Akcelik & Associates Pty Ltd

On the estimation of lane flows for intersection analysis

September 2001

On the estimation of lane flows for intersection analysis

Rahmi Akçelik

Background

This technical note has been prepared in response to a question raised by Pretty (1989) regarding the lane flow estimation method used in the SIDRA program for signalised intersection analysis (Akçelik 1984, 1986). The question, which has been raised in discussing a recent paper by the author (Akçelik 1988), is relevant to the analysis of all types of intersection.

Simply stated, the question is: "Which principle should be used as a basis for estimating lane flows at the approaches to an intersection, and how does this principle relate to the lane choice decisions of individual drivers?"

Related to this is the question of lane underutilisation, that is, unequal lane utilisation. This may manifest itself as a behavioural issue even when alternative lanes available to drivers have equal performance characteristics. On the other hand, a shared lane used effectively as an exclusive lane is a case of unequal lane utilisation that results from substantially different lane performance characteristics. For example, no through vehicles may share a lane with opposed (permitted) turns because of substantially inferior performance characteristics of the opposed turn movement (the case of a *de facto* turn lane).

Whether the definition of lane underutilisation based on relative degrees of saturation (as in SIDRA) should be used irrespective of the lane flow estimation principle, or lane underutilisation should be defined according to the lane flow estimation principle used, is another issue to resolve.

This note also discusses a recent paper by Fisk (1988), which raises the issue of the relationship between lane choice and route choice principles in a more general context.

The Equal Degree or Saturation Method

The lane flow estimation method used in SIDRA (Akçelik 1984, 1986, 1988) is based on the principle of *equal degree of saturation* (flow / capacity ratio, or x value) as discussed in detail in Akçelik (1980, 1981). Equal lane degrees of saturation mean equal utilisation of available lane capacities.

If the lane capacities are not equal, then unequal lane flows will result. This is seen for the lanes of the East approach road (Movements 3 and 4) and the West approach road (Movements 5 and 6) in *Table 1* for the SIDRA example shown in *Figure 1* (full example is given in Akçelik 1988).

Correspondence: Rahmi Akçelik
Director, Akcelik & Associates Pty Ltd,
P O Box 1075 G, Greythorn Victoria, Australia 3104
Tel: +61 3 9857 9351, Fax: + 61 3 9857 5397
rahmi@akcelik.com.au

www.aatraffic.com

Table 1

Equal Degree of Saturation solution for the East and West approach road lanes in Figure 1

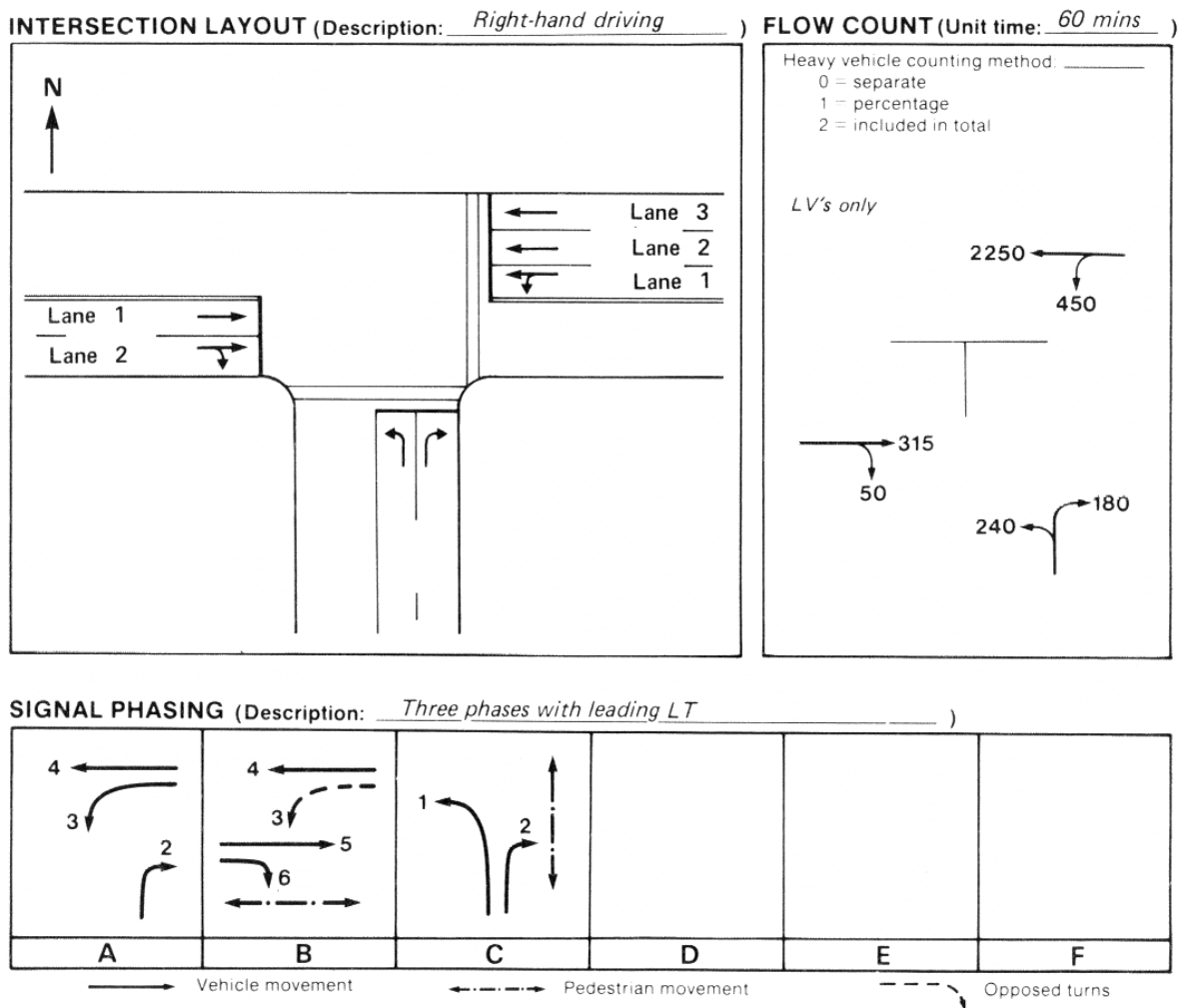| Lane No.	Flow (veh/h)	Capacity (veh/h)	Degree of Saturation	Lane Util. Ratio	Delay (s/veh)	Queue (veh)
East approach road (Movements 3 and 4)						
1	629	765	0.822	1.00	18.7	11.9
2	1036	1260	0.822	1.00	12.8	21.1
3	1036	1260	0.822	1.00	12.8	21.1
West approach road (Movements 5 and 6)						
1	215	450	0.477	1.00	31.9	5.1
2	150	315	0.477	1.00	34.7	3.7

Figure 1 - Example from Akçelik (1988)

The degree of saturation is a direct measure of *congestion level,* and the use of this criterion for lane flow estimation implies that individual drivers choose the lanes with minimum congestion levels. The relation of this measure to other possible lane choice criteria is discussed in the following sections.

In SIDRA, lane underutilisation means unequal use of available lane capacities resulting in unequal degrees of saturation. The cases of effective (de facto) exclusive lanes are found as a result of substantially lower lane capacities (hence higher degrees of saturation).

When effective green times are equal for all lanes, the equal degree of saturation method is equivalent to the *equal flow ratio* (flow / saturation now, or y value) method.

The traditional methods that treat movements in lane groups and use opposed turn adjustment factors for capacity estimation assume the same effective green time for all lanes in the lane group. Therefore, the equal flow ratio method is adequate in this case. Several recent capacity estimation techniques that have adopted the lane group method use the equal flow ratio principle for lane flow estimation implicitly, e.g. the US Highway Capacity Manual (Transportation Research Board 1985) or the OSCADY program (Burrow 1987).

In the Swedish (Peterson, Hansson and Bang 1978) and the Canadian (Teply 1984) capacity manuals, the equal flow ratio method is used in association with a lane-by-lane method of capacity estimation.

The lane-by-lane method used in SIDRA allows for different effective green times for different lanes of the same approach road for better accuracy in performance prediction. Different effective green times for different lanes may result from opposed turns, lane blockages, pedestrian interference, and so on. This method requires the use of the equal degree of saturation method for lane flow estimation as a more general form of the equal flow ratio method (Akçelik 1980, 1981).

The Equal Average Delay Method

The equal delay method for lane flow estimation uses minimum average delay (seconds per vehicle) as the criterion for lane choice.

In the literature, the *equal flow ratio* and the *equal average delay* methods are often stated as if they are the same method (e.g. Bonneson, Messer and Fambro 1988; White and Wallace 1988).

One of the conditions for the two methods to give the same result is that the lanes must have equal capacity (e.g. for lanes 2 and 3 of the East approach road as seen *in Figure 1* and *Table 1*). With shared lanes, the capacities of approach lanes will almost always differ because of different capacity conditions for different movements through the intersection.

The fact that *the equal flow ratio,* or *the equal degree of saturation,* and the *equal average delay* methods are different was discussed in earlier papers by the author (Akçelik 1978, 1980).

The *equal average delay* results for the example in *Figure 1* are given in *Table 2,* and can be compared with those for equal degrees of saturation given in *Table 1*.

Table 2

Equal Average Delay *Solution for the East and West approach road lanes in Figure 1*

Lane No.	Flow (veh/h)	Capacity (veh/h)	Degree of Saturation	Lane Util. Ratio	Delay (s/veh)	Queue (veh)
East approach road (Movements 3 and 4)						
1	566	744	0.760	0.90	14.4	9.3
2	1067	1260	0.847	1.00	14.3	23.0
3	1067	1260	0.847	1.00	14.3	23.0
West approach road (Movements 5 and 6)						
1	296	450	0.658	1.00	33.7	7.4
2	69	266	0.259	0.39	33.7	1.6

The Minimum Travel Time Principle

Pretty (1989) refers to a *minimum travel time* principle as a basis for the lane choice decisions of individual drivers. This is equivalent to the use of *equal travel times* for estimating lane flows. Firstly, it should be noted that, in the literature, this principle is often equated to the principle of *equal average delay.* This is based on the assumption that link travel times are constant. However, this assumption does not allow for any queue interaction. In reality, drivers may adjust their link (or *cruise)* speeds as a function of the queue length, and therefore the assumption of constant link travel times may not be satisfactory .

In a recent paper, Fisk (1988) argues that the minimum travel time principle should be used for traffic assignment (for route choice) and intersection calculations (lane choice) so as to achieve consistency between these two levels of modelling from a behavioural viewpoint.

Against this argument, it may be postulated that drivers use different criteria in their *strategic* trip decisions (e.g. travel through a network) and in their *tactical* trip decisions (e.g. manoeuvres through an intersection). Even at the route choice level, individual drivers may (and some certainly do) prefer a more *direct,* or major (i.e. high-capacity) route rather than the *minimum travel time* route.

While being analytically elegant, most traffic assignment models are based on insufficient knowledge about drivers' route choice behaviour, and the minimum travel time principle for route assignment may be too simplistic. For example, in a recent study of drivers' perceptions of different levels of service and route choice decisions, Fulton (1988) has concluded that congestion and the related unpredictability of delay (travel time) play a very important role in route choice decisions of drivers, and when the extent of congestion and delay is unpredictable, drivers may choose a more reliable but longer/slower route.

Another aspect of route choice (and lane choice) which should be considered is the difference in the behaviour of drivers of light and heavy vehicles. For example, in a recent study, Bowyer and Ogden (1988) found that truck movements in an urban area are determined by decisions made by a large number of parties - while truck operators made route choice decisions on a wider network

level from a productivity viewpoint, truck drivers were allowed to make route choice decisions at a local level. Furthermore, *fewer stops* and *less hassle* were found to be important criteria in truck drivers' route choice decisions.

Therefore, before adopting the minimum travel time principle of traffic assignment to lane flow estimation, alternative route and lane choice principles should be considered, the possibility of different driver behaviour at strategic and tactical levels should be explored, and the issues of model accuracy, interaction with traffic control (e.g. signal timings) and relation to the level of service criteria used in intersection analysis should be studied carefully. Similarly, these considerations should be applied to traffic assignment methods with a view to the improvement of existing methods.

For example, another way to achieve modelling consistency would be to apply the equal degree of saturation principle to traffic assignment rather than applying the minimum travel time principle to lane flow estimation! According to this principle, drivers would choose the routes which offer minimum congestion levels. Certainly, allowing for congestion on a section of a route as a factor in route choice could offer an improvement to existing traffic assignment methods.

Alternative Principles

Criteria other than equal degree of saturation, equal delay or equal travel time can be used, and have been used, for lane flow estimation. A method which is reasonable, at least at first consideration, is the use of *equal queue length* principle. This principle, which has been used, for example in the microscopic simulation model, INSECT (Cotterill, Moore and Tudge 1984), is based on the assumption that drivers choose the lanes with the shortest queue.

This principle may give similar results to the *equal average delay* principle only when the lane capacities are equal. Essentially, it differs from both the *equal average delay* and the *equal degree of saturation* principles. In other words, equal lane queue lengths do not necessarily correspond to equal delays or equal degrees of saturation.

Table 3

Equal Queue Length *Solution for the East and West approach road lanes in Figure 1*

Lane No.	Flow (veh/h)	Capacity (veh/h)	Degree of Saturation	Lane Util. Ratio	Delay (s/veh)	Queue (veh)
East approach road (Movements 3 and 4)						
1	724	783	0.925	1.00	34.6	18.7
2	988	1260	0.784	0.85	11.1	18.6
3	988	1260	0.784	0.85	11.1	18.6
West approach road (Movements 5 and 6)						
1	188	450	0.418	0.77	31.4	4.4
2	177	328	0.540	1.00	34.5	4.4

akcelik & associates

The equal queue length results for the example in *Figure 1* are given in *Table 3,* and can be compared with the results given in *Table 1 and 2.* It should be noted that the queue lengths given in *Tables 1 to 3* are estimates of the *average back of the queue* (see Akçelik 1988).

The *equal queue length* principle is not always appropriate, since drivers may not choose a shorter queue because of their knowledge that vehicles in that queue will have longer delays (typically, due to turning vehicles delayed by pedestrians or heavy opposing flows). Very low flow rate in a lane means a very short queue, but may be associated with very long delays.

Another possibility for lane flow estimation is to use a method that equalises *the end of the saturated portion of the green period* on alternative lanes. This would mean that the queues in alternative lanes are cleared at the same time. The end of the saturated portion of the green period equals the maximum back of the queue divided by the saturation flow rate. Thus, according to this method, individual drivers will choose the lanes with minimum queue length and the maximum departure rate. Low saturation flows of opposed turns would then result in lower flows in such lanes in spite of short queue lengths.

Other possibilities for lane flow estimation are to use an *equal stop rate* method, or to use any combination of this and the other measures mentioned above.

The SIDRA Solution

The *equal degree of saturation method* used in SIDRA gives lane flow estimates which fall between estimates from the *equal average delay* and the *equal queue length* methods as seen in *Table 1 to 3.*

According to the *equal degree of saturation method,* the drivers choose the lanes with minimum level of congestion, or maximum throughput. This could imply a driving tactic of clearing the intersection in a minimum number of cycles *(quickly).* Thus, it is related to the principle of the minimum end of the saturated portion of the green period.

Which Principle?

Without extensive surveys, it is not easy to suggest that one method or the other represents the reality better. The answer is difficult to find since this type of lane flow estimation method is a *steady-state approximation to a dynamic process.*

For example, it is possible that drivers' lane choice tactics are different during the red and green intervals, and vary with the congestion level (and hence, with the time of day).

It is easy to observe that the existence and position of heavy vehicles in a lane affect the lane choice significantly. Maybe there are differences between high-speed and low-speed roads (or between good and poor driving environments).

Lane underutilisation effects due to, for example, *destinations* of drivers well beyond the intersection, and such *driving policies as* 'always use the middle lane since it is safer', or 'always use the left lane because it is quicker' complicate the issue substantially. For example, the US Highway Capacity Manual (Transportation Research Board 1985) suggests the use of unequal lane utilisation by default.

akcelik
& associates

Driver variability in terms of lane choice tactics is certainly an important factor to consider. This also means that the *transferability* of findings from one environment to the other becomes an important issue.

A further consideration is that the equal delay (or equal travel time) method does not distinguish between delays to *stopped* and *unstopped* vehicles (they use an overall average delay value to all vehicles). The lane choice mechanism may very well be different for these two types of vehicles. Such dynamic real-life processes are not reflected by the equal delay (or equal travel time) method.

Model Accuracy

An important issue often overlooked in traffic signal analysis generally is the question of *model accuracy.* Usually, it is forgotten that the basis of estimation or optimisation may be an estimated parameter, and too much reliance is attached to the use of a parameter which is estimated on the basis of various assumptions. This is certainly the case with the *delay* and *queue length* parameters.

Firstly, delay and queue length are estimated using capacity and degree of saturation as input parameters, therefore it is one more step away from reality compared to the capacity and degree of saturation parameters. Therefore, the use of capacity and degree of saturation will probably give more reliable results than the use of delay or queue length for lane flow estimation, or at least, the results from the former method may fall within the confidence interval of the results from the latter.

Secondly, the modelling of delay and queue length has various weaknesses. The methodological issues related to delay estimation have not all been resolved, for example, estimation of delay and queue length on a lane-by-lane basis against on a lane group basis makes a lot of difference to the results (Akçelik 1984).

As discussed in Akçelik (1988), capacity estimation methods which employ a lane group approach and use adjustment factors for opposed turns in shared lanes again lead to some serious inaccuracies in both capacity and delay estimation. The question of the principle used in lane flow estimation is of secondary importance in relation to these types of model inaccuracies. Similar considerations apply to the delay-minimising signal settings favoured by theoreticians against the practical timing method used in SIDRA.

The relation between lane flow estimation and signal timing calculations is now discussed.

Interaction with Signal Settings

The implications of alternative lane flow estimation principles on the signal timing results should also be considered. Compared with the equal degree of saturation method, the equal delay method results in larger degrees of saturation in some lanes (lane underutilisation according to the SIDRA method).

When the *equal average delay* results given in *Table 2* for the example in *Figure 1* are compared with those for *equal degrees of saturation* given in *Table 1,* it is seen that Lane 1 of the East approach road and Lane 2 of the West approach road are underutilised (lane utilisation ratios of 0.90 and 0.39, respectively). As a result, the equal average delay method results in longer queues

in some lanes, and these critical lanes require longer green times. This would lead to longer cycle times and longer queue lengths and delays.

Similarly, the *equal queue length* method tends to result in larger degrees of saturation in some lanes and this leads to increased required green times and cycle times (compare the results given in *Table 3* with those in *Table 1 and 2*). The method allocates higher flows to the lanes with lower saturation flows and these queues take longer to clear. It is seen that Lanes 2 and 3 of the East approach road and Lane 1 of the West approach road (which are high-capacity lanes) are underutilised (lane utilisation ratios of 0.85 and 0.77, respectively).

If the drivers try to maximise their chances of clearing the intersection al the end of the green period, this would tend to shorten the required green times. This type of lane choice process approximates the principle of *equal end of the saturated portion of the green period,* and it is possible that it occurs at the end of the green period. On the other hand, the drivers who arrive during the earlier part of the red period may choose their lanes according to perceived delays. However, this process probably depends on congestion levels, and the lane choice mechanism may be very different under congested conditions where very long queues develop.

Of course, the signal timings affect the capacities, delays and queue lengths, and hence, will in turn affect the lane flow estimates. In addition, the lane flow estimates will affect the capacities of opposed turns, short lanes, shared lanes. An iterative process as used in SIDRA is required to solve the total problem, and different solutions are likely to result from the use of different criteria.

In this context, it would be interesting to compare the equal delay method for lane flow estimation combined with the use of minimum-delay settings against the SIDRA method of equal degrees of saturation for lane flows combined with the use of practical degrees of saturation for signal timing calculations.

The issue of interaction between signal settings and lane flows is important as a matter of consistency of the methodology, and also important in relation to the method of control used in practice. For example, the commonly used vehicle-actuated signals based on gap settings do not employ a delay-minimising timing strategy.

The interaction between signal settings and route flows (traffic assignment) is also important for similar reasons. This issue has been explored by several researchers but has not been resolved yet.

Relation to Level of Service Criteria

As a matter of consistency, the level of service criteria used in signalised intersection analysis is also relevant to the question of lane flow estimation. The US Highway Capacity Manual (Transportation Research Board 1985) specifies the use of *average individual stopped delay* as a measure of level of service at signalised intersections. This is based on consideration of driver perceptions of intersection conditions.

On the other hand, the equal flow ratio method is used implicitly in the capacity estimation procedures of the US Highway Capacity Manual, and this is different from the equal delay method as discussed above. Therefore, the level of service criterion (delay) and the lane flow estimation criterion (flow ratio) used in the US Highway Capacity Manual are not consistent.

However, the use of unequal lane utilisation method in the US Highway Capacity Manual would tend to compensate for this inconsistency as it would tend to approximate the equal delay solution (provided the low-capacity lanes are designated as underutilised lanes - see *Table 2).*

The fact that *delay and the degree of saturation are not well correlated* has been recognised (Roess and McShane 1987), and Berry (1987) has recommended the use of the degree of saturation to supplement delay for determining level of service so as to avoid oversaturation (x values that exceed 1.0) in some cases.

Consideration should be given to the use of x values in defining levels of service at signalised (and other) intersections from the view point of consistency between criteria used for level of service definition and lane flow estimation (there is a parallel here with the use of oversaturation as a factor in traffic assignment).

Conclusion

In conclusion, the answer to the question of individual drivers' lane choice is not straightforward (as also is the case for route choice). At this stage, the author *feels* that the equal degree of saturation principle used in SIDRA provides a satisfactory lane flow estimation method which gives a solution between the equal delay and equal queue length methods.

Fisk (1988) emphasises tile minimum travel time principle as a 'behavioural model of lane choice'. This technical note explains that the equal degree of saturation method for lane flow estimation is also a behaviourally based model, but a more complex one.

As a solution between the equal delay (or travel time) and equal queue length solutions, the equal degree of saturation solution is probably a better steady-state approximation to a dynamic real-life process. The use of the equal degree of saturation principle is also consistent with the use of practical degrees of saturation for signal timing calculations.

The equal degree of saturation method is the basis of most traditional and new capacity estimation techniques (although it has been confused with the equal average delay principle). With the traditional techniques that estimate capacities on the basis of lane groups, this is an *implicit* assumption.

On the other hand, in SIDRA that uses a lane-by- lane technique for capacity estimation, this assumption is *explicit.* In fact, a benefit of the lane-by-lane technique (especially with different effective green times allowed for different lanes) is that the implications of the equal degree of saturation in terms of delay, queue length, and So on, become obvious.

If reliable information becomes available on this subject through extensive field studies, the results can be incorporated into SIDRA easily. However, the interaction with signal timings need to be considered carefully when adopting any change to the lane flow estimation method.

REFERENCES

AKÇELIK, R. (1978). X and Y in traffic signal design. *Proc. 9th ARRB Conf.* 9 (5), pp. 45-60.

AKÇELIK, R. (1980). Lane utilisation and saturation flows. *Traffic Engineering and Control* 21 (10), pp. 482-484.

AKÇELIK, R. (1981). *Traffic Signals: Capacity and Timing Analysis*. Australian Road Research Board. Research Report. ARR No. 123 *(Fourth Reprint* 1989).

AKÇELIK, R. (1984). SIDRA does it lane-by-lane. *Proc. 12th ARRB Conf.* 12 (4), pp. 137-49.

AKÇELIK, R. (1986). *SIDRA Version 2.2 Input and Output*. Australian Road Research Board. Technical Manual. ATM No. 19.

AKÇELIK, R. (1988). Estimating the capacity of a shared lane. *Proc. 14th ARRB Conf.* 14 (2), pp. 228-41.

BERRY, D.S. (1987). Using volume-to-capacity ratios to supplement delay as criteria for levels of service at traffic signals. *Transp. Res. Rec.* 1112, pp. 23-28

BONNESON, J.A., MESSER, C.J. and FAMBRO. D.B. (1988). Examination of shared lane operations. Paper presented at the *67th Annual Meeting of Transportation Research Board*.

BOWYER, D.P. and OGDEN, K. W. (1988). *Impacts of Urban Road Investment on Road Freight Transport: West Gate Freeway Case Study*. Australian Road Research Board. Internal Report. AIR 424-2.

BURROW, I.J. (1987). *OSCADY: A Computer Program to Model Capacities, Queues and Delays at Isolated Traffic Signal Junctions*. Transp. Road Res. Lab. (UK) TRRL Research Report 105.

COTTERILL, P.J., MOORE, S.E. and TUDGE, R. (1984). Intersection simulation model: INSECT. *Proc. 12th ARRB Conf.* 12 (4), pp. 171-182.

FISK, C.S. (1988). Lane choice and intersection performance. *In: Urban Road Traffic Models for Economic Appraisal Stage III, Volume 1* (Report No. 5, Appendix I). University of Auckland, Department of Civil Engineering, Report No. 452.

FULTON, E.J. (1988). *Drivers' Perceptions of Levels Of Service*. Inst Transp. Studies, University of California, Berkeley. Research Report UCB-ITS-RR-88-15.

PETERSON, B.E., HANSSON, A. and BANG, K.L. (1978). Swedish Capacity Manual. *Transp. Res. Rec.* 667, pp.1-28.

PRETTY, R.L. (1989). Discussion. *Aust. Rd Res.* 19 (1), pp. 69.

ROESS, R.P. and McSHANE, W.R (1987). Changing concepts of level of service in the 1985 Highway Capacity Manual: some examples. *ITE Journal* 57 (5), pp. 27-31.

TEPLY, S. (Ed.) (1984). *Canadian Capacity Guide for Signalised Intersections.* ITE District 7 and the University of Alberta.

TRANSPORTATION RESEARCH BOARD (1985). *Highway Capacity Manual*. Special Report 209, Washington, D.C.

WHITE, F.J. and WALLACE, C.E. (1988). Modelling shared lane use in TRANSYT-7F. Paper presented at the *67th Annual Meeting of Transportation Research Board*.

akcelik
& associates