THE GEH MEASURE AND QUALITY OF THE HIGHWAY ASSIGNMENT MODELS

Olga Feldman
Affiliation: Transport for London, Windsor House, 42-50 Victoria Street, London, UK, SW1H 0TL; E-mail: olja.feldman@gmail.com ; Tel: +44 20 7126 3934

ABSTRACT

This paper reports on some theoretical findings as well as analysis of different ATC data sets so as to get a better understanding of the nature of the GEH measure. The goodness-of-fit measures suggested by the DRMB guidance about supplying assessments of transport model validity should take account of the variability of the count data. The GEH measure also confronts the significant issue of how to distinguish that a given level of absolute difference and of percentage difference can have very different levels of significance depending on the scale of the flows. Analysis of the different ATC data sets was carried out in order to understand variability of link flows, turning counts and enclosure flows. The research shows that the variability of traffic counts between comparable data on different survey days is significant and depends on road types, time of a day, area, and other factors. The variance-to-mean ratio can go up to 17, whilst the GEH values can go up to 67. In addition, the data suggest that some theoretical assumptions such as proportionality of variance to mean are quite strong assumptions and do not always hold.

The opinions expressed here are those of the author and not necessarily those of the Transport for London.

Keywords: GEH, Data variability, ATC data

1. INTRODUCTION

The GEH statistic[1] is used to represents goodness-of-fit of a model. It takes into account both the absolute difference and the percentage difference between the modelled and the observed flows. The GEH statistic for a link is computed as follows:

$$GEH_j = \sqrt{\frac{2(K_j - M_j)^2}{K_j + M_j}} \qquad \text{(Eq. 1)}$$

where K_j is an observed flow on a link j, M_j is the modelled flow for the same link.

It should be noted that the GEH is not unitless, but has units of (vehicles/hour)$^{0.5}$, since the K and M as used in British practice are typically hourly flows e.g. for a peak hour. It is possible to measure the GEH for a longer period such as a three hour AM peak or 6 hour interpeak, by taking an average of the hourly counts in the period.

Although its mathematical form is analogous to a chi-squared statistic, GEH is not a true statistical test but an empirical formula that is used for a range of traffic analysis

[1] The GEH formula is named after its inventor, Geoffrey E. Havers.

purposes. For a set of links, the DMRB goodness-of-fit criterion is that GEH_j should be less than 5 for more than 85% of the individual links. A similar statistic can be constructed for a screenline/cordon, and, according to the guidance, should be less than 4 in all or nearly all cases. Again, this “5” or “4” threshold is only appropriate if the units of K and M are vehicles/hour.

The author is unclear how the GEH measure was derived in detail originally and as a result, whether relaxing it implies a violation of some fairly rigorous statistical tests. MVA have suggested the following derivation of the GEH statistic (QUORAM, 2008) based on the Poison index of dispersion for a set of observations x_i derived from a Poisson distribution[2]:

$$ID = \frac{\sum_i (x_i - \overline{x})^2}{\sigma^2} = \frac{\sum_i (x_i - \overline{x})^2}{\overline{x}}, \quad (\text{Eq. } 2)$$

which for two observations x_1 and $x_2$is equivalent to

$$ID = \frac{(x_1 - x_2)^2}{2\overline{x}}, \qquad (\text{Eq. } 3)$$

where $\overline{x}$ is the mean of the observations.

The Poisson distribution is used to model the number of events occurring within a given time interval and can be applied to systems with a large number of possible events, each of which is rare. The Poisson distribution is defined over the set of integers (including 0), and depends only on the single positive parameter λ, the expected number of occurrences in this time interval. The probability that there are exactly k occurrences (*k* being a non-negative integer, *k* = 0, 1, 2, ...) is equal to

$$f(k, \lambda) = \frac{\lambda^k \exp(-\lambda)}{k!} (\text{Eq. } 4).$$

The Poison index of dispersion (Eq.2) is distributed approximately as χ^2 variable with one fewer degrees of freedom than the number of observations. With two observations the critical value of the χ^2 variable at the 95% confidence level is 3.84, which is the square of the corresponding critical value of the normal distribution.

If we assume that the variance of the observed values is proportional to their mean the equivalent index of dispersion for observations derived from a “quasi-Poisson” distribution would be

$$ID^* = \frac{(x_1 - x_2)^2}{2\lambda\overline{x}}. \quad (\text{Eq. } 5)$$

Suppose that x_1 is our observed flow on a link K_j, and x_2 is the modelled flow for the same link M_j and they are tested as to whether they could have arisen from the same Poisson-like distribution. Then, from (Eq.5) it follows that

$$ID^* = \frac{\left(K_j - M_j\right)^2}{2\lambda\frac{(K_j + M_j)}{2}} = \frac{GEH_j^2}{2\lambda} \qquad (\text{Eq. } 6).$$

[2] Note that in the case of the Poisson distribution, the variance is equal to the mean

The critical value for the normal distribution at the 85% level of confidence is 1.44, and therefore (note that GEH_j should be less than 5 for more than 85% of the individual links)

$$\lambda = \frac{GEH_j^2}{2ID^*} = \frac{25}{2 * (1.44)^2} \approx 6.03. \quad (\text{Eq. } 7)$$

Similarly to (Eq.7), an equivalent value of λ for screenline flows can be calculated, and at the 95% level of confidence and the GEH statistic equal to 4, it is approximately equal to 2.08. Therefore, the GEH statistic seems to be testing whether observed and modelled flows can be assumed to arise from a Poisson-like distribution with the variance six times the average for individual links and twice the average for screenlines.

There is a body of research about the distribution of counts on a road link (see for example, Oliver, 1962; Vaughan, 1970; Yi et all, 2005), and the following points are worth noting:

- On a lightly-loaded dual carriageway or motorway, cars arriving at a counter are known to follow a Poisson distribution fairly closely even at the minute-by-minute level, since individual cars can travel independently. This gives confidence that an aggregate hourly count should be Poisson distributed in most cases.
- On a single carriageway or urban situation with traffic signals there can be platoon formation even at quite low loads so that the counts are not Poisson (independent) minute-by-minute. Nevertheless at hourly level a Poisson assumption would seem plausible provided that there are not significant upstream or link capacity effects censoring the distribution.
- On a congested inter-urban motorway, counts are in general not Poisson; the distribution is heavily censored in that there is a maximum capacity on the link, and on “good” days the volume varies within a relatively narrow band near capacity. Such links also have “bad” days where accidents or poor weather cause significant capacity reduction, which causes the count distribution to skew towards 0 which cannot be matched by correspondingly high volume. (If one could measure the demand rather than the actual flow which can pass in the hour, it might have a more Poisson distribution). It is normal in modelling “typical” days to try to omit known roadwork days and large outliers from the counts but this may not be sufficient to completely mitigate the effect. However normal distribution might represent a suitably good approximation to such a distribution in practice despite the censoring and skew. This could be relevant to Regional modelling in London, for example if considering motorway links across the Outer London cordon.
- In congested urban areas there is often a flow metering effect, where the actual load on links is constrained by the capacity of a junction or cordon of junctions upstream. The amount of traffic which can then be passed into the link is again variable within a band around the mean. Rather than a Poisson model, which is equivalent to a queue with independent Markov-distributed arrivals, this might be better modelled as a queue with deterministic arrivals from the upstream junction plus some random local variation, which also suggests a normal distribution.

- Looking at totals across a screenline, it is even less clear why this should be considered Poisson. Certainly the variance across a cordon of links can actually be quite small, even while the mean is very large, as it cancels out much of the day to day route choice variation. While the GEH might be useful at the link level, it may be inappropriate at the screenline level. Certainly we believe it is possible to analyse and defend the GEH at link level without needing to defend its use for aggregate counts.

2. GEH: MODELLED, OBSERVED AND THE LONG-RUN AVERAGE FLOWS

In this section we investigate the relationship between the GEH statistics based on observed versus modelled, observed versus the long-run average and modelled vs. long-run average flows assuming that modelled and observed values come from the same distribution.

Let

$$GEH_K = \sqrt{\frac{2(K-\mu)^2}{K+\mu}}, \quad \text{(Eq. 8)}$$

$$GEH_M = \sqrt{\frac{2(M-K)^2}{M+K}}, \quad \text{(Eq. 9)}$$

$$GEH_* = \sqrt{\frac{2(M-\mu)^2}{M+\mu}}, \quad \text{(Eq. 10)}$$

where K is an observed flow, M is the modelled flow and μ is the long-run average for the same link; GEH_K, GEH_M and GEH_* are GEH statistics associated with the observed vs. the long-run average, modelled vs. observed, and modelled vs. the long-run average respectively.

Then, substituting (Eq.9) and (Eq.8) into (Eq.10):

$$GEH_* = \sqrt{\frac{(M-\mu)^2}{0.5(M+\mu)}} = \sqrt{\frac{(M-K+K-\mu)^2}{0.5(M+\mu)}} =$$
$$= \sqrt{\frac{(M+K)GEH_M^2 + 4(M-K)(K-\mu) + (K+\mu)GEH_K^2}{M+\mu}}. \quad \text{(Eq. 11)}$$

Assume that both observed and modelled flows are proportional to the long-run average: $M = \alpha\mu$, $K = \beta\mu$, where α and β are proportionality constants, and GEH_M is the DMRB GEH statistic while GEH_K is estimated from data. Therefore, the following relationship holds:

$$GEH_* = \sqrt{\frac{(\alpha+\beta)GEH_M^2 + 4(\alpha-\beta)(\beta-1)\mu + (\beta+1)GEH_K^2}{\alpha+1}}. \quad \text{(Eq. 12)}$$

If α and β are equal for a pair of observed and modelled valued, GEH_* can be calculated as

$$GEH_* = \sqrt{\frac{2\alpha}{\alpha+1}GEH_M^2 + GEH_K^2}\,. \quad \text{(Eq. 13)}$$

3. GEH: INDIVIDUAL OBSERVED VALUES VS. AN AVERAGE OBSERVED OVER A TIME PERIOD

In this section we look at the relationship between the GEH statistic for an individual observation versus the GEH statistics calculated for an average flow of a sample of observations. As before, GEH_K can be calculated using formula (Eq.8).

Assume that our error in observations is proportional to the long-run average flow, namely:

$$K - \mu = \xi\mu, \quad \text{(Eq. 14)}$$

where ξ is a coefficient of proportionality.
Then

$$GEH_K = \sqrt{\frac{2(K-\mu)^2}{K+\mu}} = \sqrt{\frac{2\xi^2\mu}{\xi+2}} \ \text{(Eq. 15)},$$

and the long-run average can be computed as

$$\mu = \frac{(\xi+2)GEH_K^2}{2\xi^2}. \quad \text{(Eq. 16)}$$

Suppose that we have a set of observations K_j for the same link over a particular time period. Let $\overline{K}$ be an average of K_j. Then the corresponding GEH statistic is

$$GEH_{\overline{K}} = \sqrt{\frac{2(\overline{K}-\mu)^2}{\overline{K}+\mu}}\,. \quad \text{(Eq. 17)}$$

Similar to (Eq.14), let $\overline{K} - \mu = \varphi\mu, \quad \text{(Eq. 18)}$
which gives

$$GEH_{\overline{K}} = \sqrt{\frac{2\varphi^2\mu}{\varphi+2}} = \frac{\varphi}{\xi}GEH_K\sqrt{\frac{\xi+2}{\varphi+2}}. \quad \text{(Eq. 19)}$$

Normally we would expect the value of φ being smaller than the value of ξ, and therefore $GEH_{\overline{K}}$ being smaller than GEH_K.

An alternative approach can be derived to estimate the relationships between GEH_K and $GEH_{\overline{K}}$. Let $f(x)$ be a function of observed x and long-run average flow:

$$f(x) = \frac{2(x-\mu)^2}{(x+\mu)} \quad \text{(Eq. 20)}$$

And let $f_\mu(y) = f(y-\mu)$ be a function of $y = x + \mu$, i.e. $f(x) = f_\mu(x+\mu) = f_\mu(y)$:

$$f_\mu(y) = \frac{2(x-2\mu)^2}{y}. \quad \text{(Eq. 21)}$$

The first derivative of $f_\mu(y)$ is then

$$f_\mu^{'}(y) = 2 - \frac{8\mu^2}{y^2}, \quad \text{(Eq. 22)}$$

whilst the second derivative is

$$f_\mu^{''}(y) = \frac{16\mu^2}{y^3}. \quad \text{(Eq. 23)}$$

Using Taylor's series expansion of a function with the Lagrange reminder, we get

$$f(x) = f(x-\mu+\mu) = f_\mu\big(2\mu + (x-\mu)\big)$$
$$= f_\mu(2\mu) + f_\mu^{'}(2\mu)(x-\mu) + \frac{f_\mu^{''}(\theta)}{2}(x-\mu)^2 = \frac{f_\mu^{''}(\theta)}{2}(x-\mu)^2. \quad \text{(Eq. 24)}$$

where $\theta\epsilon[2\mu, x]$.

Let us assume that the values of x can vary in a range of $x\epsilon(\mu - c\sigma, \mu + c\sigma)$, where c is a constant, in which case $y\epsilon(2\mu - c\sigma, 2\mu + c\sigma)$, and where σ denotes the standard deviation of x.

The expectation of f(x) is

$$\mathbb{E}\big(f(x)\big) = \mathbb{E}\left(\frac{f_\mu''(\theta)}{2}(x-\mu)^2\right) \geq \frac{\min_\theta f_\mu''(\theta)}{2}\sigma^2 = \frac{8\mu^2}{\theta^3}\sigma^2 \geq \frac{8\mu^2}{(2\mu + c\sigma)^3}\sigma^2, \quad \text{(Eq. 25)}$$

For the reasons described in the previous paragraph, and especially for a large sample, it seems better to use Normal approximation rather than Poisson. Note that for a Normal distribution, nearly all values (99.7%) lie within 3 ($c = 3$) standard deviation of the mean, 95% lie within 2 standard deviations ($c = 2$) and 68% lie within 1 ($c = 1$) standard deviation from the mean. These confidence intervals do depend on the normal approximation; for a skewed count with a heavy tail of "low" values the above limits might not quite be true.

Similarly, if we assume that $2\mu - c\sigma > 0$,

$$\mathbb{E}(f(x)) \le \frac{8\mu^2}{(2\mu - c\sigma)^3}\sigma^2, \quad \text{(Eq. 26)}$$

resulting in

$$\frac{8\mu^2}{(2\mu + c\sigma)^3}\sigma^2 \le \mathbb{E}(f(x)) \le \frac{8\mu^2}{(2\mu - c\sigma)^3}\sigma^2. \quad \text{(Eq. 27)}$$

When $2\mu - c\sigma < 0$, the boundaries of (Eq.28) are reversed.

For a series of n observations with the standard deviation, a comparable estimate is

$$\frac{8\mu^2}{(2\mu + c\sigma_n)^3}{\sigma_n}^2 \le \mathbb{E}\left(f\left(\frac{\sum x}{n}\right)\right) \le \frac{8\mu^2}{(2\mu - c\sigma_n)^3}{\sigma_n}^2. \quad \text{(Eq. 28)}$$

Again, when $2\mu - c\sigma_n < 0$, the boundaries of (Eq.28) are reversed.

Then, taking into account that ${\sigma^2}/{{\sigma_n}^2} = n$

$$\frac{\frac{8\mu^2}{(2\mu + c\sigma)^3}\sigma^2}{\frac{8\mu^2}{(2\mu - c\sigma_n)^3}{\sigma_n}^2} \le \frac{\mathbb{E}(f(x))}{\mathbb{E}\left(f\left(\frac{\sum x}{n}\right)\right)} \le \frac{\frac{8\mu^2}{(2\mu - c\sigma)^3}\sigma^2}{\frac{8\mu^2}{(2\mu + c\sigma_n)^3}{\sigma_n}^2}, \qquad \text{(Eq. 29)}$$

$$\frac{n(2\mu - c\sigma_n)^3}{(2\mu + c\sigma)^3} \le \frac{\mathbb{E}(f(x))}{\mathbb{E}\left(f\left(\frac{\sum x}{n}\right)\right)} \le \frac{n(2\mu + c\sigma_n)^3}{(2\mu - c\sigma)^3}, \qquad \text{(Eq. 30)}$$

And therefore, under the assumptions made, the ratio of $\mathbb{E}(f(x))$ to $\mathbb{E}\left(f\left(\frac{\sum x}{n}\right)\right)$ can be estimated as

$$n\frac{\left(1 - \frac{c\sigma}{2\mu\sqrt{n}}\right)^3}{\left(1 + \frac{c\sigma}{2\mu}\right)^3} \le \frac{\mathbb{E}(f(x))}{\mathbb{E}\left(f\left(\frac{\sum x}{n}\right)\right)} \le n\frac{\left(1 + \frac{c\sigma}{2\mu\sqrt{n}}\right)^3}{\left(1 - \frac{c\sigma}{2\mu}\right)^3}. \quad \text{(Eq. 31)}$$

4. VARIABILITY OF THE DFT ATC DATA

This section reports on the analysis of the Department for Transport's (DfT) automatic traffic counter (ATC) data which was carried out in order to investigate daily variability of the data. The dataset consists of hourly counts from 54 sites across London. Only 45 sites were used in the analysis due to either missing or negative counts[3]. The analysis was done on a sample of Monday to Friday counts of

[3] It is unknown to us how the database was processed and what lead to negative counts

light goods vehicles over the period February 2008 to June 2008 for one hour periods commencing 08.00 to 18.59, excluding major holidays. Traffic estimates are calculated for each link of Britain's major road network, with links' start and end points defined as where links join a major road junction. The roads are split into 6 different DfT classifications: Motorways, Principal roads, B roads, C roads, Trunk roads, Unclassified roads.

Table 1. DfT automatic traffic sites within London included in analysis: summary

ATC site	**Central**	**Inner**	**Outer**	**Total**
Motorway	0	0	1	*1*
Trunk	0	3	6	*9*
Principal	3	9	6	*18*
B	1	2	1	*4*
C	0	0	1	*1*
Unclassified	2	5	5	*12*
All	*6*	*19*	*20*	**45**

Table 1 summarises the number of sites used in the analysis by road type and area: Inner; Outer; and Central London. Taking into account that the overall number of sites is not very large (45 sites) and the presentation of some road types is very limited (e.g. only 1 motorway and 1 C-road site), the results in terms of the typical representation of a particular road type need to be treated with caution.

<u>ATC data: average flow</u>

Vehicle flows at monitoring points on the network in terms of average hourly flows between 7am and 7pm are presented in Table 2. The hourly flows varied from just about 80 vehicles during this time period on unclassified roads (with the highest of about 180 vehicles per hour on unclassified roads in Central London and the lowest of about 60 vehicles on unclassified roads in Inner and Outer London) to about 2100 vehicles per hour on the M1 (junctions 2-4) at Barnet and on trunk roads in Outer London. Flows of over 500-800 vehicles per hour were recorded on principal roads in all areas of London, as well as on trunk roads in Inner London and B-roads in Outer London.

Table 2. DfT ATC data: area average hourly flows, 7am-7pm

Road type	**Central London**	**Inner London**	**Outer London**	**Average**
Motorway			2091	2091
Trunk		734	2119	1607
Principal	611	514	705	593
B	247	203	751	372
C			363	363
Unclassified	183	62	62	82

Table 3. DfT ATC data: day of week average hourly flows, 7am – 7pm

Road type	Monday	Tuesday	Wednesday	Thursday	Friday
Motorway	1994	2064	2102	2115	2172
Trunk	1574	1591	1604	1624	1640
Principal	579	589	596	598	603
B	357	370	376	380	377
C	351	359	365	367	371
Unclassified	78	83	83	83	83

Table 3 and Table 4 show average hourly vehicle flows by road type, during 7am-7pm, for day of the week; and area respectively. The variation of traffic throughout weekdays is relatively small, about 1%, with the highest variation on Mondays (up to 5%). Wednesday's hourly flows are the closest to the 5-day average hourly flows for different road types and areas. Typical hourly flows in Central and Inner London are similar at about 400 vehicles per hour, whist the average hourly flows in Outer London are about 2.3 times higher.

Table 4. DfT ATC data: average hourly flow by area and days of the week (7 am – 7pm)

Area	Monday	Tuesday	Wednesday	Thursday	Friday	average
Central	387	400	402	402	402	399
Inner	391	400	405	408	409	403
Outer	912	929	932	950	953	936

ATC data: variance to mean ratio

The variance-to-mean ratio, or the index of dispersion, is a measure used to quantify whether a set of observed occurrences are clustered or dispersed compared to a standard statistical model. As it was mention above, the Poisson distribution has equal variance and mean, and therefore the variance-to-mean ratio is equal to 1. If the coefficient of dispersion is less than 1, the dataset is assumed to be under-dispersed (not dispersed when the index is equal to 0), and this often relates to some more regular patterns of occurrence compared to a random pattern. If the coefficient of dispersion is greater than 1 (over-dispersion) this often means that a dataset is clustered. As shown in Section 1, the GEH statistic appears to be testing whether observed and modelled flows can reasonably be assumed to be drawn from the same Poisson-like distribution with the variance six times the mean for individual links. The purpose of this section is to understand average hourly variance-to-mean ratios for observed data and the variation between different geographies, road types, days of week and time periods.

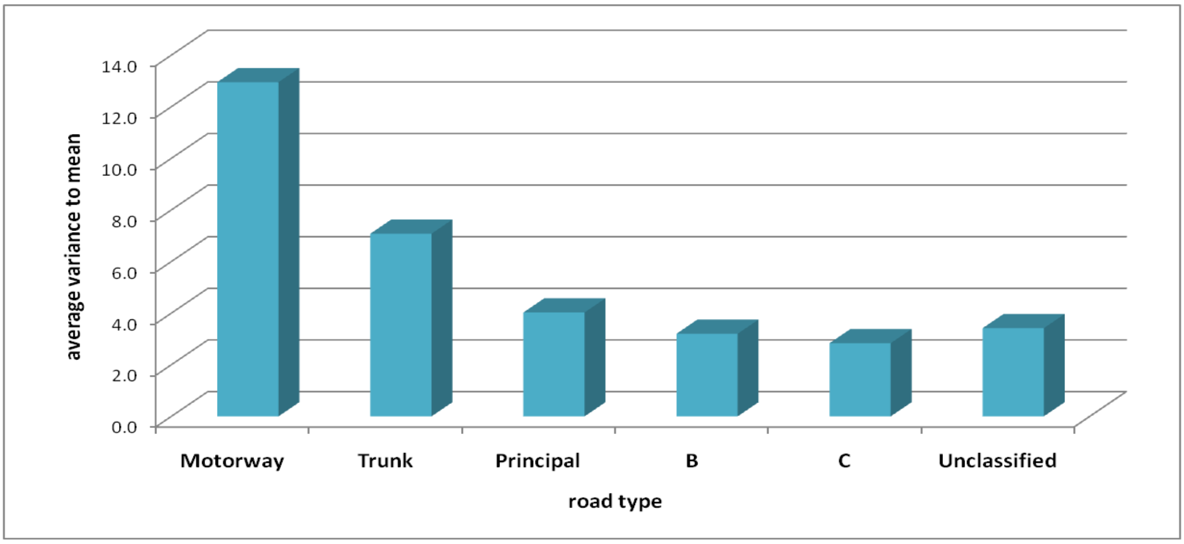

Figure 1. DfT ATC data: hourly variance-to-mean ratios for different road types (7am-7pm)

Figure 1 illustrates the average hourly variance-to-mean ratios calculated for the 7am-7pm time period from the DfT ATC data for different road types. The highest value of 13.0 is reported for the M1. The coefficient for trunk roads is the second highest (7.1), with the coefficients for the other road types in the range of around 3-4, with the lowest of 2.8 for C-roads. Outer London variability is the highest among the three areas with the variance-to-mean coefficient of 5.5 compared to 3.6 for Central London and 3.7 for Inner London. The am peak (8:00-9:00) and mid-morning time period (11:00-12:00) have greater fluctuations and the other time periods. The variance-to-mean ratios for those time periods are about 1.5, 6.6 and 6.3 times higher respectively, compared to the other time periods under consideration which has an average of about 4.

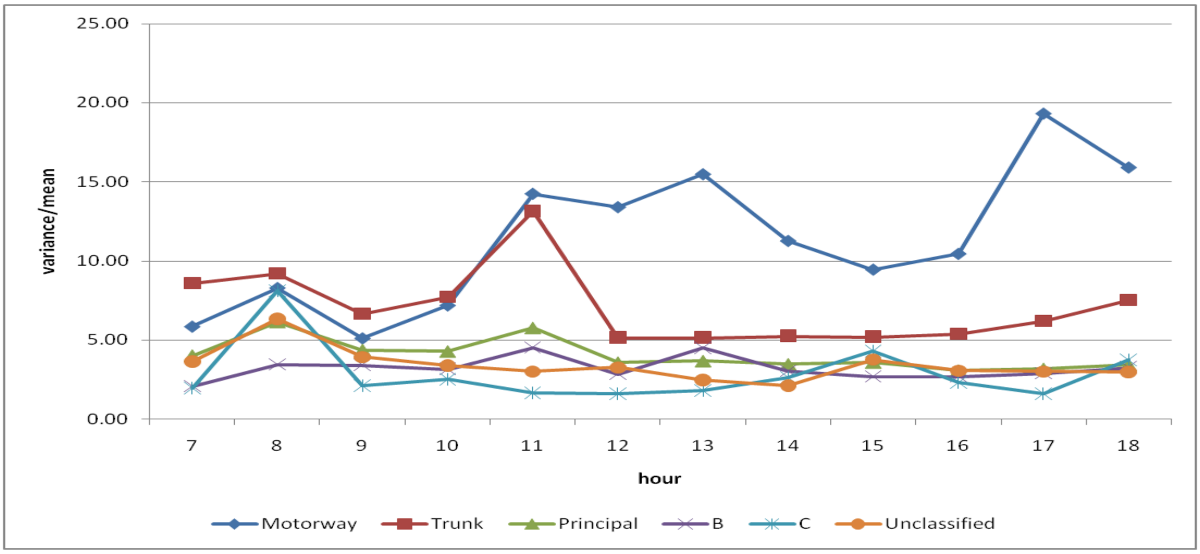

Figure 2. DfT ATC data: variance to mean ratios for different road types and time periods

Further disaggregation by time period and road type is presented in Figure 2. Different road types demonstrate different patterns for different time periods. For example, the M1 variance-to-mean ratios are the highest for the inter-peak and the evening peak time periods whilst the C-road type site (Harmondsworth Road, Hillingdon) shows the highest coefficient of variation of 8.1 for 8-9 time period and a more stable pattern throughout the rest of the day. The traffic flow variability is the highest in Outer London during the morning and inter-peak especially between 8-9 and 11-12. The most variable hour in Inner London is 8am-9am with the average coefficient of variation going up to 6.4.

One might expect two different sources of variation: an underlying variance in demand (especially for discretionary/inter-peak trips, or due to weather, etc.) and an underlying variation in supply characteristics, e.g. due to accidents or ongoing queues from earlier time periods. Particularly where the M25 is concerned, issues earlier in the day can cause knock-on effects well into the inter-peak, which seems a possible explanation for the pattern observed above. Perhaps this does not matter in as much as we are interested in magnitude rather than the cause, but it does raise questions about correlation of errors and outliers, and whether an outlier that is discarded in the AM peak should also be considered for discarding in other periods. It would also be interesting to know how much of this variance is below the mean and how much is above, and whether the indices of dispersion should be asymmetric in order to address the censoring introduced by capacity constraint.

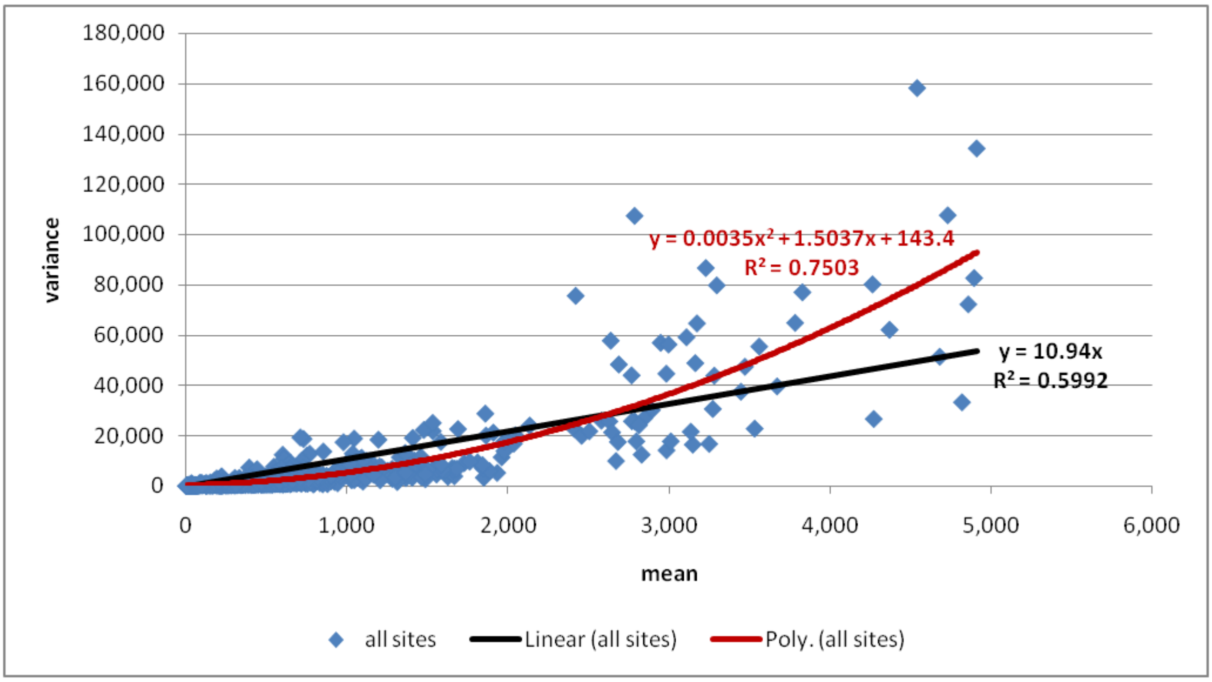

Figure 3. DfT ATC data: the average hourly flow mean-variance correlation and trend lines: all sites

One of the assumptions made in evaluation of the GEH statistic is the proportionality of the flow variance to the mean. The strength of the association between these two statistics is presented in Figure 3, and it can be seen that the linear proportionality assumption is a strong assumption and it does not hold for the ATC hourly flow variability. A linear proportionality and a quadratic trend lines are fitted. The coefficient of determination[4] R^2 associated with the quadratic trend is 0.75 compared to that of the linear proportionality trend of 0.6, which means that 15% more of the variation in the variance of traffic flows can be explained, or accounted for, by the variation in the flow mean. The estimated variance-to-mean ratio for all sites is 10.94.

ATC data: GEH

Calibration and validation of a model forms a crucial element of the model development process through which confidence in the model results can be ascertained. Variations between the model and observed data are normally expected and the responsibility is upon the model developer to establish a desired reliability level and the validation effort required to achieve it. GEH statistic (see Eq.1) is often used for this purpose. The aim of this section is to look at the GEH statistic calculated for observed vs. the long-term average observed values (as in Eq.8) in order to understand the GEH-variability of observed data. As before, the analysis is based on the DfT ATC dataset.

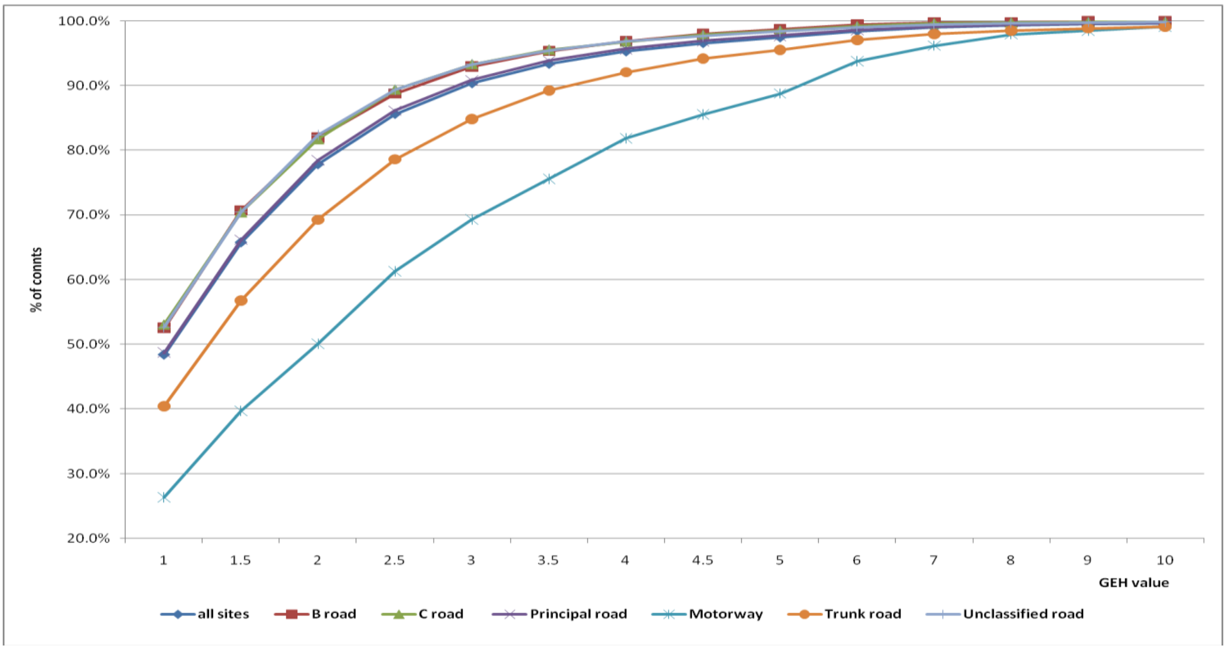

Figure 4. DfT ATC data: GEH values for different road types and time periods: cumulative %

The analysis shows that overall the GEH values calculated for individual hourly link flows relative to the long-term average vary enormously, namely, between 0 and 67, but values of GEH higher than 10 occur in less than 1% of cases. The range of values for all sites is between 1.33 (for 14-15 time period) and 1.79 (for 8-9 morning peak). The data calculated for the 3 time periods also show the highest GEH value of 1.55 for the am peak compared to the average of 1.37-1.38 for the inter-peak and evening-peak periods. M1-based average GEH values are the highest among all road types with GEH average for truck roads being the second highest. The

[4] The coefficient of determination is the proportion of the total variation in the dependent variable (the variance) that is explained, or accounted for, by the variation in the independent variable (the mean).

distribution patterns are also different for different road types. The cumulative distribution of GEH values by road type is presented in Figure 4. It can be seen the B-, C- and unclassified roads have the lowest variability with about 90% of counts lying within GEH of 2.5. The GEH variability patterns are very similar for different days of the week with Mondays' values being slightly higher. About 90% of counts have the GEH value of less than 3. The 8-9 morning peak is usually one of the most variable with about 85% of counts having the GEH value of less than 3 and 94% of less than 5. As with the variance-to-mean ratios, Outer London has higher GEH values compared to other London areas, whereas the variability in Inner London is the lowest in most time periods.

Further analysis can be done to see how the GEH variability relates to the mean and the variance of the counts. It would be interesting to try to tease this out: is the GEH distribution more to do with absolute load (which is plausible) or does the fact of being a dual carriageway change the distribution in itself even at the same level of load?

5. VARIABILITY OF ENCLOSURE DATA

This section reports on an analysis of the ATC datasets for three enclosures, namely in the Bexley, Hornchurch and Wembley areas. The survey was carried out during a period of three weeks covering 22 February to 12 March inclusive in order to maximise the chance of obtaining two full weeks of weekdays survey data between 7am and 7pm. Unfortunately, a number of sites in each enclosure were affected by incomplete data at different points during the three weeks of the survey, and therefore the numbers of days included in whole-enclosure analysis for the Wembley and Hornchurch enclosures were reduced. In addition, due to substantial roadworks, whole enclosure analysis for Bexley was unavailable.

The purpose of the current analysis was to look at the day-to-day variability of the data by calculating the average variance –to–mean rations and GEH statistics for the three enclosures. The results need to be treated with caution as the sample sizes are small for all enclosures.

Table 5, Table 6 and Table 7 present the results for the Wembley, Bexley and Hornchurch enclosures respectively based on the total flows for both in and out directions. The 8am-9am average flow level for the Wembley and Bexley enclosures are similar and just over 30,000 vehicles. The day flows for this time period at Hornchurch enclosure are 1.7 times higher and is about 52,000 vehicles.

A higher level of variability is observed in Wembley and Hornchurch enclosures with the variance-to-mean rations of about 19 and 13, compared to that of about 5 for the Bexley enclosure. Separating inbound and outbound enclosure directions reduces the variability of data by not considerably. For example, the variance-to-mean ratio for Wembley enclosure for the outbound direction is reduced to 17 and that for the inbound direction is reduced to 10.

Table 5. Enclosure counts: Wembley enclosure summary, 8am-9am

Wembley Enclosure				
average flow	30805			
median	30985			
average GEH	3.48		**variance**	595547
max GEH	10.91		**variance/mean**	19.33
85%tile	4.77			

GEH	Frequency	Cumulative	Cumulative %
0.5	0	0	0%
1	1	1	7%
1.5	2	3	21%
2	0	3	21%
2.5	1	4	29%
3	4	8	57%
3.5	1	9	64%
4	2	11	79%
4.5	0	11	79%
5	1	12	86%
6	1	13	93%
7	0	13	93%
8	0	13	93%
9	0	13	93%
10	0	13	93%
11	1	14	100%
12	0	14	100%
13	0	14	100%
14	0	14	100%
15	0	14	100%

The variance-to-mean rations and GEH values are correlated and therefore it is not surprising that the highest value of GEH is also associated with the Wembley enclosure (maximum GEH is about 10.91 and the 85^{th} percentile is 4.77). Bexley enclosure demonstrates smaller variability: the average GEH is 1.79, almost half the corresponding value for the Wembley enclosure, the maximum GEH is 3.79, and the 85^{th} percentile's value is 2.82. The Wembley and Hornchurch enclosures experience more variation in outbound flows than inbound flows. This trend is reversed for the Bexley enclosure.

About 80% of counts for the Bexley enclosure have GEH values less than 3 contrasting to those of 57% and 53% for the Wembley and Hornchurch enclosures respectively. The maximum observed variability of the individual links is higher than that of enclosures with the variance-to-mean values ranging up to 38.5 and the GEH values ranging up to 67, but the average variability of the enclosure data is higher (Figure **5**) with the average variance-to-mean ratio of 12.5 and the average GEH value of 2.7 compared to those for individual links of 4.4 and 1.79 respectively. 18% of the enclosure counts have GEH values less than 1 contrasting to 48% for individual links. The difference is gradually reduced and the GEH value of 4, the percentages of counts are 85% and 95% respectively.

Table 6. Enclosure counts: Bexley enclosure summary, 8am-9am

Bexley Enclosure

average flow	30771		
median	30752		
average GEH	1.79	**variance**	147471
max GEH	3.79	**variance/mean**	4.79
85%tile	2.82		

GEH	Frequency	Cumulative	Cumulative %
0.5	2	2	18%
1	1	3	27%
1.5	1	4	36%
2	3	7	64%
2.5	2	9	82%
3	0	9	82%
3.5	1	10	91%
4	1	11	100%
4.5	0	11	100%
5	0	11	100%

Table 7. Enclosure counts: Hornchurch enclosure summary

Hornchurch Enclosure

average flow for enclosure	51764		
median	52011		
average GEH	2.90	**variance**	694544
max GEH	7.82	**variance/mean**	13.42
85%tile	4.56		

GEH	Frequency	Cumulative	Cumulative %
0.5	1	1	7%
1	2	3	20%
1.5	2	5	33%
2	2	7	47%
2.5	1	8	53%
3	0	8	53%
3.5	1	9	60%
4	3	12	80%
4.5	0	12	80%
5	1	13	87%
5.5	0	13	87%
6	0	13	87%
6.5	1	14	93%
7	0	14	93%
7.5	0	14	93%
8	1	15	100%

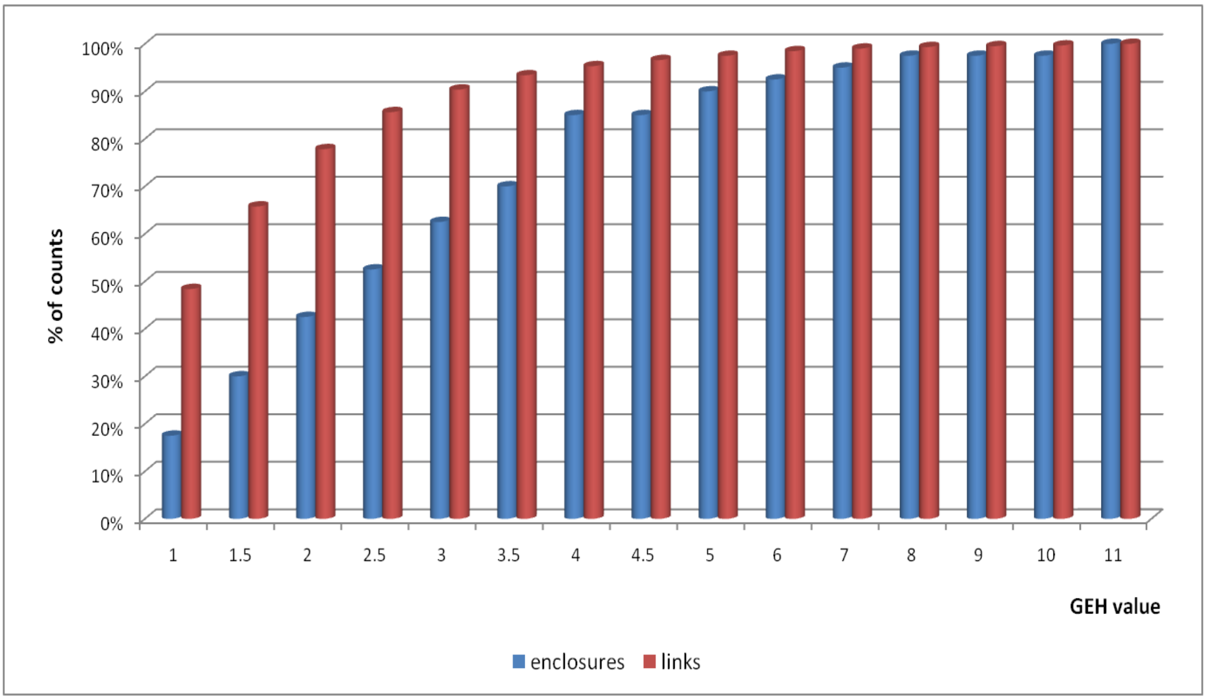

Figure 5. ATC enclosure data and DfT ATC link data: GEH distributions of the combined Wembley, Bexley and Hornchurch enclosures datasets and links (all types): cumulative %

6. VARIABILITY OF ETHE TURNING COUNT DATA

In 2009 TfL carried out collection, processing and analysis of turning movement data at six junctions. Surveys were undertaken at each junction on five consecutive weekdays from Monday 30th November to Friday 4th December with the exception at one site where due to technical problems evening peak and inter-peak surveys for Monday 30th were repeated on Monday 7th December. The turning counts were disaggregated by junction, vehicle type, day and time period (to 15 minute interval) and turning movements, and analysed at three levels: individual turning movements, single direction link flows and total junction throughput using two-hour period flows. This section summarises a follow-on analysis of the data in terms of variance-to-mean rations and GEH values. The focus of the analysis is on the variability between comparable data on different survey days. The investigation was carried out for six one-hour time periods, namely 7am-8am, 8-am-9am, 12pm-1pm, 1pm-2pm, 16pm-17pm, 17pm-18pm.

According to the data, the average index of dispersion for cars during 4pm-6pm is higher compared to the other time periods considered, with the average variance-to-mean value peaking at 3.4, about 2.6 times higher than that for 7am-8am time period. The average variability of GEH values by time of day is smaller and is in the interval [2.2, 3.0] with the top of the range associated with the 5pm-6pm time period.

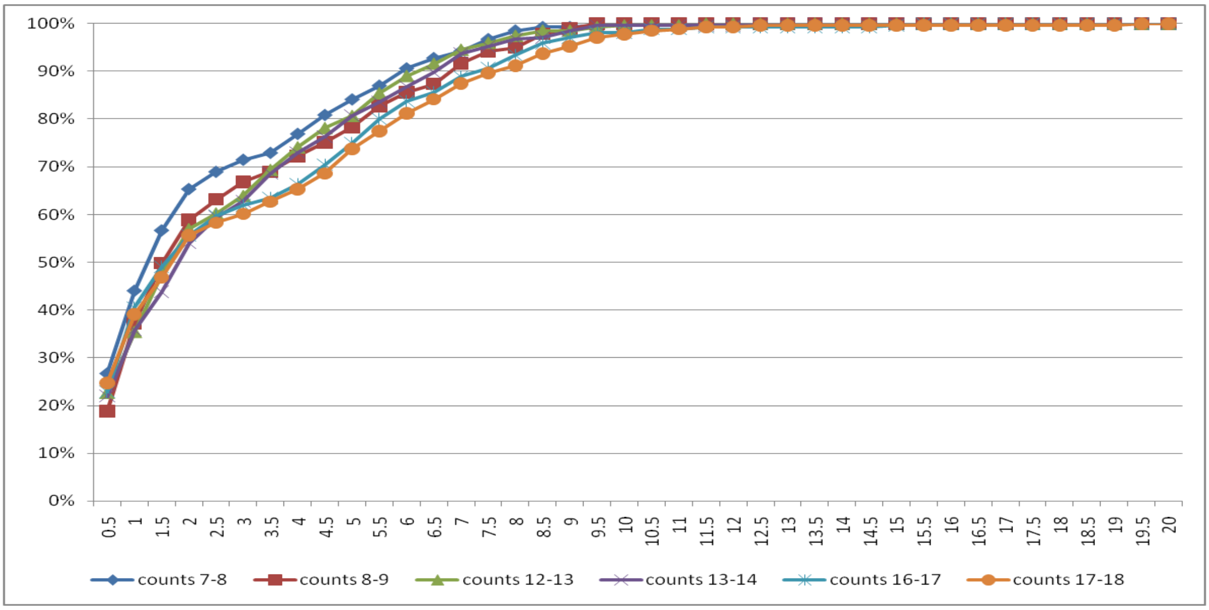

Figure 6. Turning counts data: GEH values for cars for different time periods, cumulative %

Figure 6 compares the cumulative GEH distributions for turning counts and principal road links. About 40% of counts have a GEH value of less than 1, 60% of less than 2, 90% of less than 7, and 99% of less than 10. As most of the junctions surveyed are principal road junctions, it is more appropriate to compare GEH values and variance-to-mean ratios associated with the turning count dataset for those associated with the principal road links rather than with the statistics for all links. The variance-to-mean ratio for principal road links ranges from 3.06 (during 4pm-5pm) to 6.11 (8am-9am) with the highest index during the morning peak. These values are higher than for the turning count data. In addition, the distribution throughout the day is very different, almost reversed. The GEH statistic for principal road links can go up to 44 with the average values ranging from 1.26 (4pm-5pm) to 1.75 (8am-9am).

7. CONCLUSION

The predominantly used UK validation criteria are provided by the Design Manual for Roads and Bridges (DMRB) and are strongly biased towards highway models and depend significantly on the GEH measure developed by Geoff E. Havers. The GEH statistic is widely used in traffic modelling, forecasting and engineering to compare two sets of traffic volumes one of which is an observed dataset and one is predicted by the model. Although its mathematical form is analogous to a chi-squared statistic, GEH is not a true statistical test but an empirical formula.

The inconsistency of data is often one of the main problems facing transport modellers when calibrating and validating models, and the questions of reliability of the observed data, when the data was collected, whether it was an “average” day, etc are some of the issues which a modelled needs to know. The more data is collected, the more reliable a model may be but it may make it more difficult to calibrate and validate. It is also important to note that data collection is a very expensive task both in terms of collecting the data as well as processing and analysing the data subsequently. A balance needs to be achieved between collecting enough data so that the model is fit for purpose and not exceeding the budget and timescale of the project.

This note reports on the additional theoretical findings as well as further analysis of different ATC data sets in order to get a better understanding of the nature of the GEH statistic. The research shows that the variability of traffic counts between comparable data on different survey days is significant and depends shows a discrepancy between different road types, time of a day, area, etc. The variance-to-mean ration can go up to 17, whilst the GEH values can go up to 67. In addition, the data suggest that some theoretical assumptions such as proportionality of variance to mean are quite strong assumptions and do not always hold.

The goodness-of-fit measures suggested by the DRMB guidance about supplying assessments of transport model validity should take account of the variability of the count data. This is currently not included in the GEH metric and this paper suggests that it should be considered as an integral part of any criterion. The GEH measure also confronts the significant issue of how to distinguish that a given level of absolute difference and of percentage difference can have very different levels of significance depending on the scale of the flows.

REFERENCES

Oliver, R.M. (1962). A note on a traffic counting distribution, Operational Research Quarterly, 13 (2), 171-178.

QUORAM (2008). Quality of Regional Assignment Models, MVA report to TfL.

Vaughan, R.J. (1970). The distribution of traffic volumes, Transportation Science, 14 (1), 97-110.

Yi, H., Mulinazzi, T.E., and Lee, J. (2005). Traffic-count based distribution model for size impacts studies, Journal of Transportation Engineering, 31 (4), 286-293.