VARIATIONS IN TRAFFIC COUNT DATA

V Chauke

ARUP, Johannesburg, South Africa
Postnet Suite No 93, Private Bag X1, Melrose Arch, 2076
Tel: +27 11 218 7625, Email: vululami.chauke@arup.com / vululami@gmail.com

ABSTRACT

Traffic counts are the most basic input into transport planning studies, yet there is the potential for errors to propagate within the data collection process that can result in uncertainty in the final collected data. If left unaccounted for, these errors have the potential to skew transport planning and traffic engineering decision making. Multi-million Rand infrastructure projects are born out of studies based on this data.

This paper explores variations in traffic flow data observed from a variety of traffic count surveys. It considers measurement errors that are inherent in manual and automatic counts, and explores the impact of outliers and seasonality adjustments, as well as the concept of an average day in traffic data summaries. Local and international data sources have been applied in a case study to demonstrate the likely outcome of variations in analytical results when using erroneous traffic count data.

The paper also draws conclusions about the potential implications to policy and decision-making and offers suggestions about the role that planning authorities in all spheres of government should play in controlling the quality of data used in transportation studies.

1 Introduction

Traffic count surveys are typically carried out in order to determine baseline traffic flows that are used in transport studies or calibration of traffic models. Based on these counts, baseline demand of the transport network are estimated. Historical counts are also used in determining traffic growth rates within a study area. The growth rates are then used in the estimation of future traffic demand forecasting, using a combination of growth and build-up methods.

However, there are natural variations in traffic flow that can have a major impact in the outcome of the studies and designs that are based on the collected traffic count data. These variations can be for the time of the day, day of the week, or a season. In addition to the natural variations, there are measurement errors associated with human cognitive issues in case of manual counts, and equipment errors / limitations in case of automatic counts. These sort of variations are often overlooked in most traffic studies as insufficient data is obtained to properly assess them.

If the variations or errors are not properly considered at the early stages of transport studies, they are most likely going to be propagated in downstream analysis and decision-making. The UK's Department for Transport provides guidance on 95% confidence intervals for various types of count surveys (Department for Transport (UK), 2014 (1)). The inclusion of these limits illustrate that different survey methods has some degree of inherent measurement errors, which would need to be considered early on in a transport study.

The purpose of this paper is to explore traffic flow variations and measurement errors inherent in traffic count data. Both local and international data sources are used in the analyses. The paper draws conclusions on the potential implications to policy decision making and offers suggestion on the role that planning authorities could play in controlling the quality of data used in transportation studies.

2 Data sources

The analyses carried out for this paper was based on data that was collected for the following transport studies:

- Waterfall City Model Update (2013) (SA); and
- London count dataset for a strategic transport model.

For the Waterfall City model update, the data were sourced from the SANRAL Continuous Traffic Observation (CTO) database. The locations of the SANRAL count sites used are indicated in Figure 1.

Figure 1: SANRAL's Continous Traffic Observation (CTO) station locations

For the London study, data were collected, which comprised of both the Automatic Traffic Counts (ATCs) and Manual Classified Counts (MCCs). The locations of the London count sites are shown in Figure 2. ATCs were undertaken over a two-week period, using tube based counters. MCCs were undertaken over a 12 hours, for a duration of between one and four days per site. Video survey methods were used for the MCCs.

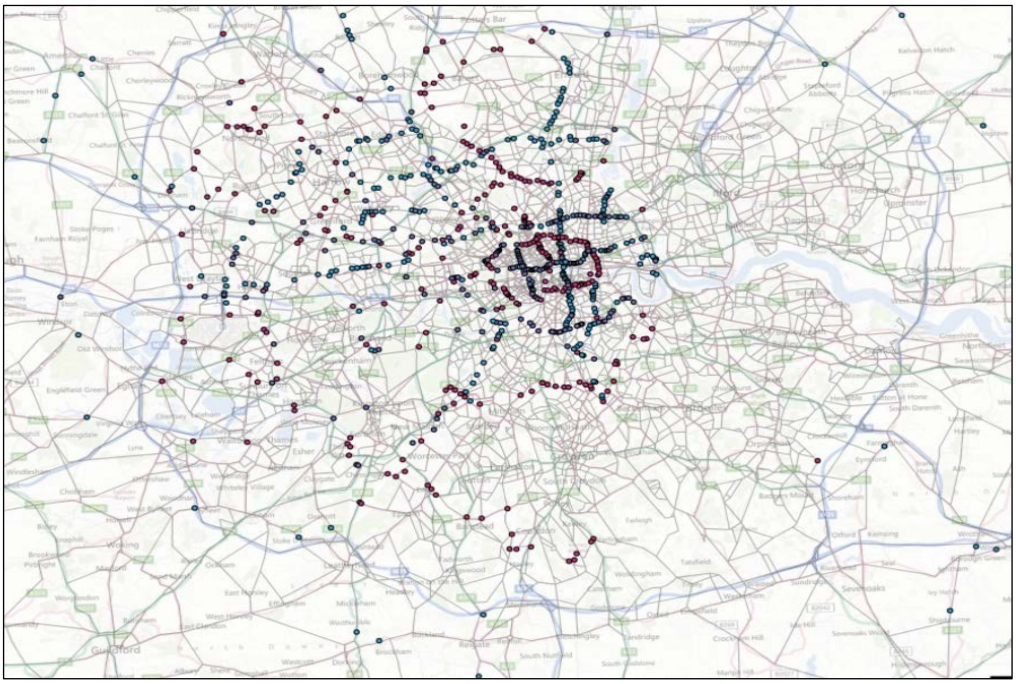

Figure 2: Locations of London Count sites.

3 Analyses

The analyses carried out were aimed at studying the following:

- Weekday flow variations, and outlier identification; and
- ATC / MCC comparison;

3.1 Weekday flow variations

An example of weekday flow variations at one of the SANRAL CTO station is given in Figure 3. The data used in the analysis presented in Figure 3 was collected in 2012 and 2013 using buried loops (ATC), in February, May, July and November, therefore representing both summer and winter periods for South Africa. Weekends, holidays and days before and after long weekends were excluded to represent normal days in which counts are typically carried out.

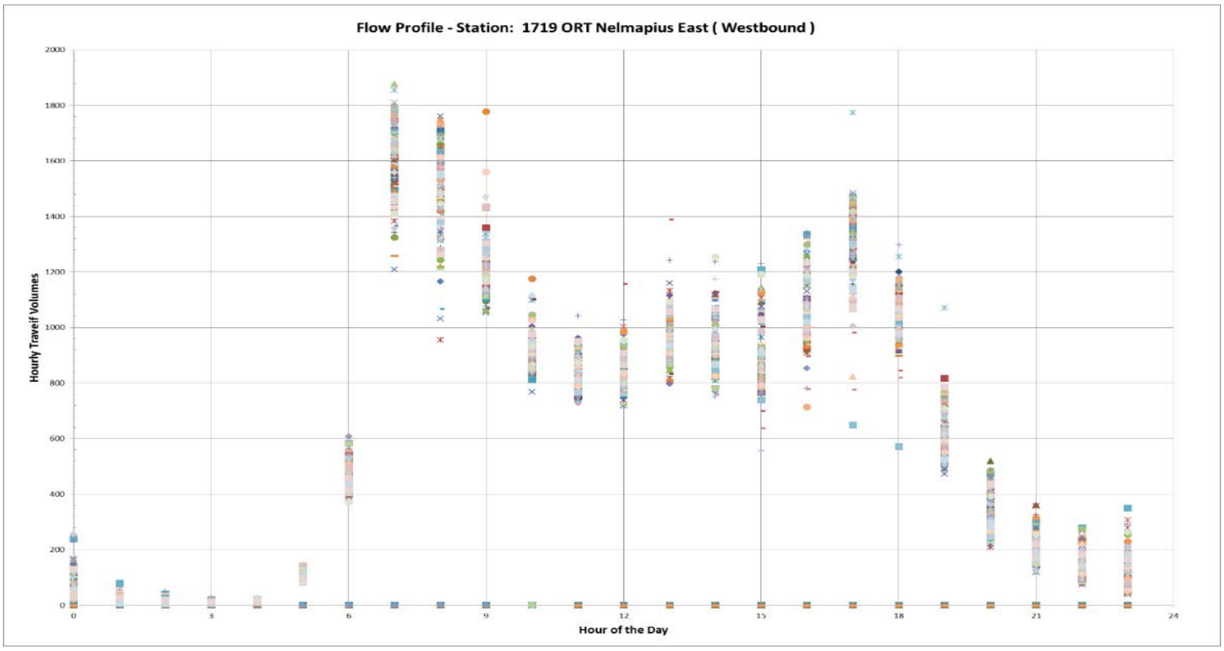

Figure 3: Weekday flow variation – SANRAL CTO Station 1719 Nelmapius Rd (Westbound)

In the figure above, it is evident that traffic flow volumes change between days. The flow patterns are generally similar for the various days. However some outliers are visibly notable from the plot above. In the AM peak hour (07h00 – 08h00), traffic volumes range between 1,200 and 1,850, and the average hourly flow is approximately 1,550 vehicles. This result in an error of between 19% and 23% for the low and high volume as compared to average.

In terms of the UK's Transport Analysis Guideline M3.1 (Department for Transport (UK), 2014 (2)), traffic modelling flows for the flow range between 700 and 2,700 needs to be within 15% of the observed flow. If this criteria was to be used on the minimum and maximum observed AM peak hour flows, using the average as representative of the observed, the minimum and maximum observed flows would be outside the tolerable limit.

The implication to the above is that if data is collected on a single day, and no historical data is available for the area, it would be impossible to determine if the observed data is for an average day, an extreme low or an extreme high flow day. The service levels calculations could therefore be grossly over or under stated, with implications to the recommendations given within the study.

For the study where the above count data was used, outliers were identified and removed from the final version of the count database. Figure 4 show the profile of the same count site, as above, but with outliers removed.

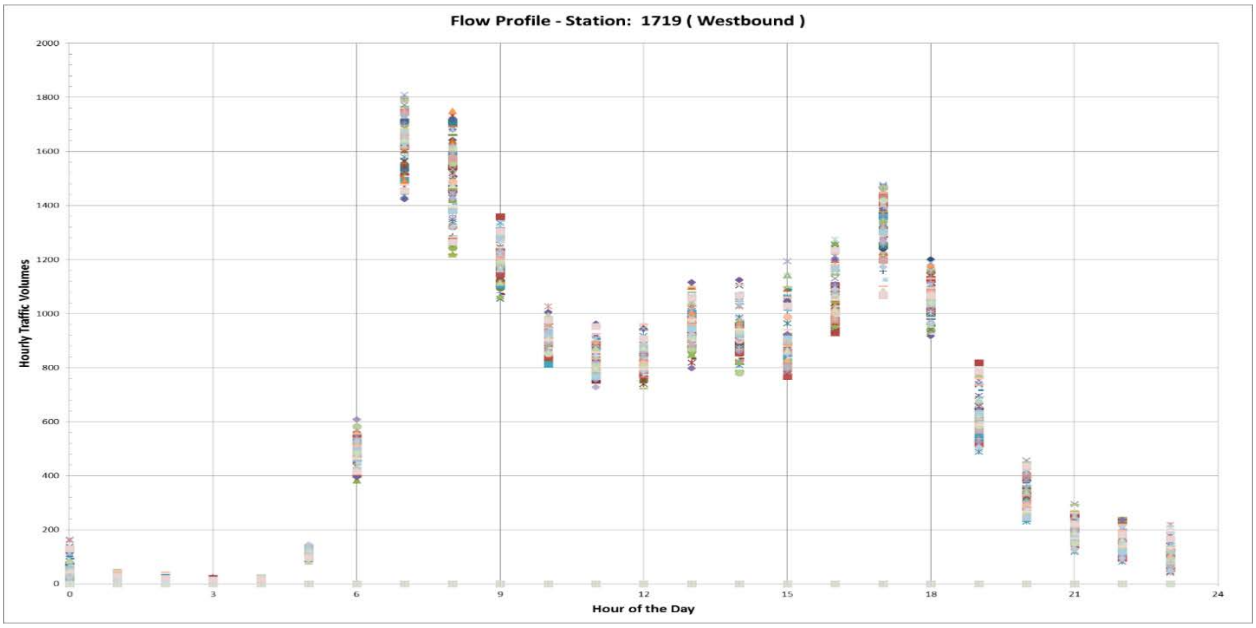

Figure 4: Weekday flow variation – SANRAL CTO Station 1719 Nelmapius Rd (Westbound) | Outliers Removed

Considering the same AM peak hour, the average flow would now be approximately 1,600. The minimum and maximum values will be approximately 1,400 and 1,800 respectively. This would result in minimum and maximum hourly volumes of about 12.5%. This percentage difference would be within the acceptable error range for vehicle flows in the flow category of 700 to 2,700 vph.

A similar analysis to the above is carried out using one of the count station from the TfL database. The data used in the analysis below was collected in October 2012, which was during the UK Autumn season. The analysis results are summarised in Figure 5.

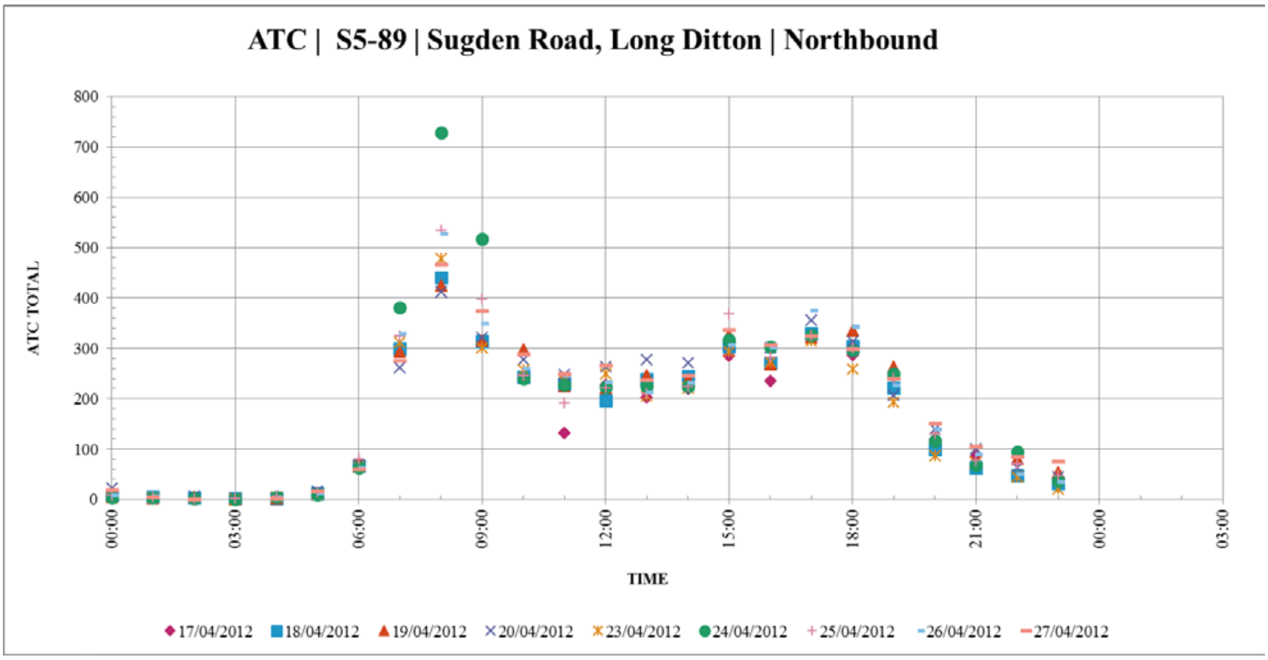

Figure 5: Weekday Flow variation – TfL Count Site S5-89 – Sugden Road, Long Ditton (Northbound)

From Figure 4, it can be observed that the flows profile for all the days, was about the same. However outliers were identified on the 17th and 24th of April 2012 counts. Data for these days were subsequently removed from further analyses of the count data. The consistency of the flow profile is be expected for this example as there is no seasonality variation at play for the duration of the count period. However, as would be illustrated later in the paper, seasonality still need to be kept in mind when analysing count data for model development purposes.

3.2 ATC / MCC comparison

In order to understand the relationship between ATC and MCC survey data, a comparative analysis was carried out on the TfL count surveys, where both ATC and MCC surveys were carried out on the same day, for the same count site.

The objectives of this comparative assessment are listed below:

- To identify count sites with large differences between ATC and MCC hourly flows;
- To classify count sites based on ATC / MCC variation; and
- To calculate factors for correcting large flow variation.

The ATC surveys were carried out using pneumatic tubes and the MCC surveys were carried out using video count methods. 320 count sites, or about 7,680 hourly records, were evaluated. The results of this analyses are summarised in Table 1 and 2 below:

Table 1: ATC / MCC differences - Summary using absolute flows

ATC_MCC CASE Classification	Absolute Differences (%)				
	<5%	5-10%	10-15%	>15%	Total
Rule 1 (ATCs within 5% of MCCs)	3,966	18		0	**3,984**
Rule 2 (MCCs > 5% if ATCs)		1,732	725	977	**3,434**
Rule 3 (ATCs > 5% of MCCs)		147	39	79	**265**
Total	**3,966**	**1,897**	**764**	**1,056**	**7,683**

From table 1, it can be concluded that:

- About 52% of the hourly records had ATC/MCC flow differences of less than 5%;
- 24% had flow differences ranging between 5 and 10%; and
- 24% had flow differences greater than 10%.

An alternative comparison using GEH statistics, which is commonly used in highway assignment traffic models, was carried out. GEH is defined as a form of Chi-square statistic that incorporates both relative and absolute errors. It was first used Geoffrey E Havers in the 1970s, and since been used to measure the comparison between observed and modelled flows. The GEH formula, as given in the DMRB ((DfT, 1997)) is given below:

Equation 1: GEH statistic formula

$$GEH = \sqrt{\frac{(M-C)^2}{\frac{1}{2}(M+C)}}$$

Where:

GEH is the GEH statistic;

M is the modelled flow; and

C is the observed flow

A GEH of less than 5 is desirable and is considered appropriate for day-to-day variation in traffic flows. A GEH 5 – 7 is not ideal, but could be accepted under certain condition. A GEH of greater than 7 illustrate significant differences in the compared count sets.

Table 2 contain a summary of the comparison of the ATC and MCC counts, where both type of survey were undertaken on the same day for each of the compared site. The modelled and observed flows in the formula are swapped for ATC and MCC counts for purpose of comparing the two datasets.

Table 2: GEH Comparison of ATC / MCC counts

ATC_MCC Classification	**GEH**							
	0-5	**5-10**	**10-15**	**15-20**	**20-25**	**25-30**	**30-35**	**Total**
Rule 1 (ATCs within 5% of MCCs)	3,984							**3,984**
Rule 2 (MCCs > 5% if ATCs)	3,022	305	73	22	7	3	2	**3,434**
Rule 3 (ATCs > 5% of MCCs)	240	10	11	2			2	**265**
Total	**7,246**	**315**	**84**	**24**	**7**	**3**	**4**	**7,683**

The GEH analysis indicate that about 94% of the hourly records had GEH values of less than 5, which is indicative of a relatively good comparison of the two datasets. However the 6% that have poor GEH (GEH >5) may have a significant impact in the AM and PM peak hours, as the hourly volumes are significantly higher that the off-peak and inter-peak periods. Figure 6, below, illustrate the potential significance of ATC/MCC differences at higher flow volumes

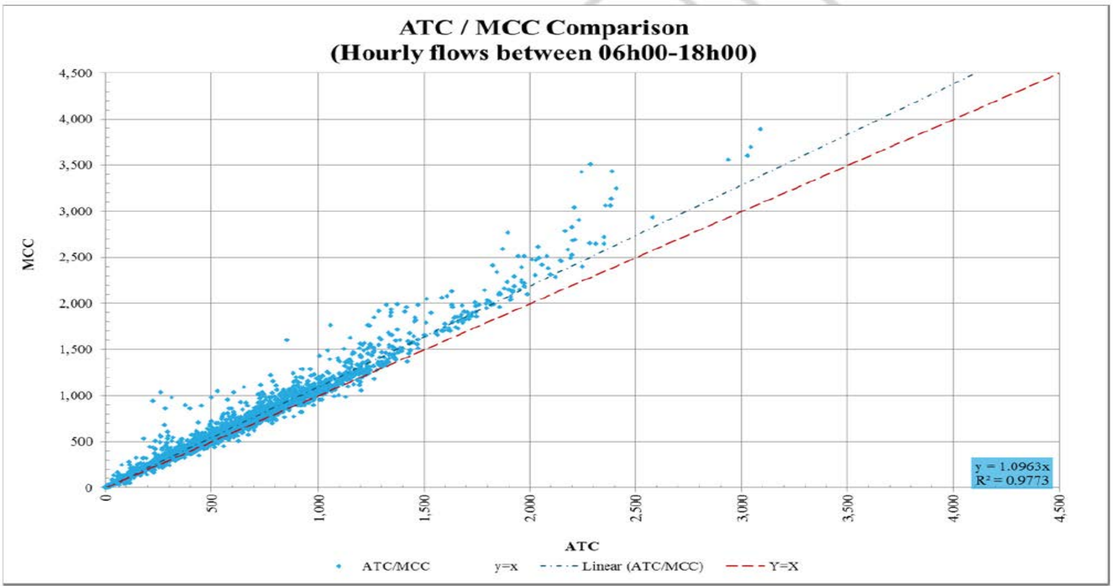

Figure 6: ATC / MCC comparison – Daily flows

Figure 6 illustrate that on average the MCC volumes are about 9.6% higher than the ATC volumes. Generally low traffic volumes, with hourly traffic volumes less than 1,400, have better correlation between ATC and MCC datasets. The variation is greater as the volume of traffic increases. During the AM and PM peak periods, when the volumes are higher, it was established that the error was approximately 11.9% and 11.5% respectively. The AM peak hour comparison is illustrated in Figure 7.

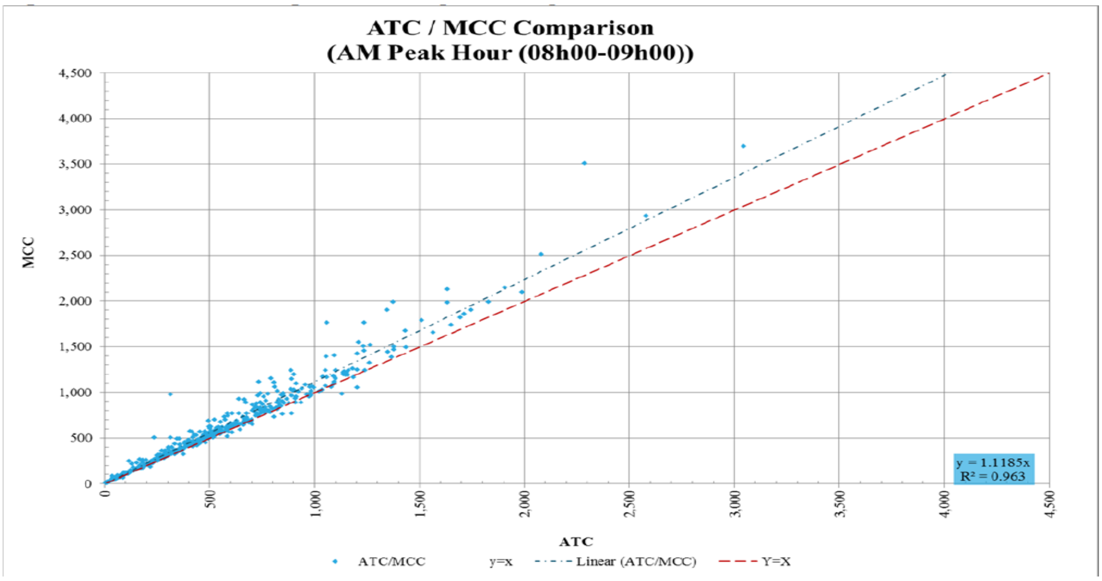

Figure 7: ATC / MCC comparison - AM peak Hour

There was no sufficient local data to study the ATC / MCC flow variation in the South African context. Although it can be widely expected that the local comparison could be similar to the London conditions given similar data collection techniques, more research is still required to determine local measurement errors between automatic counts and manual counts. In South Africa, video counts are not widely used due to the cost of data collection, and human cognitive issues could therefore have a major impact in local traffic count measurement errors.

3.3 Seasonality adjustment

In the London study, seasonality factors were sourced from the "Road Network Performance & Research Team" technical note (RNPR, 2008). November was identified as a neutral month, and count data from different months were factored to "November" flows, as such the resulting model was to be a November model. These seasonality factors are given in Table 3 below.

Table 3: Seasonality adjustment factors - London screenlines

Month	Central Screenline	Inner Screenline	Outer Screenline	Radial Screenline
January	0.943	0.997	0.982	0.986
February	0.974	1.011	1.009	1.011
March	0.982	1.035	1.042	1.043
April	0.980	1.024	1.023	1.024
May	1.001	1.000	1.036	1.016
June	0.994	1.030	1.037	1.039
July	0.981	1.018	1.034	1.041
August	0.942	0.979	0.998	0.990
September	0.985	1.029	1.045	1.038
October	0.974	0.993	1.023	1.011
November	1.000	1.000	1.000	1.000
December	0.961	0.991	0.966	0.985

Seasonality consideration is key when comparing model baselines. In case of modelling, it is important to understand the season for which a model is developed for. The best practice approach is to develop models for a neutral month, where flows are not influenced by school holidays, special events and other activities that do are not typical for an average or normal season.

The global best practice is to convert counts collected in various months to a neutral month, in order to factor out the issue of seasonality. The seasonality factors are therefore useful established a common base for comparing traffic model results.

Locally, SANRAL collect traffic count data at permanent locations on the national highways and other key corridors, which may be used in generating seasonality factors on these high order roads. However there is less data available on low order roads as counts are not collected consistently by the local transport authorities and stored centrally for ease of retrieval and analyses.

3.4 Comparison with journey time surveys

A comparative analysis using Journey Time Surveys (JTS) was also conducted in Midrand, during April and May 2015. Four JTS runs were undertaken, with each run being carried out at a different day, at more or less the same start time as the first run, using the same vehicle and driver. Run 5 (Green) is a control run that was carried out during off-peak, it only included to provide context to the interpretation of the results and was not used in determining average travel times. A summary of this analysis is given in Figure 8 below:

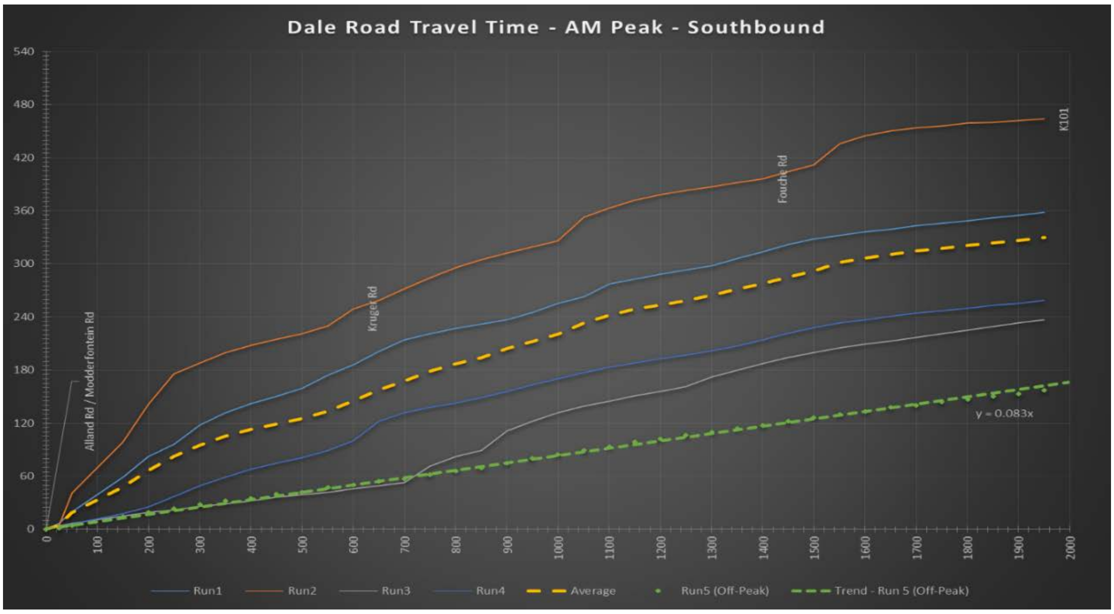

Figure 8: Journey time survey on Dale Road, Midrand, South Africa

From the above, it can be observed that travel times on the section of the road above varied during different days. In a congested network environment, the section delays can influence the number of vehicles that can pass through the section during the peak. A sharp variation in traffic count can be expected on a road with the journey time profile such as the one above.

4 Impact of traffic flow and survey data variation on planning decisions

Supposing that traffic surveys were carried out on a day with considerably low volumes of traffic, as part of a Traffic Impact Assessment (TIA) study for a major development, the road infrastructure provisions may prove to be insufficient when the development is fully operational.

On the other extreme, if surveys were carried out on an extreme high flow day, we may result with an expensive infrastructure which affect the viability of the development, and may result in higher lease rates being charged to the tenants which will eventually be passed on to the general public.

However, if the variations and errors are well understood, necessary sensitivity analysis would be carried out and optimal infrastructure provision and phasing would be determined. This would result in economical designs that would be adequate for most days of the year and not just the isolated extremes that would not even have been picked up with single day counts and observations. It should be noted, though, that issues of cost, equipment security and other factors, play a role in deciding the extent of the surveys. These should be weighed against accuracy requirements when deciding the scope of surveys.

5 Recommendations

It is recommended that data collection guidelines be developed by the South Africa planning authorities, which will set the minimum data collection periods for the various type of studies. This will ensure that all the practitioners in the transport planning and traffic engineering sector will be assessed on the same standards. The requirements for carrying out counts over multiple of days or weeks should be weighed against accuracy requirements, and the in context with the size of the study being carried out or proposed.

It is also recommended that the planning authorities collect and store traffic count data that may be used for seasonality adjustment. Third parties should have access to this database in order to allow them to carry out the required sensitivity analyses as part of transportation studies and infrastructure designs. Similarly private individuals and organisations that collect data over longer periods should be encourage to share their data with planning authorities, for inclusion in (open) central databases.

It is recommended that more research be carried out in order to understand the differences between manual and automatic counts in the South African context.

6 Conclusion

It has been demonstrated above that there are both natural variations and measurement errors in traffic flow data measurements. There is potential to propagate these errors onto future analyses and transportation decision-making.

These variations need to be considered carefully when using traffic data to determine baseline traffic demand and when the data is used for future demand forecast. The cost of designing and providing infrastructure for very low or high flow days/peak, as compared with the average flow, range from the economic viability of land-use development projects, through to human cost associated with congestion-induced frustrations on our roads.

It is therefore important that these variations and potential errors be understood better, and accounted for in future transport studies. Specifying data collection over extended periods could improve the quality of the input data into transport studies and minimise error propagation in future traffic demand forecasts. However cost, safety, security and other factors should be considered together with accuracy objectives of transportation studies.

Greater research is required to understand the differences between manual and automatic counts in the South African context. This will be important in situations where multiple data sources are used for establishing baseline traffic demand.

Sensitivity analyses should be carried out when high-capital cost upgrades and new transport infrastructure interventions are proposed.

7 References

RNPR, 2008. *Factors for road traffic counts in London,* London: Transport for London (TfL).

Department for Transport (UK), 2014 (1). *WebTAG: TAG unit M1-2 data sources and surveys, January 2014.* [Online]
Available at: https://www.gov.uk/government/publications/webtag-tag-unit-m1-2-data-sources-and-surveys
[Accessed 18 May 2015].

Department for Transport (UK), 2014 (2). *WebTAG: TAG unit M3-1 highway assignment modelling, January 2014.* [Online]
Available at: https://www.gov.uk/government/publications/webtag-tag-unit-m3-1-highway-assignment-modelling
[Accessed 18 May 2015].

DfT, 1997. *Design Manual for Roads and Bridges | Volume 12 Section 2,* London: Highway Agency (UK).