Pedestrian Demand Forecasting Methods Guidance

Technical Report 2: STREAMS validation

Prepared for Department of Transport and Main Roads

CDM RESEARCH

Contents

Executive Summary iii

1 Introduction 1

1.1 Purpose 1

1.2 Background 1

1.3 Methodology 2

2 Descriptive statistics 6

3 Expansion factors 15

3.1 Methodology 15

3.2 Results 15

3.3 Conclusion 19

Document history and status

Revision	Date issued	Author	Revision type
1	18/9/2020	C. Munro	Issue-1

Distribution of Copies

Revision	Media	Issued to
1	PDF	Department of Transport and Main Roads

Printed:	18 September 2020
Last saved:	17 September 2020 10:21 AM
File name:	0171 Pedestrian Demand Forecasting - STREAMS validation - Technical Report (Issue-1).docx
Project manager:	C. Munro
Name of organisation:	Department of Transport and Main Roads
Name of project:	Pedestrian Demand Forecasting Methods Guidance – Technical Report 2
Project number:	0171

Executive Summary

This technical note summarises an exercise to develop expansion factors for STREAMS signal data to estimate pedestrian demand. The exercise compared observed pedestrian demand at seven midblock signalised pedestrian crossings in Brisbane with the STREAMS data across three weekdays between 6 am and 8 pm.

The data was analysed across 15-minute and hourly bins. Models were developed using four methods where the STREAMS count was the predictor and the observed count was the outcome variable across a pooled dataset from the seven sites. The four methods were:

- Average expansion factor for each STREAMS count
- Median expansion factor for each STREAMS count
- Ordinary least squares (OLS) regression
- Quasipoisson generalised linear model (QP-GLM).

The model fit was assessed by comparing the mean absolute error (MAE) and root mean square error (RMSE) of the residuals.

The best fitting model was found to be a linear regression using OLS using 15-minute data. The model is as follows:

$$D_{15} = 2.30 \times STREAMS_{15} \qquad R^2 = 0.68$$

In other words, multiplying each 15-minute period STREAMS count by 2.30 provides the closest estimate of the true pedestrian demand when aggregated over a day or more.

The constant expansion factor, applied to the 15-minute data, produced an accuracy of the true pedestrian demand across the three days from -21% to +47%; two of the seven sites were accurate to within 10% and six were accurate to within 24%. For signalised intersections with modest demand accuracy to within +/-25% is likely.

1 Introduction

1.1 Purpose

This report describes an attempt to develop expansion factors to apply to STREAMS cycle analyser logs to estimate pedestrian demand at pedestrian operated signals. This analysis was motivated by:

- a recognition that knowledge about pedestrian demand is generally sparse, with generally limited and *ad hoc* short-period counts in many areas of Queensland, and
- that the STREAMS traffic controllers widely used on Queensland's road network automatically log actuations ("push button presses") and thus **may** provide a very low-cost indication of pedestrian demand at traffic signals.

The present analysis was designed to ascertain how reliably these STREAMS signal logs may be at predicting pedestrian crossing demand, and specifically to assess whether a single expansion factor, or time- or demand-varying factor, may be able to adjust the log "count" to obtain a reasonable indication of the true pedestrian demand.

In assessing how reliably such a factor may forecast demand it is emphasised that the objective here is not to be able to perfectly measure demand using the STREAMS logs but rather to assess how such an approach could be used to provide a reasonable and unbiased indicator of demand while avoiding the cost of having to commission manual counts.

1.2 Background

Most traffic signals in Queensland use the STREAMS integrated intelligent transport system. When a pedestrian calls the pedestrian phase at a STREAMS-enabled intersection by pressing the pedestrian push button, the time at which this occurs is recorded. This time, along with much other data about the operation of the signals, is recorded in a log which can subsequently be queried using the Intersection Cycle Analyser (ICA) application.

The frequency of pedestrian phase call-ups within these logs will be at least weakly correlated with the pedestrian demand; if there are no call-ups in the log there would be expected to be fewer pedestrians than during periods with a high number of call-ups. However, the correlation will not be perfect and nor will there be a one-to-one correlation between the logs and true pedestrian demand:

- The ICA records only the *first* pedestrian call-up within a phase; if a pedestrian arrives at the intersection during a red pedestrian phase at 11:32.48 and presses the button and then another arrives at 11:33.03 and also presses the button only the first call at 11:32.48 will be recorded.
- The logs provide no insight into *how many* pedestrians are waiting to cross in a cycle – groups of pedestrians, and cycles where there are more than one pedestrian waiting to cross (on either side of the street) will count only as *one* event in the log.

- For single stage crossings there will be no indication of direction of travel – if a pedestrian arrives and calls the signal at 11:32.48 heading north and another arrives at 11:33.03 heading south (i.e. are on the opposite side of the crossing) the log only records that the pedestrian phase was called at 11:32.48.
- Where a pedestrian arrives at or near the intersection, observes no vehicles (or a gap in traffic) and chooses to cross against the pedestrian signal and without pressing the button, no record of their presence will be recorded within the log.
- The logs can provide no insight into pedestrian demand when the pedestrian phase is automatically initiated during every cycle irrespective of whether the phase has been called by a pedestrian.

In combination these limitations (aside from auto-initialisation) will result in the logs significantly *underestimating* pedestrian demand. The degree by which the demand is underestimated is likely to vary markedly depending on the factors such as the overall level of demand *and* the time-of-day variation of this demand. Locations with high peak period demands and modest demands at other times, such as signals located near schools, are likely to exhibit far more severe undercounting at peak times than at other times. In effect, at very high pedestrian demand the STREAMS log will saturate at one call per cycle, which for a 90 second cycle time would mean a maximum of around 40 calls per hour.

While the focus in this analysis is on the potential to use STREAMS to estimate pedestrian demand, the logs also provide insight into how often the pedestrian phase is called. This can be useful when used as a comparison method to predict the frequency of pedestrian calls; this can be useful as an input to intersection modelling for the purpose of capacity analysis.

1.3 Methodology

In order to assess the correlation between the STREAMS logs and true pedestrian demand an analysis was undertaken over three typical non-holiday mid-week (Tuesday – Thursday) days at eight midblock pedestrian operated signals in Brisbane (Table 1.1). The sites were selected on the basis that they:

- are managed by TMR, thereby simplifying access to the STREAMS logs, and
- intuitively seem to represent a representative mix of sites with vary levels and types of demand.

A video camera was positioned near each crossing and a count of <u>all</u> users[1] crossing in the immediate vicinity[2] of the intersection between 6 am and 8 pm. The sites were constrained to be within Brisbane to simplify the logistics of camera installation and removal.

[1] This includes pedestrians, children sitting in strollers or prams, bicycle riders, those on skateboards or scooters and using mobility aids such as wheelchairs and mobility scooters.

[2] This was defined as those who started and/or completed their crossing of the road within the designated crossing. This will undercount "true" pedestrian crossing demand given that some pedestrians will cross entirely outside of the designated crossing. In practice, at the sites selected for this analysis the proportion crossing entirely outside the crossing within the camera field of view was negligible.

Counts were obtained in 15-minute periods (bins) from the video record by direction of travel. The timestamp assigned to each pedestrian from the video was the time at which they *commenced* crossing the road and tends to lag somewhat behind the timestamp in the signal log (i.e. when the push button was pressed). The practical significance of this difference when aggregated to 15-minute bins is very small given that very few pedestrians arrived and then crossed at the threshold between each bin.

■ **Table 1.1: Site descriptions**

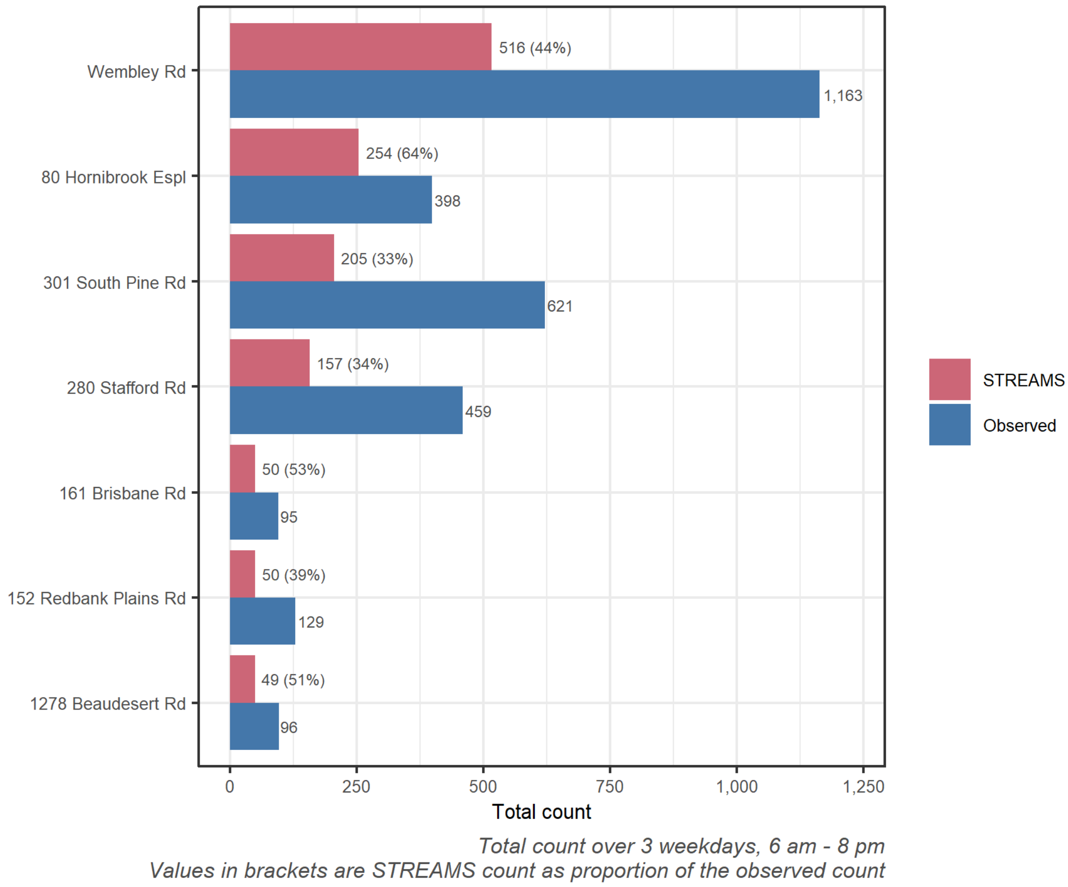
Site	Description
80 Hornibrook Esplanade, Clontarf 	Likely to be almost exclusively recreational use for access to/from foreshore across 4-lane arterial road and residential area.
280 Stafford Road, Stafford 	Crossing of 4-lane arterial road near minor retail precinct, school and suburban residential area.

301 South Pine Road, Enoggera 	Crossing of 4-lane arterial road near school and suburban residential area.
152 Redbank Plains Road, Bellbird Park 	Crossing of 4-lane arterial road near church and low density residential area.
1278 Beaudesert Road, Acacia Ridge 	Crossing of 4-lane arterial road near school and suburban residential area.

161 Brisbane Road, Booval 	Crossing of 4-lane arterial road near Cothill Street. Outer suburban residential area with big box retail precinct.
Wembley Road, Logan 	Two-stage crossing of 4-lane arterial road at entry to Logan Central Shopping Centre.

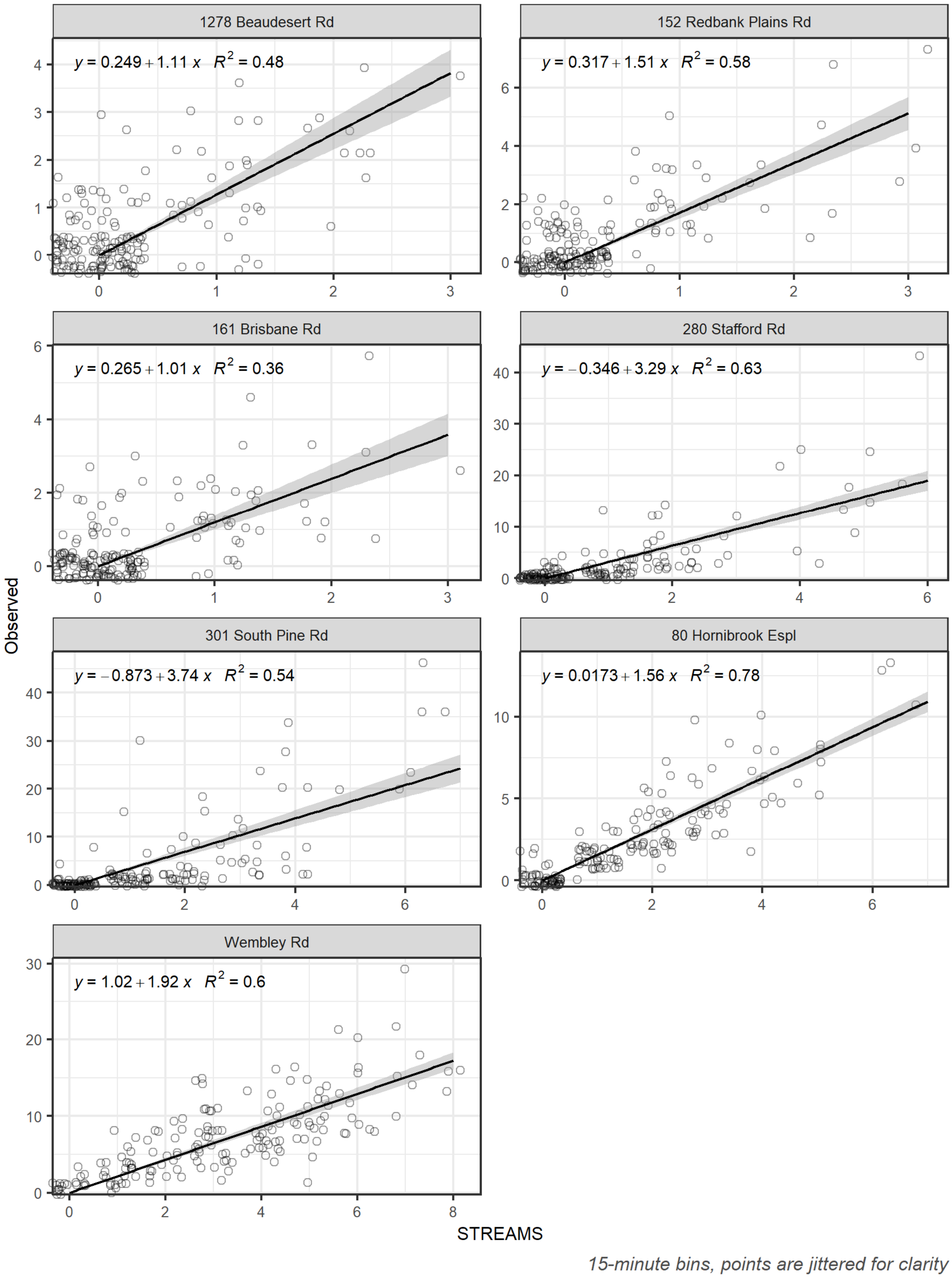
2 Descriptive statistics

The total pedestrian count observed from the video is compared to the number of pedestrian calls in the STREAMS logs over the three days in Figure 2.1. The number of pedestrian calls varied from 34% at 280 Stafford Road to 95% at 161 Brisbane Road, implying an expansion factor from 1.6 to 3.0. A naïve overall expansion factor in this range would risk severely over- or underestimating true pedestrian demand. In other words, a single aggregate expansion factor does not seem to be appropriate. Instead, a factor that adjusts smaller period counts is warranted.

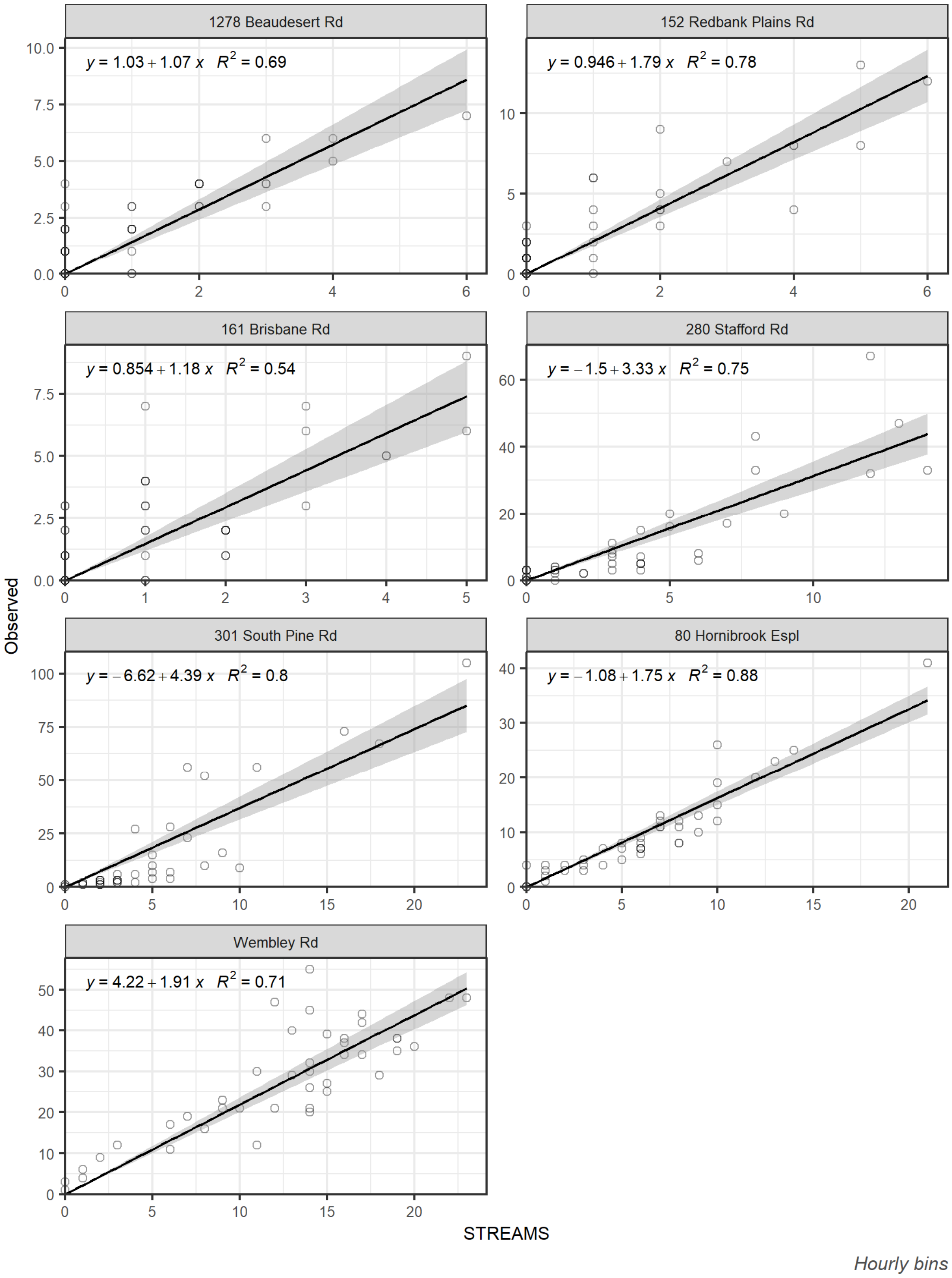


■ **Figure 2.1: Total count comparison**

The correlation between 15-minute and hourly binned counts is shown in Figure 2.2 and Figure 2.3. The coefficient of determination (R^2) is generally higher for the hourly binned data but varies from modest correlation (e.g. 161 Brisbane Road, with an R^2 of 0.54) to strong correlation (e.g. 80 Hornibrook Esplanade, R^2 of 0.88). However, the 15-minute data is especially affected by influential outliers such that the predictive validity will be lower than the R^2 would imply, particularly during low demand periods.

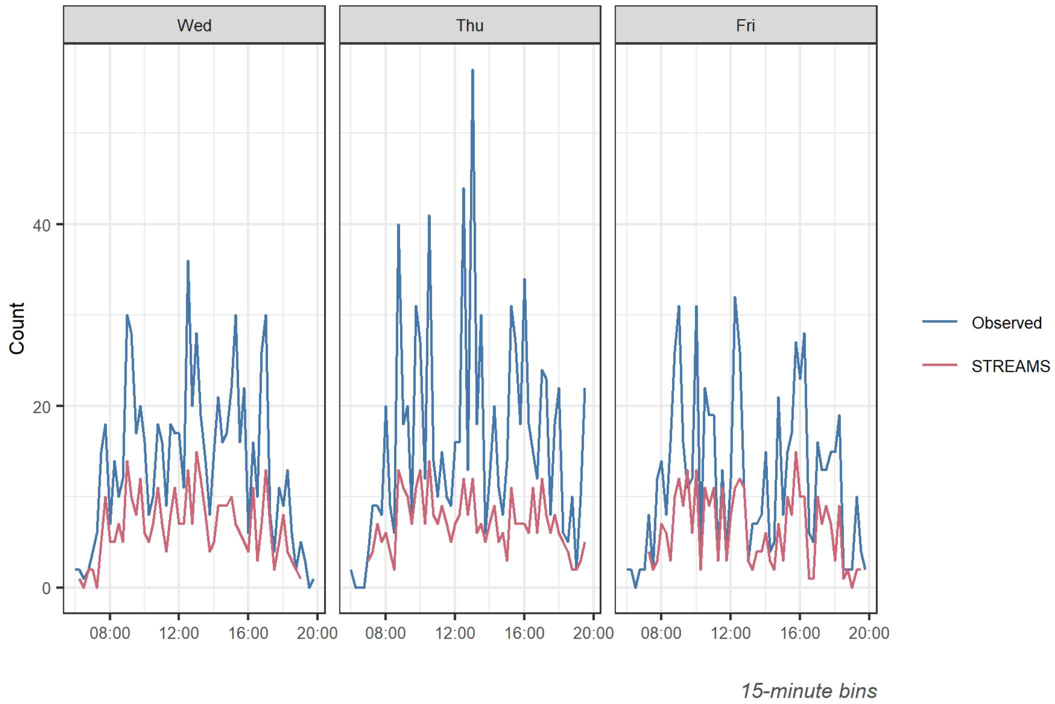


- **Figure 2.2: Relationship between STREAMS counts and observed counts using 15-minute data (line is linear regression and shaded area is 95% confidence interval)**



■ **Figure 2.3: Relationship between STREAMS counts and observed counts using hourly data (line is linear regression and shaded area is 95% confidence interval)**

The time-of-day profiles for each site are shown in Figure 2.4 to Figure 2.10. There is at most a weak trend between increasing pedestrian demand and STREAMS call-ups (Figure 2.11). The maximum number of STREAMS call-ups in a 15-minute period was 15 calls, or around one per minute.



■ **Figure 2.4: Time-of-day comparison - Wembley Road**

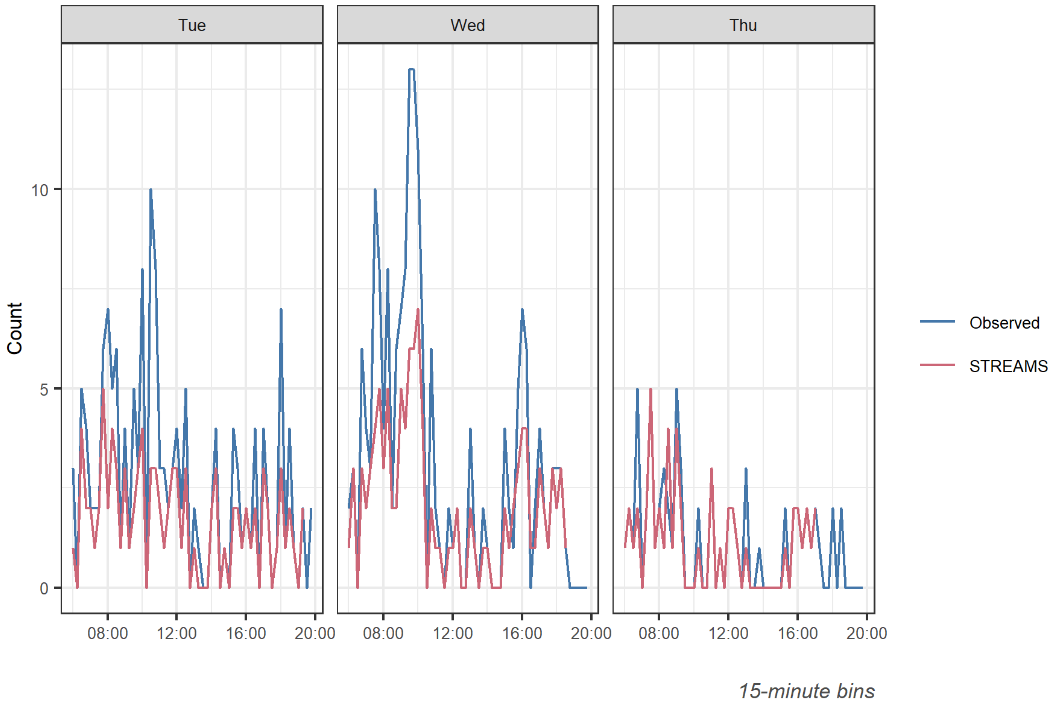


Figure 2.5: Time-of-day comparison - 80 Hornibrook Esplanade

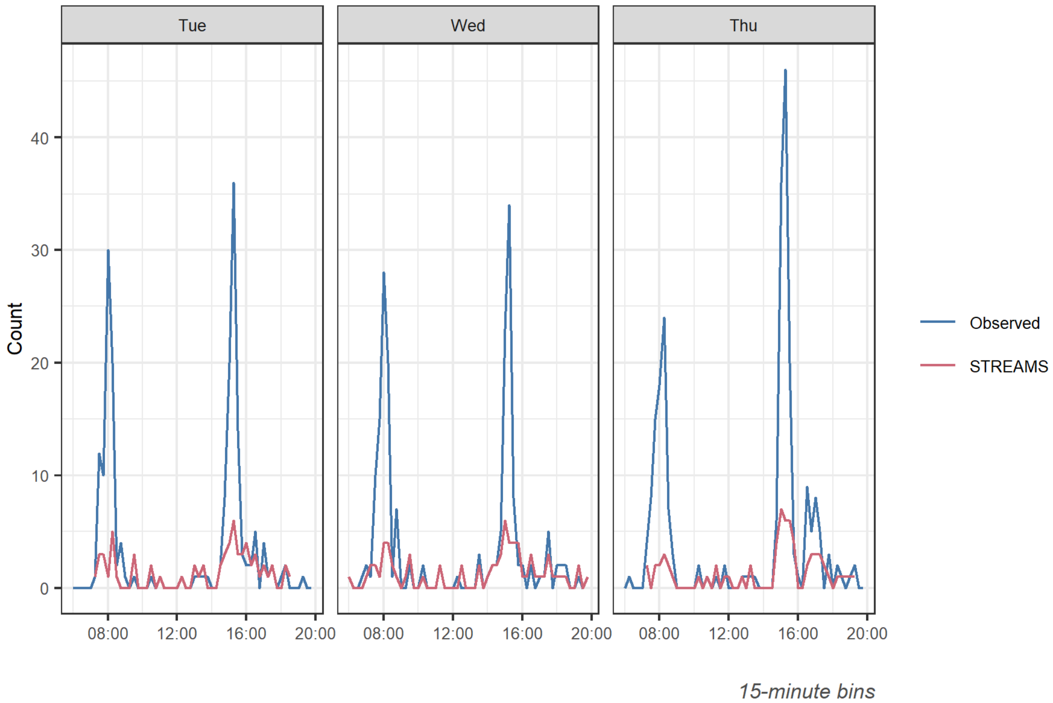
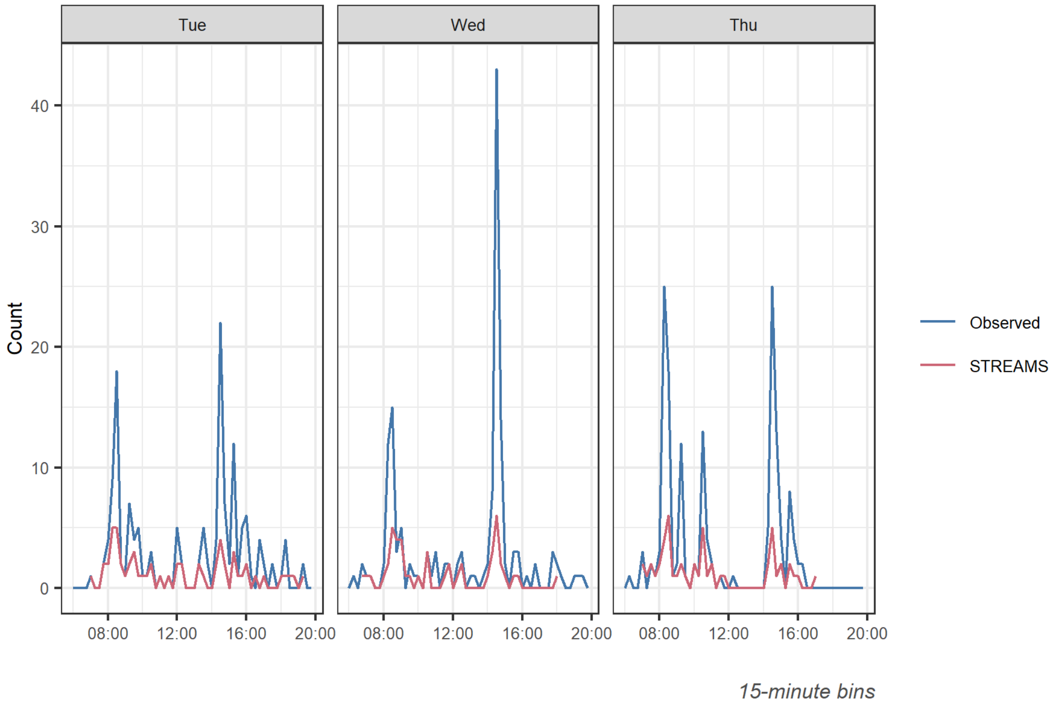
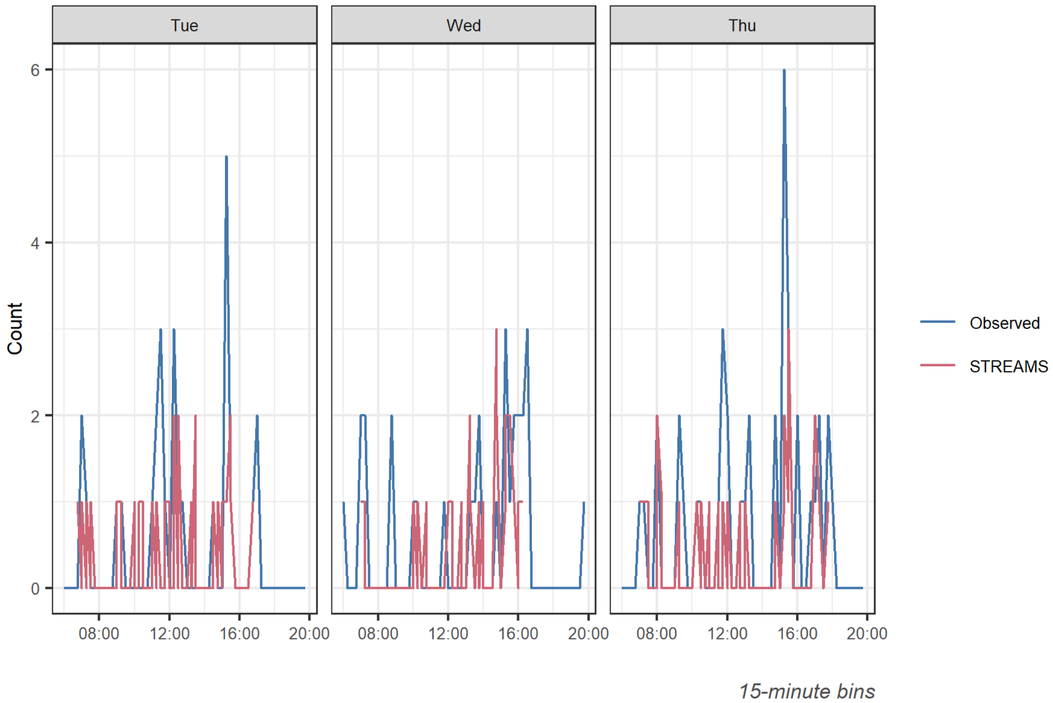


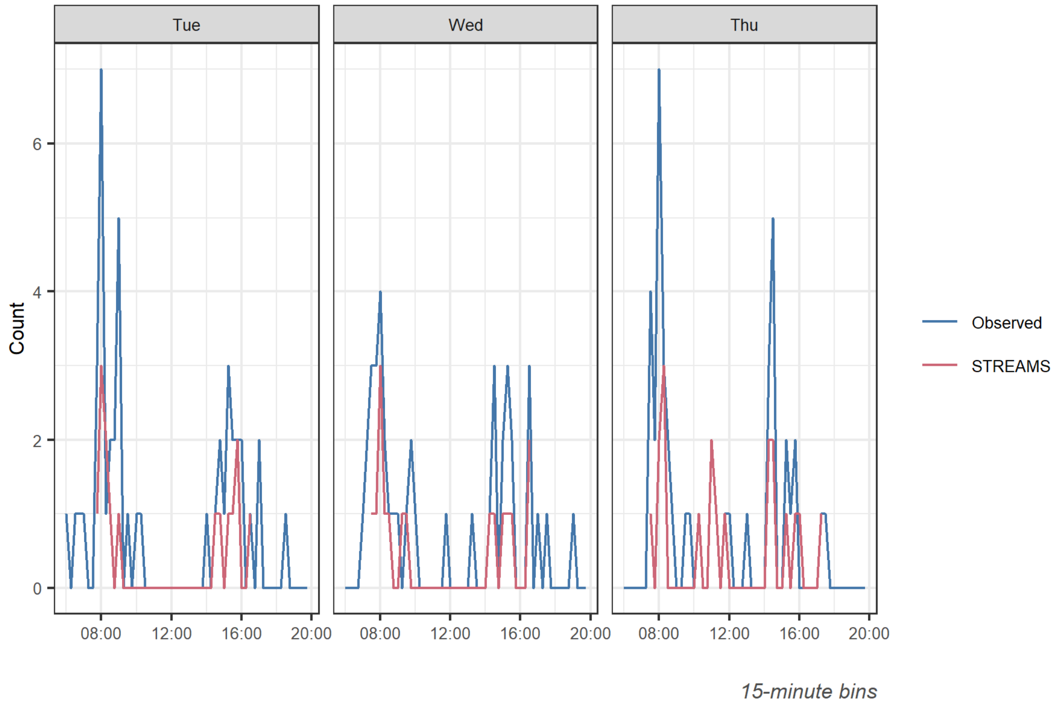
Figure 2.6: Time-of-day comparison - 301 South Pine Road



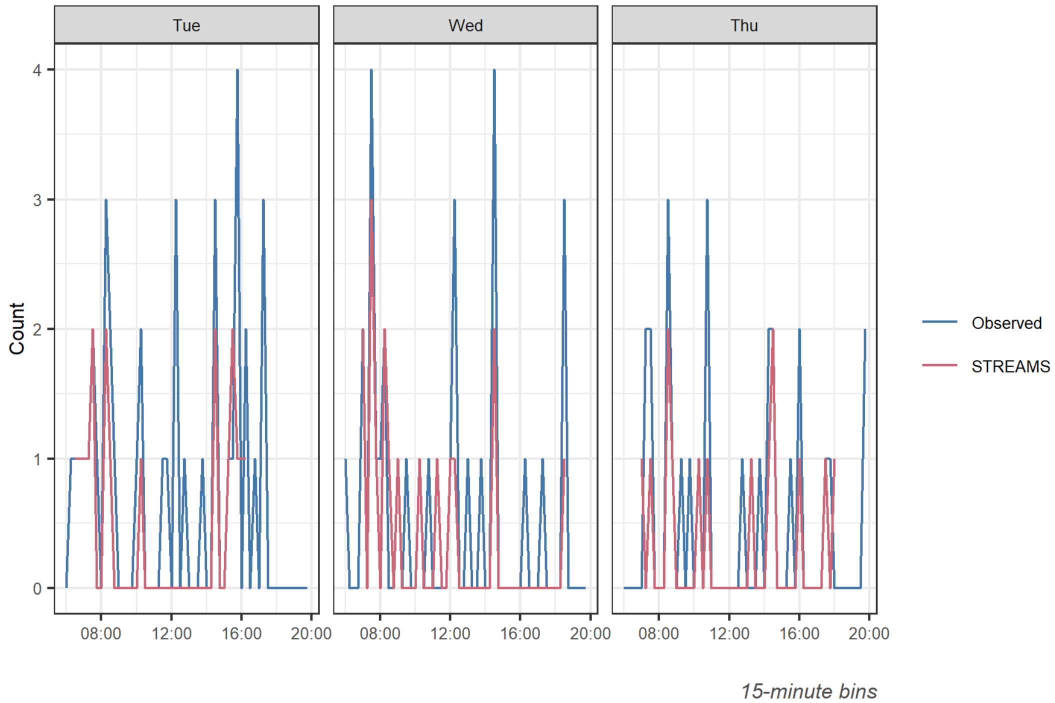
■ **Figure 2.7: Time-of-day comparison - 280 Stafford Road**



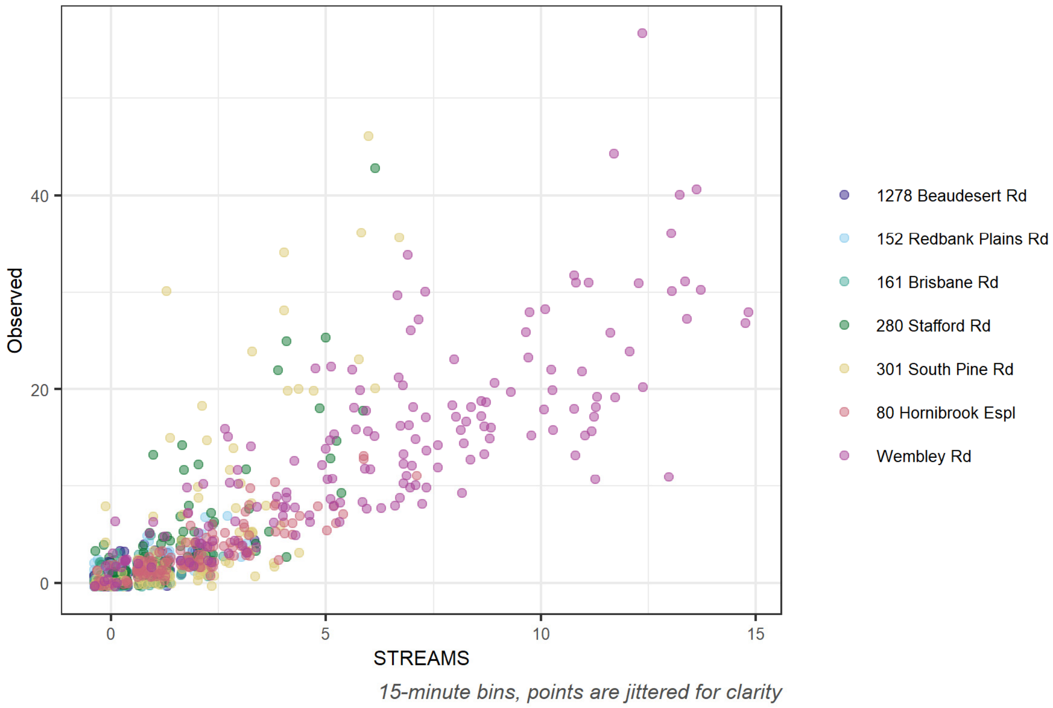
■ **Figure 2.8: Time-of-day comparison - 161 Brisbane Road**



■ **Figure 2.9: Time-of-day comparison - 152 Redbank Plains Road**

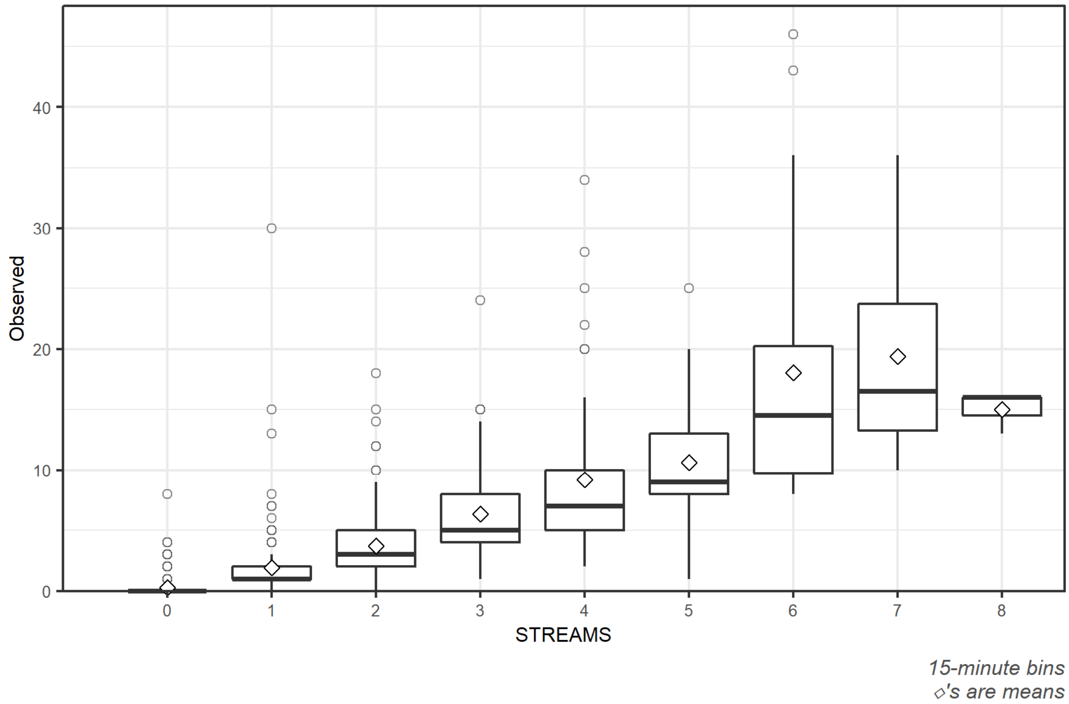


■ **Figure 2.10: Time-of-day comparison - 1278 Beaudesert Road**

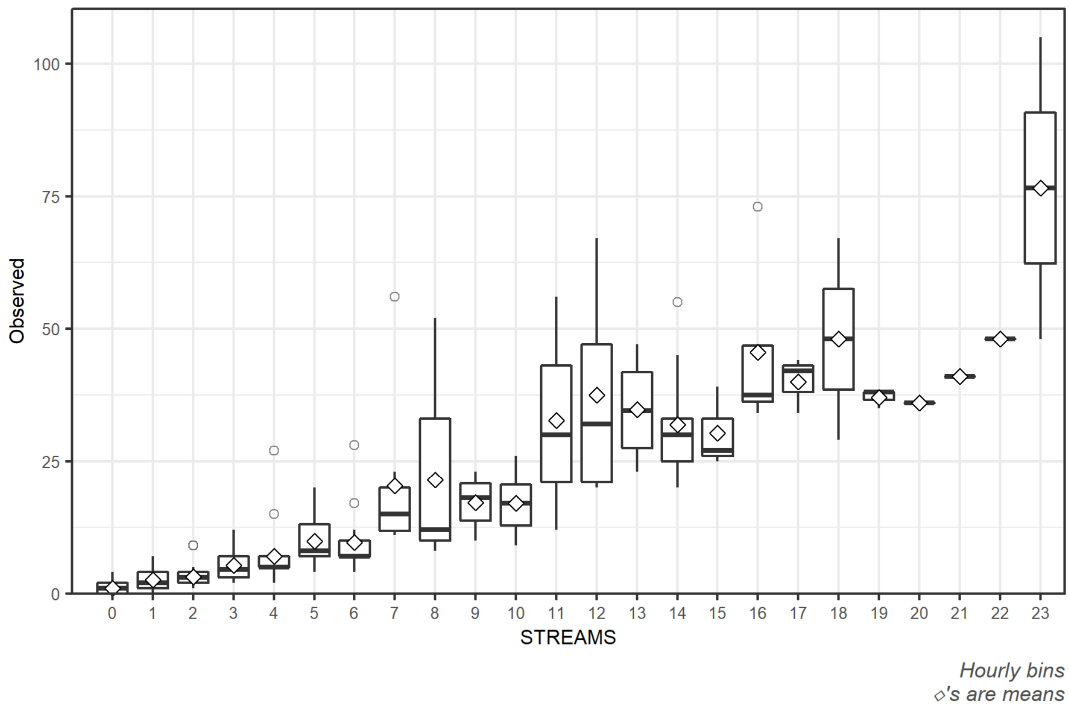


■ **Figure 2.11: Relationship between STREAMS logs and observed pedestrian demand by location**

The distribution of observed pedestrian demand at each STREAMS count in 15-minute bins is shown in Figure 2.12 and for hourly bins in Figure 2.13. There appears to be increasing dispersion in the observed count at higher STREAMS counts and a breakdown in the relationship between the STREAMS count and observed count at the highest STREAMS counts (i.e. the median observed counts above six STREAMS counts in the 15-minute binned data do not change, and there is no clear correlation in the hourly data above 11 STREAMS counts).



■ **Figure 2.12: Distribution of pedestrian demand at each STREAMS count in 15-minute bins (e.g. STREAMS count of 4 indicates 4 pedestrian call-ups in the period)**



■ **Figure 2.13: Distribution of pedestrian demand at each STREAMS count in hourly bins (e.g. STREAMS count of 4 indicates 4 pedestrian call-ups in the period)**

3 Expansion factors

3.1 Methodology

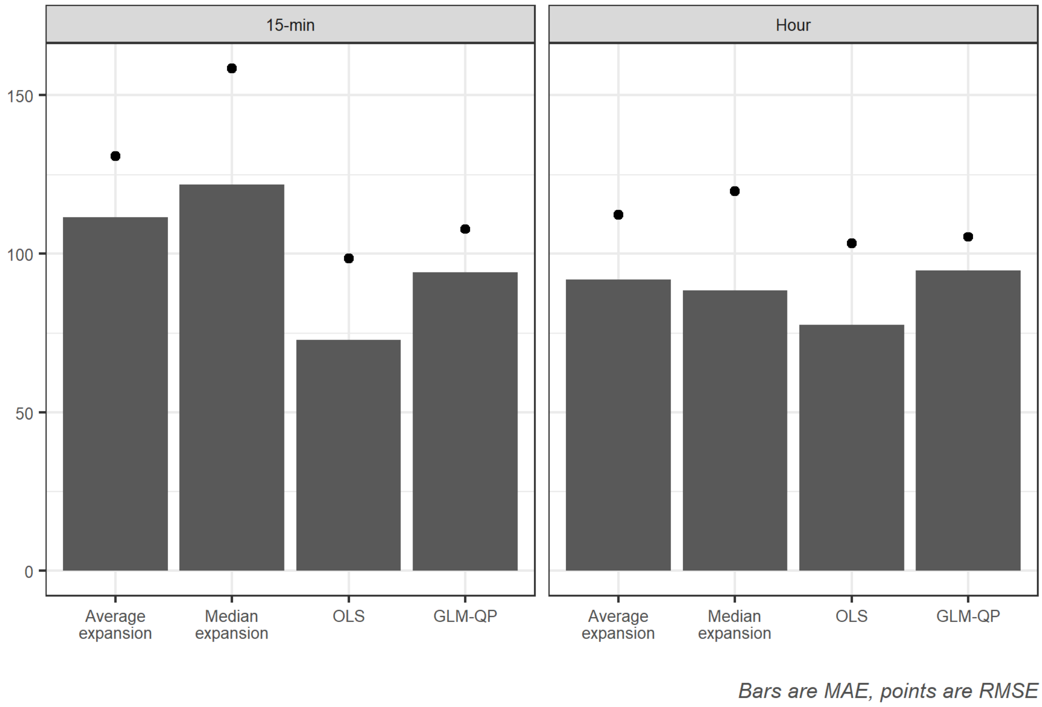
The STREAMS data and observed counts were aggregated into 15-minute and hourly bins. The relationship between the STREAMS count and observed pedestrian demand was estimated using four methods:

1. Average expansion for each STREAMS count:
    a. The *average* expansion factor was calculated for each STREAMS count (i.e. 0 to 8 for the 15-minute data and 0 to 23 for the hourly data)
    b. A weighted linear regression was performed between the STREAMS count (predictor) and average expansion factor (dependent variable) where the weighting was the count of periods in which each STREAMS count was observed.
2. Median expansion for each STREAMS count:
    a. Identical to the average expansion factor method above, except that the *median* expansion factor was used as the dependent variable.
3. Ordinary least squares (OLS):
    a. A standard linear model was estimated between the STREAMS count (predictor) and observed pedestrian count (dependent variable).
4. Quasi-poisson Generalised Linear Model (QP-GLM):
    a. The counts data was over-dispersed (dispersion parameter of 2.8), suggesting a quasi-poisson error structure is more appropriate than a poisson model. Moreover, the use of a poisson model will be more appropriate for counts data than OLS given that it will constrain the estimates to be positive values.
    b. A linear model was estimated in the same way as for OLS except that QP-GLM was used.

The model performance was assessed based on the overall mean absolute error (MAE) and root mean square error (RMSE) of the residuals and on visual inspection of the overall accuracy across the three days of counts at each site. Given there were only seven sites available all sites were used in the estimation; holdout cross-validation was not undertaken.

3.2 Results

The overall MAE and RMSE of each of the four methods, in each of the two time bins, are shown in Figure 3.1. The OLS method has the best fit across both MAE and RMSE measures and is marginally superior when implemented across the 15-minute data (MAE of 72.9 for 15-minute bins versus 77.6 for hourly bins).

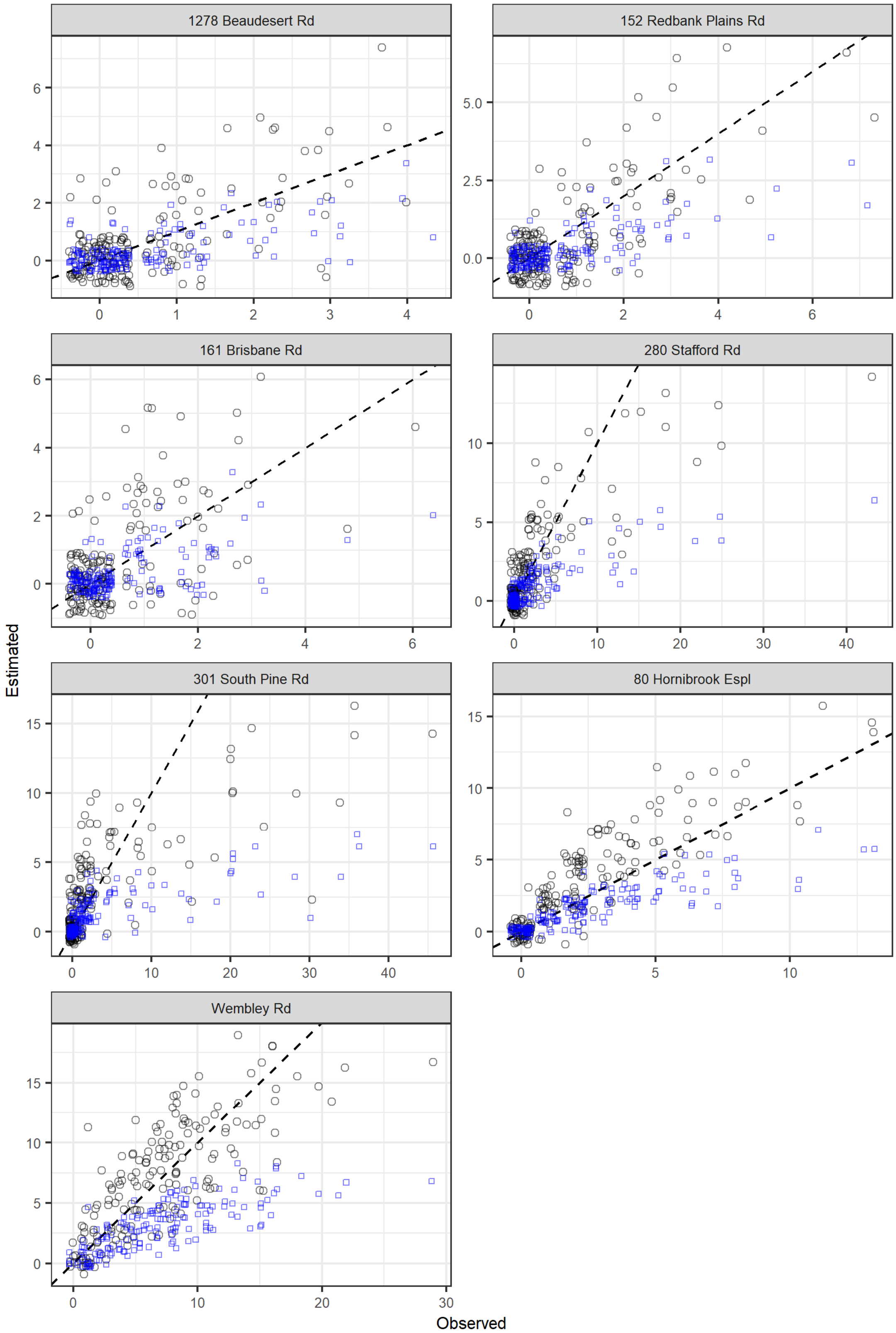


■ **Figure 3.1: Overall model fit statistics**

The accuracy of each method when applied to the total count across each of the seven sites is shown in Figure 3.2. For the preferred model (15-minute bins OLS) the accuracy ranges from undercounting by 24% at 301 South Pine Road to overcounting by 47% at 80 Hornibrook Esplanade. The count was estimated to within 9% for two sites and 24% for six sites. There is no clear pattern between the characteristics of the land uses near the sites and the over- or under-estimation of demand.

The effect of the expansion factor on the individual binned counts is illustrated in Figure 3.3. As would be expected, the expansion factor tends to bring the counts closer to the observed values. However, there remain time periods where the estimated count is significantly in error.

■ **Figure 3.2: Model accuracy by site**



15-minute bins, points are jittered for clarity
Blue squares are STREAMS counts and black circles are estimated (expanded) counts

■ **Figure 3.3: Effect of expansion factor on binned counts by site**

3.3 Conclusion

Given the residuals are lower for the 15-minute period and this aggregation period generally accords with manual count periods this aggregation period is recommended. As the intercept is very small (0.013) it is fixed to zero for simplicity, giving the final model:

$$D_{15} = 2.30 \times STREAMS_{15} \qquad R^2 = 0.68$$

Where D_{15} is the estimated pedestrian demand in any 15-minute period and $STREAMS_{15}$ is the number of pedestrian calls in the 15-minute period.

This model has the benefit of simplicity – multiplying the 15-minute counts by a fixed factor of 2.30 is trivial. Noting the caveat that this was estimated from a fairly small sample of sites, this expansion factor should facilitate estimation of pedestrian demand at a signalised crossing to within +/-25% in most cases. However, it is unlikely this level of accuracy will be maintained for very busy sites.