VEHICLE PROBES AS REAL-TIME ATMS SOURCES OF DYNAMIC O-D AND TRAVEL TIME DATA

M. Van Aerde, B. Hellinga, L. Yu, and H. Rakha
Department of Civil Engineering
Queen's University, Kingston, Canada
Fax.: 613 - 545 - 2128

Abstract

The IVHS field operation studies, that are being carried out at present, have demonstrated that the communications technology, on-board electronics and user interface of in-vehicle route guidance systems are rapidly approaching a state of development which, while not completely mature, could certainly support the deployment of several types of practical systems. Lagging considerably behind in this development is the availability of adequate real-time data to support such systems, as the data that is provided from UTCS and FTMS sources have clearly been shown to be deficient in terms of providing reliable link travel time data (especially on surface streets), and estimating real-time O-D data for the entire driver population. The proposed solution, to the latter traffic engineering problem, is expected to be the deployment of sufficient numbers of vehicle probes to directly provide such data when drivers start their trips (O-D) and traverse the various links (link travel time). Unfortunately, in view of the fact that sparse vehicle probe data has also been shown to have a considerable error component associated with it, an assessment of exactly what level of market penetration is necessary to provide adequate probe samples has become critical. This paper provides a combination of analytical development and simulation results to answer this question. Both analytical methods are developed for a general network, and are illustrated in detail for a specific example arterial/freeway network.

The analytical travel time work is based on an assessment of how the variance, as opposed to the mean travel time, of surface streets and freeways varies as a function of traffic demand levels. From this variance, sampling statistics is utilized to determine what fraction of the total probe population is likely to be observed on each link during a given time slice, and how this number of probes will decrease the confidence limits about the estimate of the mean travel time. The analytical vehicle probe work presents a similar examination of how many vehicles are likely to depart for each O-D pair within each time slice and determines when the associated sampling statistics permit statistically significant inferences to be drawn about the dynamic structure of the O-D demand pattern

The simulation work is introduced to provide a graphical illustration of how the above analytical work can be applied for a specific network. Furthermore, the explicit use of the simulation model's ability to generate real-time travel time and O-D data for a wide range of different demand patterns permits an overlaying of simulated experimental data on top of the above analytical projections. The agreement between the analytical and the simulation results is then put forward as a basis for utilizing the simulation model for performing probe assessments in networks which are too complex to permit an analytical evaluation. In addition, the overall results will be useful as a projection of the expected field performance of further operational field tests or complete systems which will rely on probe data as their main source of reliable real-time travel time and O-D data.

1.0 Introduction

With the current field trials and anticipated implementation of Route Guidance Systems (RGS) much effort has been expended considering system and user benefits that can be obtained when RGS equipped vehicles receive traffic information in near real-time. However, much less research has been conducted investigating how these near real-time data might be acquired and how reliable these data must be in order to be of benefit. The provision of two-way communications between the Traffic Management Center (TMC) and RGS equipped vehicles makes it possible to very efficiently assemble data describing traffic conditions on the network in near real-time. This paper examines the use of RGS equipped vehicles as sources of near real-time traffic information.

Two types of data that can be obtained from probe vehicles are; link travel time data and origin-destination (O-D) data. As each probe vehicle traverses a link, its travel time is transmitted back to the TMC, providing a sample of the current travel time experienced by all drivers utilizing that link at that time. When drivers of RGS equipped vehicles initiate a trip, the intended trip destination is entered into the on-board computer. The vehicle's current location (trip origin) and intended destination is then transmitted to the TMC, providing a sample of the population O-D demand matrix for that time period.

Two critical issues must be addressed when using probe vehicles as sources of traffic information. First, is there a sufficient quantity of data to provide for adequate coverage of the network. Second, are the data of sufficient quality to be of benefit to RGS equipped vehicles. This paper describes two different analysis approaches that assess both of these issues. The first approach is analytical, and examines the quantity and expected quality of probe data using statistical sampling theory. The second is empirical, and makes use of a network traffic simulation model capable of modeling probe vehicles. Both approaches are developed for application to general networks, and are illustrated in detail using a simple arterial/freeway network.

Previous research has investigated the potential of using probes as sources of O-D information [1] and link travel times [2]. However, much of this previous research has tended to only perform an analytical analysis. Unfortunately, analytical techniques often can not be practically utilized for realistic networks.

This paper has three objectives. First, the process of obtaining network traffic data from probe vehicles is examined. Second, statistical sampling equations are developed that provide estimates of the reliability of these probe data. Third, using an example network, the traffic simulation model results are verified against the analytical results. The validation of simulation based results permits the simulation model to be used for more complex situations for which statistical analyses are no longer practical.

Section 2 of this paper describes the example arterial/freeway traffic network and the process used to obtain realistic O-D demands and routes. Section 3 discusses the practical problem of estimating the prevailing level of market penetration of RGS equipped vehicles. Section 4 examines the ability of probe vehicles to provide link travel time information. Results for the example network from both the analytical and simulation analyses are provided. Section 5 investigates the accuracy of estimating O-D demands from probe data. Again results are provided for the example network.

2.0 Network Description

In order to illustrate a simulation analysis of the potential of vehicle probes to provide link travel time and O-D information, an example network is required. To permit an analytical analysis to also be performed for the same network, a simple arterial/freeway network was defined. However, in order to provide meaningful results, the network characteristics were developed to be as realistic and objective as possible.

2.1 Network structure:

Figure 1 illustrates the example network configuration. The network consists of 7 zones at the periphery of the network, 5 traffic signals, and 40 directional links. Fourteen of the 40 links are zone connectors, 8 are single lane arterials, 10 are dual lane ramps, and the remaining 8 links are 2 lane freeways.

A generalized non-linear speed-density relationship was utilized for all links. The associated speed - flow relationship can be described by the free speed, saturation flow rate, speed at capacity, and jam density [3]. The parameters chosen for the arterial, ramp, and freeway links are provided in Table 1. The freeway link parameters were obtained from a simulation study conducted on a multi-lane freeway in Toronto, Ontario [4].

Platoon dispersion was modeled using the same method as the Transyt model for the arterial and ramp links. For these links, a platoon dispersion factor of 0.35 was used. For the freeway links, platoon dispersion was assumed to follow a gaussian distribution. Examination of data from Highway 401 in Toronto indicated that a coefficient of variation of 10% was appropriate.

Since signal timings contribute significantly to variations in arterial link travel times, realistic signal behavior was required. Cycle lengths utilized by the model were pre-specified to be 60 seconds prior to 7 AM and 80 second after 7 AM. The simulation model optimized the phase splits of each signal every minute using the Canadian Capacity Guide's method [5] of balancing the degree of saturation on the approaches to the intersection.

2.2 O-D demands and routes:

To sensibly analyze link travel time variations, realistic commuting peak growth and decay dynamics must be present within the O-D demands. Manually generating an O-D which reflects these dynamics and is unbiased is rather difficult to do, even for a simple 7 zone network. A three stage approach, illustrated in Figure 2, was devised to determine unbiased O-D demands that would represent the growth and decay characteristics typically found in a commuting corridor during the 8 hour period from 4 AM to 12 PM.

First, a time series of average 15 minute link flows was determined for a typical arterial and a typical freeway link. The time series for a typical arterial link was obtained from the literature [6]. The link flows for a typical freeway link were observed on a segment of Highway 401 in Toronto, Ontario.

It was assumed that the dominant direction of travel was to the east and south. It was also assumed that the split of total flow on all links of dominant to non-dominant flow was 60% to 40%, respectively. Ramp lanes were assumed to carry 40% of the associated arterial flow. Based on these assumptions, a time series of 15 minute target link flows was determined for each non-zone connector link.

These target link flows, and a set of initial routes were then used as input to the O-D estimation module [7]. Based on an information minimization formulation, the time series of static 15 minute O-D demands was estimated for the 8 hour period. The estimated O-D matrix was used as input to a user-equilibrium assignment module, such that a better estimate of the expected time series of multipath routes was obtained. These new routes, along with the original time series of target link flows were input to the O-D estimation module in order to provide a final estimate of the time series of static O-D demands.

Thus a time series of static O-D demands, as well as a time series of user-equilibrium multi-path routes was derived. The network structure, time series of O-D demands, and the time series of multi-path user-equilibrium paths were all used as input to the network traffic simulation model INTEGRATION. Detailed descriptions of the model are available in the literature [8,9].

A time series of link flows had been obtained for both a typical freeway and a typical arterial link. These were used as input to the O-D estimation process. However, it is desirable to determine how well the O-D estimation process preserved these flow profiles. Figures 3 and 4 compare the time series of link flows provided to the O-D estimation process with the time series of flows resulting from the simulation model for a selected arterial and freeway link, respectively (link 16 and 33) . The freeway link flow pattern input to the process and output from the simulation model are very similar. For the arterial link, a significant deviation in the flow pattern exists from 7 to 8 AM. During this period, the flow profile resulting from the simulation model is significantly lower than that provided to the O-D estimation process. Between 8 and 9 AM, simulation model flows are considerably higher than those provided to the O-D estimation process.

Two causes for this shift can be identified. First, the O-D estimation technique attempts to determine a time series of static O-D demands which, when applied to the network, minimize the squared link flow errors. Clearly, this objective is more easily attained when freeway links, which carry significantly more volume than arterial links, are more closely replicated. Since the flow patterns input to the O-D estimation process were independently obtained and did not exhibit the same peaking characteristics, discrepancies tended to be accommodated by changes to the arterial link flows. Second, over-saturation effects at both the signal controlling the arterial link outflow, and the signal controlling flow into the arterial link, cause a shift in the flows observed on the arterial link.

The INTEGRATION simulation model has the capabilities of modeling probe vehicles. Each probe vehicle's O-D and experienced link travel time is recorded. Figure 5 illustrates a sample of typical probe information obtained for the example network.

The network was simulated considering all vehicles as RGS equipped. However, though the probe vehicles reported information, they did not receive information with which to alter their routes. All routes were pre-defined, as illustrated in Figure 2. To evaluate the effect of level of market penetration, the probe logs were randomly sampled to provide levels of market penetration from 5% to 100% in 5% increments. It should be noted that in this analysis, an aggregate level of market penetration was applied. The level of market penetration did not systematically vary temporally or spatially.

Based on these random probe samples, an analysis was conducted to determine how closely estimates of link travel time and O-D demands matched the *true* data. Within the simulation analysis, the truth was defined when data for 100% probe vehicles existed. Thus, in this case, the entire finite population of vehicles was measured. In practice, this approach can be used when considering a single day, for which all vehicles could be observed. However, no two days are exactly the same, and it is often desired to be able to predict future conditions based on measured data. In this case, future conditions are part of the entire population, yet by definition, can not be measured as they have not yet occurred. Thus, for predictive purposes, the population must be considered infinite. Analytical analyses are carried out under both assumptions and are compared to simulation results for the sample network.

3.0 Level of Market Penetration

An important and essential factor in estimating the value and quality of the link travel time and the O-D demand from vehicle probes, is the level of market penetration of these vehicle probes. In estimating the level of market penetration, four levels of aggregation can be considered. First, and most aggregate, is the assumption that the level of market penetration is constant across time and across all O-D pairs (M). Second, temporal variations in the level of market penetration can be considered (M_t). Third, spatial variations may exist as different O-D pairs might exhibit different levels of market penetration (M_{ij}). Lastly, spatial and temporal variations may exist concurrently ($M_{t,ij}$). The focus of this section is to demonstrate how an estimate of the level of market penetration can be made for each of these four assumptions.

3.1 Aggregate level of market penetration:

The overall number of vehicle probes that exist within a network area can be obtained from an estimate of the number of RGS sales. However, as not all vehicles equipped with RGS would travel everyday, a better estimate could be computed based on the ratio of the number of probe vehicles on each link to the total number of vehicles on each link.

$$M = \frac{\sum a_k}{\sum b_k} = \frac{A}{B} \tag{1}$$

where: a_k = total number of probe vehicles that traversed link k
b_k = total number of vehicles observed on link k

It should be noted that A and B represent vehicle counts on links, not trips. This distinction is important as the level of market penetration is defined as the proportion of the number of RGS equipped vehicles to the total number of vehicles in the network. However, Equation 1 estimates M from the number of vehicle counts on links. If, the trip length distribution between RGS equipped and non-equipped vehicles is different, then Equation 1 introduces a bias.

3.2 Temporal variations in the level of market penetration:

It is possible, if not likely, that the level of market penetration will vary throughout the day. In order to capture these temporal variations, the level of market penetration for a certain time period, t, can be estimated based on the ratio of the number of probe vehicles to the total number of vehicles observed on a series of links during period t. This formulation is similar to Equation 1 except that link counts are made only over the previous period.

$$M_t = \frac{\sum a_{k,t}}{\sum b_{k,t}} = \frac{A_t}{B_t} \tag{2}$$

where: $a_{k,t}$ = total number of probe vehicles that traversed link k during period t
$b_{k,t}$ = total number of vehicles observed on link k during period t

The ratio of $a_{k,t}$ to $b_{k,t}$ on a given link provides an estimate of the link-specific level of market penetration for time period t ($M_{k,t}$).

Using Equation 2, it is possible to estimate the level of market penetration for the simulated example network. For an overall level of market penetration (M) of 20%, Figure 6 illustrates how $M_{k,t}$ varies over time for a typical arterial and freeway link. It can be noted that during low flow conditions (4 - 7 AM), the level of market penetration on the arterial link varies considerably. These variations from the overall level of market penetration are simply due to the randomness of the probe sample, not due to any systematic change in M. During the periods of high flow (7 - 9 AM), the variation in level of market penetration is reduced due to the increase in the number of vehicles in the sample. The freeway link experiences the same trend in variations in the level of market penetration as the arterial link.

Figure 7 illustrates the level of market penetration ($M_{k,t}$) for all of the 40 links in the network for each 15 minute time period from 4 - 12 AM. Again, the overall level of market penetration is 20%. It is evident from Figure 7, that for the initial 2½ hours of simulation (4 - 6:30 AM), the variability in the level of market penetration is quite high, due to the low flows existing during this period. Specifically, this variability ranged from $M_{k,t}$ = 0% to $M_{k,t}$ = 100%. Alternatively, during the period of peak demand (7 - 9 AM), this variability is reduced to range from $M_{k,t}$ = 10% to $M_{k,t}$ = 30%. Thus, estimating the overall level of market penetration from a single link can, depending on the flow on the link, produce highly erroneous results. Furthermore, these variations are due simply to randomness, not to any systematic change in the level of market penetration. If one desires to estimate a level of market penetration that may vary by time (M_t), an average of several different representative links across the network should be used. Any variation in the fraction of RGS equipped vehicles in the traffic stream during the day could be captured by computing the value of M_t dynamically over time. Consequently, differences in the proportion of trips made by RGS equipped vehicles during various portions of the day could be traced.

In practice, the measurement of the level of market penetration would lag by one time period, as the observed data for the time period from 8:00 to 8:15 would be utilized as an estimate of the level of market penetration for time period 8:15 to 8:30. If it is found that levels of market penetration change significantly from one period to the next, exponential smoothing might be used to provide an estimate of the level of market penetration at each time interval.

3.3 Spatial variation in level of market penetration:

In practice, it is also possible that the level of market penetration will vary by geographical area. Using a historical survey data base, it may be possible to relate selected socio-economic factors with RGS ownership. Multiplier factors for each origin zone could be computed and used to scale the overall level of market penetration. Unfortunately, this procedure could not be carried out in real-time.

In order to overcome this shortcoming , it may be possible to estimate the level of market penetration for a specific O-D pair (M_{ij}) by only estimating the percent vehicle probes on the minimum path trees for the specified O-D pair. These minimum path trees could be derived from the vehicle probe data logs. Then M_{ij} could be computed from Equation 1 except that for each O-D pair, the link subscript k refers only to those links that lie along the route from origin i to destination j. However, if different paths share common links, the number of probes on a specific link would be the summation of probe vehicles from the various O-D pairs that contribute to the flow on this link.

3.4 Spatial and temporal variation in level of market penetration:

The level of market penetration for a specific O-D pair (ij) and time interval (t) ($M_{t,ij}$), can be estimated by averaging the percent probes over the links that constitute the minimum path trees for the specified O-D during time period t. Again, these minimum path trees can be derived from the vehicle probe data logs. However, a limitation to this method is that vehicles experiencing trip durations which are longer than the specified time period will not have traversed, and thus not be counted on, all links on their path.

4.0 Link Travel Times

4.1 Analyzing the standard deviation for the individual vehicles:

In the analysis of the total link travel time, Yu, *et al* [10] separated the total time incurred while traversing a link into two independent components; the travel time on the link and the delay experienced at the downstream intersection. It should be noted that the travel time on a link is defined to be different from the link travel time. Specifically, the link travel time includes both the travel time on the link and the delay experienced when there is a downstream signal and/or a queue present. Similarly, the standard deviation of the total link travel time can also be separated into the standard deviation for the travel time on the link and the standard deviation for the delay (Equation 3).

$$s_T = \sqrt{s_L^2 + s_W^2} \tag{3}$$

where: s_T = standard deviation of the total link travel time for all links (sec)
s_L = standard deviation of the travel time on a single link (sec)
s_W = standard deviation of the delay incurred at the downstream intersection (sec)

Details for the derivation of the standard deviation for the travel time on a link (s_L) and the standard deviation for the delay at the downstream intersection (s_W) are available in the literature [11]. The resulting equations are presented below.

$$s_L = \left[abT_L\left(1+abT_L\right)\right]^{0.5} \tag{4}$$

$$s_W = \left[s_{Wu}^2 + s_{Wor}^2\right]^{0.5} \tag{5}$$

$$s_{Wu}^2 = \left(\frac{C^2(1-l)^2(2-PI)}{16(1-lU)^2}\right)\left(\tfrac{2}{3}(3l-4U+1)+PI(1-l)\right) \tag{6}$$

$$s_{Wor}^2 = \sqrt[3]{5}\,\Delta t\left(U-1+\sqrt{(U-1)^2+\frac{240U}{C\Delta t}}\right) \tag{7}$$

where: α = platoon dispersion factor of the link
β = travel time factor of the link, and ß=1/(1+α)
T_L = the average travel time on link (sec)
$s_{Wu}{}^2$ = standard deviation for the uniform delay (sec)
$s_{Wor}{}^2$ = standard deviation for the over-saturation and random delay (sec)
g = green time of downstream signal (sec)
c = cycle time of downstream signal (sec)
l = signal split (λ=g/c)
PI = Platoon Index at the stop line of downstream signal
U = volume to capacity ratio (U=V/C)
V = volume on the link (vph)
C = capacity of the downstream intersection (vph)

For uncongested traffic conditions, only uniform delay is considered in Equation 5. Under congested traffic conditions, both uniform delay and over-saturation and random delay should be considered.

4.2 Standard deviation for probes:

The accuracy of the link travel time obtained from the probe data is described by the magnitude of its standard deviation. Suppose that there are n_p vehicle probes traveling on a link and that each of these vehicle probes reports the time required to traverse the link. Then, the standard deviation of the mean reported link travel times from the probes is as follows:

$$s = \frac{s_T}{\sqrt{n_p}} \tag{8}$$

where: s = standard deviation of the mean link travel time reported by the probes
n_p = number of probe vehicles

As explained in Section 4.1, the standard deviation of link travel times for individual vehicles is a function of the platoon dispersion factor of the link, downstream arrival Platoon Index, downstream signal phase split, and the volume to capacity ratio. The number of probe vehicles is proportional to the flow on the link, the time interval for processing the probe data and the level of market penetration (Equation 9).

$$n_p = V \Delta t \, M \tag{9}$$

where: M = the level of market penetration

Equation 8 is derived based on the assumption that the total vehicle population is infinite. If the total vehicle population is known (N), then Equation 8 should be adjusted by the standard finite population adjustment factor, producing Equation 10.

$$s = \frac{s_T}{\sqrt{n_p}} \sqrt{\frac{N - n_p}{N - 1}} \tag{10}$$

4.3 Results for example network:

a. Estimating link travel times

Consider a level of market penetration equal to 20% having no systematic temporal or spatial variations. Figures 8 and 9 show the analytical estimates of mean link travel times for an arterial link and a freeway link, respectively. The actual average link travel times experienced by all vehicles is also illustrated. The 95% prediction intervals about the analytical estimate are also provided.

Figures 8 and 9 illustrate that the actual average travel times are all within the confident intervals of the estimated link travel times. It can also be seen that the confident interval for the arterial link is wider than that for the freeway link. This is because the signal timings for the arterial links result in more variation of link travel times.

b. Comparing simulation and analytical statistics

The main objective of comparisons here is to verify whether or not the analytical approach for computing the statistics of the estimated link travel times from vehicle probes is effective. If the analytical approach can be shown to be effective then, as the standard deviation and/or coefficient of variation of estimated link mean travel times are well related to the number of vehicle probes, the number of vehicle probes required for a given expected level of reliability can be easily determined.

Since the standard deviation of the estimated link travel times alone can not reflect the relative accuracy of an estimate without mentioning the corresponding magnitude of the average link travel times, the coefficient of variation for the estimated link travel times are used in the comparisons.

Figure 10 illustrates the coefficient of variation of arterial link travel time as a function of the Volume to Saturation flow ratio (V/S) for a level of market penetration of 20%. The coefficient of variation decreases rapidly as the V/S ratio increases. This decrease occurs because as the V/S flow ratio increases, the absolute number of vehicle probes also increases. The simulation results correspond well with the analytical estimates, particularly for high V/S ratios. For conditions of low V/S ratios, a higher level of market penetration is required to maintain the accuracy of the estimated link travel times. It should be noted that a coefficient of variation equal to zero can occur when, due to randomness, no probe vehicle traverses a the link during the time period.

A graph similar to Figure 10 could be produced for a freeway link, however, since the same trend exists, albeit at a lower coefficient of variation, it is not presented here.

Figures 11 and 12 show the overall averages of the coefficient of variation as a function of the level of market penetration for all the arterial links and all the freeway links, respectively. For both Figure 11 and

Figure 12, the coefficient of variation decreases as the level of market penetration increases. This relationship is particularly strong if a finite vehicle population is assumed as the coefficient of variation becomes zero if the level of market penetration is 100%. In reality, however, link travel times may vary over different days, even if the flows are the same. For this situation, the assumption of infinite vehicle population is more realistic. Under this assumption, for a level of market penetration equal to 100%, the coefficient of variation is not equal to zero.

5.0 Origin-Destination Demands

An estimate of the prevailing O-D demands is essential for the evaluation and operation of a route guidance system. O-D demands are required in order to evaluate the impacts of potential re-routing strategies and to provide estimates of future link travel times. However, deriving only an O-D estimate is not sufficient. It is also necessary to obtain some measure of the reliability of the estimate.

In this section, the derivation of an expression for the expected mean and prediction interval for the O-D demand between some origin i and destination j is presented first. Subsequently, this expression is applied to the example network and conclusions are drawn.

5.1 Derivation of an expression for mean and reliability of O-D demands:

An estimate of the prevailing demand for a specific O-D pair (ij) can be made, for some period t, based on an estimate of the level of market penetration and on the number of probe calls received during the previous time period for this O-D pair (Equation 11).

$$T_{ij} = \frac{c_{ij}}{M} \tag{11}$$

where: c_{ij} = number of probe calls received during previous time period for O-D (ij)
M = level of market penetration
T_{ij} = estimate of total O-D demand between origin i and destination j

The number of probe calls during the previous time period is known, however, the level of market penetration, is not known *a priori* and must be estimated. Section 3 discussed the difficulties of estimating level of market penetration under various assumptions. For this analysis, it is assumed that level of market penetration does not systematically vary spatially. It is assumed that temporal variations may exist and as such Equation 2 from Section 3.2 is used to provide an estimate of M_t.

Based on Equations 2 and 11 an estimate of the expected total demand between origin i and destination j can be made. However, as stated previously, it is equally important that a measure of the reliability of the estimate be made. Vehicles that are sampled are either RGS equipped, or non-equipped. The probability that a vehicle is RGS equipped is equal to M. Thus, for a finite number of trips (T), the number of probe vehicles (n_p) that exist within a sample of vehicles from the network can be considered to follow the hypergeometric distribution. From the hypergeometric distribution, the expected number of probe vehicles is,

$$n_p = T\,M \tag{12}$$

and the standard deviation of the number of RGS vehicles within a given sample size is,

$$s_{n_p} = \sqrt{T\,M(1-M)}\sqrt{\frac{B-A}{B-1}} \tag{13}$$

where:

$$T = \frac{B \cdot \bar{t}_L}{t_p} \tag{14}$$

T	=	total number of trips in the network during the previous period
$\bar{t}_L$	=	average link travel time as measured by probe vehicles (seconds)
t_p	=	length of time period (seconds)

However, it is desired to determine the standard deviation of the level of market penetration, not the number of probe vehicles. Based on an equivalent expression for the variance of a continuous random variable [12]

$$\mathrm{VAR}(n_p) = E(n_p{}^2) - [E(n_p)]^2 \tag{15}$$

it can be shown that the standard deviation of the level of market penetration is,

$$s_M = \sqrt{\frac{M(1-M)}{T}}\sqrt{\frac{B-A}{B-1}} \tag{16}$$

If it is further assumed that the number of probe calls (c_{ij}) received for a particular O-D follows the hypergeometric distribution, then the standard deviation of c_{ij} is represented as

$$s_{c_{ij}} = \sqrt{n_p p_{ij}\left(1 - p_{ij}\right)}\sqrt{\frac{B-A}{B-1}} \tag{17}$$

where: p_{ij} = probability that a trip will be between origin i and destination j (c_{ij} / n_p)

Then, the 95% prediction interval for the total number of trips expected to exist between origin i and destination j during a given period can be expressed as

$$T_{ij}^{95\%} = \frac{c_{ij} \pm 1.96 s_{c_{ij}}}{M \mp 1.96 s_M} \tag{18}$$

5.2 Results for example network:

The simulation model provides the number of probe calls (c_{ij}) received from the probe RGS equipped vehicles during each time period for each O-D pair. In addition, the number of probe vehicles on each link and the total number of vehicles on each link is available each period. In practice, the number of probe vehicles on a link can be determined from the probe logs. The total number of vehicles on a link would be obtained from loop detector counts. The average link travel time is the average of the travel time experienced by probe vehicles that traversed the link during the previous period.

An estimate of the expected O-D interaction for the origin i and destination j can be made using Equation 11. The 95% prediction interval about this mean can be computed using Equation 18. The denominator can be computed using Equations 2 and 16. The numerator can be computed from the number of probe calls received and Equation 17.

Figure 13 illustrates the estimated mean and 95% prediction intervals for the demand between origin 1 and destination 3 by time of day. The average overall level of market penetration is 20%. When one or fewer probe calls are received, it is not possible to compute the standard deviation of the level of market penetration or the 95% prediction intervals. In practice however, one can place some finite upper bound on the prediction interval. As well, physically, the lower bound must be zero. Therefore, for Figure 13, if the prediction intervals could not be computed, a lower bound of zero and an upper bound of 180 was chosen.

From Figure 13 it is clear that during periods of low flow, the 95% prediction intervals become quite large. This is consistent with intuition, as during periods of low flow, sample sizes are smaller, and thus confidence intervals must also be larger. It is also clear, that even for a 20% level of market penetration, the O-D estimates that can be made based on the probe information are rather unreliably. Furthermore, it can be safely assumed that levels of market penetration of RGS implementations will be, at least initially, substantially less than 20%.

It is instructive to examine the change in the prediction interval with changes in level of market penetration. Figure 14 illustrates the mean and prediction interval for O-D demands between origin 1 and destination 3 during the period from 7:15-7:30 AM. The estimated mean is of limited meaning as it is a function of the number of probe calls received, which is in-itself a random variable. However, the prediction intervals are based on the expected variance of the level of market penetration and of the number of probe calls. Clearly, for average overall levels of market penetration of less than 20%, the prediction intervals are so large as to lend little credibility to any estimates.

The previous discussion has been limited to the examination of a specific O-D pair. However, aggregate results for all O-D pairs in the network and for all time periods can also be examined. Figure 15 illustrates the aggregate accuracy of O-D estimates made based solely on probe information as a function of the average level of market penetration. The accuracy of the O-D estimates is measured by the root-mean squared (RMS) difference between estimated and actual O-D demands represented as a fraction of the average actual O-D demand. For an average level of market penetration of 10%, the average RMS O-D error is 60% of the average true O-D demand. If it is desired to estimate O-D demands to such a level of accuracy that the RMS error does not exceed 20% of the average observed O-D demand, Figure 15 indicates that an average level of market penetration of 40% is required.

6.0 Conclusions

The paper has attempted to quantify the impact of a number of factors which are qualitatively known to affect the quality and reliability of vehicle probe travel time and O-D data. The combined analytical/simulation approach leads to the following main general conclusions:

1. While considerable data are generated by vehicle probes on a daily basis, these data become rather sparse when the daily data are parsed in time and in space, as is required in a dynamic analysis which considers temporal and spatial differences in level of market penetration, departure rates and link travel times.
2. Relatively low levels of market penetration are likely to yield sufficient probe data to estimate the mean level of market penetration, the mean link travel time and the mean O-D departure for a traffic network whose dynamic attributes need to be determined.
3. Considerably higher levels of market penetration are required when the above mean estimates are to be reliable and robust and therefore need to be obtained with adequate levels of confidence.
4. The general agreement between the analytical and simulation results provides for an opportunity to utilize the analytical methods for more aggregate planning purposes and the simulation approach for more site-specific operational analyses, with a known level of consistency between the two approaches when their areas of application overlap.

In addition, the actual network analysis also yielded the following specific conclusions regarding the value of vehicle probes as sources of dynamic travel time and O-D demand data:

1. It is very difficult to reliability estimate the travel time on arterials for low levels of market penetration due to the high variability in the percent of probes in the traffic stream that results from the low flow conditions coupled with the interrupted flow nature of signalized links.
2. The presence of non-interrupted flow makes the estimation of link travel times for freeway links intrinsically much more reliable. In addition, the increased traffic volumes on these higher capacity links increases the likelihood of observing many more probes and therefore also simplifies the task of obtaining reliable freeway link travel times from probes.
3. Dynamic O-D demand rate estimates are highly dependent on the accuracy of the dynamic level market penetration estimate. Such an accurate estimate is relatively easy to obtain if a single network-wide value is utilized, but is much more difficult to obtain when an O-D specific value is needed which fully reflects the lag between trip departure time and the subsequent observation of these trips on detectorized links.
4. It is clear that any benefit assessment of IVHS architectures or of simply the routing strategy within such an architecture, must directly consider the reliability and variability of vehicle probe data. Based on the results provided within this paper, it must be concluded that an assessment which is based on the availability of continuous and accurate probe input data would be seriously flawed for low to medium levels of market penetration.

References

[1] Van Aerde, M., MacKinnon, G., and Hellinga, B. *The Generation of Synthetic O-D demands from Real-Time Vehicle Probe Data: Potential and Limitations.* Proceedings of the VNIS-91 Conference, Dearborn, Michigan, 1991.

[2] Boyce, D.E., Hicks, J. and Sen, A., *In-Vehicle Navigation Requirements for Monitoring Link Travel Times in a Dynamic Route Guidance System.* Presented at the 70th Annual Meeting, Transportation Research Board, Washington, D.C. January, 1991

[3] M. Van Aerde. *A Generalized Version of Greenshields' Speed-Flow Relationship for Highways*. Working paper. Transportation Systems Research Group, Queen's University , Kingston, Ontario. 1993

[4] Hellinga, B., and Van Aerde, M. *An Overview of a Simulation Study of Highway 401 FTMS*. CITE '93 Conference Compendium. Edmonton, Alberta. June 1993.

[5] S. Teply. *Canadian Capacity Guide for Signalized Intersections.* Institute of Transportation Engineers. 1984.

[6] Stover, V.G. and Kepke, F.J. *Transportation and Land Development*. Institute of Transportation Engineers. 1988. page 43.

[7] M. Van Aerde, B. Hellinga, G. MacKinnon, and J. Wang. *QUEENSOD: A Practical Method for Estimating Time Varying Origin-Destination Demands for Freeway Corridors/Networks*. Presented at the Annual Transportation Research Board Meeting, Washington , D.C. January 1993

[8] Van Aerde, M., Rilett, L., and Krage, M., *INTEGRATION Model Simulation of the Traffic Responsive Route Guidance Functions of the TRAVTEK Experiment*. Proceedings of the VNIS-91 Conference, Dearborn, Michigan, 1991.

[9] M. Van Aerde and S. Yagar. *Dynamic Integrated Freeway/Traffic Signal Networks: A Routing-Based Modelling Approach.* Transportation Research A, Volume 22A, Number 6, pp. 445-453. 1988

[10] Yu, L. and Van Aerde, M. *Implementing Platoon Dispersion in Microscopic Simulation Model*, Transportation Research Group, Department of Civil Engineering, Queen's University, Canada. 1991

[11] Yu, L. and Van Aerde, M. *Level of Market Penetration Requirements for the Derivation of Link Travel Times from Vehicle Probe Data*, Transportation Research Group, Department of Civil Engineering, Queen's University, Canada. 1992

[12] Lapin, L.L. *Probability and Statistics for Modern Engineering*. Brooks/Cole Publishing Company. 1983.

Table - 1: Speed - flow characteristics of links in example network

Link Type	Free Speed (km/h)	Saturation Flow (vphpl)	Speed at Capacity (km/h)	Jam Density (v/km)
Arterial	60	1600	30	125
Ramp	50	1600	25	125
Freeway	110	2200	70	125

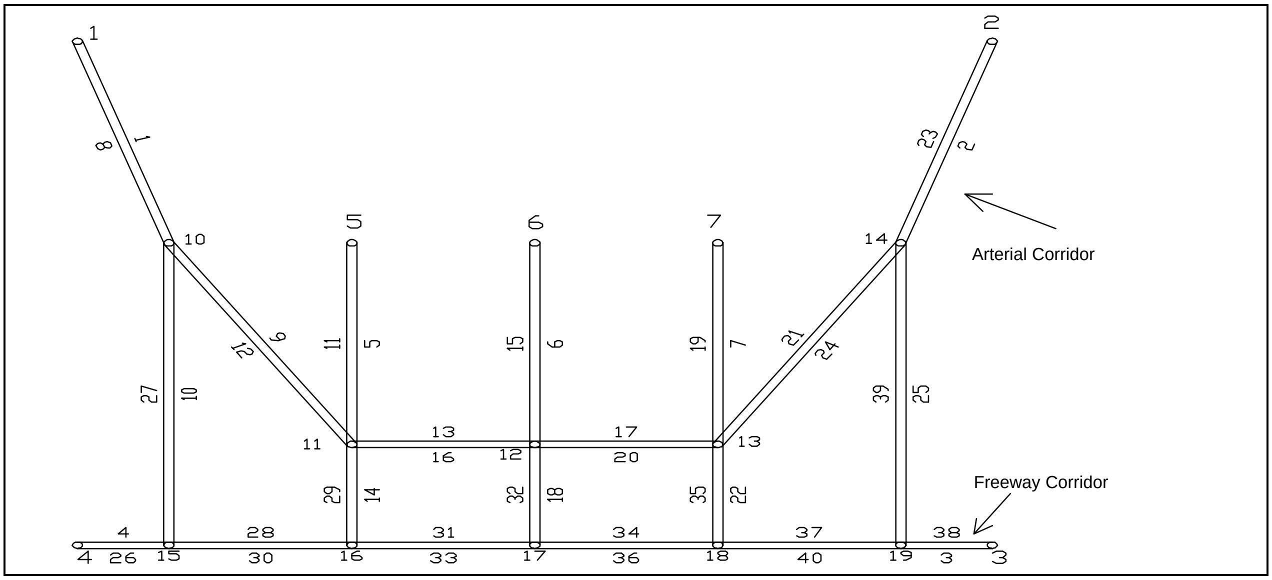

Figure - 1: Configuration of example traffic network

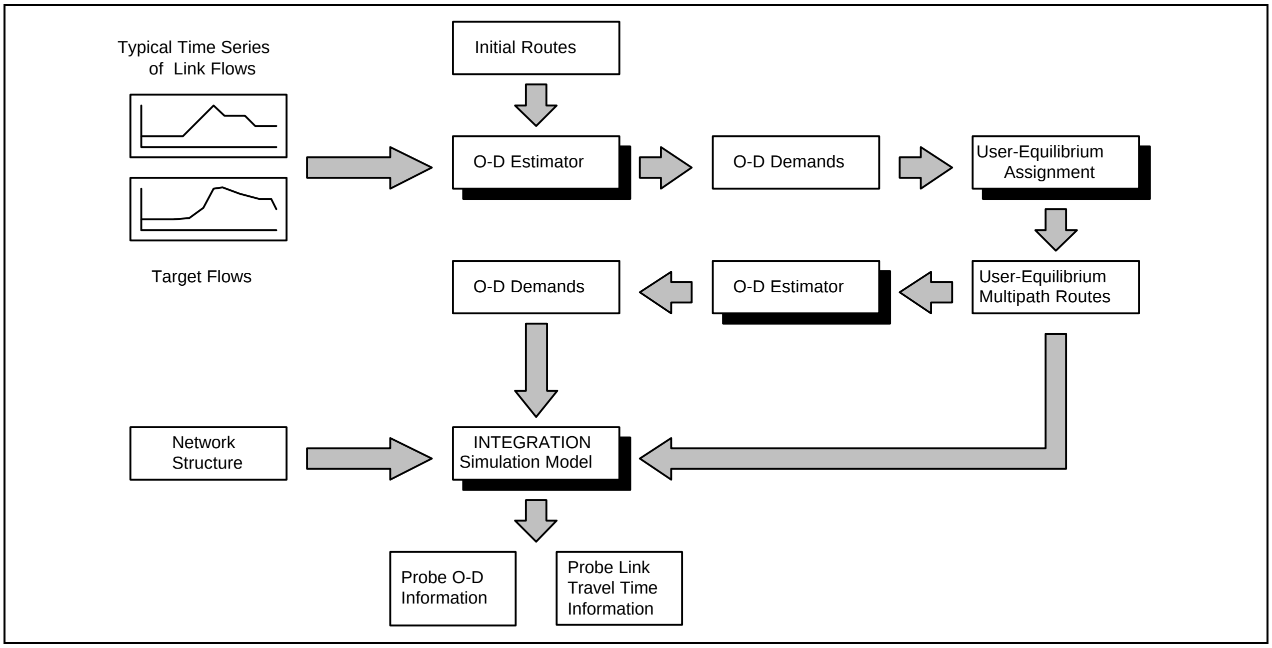

Figure - 2: Process used to produce realistic traffic conditions on example network

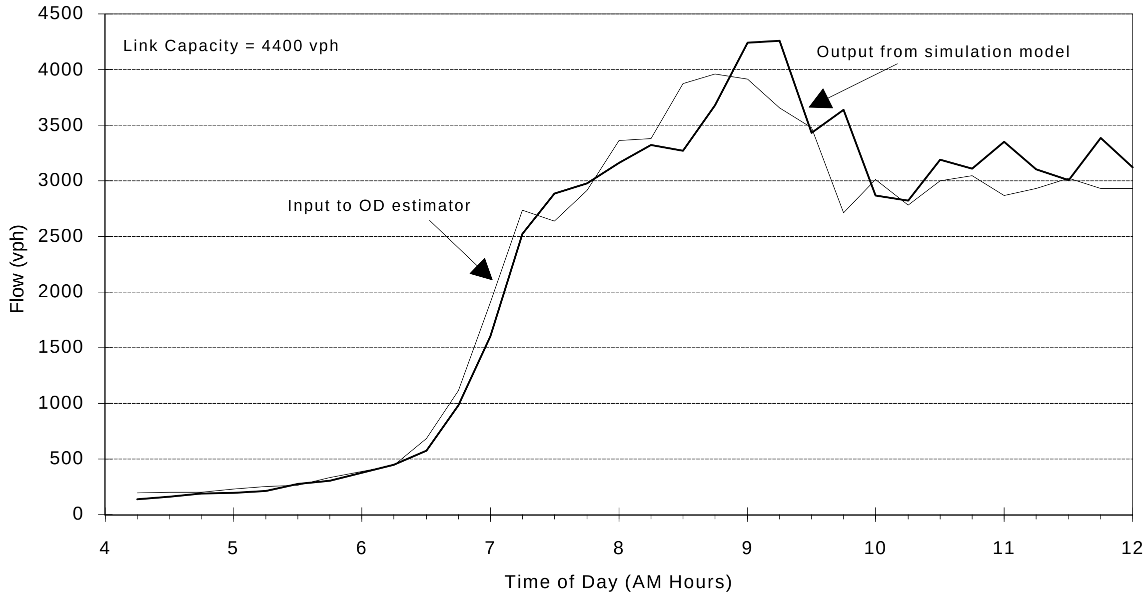

Figure - 3: Comparison of freeway link flow profile input to the O-D estimation process with the one resulting from the simulation model (Link 33)

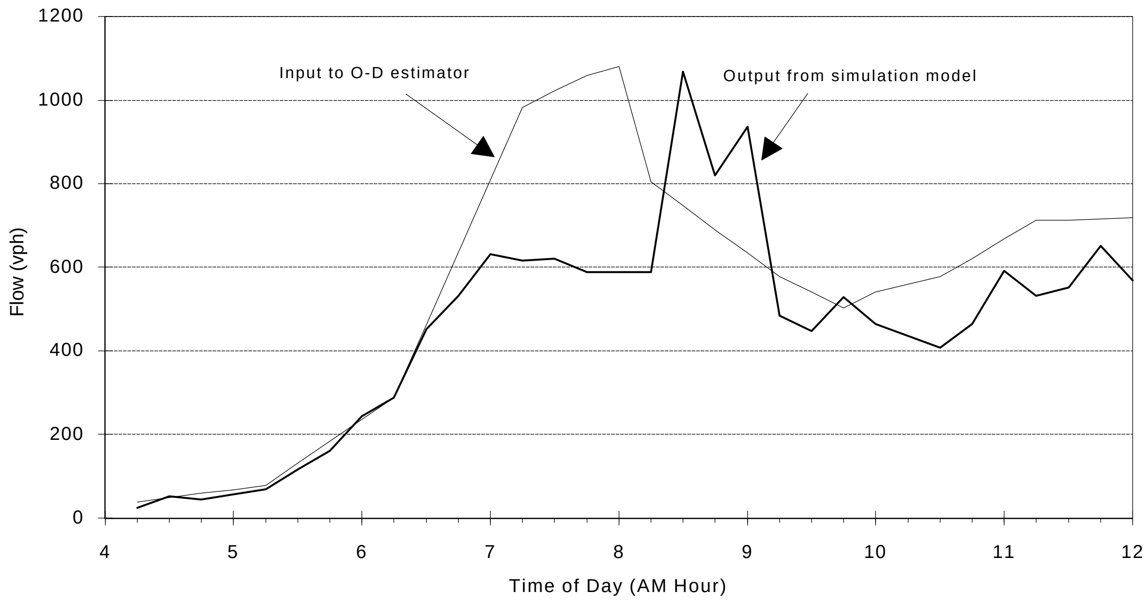

Figure - 4: Comparison of arterial link flow profile input to the O-D estimation process with the one resulting from the simulation model (Link 16) probe data log file

```
  1     4     2     3     4     2
              .
              .
              .
 11     4    38    38    36 0.00
              .
              .
              .
 11     4   103    37    65 0.00
              .
              .
              .
 11     4   169    34    65 0.00
              .
              .
              .
 2      4   426   4   3   424
```

Data Legend:

Column	Description
1	Record Type Identifier type 1 = departure record type 2 = arrival record type 11 = link completion record
2	unique vehicle I.D. number
3	simulation clock time (seconds)
4	for record type 1 and 2: origin zone number for record type 11: link number
5	for record type 1 and 2: destination zone number for record type 11: link travel time
6	for record type 1: scheduled departure time for record type 2: total trip time for record type 11: stop fraction

Figure - 5: Sample of typical probe information provided by simulation model

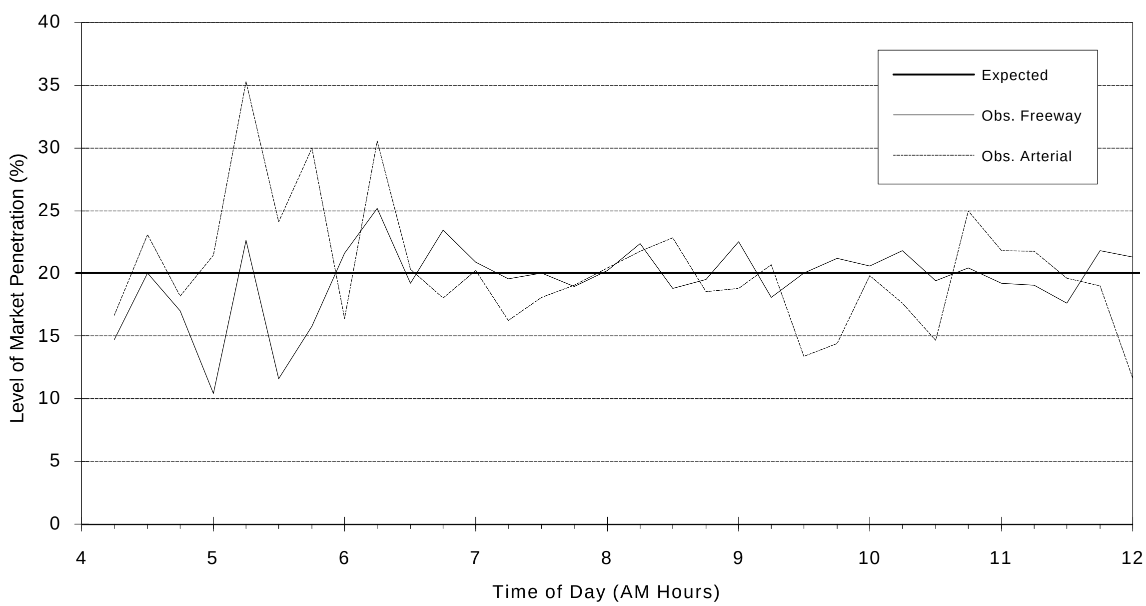

Figure - 6: Random variation in level of market penetration (Arterial: link 16, Freeway: link 33)

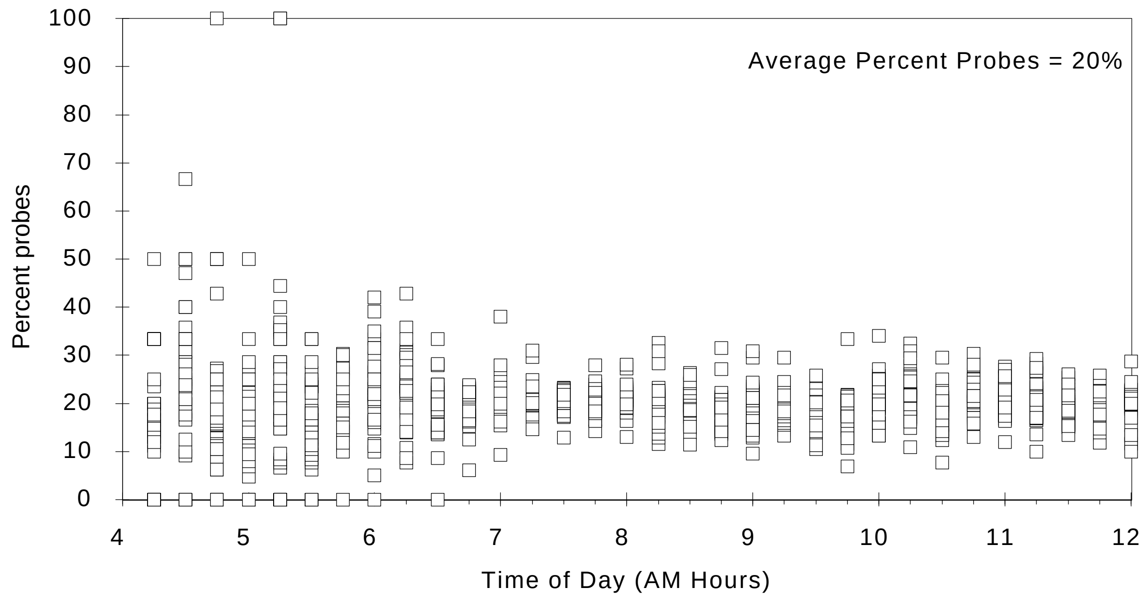

Figure - 7: Random variation in level of market penetration for all links over all time periods

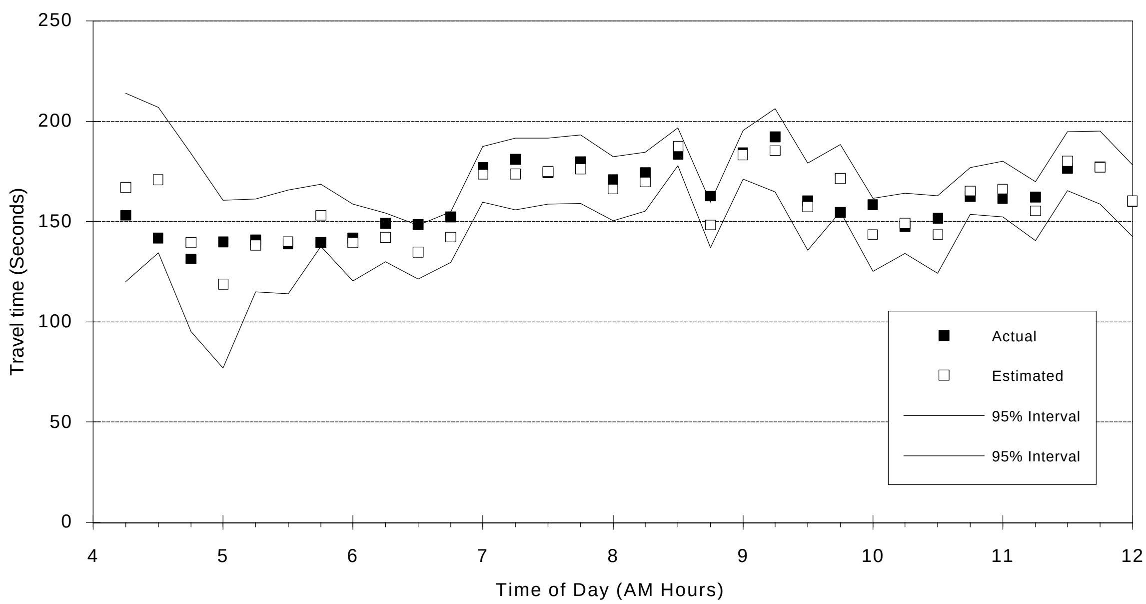

Figure - 8: Analytical estimates and actual simulation arterial link travel times for a level of market penetration of 20% (Link 16)

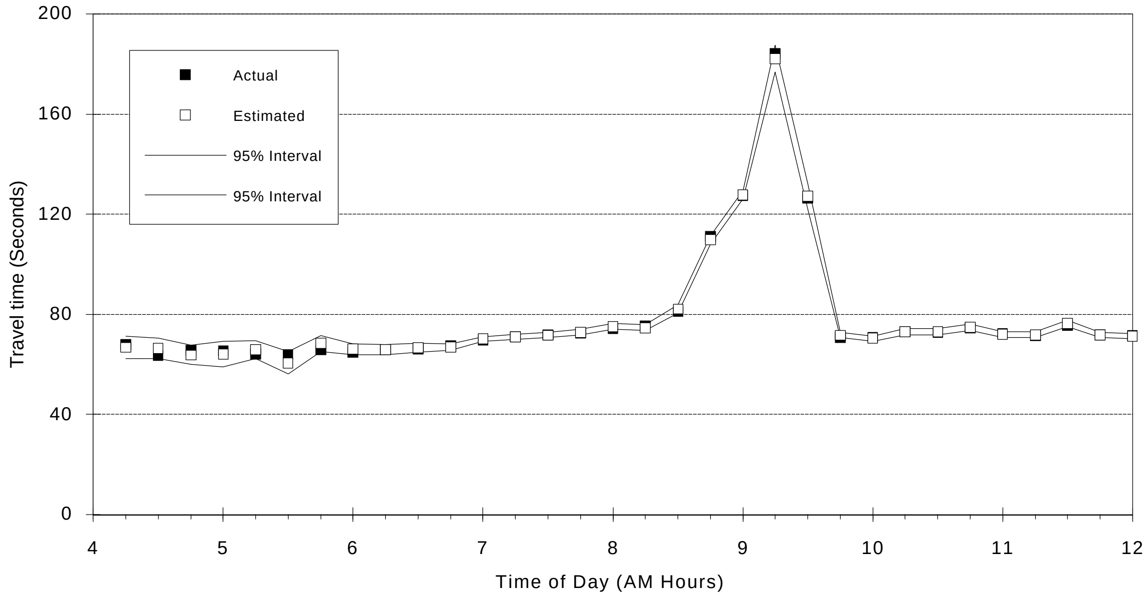

Figure - 9: Analytical estimates and actual simulation freeway link travel times for a level of market penetration of 20% (link 33)

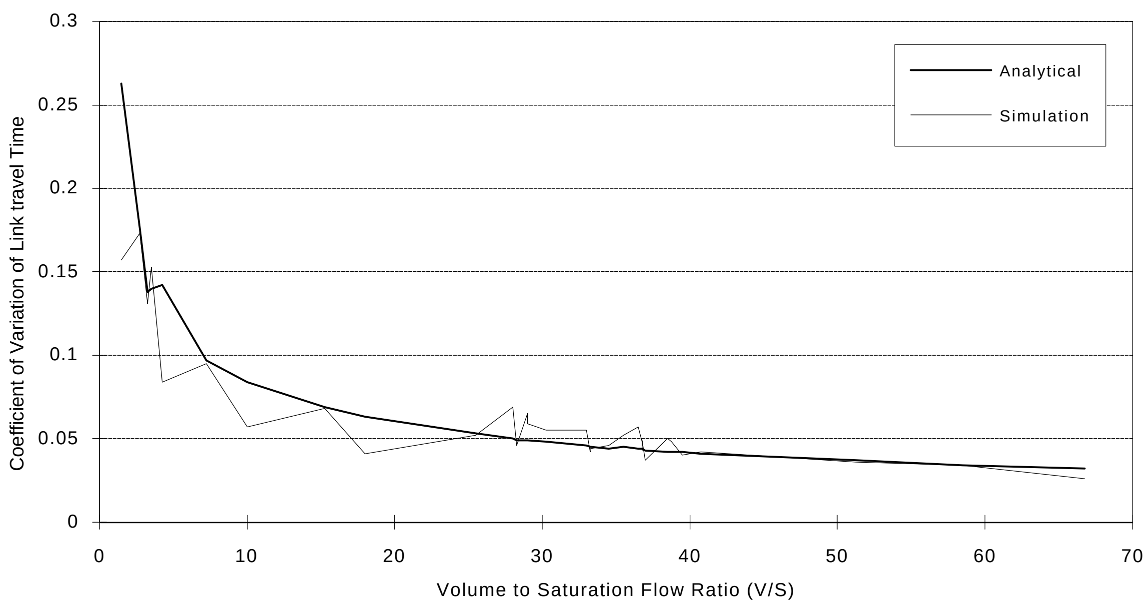

Figure - 10: Analytical estimates and actual simulation coefficient of variation of arterial link travel times for a level of market penetration of 20% (link 16)

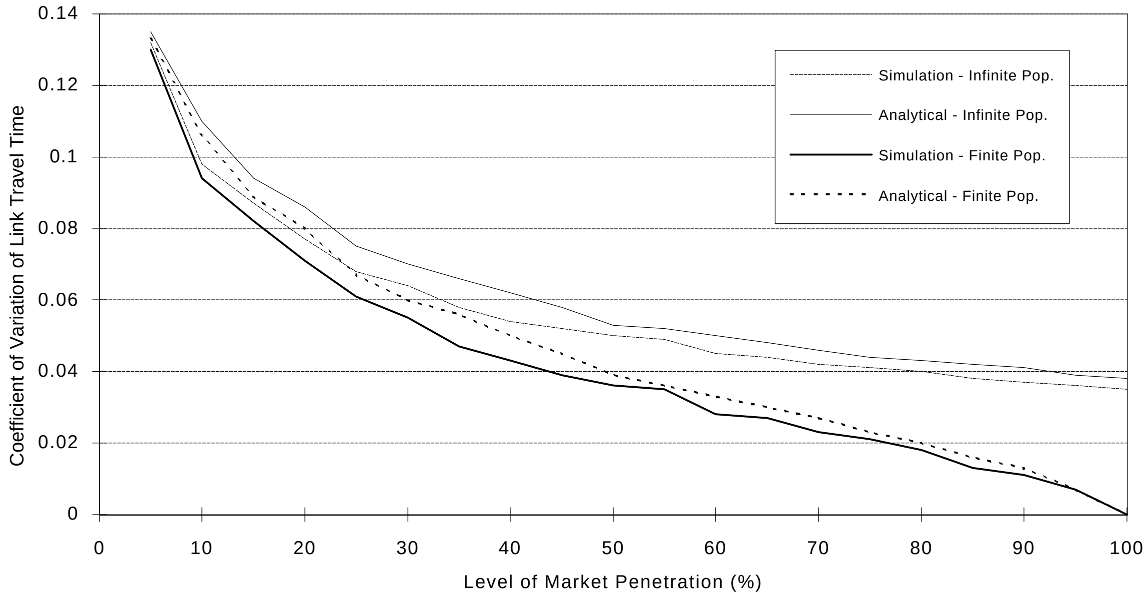

Figure - 11: Analytical estimates and actual simulation aggregate coefficient of variation of all arterial link travel times assuming both finite and infinite populations

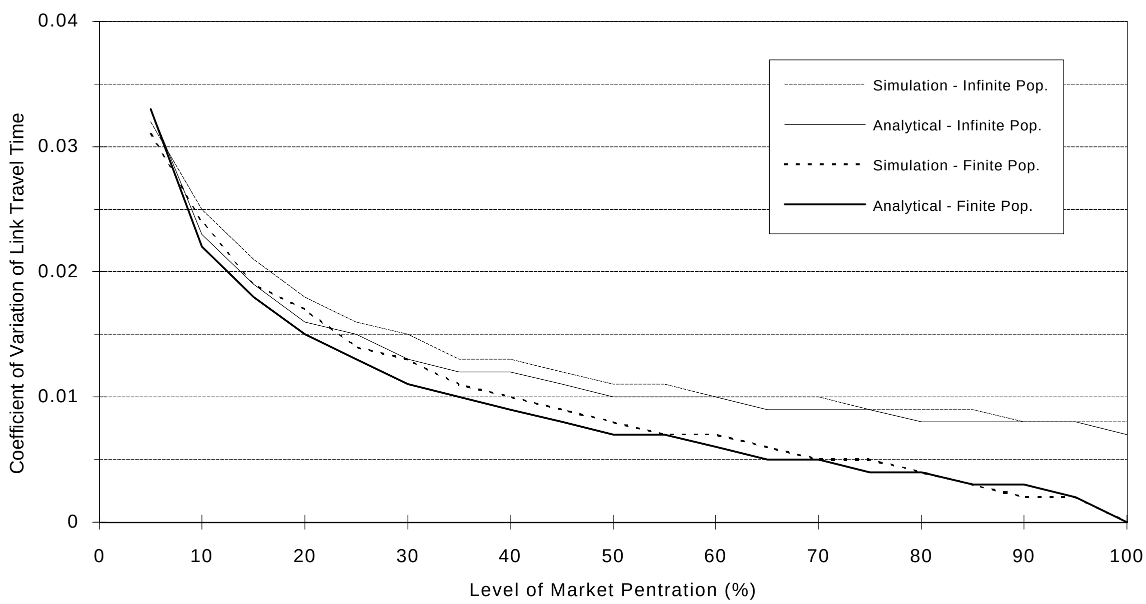

Figure - 12: Analytical estimates and actual simulation aggregate coefficient of variation of all freeway link travel times assuming both finite and infinite populations

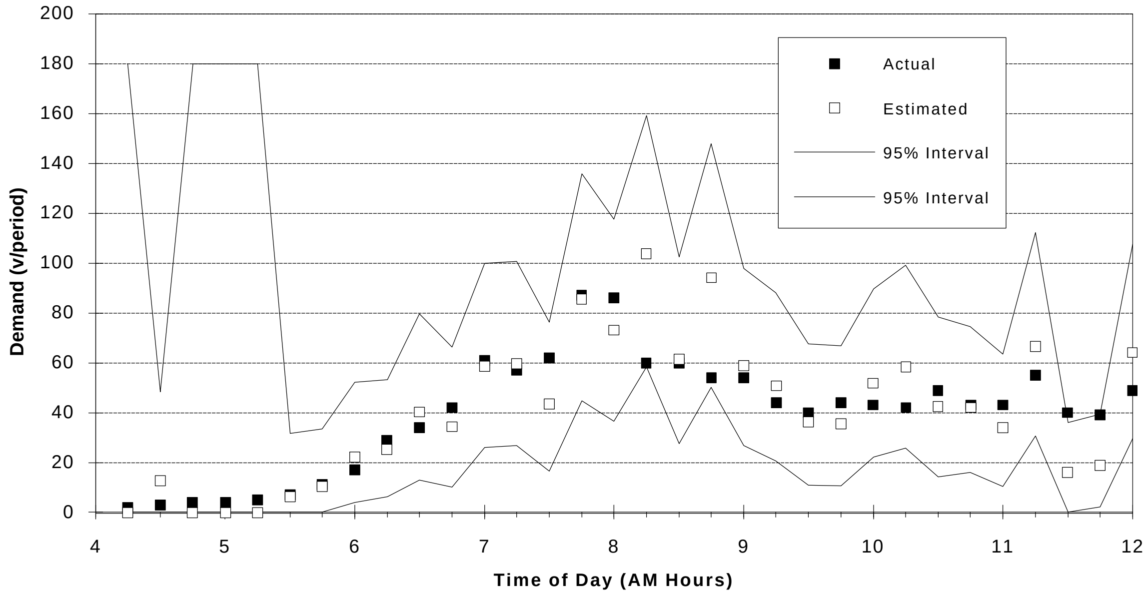

Figure - 13: Analytical estimates and simulation results for O-D demand over time of day for level of market penetration of 20% (O-D pair 1-3)

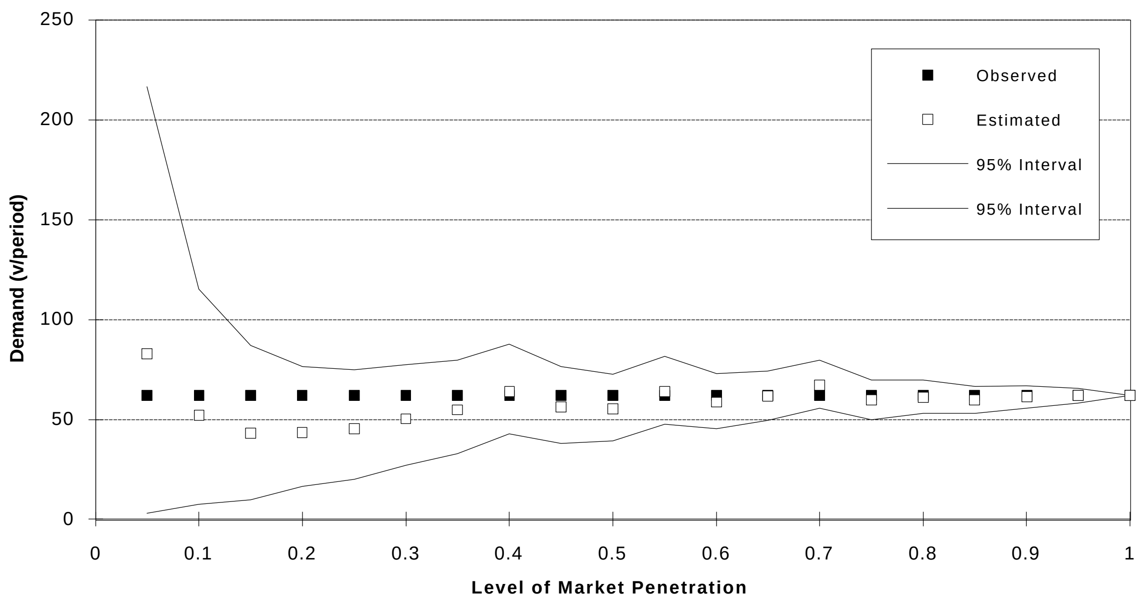

Figure - 14: Analytical estimates and simulation results for O-D demand for various levels of market penetration (O-D pair 1-3 for time period from 7:15 - 7:30 AM)

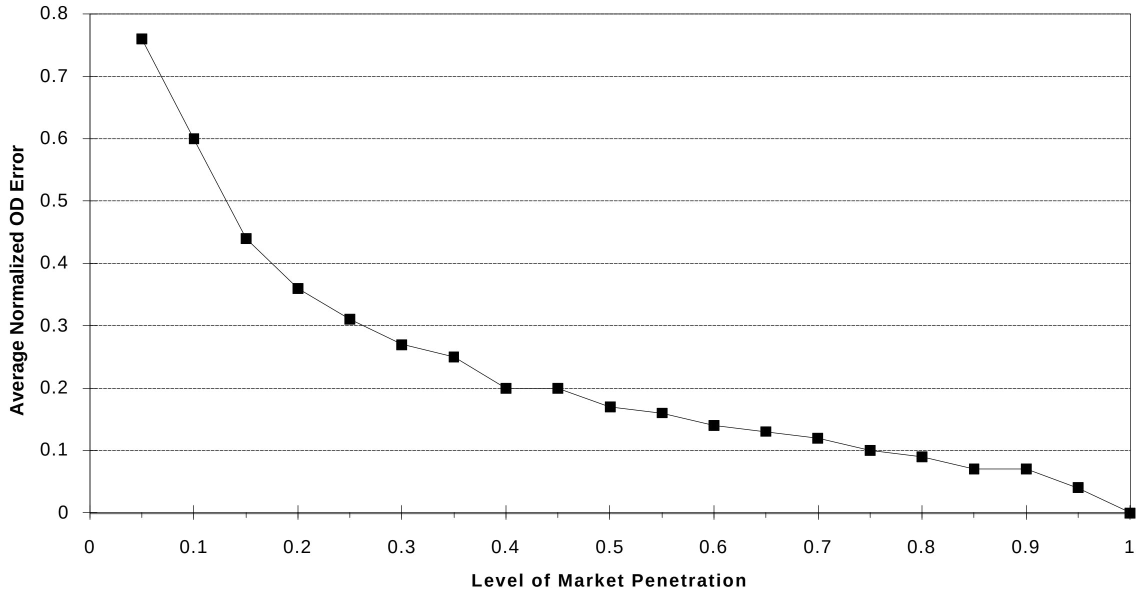

Figure - 15: Effect of level of market penetration on accuracy of O-D estimation