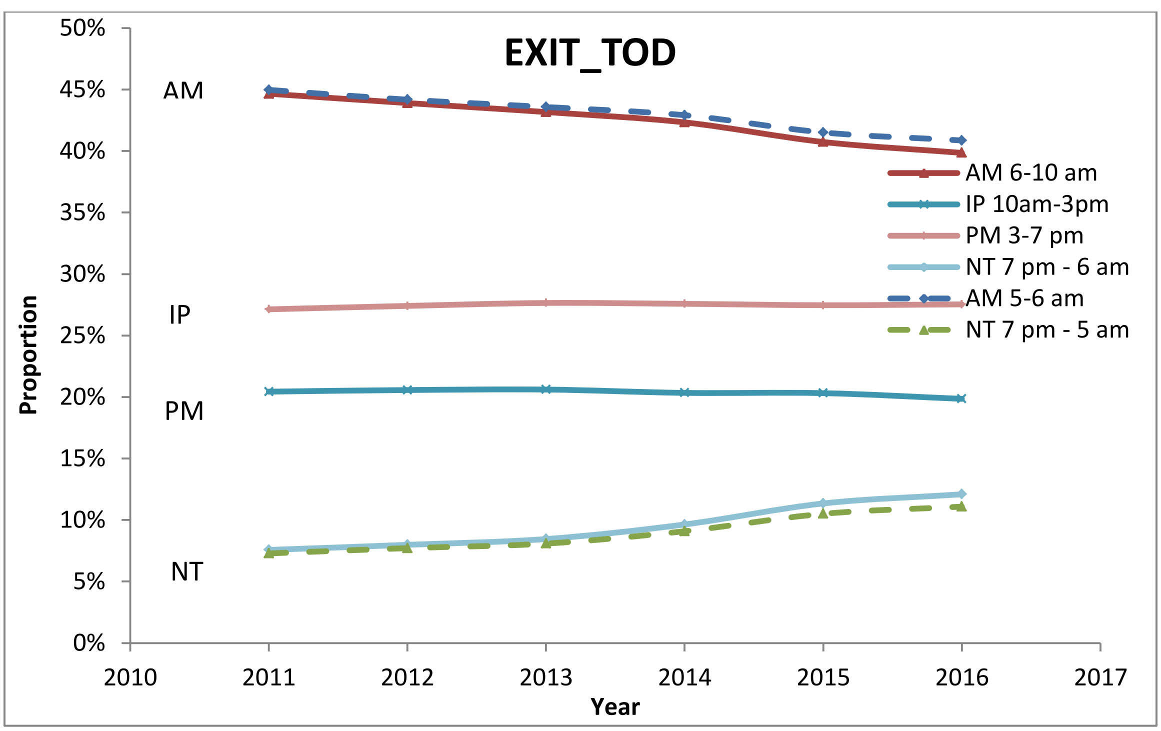


3. Peak spreading analysis

This section presents the results of patronage data analysis to provide insights into the peak spreading behavior and travel patterns of rail passengers in Sydney. The share of each time band in weekdays was calculated by month and by year for all gated stations using MST and Opal card entry/exit data,. Figure 2 shows the proportions of entries and exits by time periods from 2011 to 2016. It can be seen that the share of growth in trips in the Interpeak remained relatively steady for both entries and exits. The share of the AM Peak dropped from 45% to 41% from 2011 to 2016 for passengers entering stations. Exits experienced a similar reduction in AM Peak share. There was a marginal decrease in the PM Peak share for entries, while the share remained steady at 27% for exits. It can also be seen that the share of Night-time travel grew faster off a low base. We then extended the AM Peak from 6 -10am to 5-10am. The analysis in Figure 2 shows that the share of adjusted Night-time in trips grew at a lower rate.

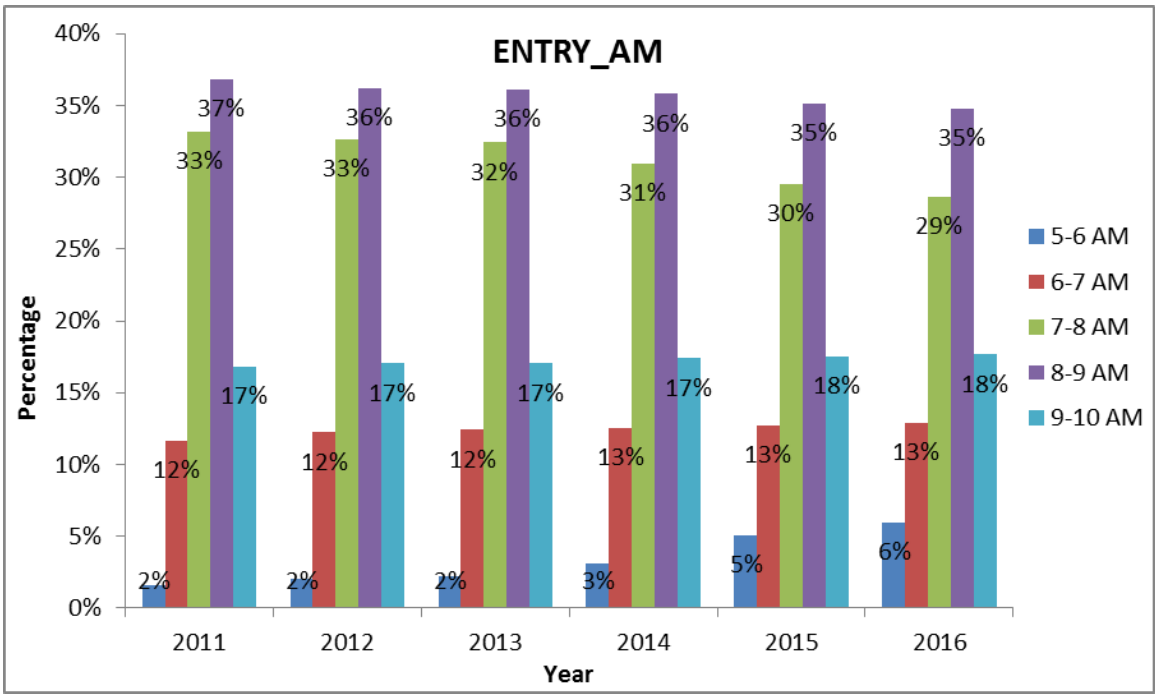
Figure 2: Proportion of time periods by entry and exit





We took a deeper dive into the proportions of entries for the peak hours and shoulder peaks from 2011 to 2016 to further investigate peak spreading behaviour. The results in Figure 3 show a 5% increase in the share of trips from 5-7am and a 7% decrease from 7-9am. The PM Peak analysis showed a marginal increase in share of trips of 3% from 6-8pm and a 2% decrease from 4-6pm. Despite the shift of patronage from peak hours to the peak shoulders, the busiest hours in the AM Peak and PM Peak were still 8-9am and 4-5pm respectively. The share of trips from 7-8am displayed the largest drop from 2011 to 2016, especially after 2013. The Opal card system provided a 30% discount on train fares for journeys in the offpeak. Standard peak hours for the Sydney Trains network area were 7-9am and 4-6.30pm on weekdays. The cheaper fares appeared to be an incentive for train passengers to leave earlier for work.

Figure 3: Proportion of entries for the AM and PM Peak by hour



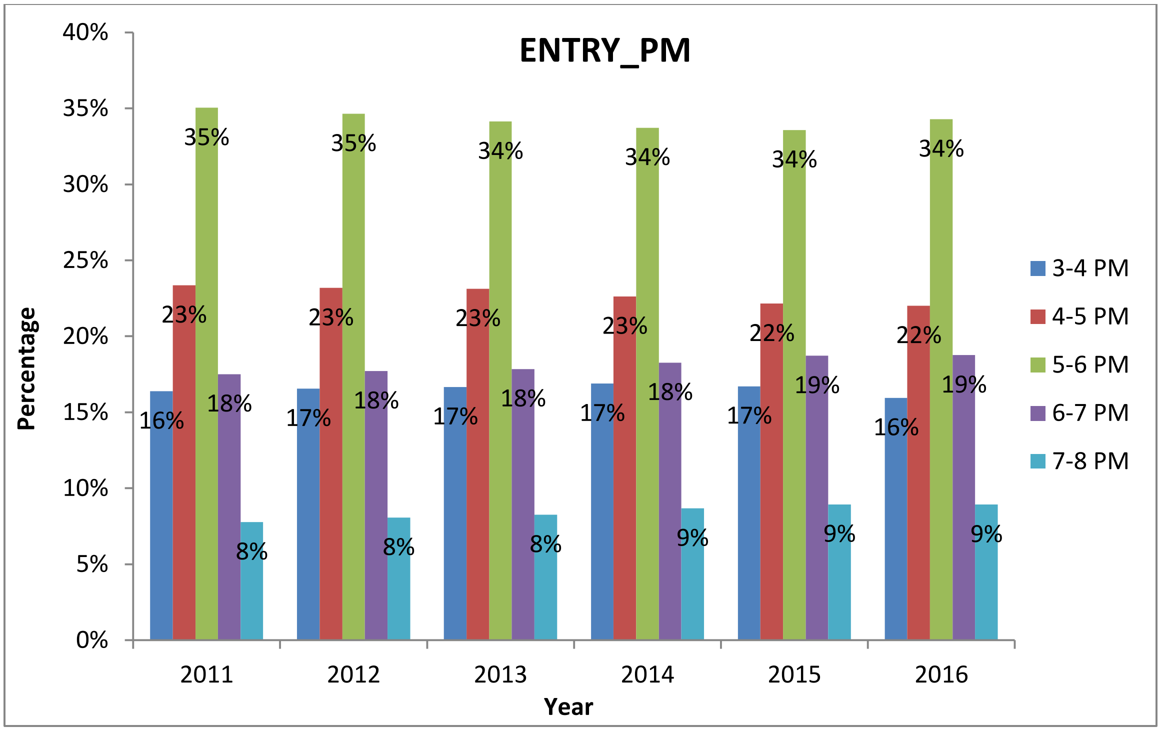
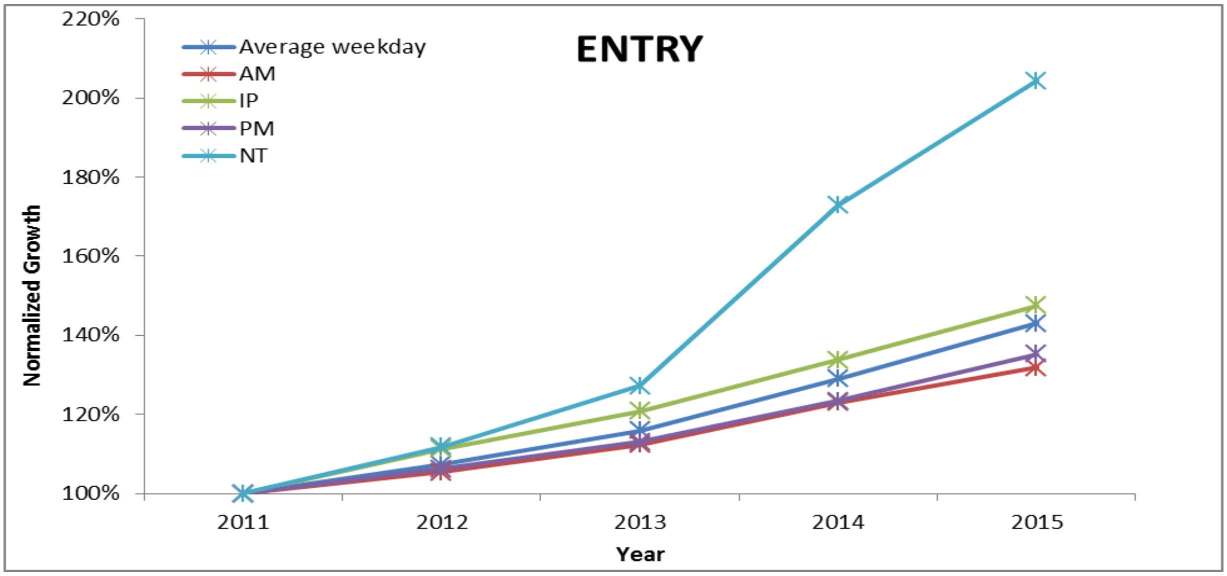
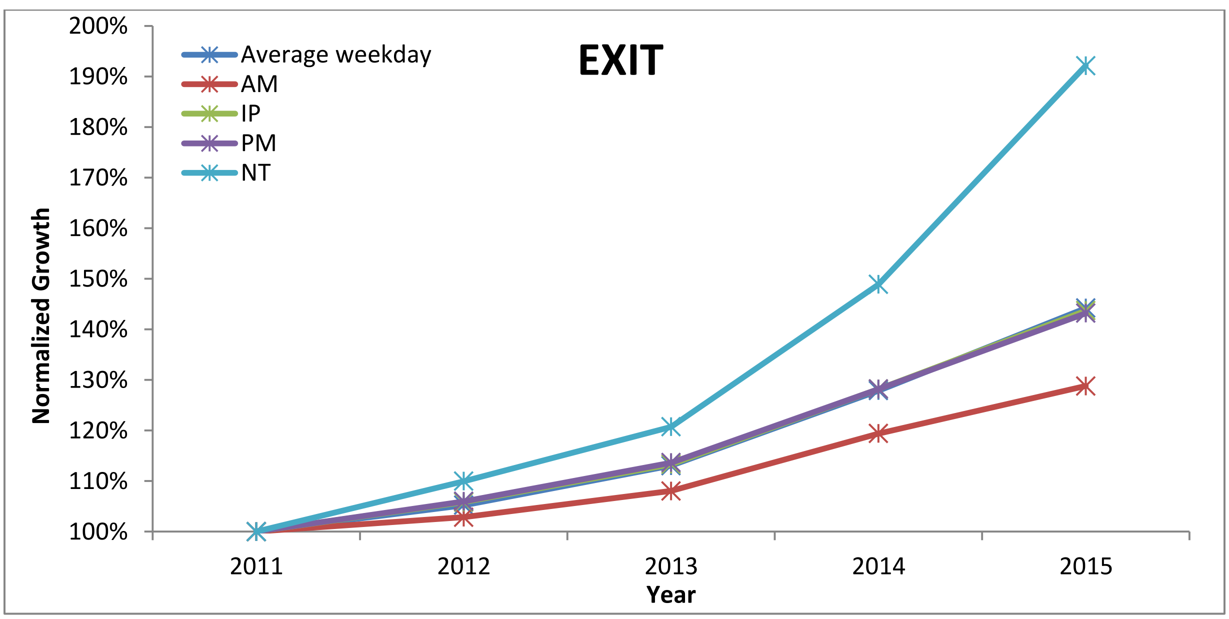


Figure 4 compares the growth rate of trips in each time period from 2011 to 2015. It can be seen that the rate of growth in Night-time trips was highest relative to other time bands, increasing by more than 100% from 2011 to 2015 for passengers entering and exiting stations with barriers. Night-time trips increased around 70% from 2013 to 2015 compared with an increase of 23% from 2011 to 2013 as a result of the rollout of the Opal system from 2013. The rate of growth in trips was lowest for the AM Peak relative to other time periods, an increase of 30% over five years.

Figure 4: Normalized growth rate by entry and exit





4. A case study

The analysis in the preceding section was based on total entries/exits for all gated stations. This section analyses hourly shares in peak and peak shoulders for selected stations in the network to further explore peak spreading characteristics. Auburn and Chatswood stations were selected for the analysis because of their varying demographic characteristics. Figure 5 shows the locations of the two stations. Chatswood Station is located in the Lower North Shore of Sydney, while Auburn Station is located west of the Sydney Central Business District (CBD). According to Australian Bureau of Statistics (ABS) 2011 census data, the median weekly income of households with children in Auburn is $2,135 compared with $1,988 for households without children. In Chatswood, the median household income is $3,341 and $2,987 respectively for households with and without children.

Figure 5: Maps of Auburn and Chatswood

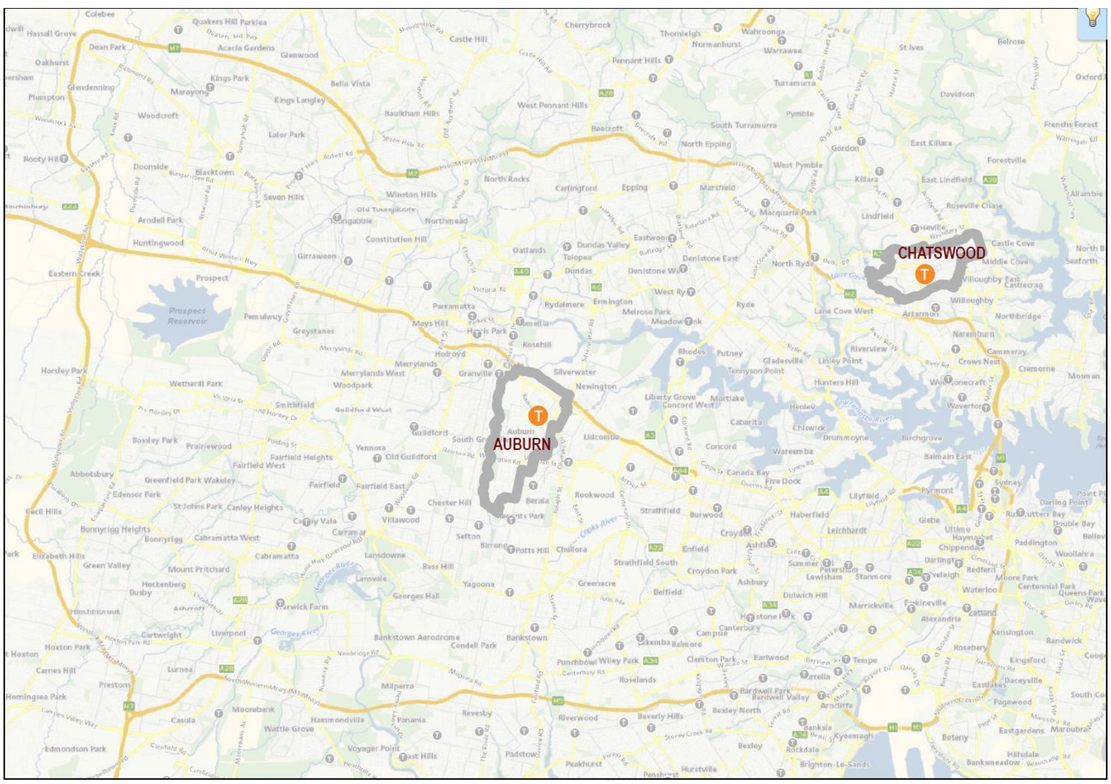
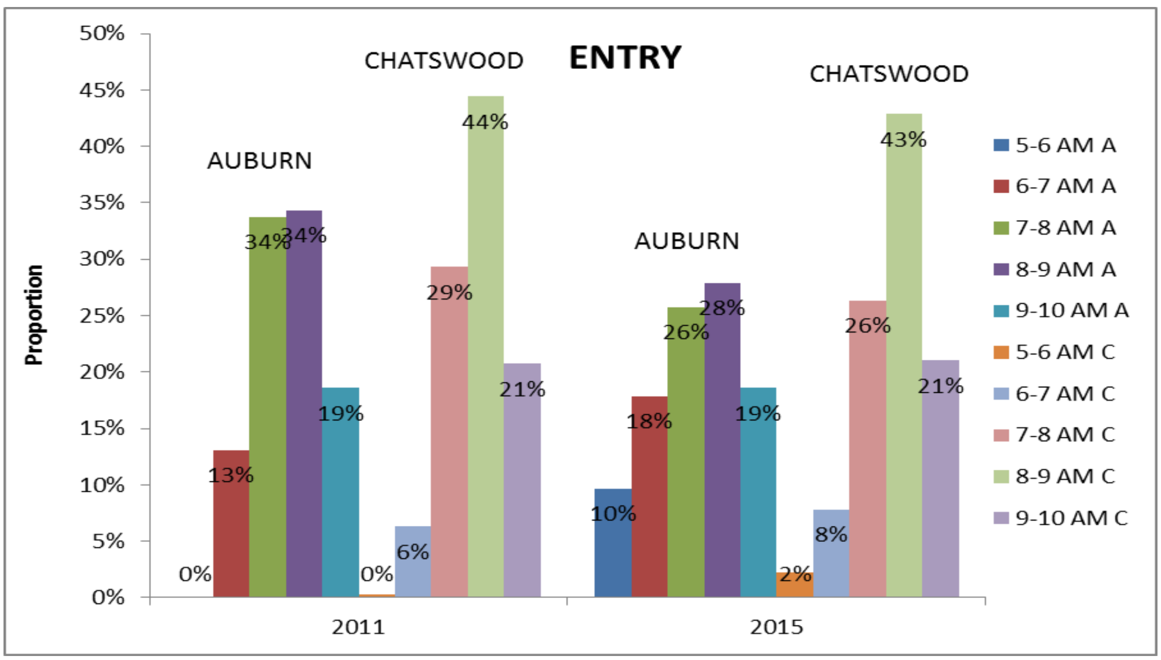


Table 1 shows the total median number of entries for Auburn and Chatswood stations between 5-10am in 2011 and 2015 on a weekday. Patronage of Chatswood Station in the study period grew at a faster rate than at Auburn Station between 2011 and 2015. The peak spreading patterns in Figure 6 show that the extent of peak spreading is not consistent across the network. From 2011 to 2015 there was a 15% increase in share of trips from 5-7am for entries and a 14% decrease from 7-9am at Auburn Station. At Chatswood Station there was a 4% increase and a decrease in the share of trips from 5-7am and 7-9am respectively for entries. These results indicate that departure time choice may be associated with socio-economic status. Further analysis found that spreading to the peak shoulder at Auburn Station grew significantly faster following the rollout of the Opal system in 2013. For example, the share of trips from 5-6 am at Auburn Station increased from 3% in 2014 to 13% in 2016. This suggests that commuters from catchment suburbs with lower household incomes are more likely to change their departure time in response to cheaper fares in the offpeak.

Table 1: Total number of entries for Auburn and Chatswood stations in 2011 and 2015 between 5-10am on a weekday (median)

Station	2011	2015
Auburn	3,381	4,063
Chatswood	4,410	6,057

Figure 6: Peak spreading pattern comparison between Auburn and Chatswood stations



5. Peak spreading forecasting methodology

Reliable forecasts of train patronage by peak and offpeak periods are a key input for operators to plan services and allocate resources. This section introduces methodologies for forecasting peak spreading in rail patronage. Peak spreading is indicated as a change in offpeak proportions, with ‘offpeak’ defined as the time periods 10am-3pm and 7pm-6am. Figure 7 shows the percentage growth in offpeak proportions across years for gated stations in the Sydney Trains network. The analysis suggests that the monthly offpeak share is impacted by seasonality, as the offpeak share is the highest in December, with other months remaining relatively flat. This might be partially explained by the holiday effect, in which passengers are more likely to have flexible work arrangements during the holiday season. Figure 7 also shows that the offpeak share grew faster after the rollout of the Opal card system because of cheaper fares in the offpeak period.

Figure 7: Percentage growth in offpeak proportion (de-seasonalised)

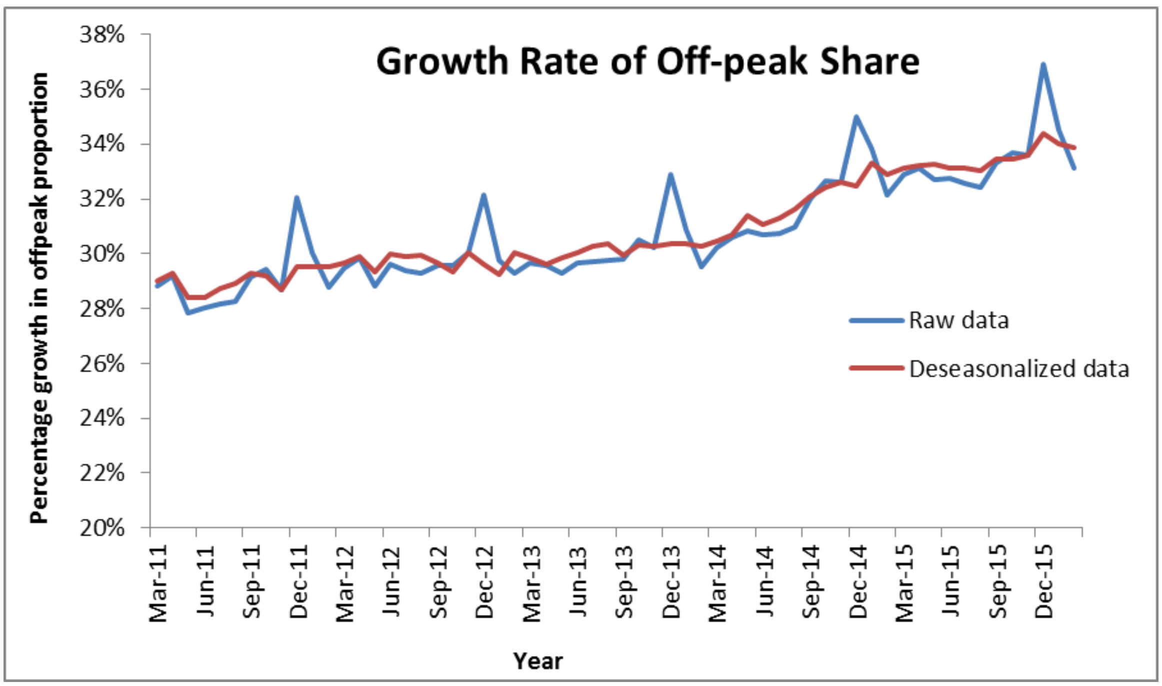


Table 2 shows the correlations of the de-seasonalised offpeak shares with past values at different lags. It indicates there is a strong correlation between the current and previous values of peak spreading. The autocorrelation factor decays very slowly, as can be observed from the results in Table 2, which suggest that the series is not stationary. The correlation test also confirms that the rate of growth in the offpeak trip share is strongly correlated with the Opal take-up rate (0.96). These results suggest that a time series model which allows for the inclusion of other relevant independent variables could be used to predict peak spreading. This kind of model is a dynamic regression model that combines an Autoregressive Integrated Moving Average Model (ARIMA) and a general regression model. It analyses the correlations between current and lagged values of responses, independent series and errors to accurately make predictions.

Table 2: Autocorrelations

Lag	0	1	2	3	4	5	6
Correlation	1	0.94	0.89	0.83	0.77	0.72	0.67

The basic model can be written as follows:

$Y^* = \mu + \sum_i \beta_i X_{t-i} + \phi_1 Y^*_{t-1} + \phi_2 Y^*_{t-2} + \ldots + \phi_p Y^*_{t-p} + \Theta_1 \mathcal{E}_{t-1} + \Theta_2 \mathcal{E}_{t-2} + \cdots + \Theta_q \mathcal{E}_{t-q} + \mathcal{E}_t$ (5.1)

Where: Y^* = Differenced series from non-stationary series Y for stationarity

$\mathcal{E}$ = Mean term

X = Independent series, in this case, Opal take-up rate

p = The order of the autoregressive part

$\mathcal{E}$ = Error term

q = The order of the moving average process

It should be noted that the independent series, X in Equation 5.1, can also be lagged and differenced.

Simple differencing can be made using the following equation:

$Y^*_t = (1 - B)^d Y_t$ (5.2)

Where d is the order of differencing and B is the backshift operator; that is BY_t = Y_{t-1}.

In applying the dynamic regression model in Equation 5.1 to the observed offpeak spreading, the time series was differenced at lag 1 and 12 because the series was not stationary and showed strong seasonality. That is, instead of modelling the offpeak share series itself, the change in offpeak share was modelled from one period to the next and the change 12 periods before. A total of 60 observations were available for this study. We used 50 observations for the training dataset and 10 observations for the validation set. Candidate models were compared and the final model structure was identified based on the Goodness-of-Fit statistics. The model was developed according to the following steps:

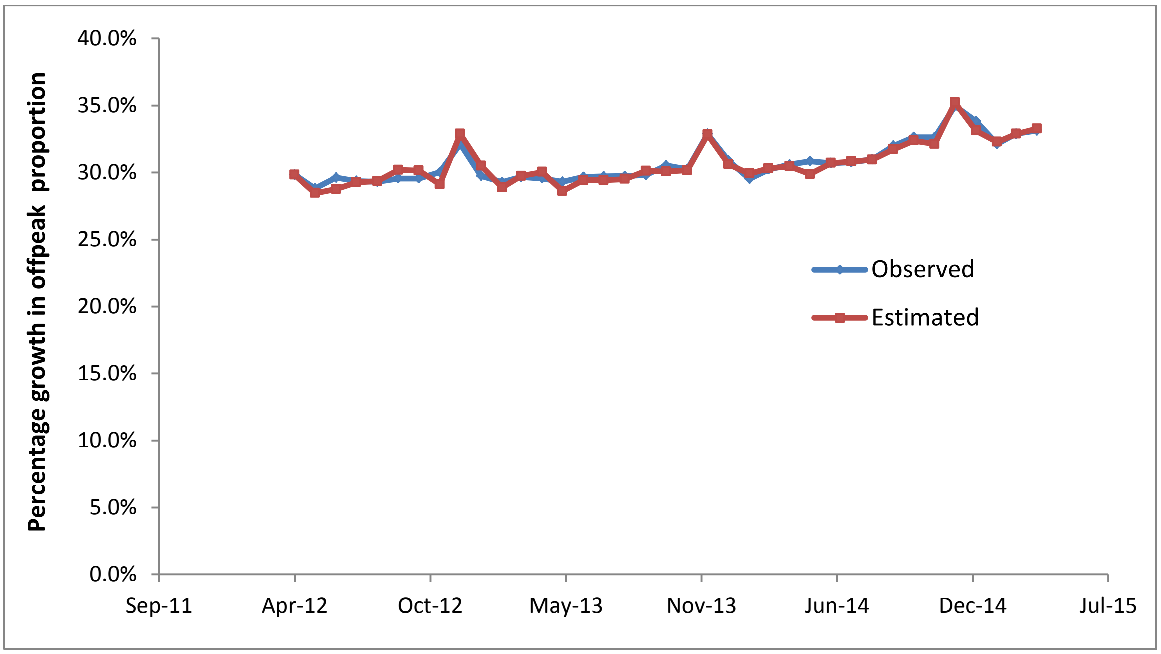
1. The dependent peak spreading series, Y, was differenced at lag 1 and lag 12.

2. The independent peak spreading series, X, was differenced at lag 1.

3. The following processes were used to fit the twice-differenced predictor series: (i) a first-order autoregression process combined with a seasonal autoregression process with lag 12; (ii) a first-order moving average model combined with a seasonal moving average model with lag 12 (ARMA $(1,1)(1,1)_{12}$); and (iii) the differenced independent series.

Figure 8 compares the observed and estimated offpeak shares. The pattern shows that the proposed model appears to fit the observed offpeak shares quite well. We also quantitatively evaluated the prediction results using measure of mean absolute error (MAE) and mean absolute relative error (MARE). The MAE value of 0.007 and MARE value of 2.2% also suggest that the proposed dynamic regression model is a reliable tool to predict peak spreading.

Figure 8: Observed and estimated offpeak shares



6. Conclusion

Peak spreading forecasting has proved a popular research topic for many years. In the literature different methodologies were reportedly applied on wide range of data for transport modes in different regions around the world. This study provides insights into train passenger peak spreading behavior in Sydney by synthesizing MST and Opal card entry/exit data over approximately five years. The analysis shows that the gradual discontinuation of MST and rollout of the Opal card introduced a noticeable trend to passenger peak spreading. We then formulated a dynamic regression model to predict the peak spreading by incorporating the Opal card take-up rate series and the ARIMA process. The results demonstrated the reliability and effectiveness of the developed model.

References

1. Palm, H & Meulen, M 2014, '*Rewarding for avoiding the peak period in Rotterdam*', Proceeding of 93rd Annual Meeting of The Transportation Research Board, Washington, D.C.
2. Barnes, I, Deakin,E, Frick, K & Skabardonis, A 2012, ' *Impact of peak and off peak tolls on traffic in the San Francisco-Oakland Bay Bridge Corridor*', Proceeding of 91st Annual Meeting of The Transportation Research Board, Washington, D.C.
3. Ben-Elia, E & Ettema, D 2010, '*Commuters' choice-behaviour with rewards for avoiding peak-hour driving*', Proceeding of 89th Annual Meeting of The Transportation Research Board, Washington, D.C.
4. Nelson, J, Zuyeva, L, Wood, N & Guensler, R 2010, '*A sociodemographic analysis of Northeast Atlanta I-85 peak period commuters likely to be affected by implementation of value pricing along the corridor*', Proceeding of 89th Annual Meeting of The Transportation Research Board, Washington, D.C.
5. Currie, G 2011, '*Free fare incentives to shift rail demand peaks – medium term impacts*', Proceeding of 90th Annual Meeting of The Transportation Research Board, Washington, D.C.
6. Liu, Y & Charles, P2013, '*Spreading peak demand for urban rail transit through differential fare policy: A review of empirical evidence*', Australasian Transport research Forum 2013, Australia, Brisbane.
7. Daniels, R & Mulley, C2013, '*The Paradox of public transport peak spreading: universities and travel demand management*', International Journal of Sustainable Transportation, Vol. 7, pp.143–165.
8. Transportation Research Center, The University of Florida 2007, '*Development of time-of-day modeling procedures using FSUTMS powered by Cube Voyager*', http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PL/FDOT_BD545_65_rpt.pdf
9. Fox, J, Dunkerley, F, Patruni, B & Daly, A 2014,'Time period choice modeling – review of practice', http://www.rand.org/content/dam/rand/pubs/research_reports/RR1100/RR1127/RAND_RR1127.pdf
10. Tirachini, A, Hensher, D & Rose, J 2013, '*Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand'*, Transportation Research Part A, Vol. 53, pp. 36–52.
11. Ivan, J & Allaire, S 2001, '*Regional and area –type modeling of peak spreading on Connecticut Freeways*', Journal of Transportation Engineering, Vol. 127,No. 3, May/June.

12. Taylor, B, Miller,D, Iseki, H & Fink, C 2009, ‘Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas’, Transportation Research Part A, Vol. 43, pp.60–77.

13. Hofleitner, A, Herring, R & Bayen, A 2012, ‘Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning’, Transportation Research Part B, Vol. 46, pp.1097–1122.

14. Chiang, W, Russell, R & Urban, T 2011,*’Forecasting ridership for a metropolitan transit authority*’, Transportation Research Part A, Vol. 45, pp. 696–705.

15. Oztaysi,B, Yanik,S & Kahraman,C 2015,‘ *Forecasting passenger volumes in transit systems using support vector Machines: the Case of Istanbul’*, Journal of Multiple-Valued Logic & Soft Computing, Vol. 25, pp. 215–235