INTEGRATION:
An Overview of Traffic Simulation Features

M. Van Aerde, B. Hellinga, M. Baker and H. Rakha

Department of Civil Engineering
Queen’s University, Kingston, Canada
Phone: 613-545-2122 Fax: 613-545-2128
E-mail: vanaerde@civl6000a.civil.queensu.ca

A Paper Accepted for Presentation at
the 1996 Transportation Research Board Annual Meeting
Washington, D.C.

INTEGRATION: Overview of Simulation Features

M. Van Aerde , B. Hellinga, M. Baker and H. Rakha
Department of Civil Engineering
Queen's University, Kingston, Canada
Phone: 613-545-2122 Fax: 613-545-2128
E-mail: vanaerde@civl6000a.civil.queensu.ca

ABSTRACT

The INTEGRATION model first attempted to provide a single model that could consider both freeways and arterials, and which also could consider both traffic assignment and simulation. This ability was intended to bridge the gap between the scope of network on which planning models could be applied and the level of resolution that operational tools needed to provide. Since this time, the model has also integrated traditional Intelligent Transportation Systems considerations such as ATMS and ATIS, the coupled modeling of traffic and vehicle emissions, and more recently the combined modeling of traffic and communications subsystems. However, all of these model extensions depend upon the validity of the same traffic simulation and assignment logic that is at the core of the the model.

The objective of this paper is to describe the core traffic simulation logic and specific attributes related to the model's ability to represent the operational aspects of freeways and signalized links. The main measures of performance that can be generated and the various approaches to performing traffic assignment are presented elsewhere.

1. INTRODUCTION

The INTEGRATION model was conceived during the mid 1980's as an integrated simulation and traffic assignment model (1-4). What made the model unique was that the model's approach utilized the same traffic flow logic to represent both freeway and signalized links, and that both the simulation and the traffic assignment components were also microscopic, integrated and dynamic. In order to achieve this mix of attributes, traffic flow was represented as a series of individual vehicles that each followed macroscopic traffic flow and assignment relationships. The combined use of individual vehicles and macroscopic flow theory resulted in the model being considered mesoscopic by some.

1.1 Recent Developments

During the past decade the INTEGRATION model has evolved considerably from these original mesoscopic roots. This evolution has taken place through the addition, enhancement and refinement of various new features during the application of the model in the classroom as well as in the field. Some of these improvements have enhanced the fundamental traffic flow model, such as the addition of car-following logic, lane-changing logic, and more dynamic traffic assignment routines. However, others have extended the model's application domain, such as the inclusion of features for modeling toll plazas, vehicle emissions, weaving sections, and HOVs. In addition, some features, such as the real-time graphics animation and the extensive vehicle probe statistics, have been added to simply make the model easier to understand, validate and calibrate.

1.2 Objectives of the Paper

The objective of this paper is to provide a summary of the current status of the most important traffic simulation features of the model, and to briefly indicate why and how certain features were implemented and/or enhanced. This overview is primarily qualitative in nature, in order to focus on the fundamental modeling approach and concepts. More quantitative discussions of each main model feature are available elsewhere. In addition, the more qualitative nature of this presentation also better addresses the interdisciplinary needs of those who may not already be familiar with traffic theory and/or traffic models in general. The specific way, in which the input data files can be used to invoke the various model features, is described fully in the model's User's Guide Volume I and II and is therefore not repeated here.

1.3 Structure of Paper

The remainder of this paper is organized into six further sections. The balance of this introduction will describe the general domain of application for the model, the microscopic modeling approach, and the dynamic capabilities of the model.

In the second section the general traffic flow characteristics, that are common to all applications of the model are presented, including a summary of the car-following and lane-changing logic, as well as the basics of the routing and incident sub-models.

The third section discusses the application of these traffic flow attributes to the specifics of modeling of freeways. In particular, the treatment of speed-flow-density relationships, in order to derive shockwave and queue spill-back analysis, is discussed. In addition, the method in which merges, diverges, and weaving sections are modeled is described.

The fourth section describes the manner in which the operation of traffic signals and/or ramp meters are modeled. It includes a discussion of the automatic estimation of uniform delay, coordination impacts, stochastic queuing, as well as a summary of the estimation of random and over-saturation delay, opposed and unopposed left turn capacity and the option of right-turns-on-red.

The fifth section of the paper briefly describes the model's capabilities for assigning O-D traffic demands to the network for a range of alternative assumptions. The combined routing and traffic loading procedure is discussed first for each of the available traffic assignment features. Subsequently, the section summarizes the use of stochastic and deterministic variations to multi-path assignment.

The sixth section of the report presents the various measures of performance that can be generated by the model. These include the estimation of travel time, delay, number of stops, as well as fuel consumption and the emissions of HC, CO, NO_x. This section is concluded with an indication of how loop detectors and vehicle probes can be modeled.

The seventh and last section of the paper describes a variety of miscellaneous other features that are available in the model to represent certain much more specialized and unique scenarios or control applications.

1.4 INTEGRATION's Domain of Application

In order to appreciate INTEGRATION's intended domain of application, it is useful to view travel within an urban area as an interrelated sequence of 6 decisions that the traveler typically must make in order to complete a particular trip. Three of these decisions are made prior to drivers leaving their driveway, and usually cannot be revisited during that same trip. The three others, however, need to be revisited repeatedly, once a particular trip has been initiated.

Pre-trip Decisions - At the highest level of the trip making process, are decisions related to where a particular trip maker may decide to live and work/shop. The trip maker must therefore decide how many trips to make towards each potential destination during each particular departure time window. Once the decision, to make a particular trip to a given destination has been made, the traveler must decide whether to utilize some form of transit (if available), or whether to utilize a private car, either as a single vehicle occupant or as a car pool participant. The third set of pre-trip decisions relates to the particular time at which the trip maker may elect to start the trip. Each of these first three types of decisions may interdependent, but are usually not made more than once for a particular trip.

On-Route Decisions - In contrast, the next three types of trip decisions need to made once the trip has commenced and usually also need to be revisited several times as the actual trip progresses. Specifically, prior to starting the trip, the trip maker must select what route to take. This decision, even when the trip has commenced, is usually not fixed, as a driver may elect to change any remaining portion of the trip. Once a vehicle has entered a particular link along this route, the driver must also select the speed at which to drive at and which lane to utilize. A driver's speed and lane choice are again likely to change, at a minimum from one link to the next but usually several times along the same link. Speed and lane changes often also occur along a link as a result of interactions with other vehicles. Finally, when a driver arrives at the end of a link, the driver may be required to cross an opposing traffic stream, and must decide whether to accept or reject any available gaps and/or how to merge with a converging traffic stream..

Domain of Application - The current domain of application of the basic INTEGRATION model is represented by the latter set of on-route driver decisions. This set starts from the time when the driver has elected to depart from a particular origin to a particular destination, at a particular time, and by means of a private car. This implies that, at present, INTEGRATION does not directly model the impact of someone who elects to depart at a different time, by means of a different mode, or to an alternate destination.

However, in order to reflect the increasing interest, in being able to explore the potential traffic impacts on these latter decisions, an outer loop is being added around the current INTEGRATION model. This loop will permit estimates of the expected changes in trip mode, departure time and/or destination to be made through iterative applications of the model.

1.5 Microscopic Modeling Approach

INTEGRATION has during the past 2 years become a fully microscopic simulation model, as it tracks the lateral as well as longitudinal movements of individual vehicles at a resolution of one deci-second.

This microscopic approach permits the analysis of many dynamic traffic phenomena, such as shockwaves, gap acceptance, and weaving, that are usually very difficult or infeasible to carry out under non-steady state conditions using a macroscopic rate-based model. For example, in a dynamic network, average gap acceptance curves typically cannot be utilized at permissive left turns if the opposing flow rate varies from cycle to cycle and/or within a particular cycle. It is also not appropriate to use them if the size of the acceptable gaps varies as a function of the length of time for which a vehicle has been waiting to find a gap.

The microscopic approach of INTEGRATION is therefore considered as a means to an end, rather than as an end in itself. This choice significantly increased the memory and computational requirements of the model, but is perceived to have yielded as a result some critical improvements in the fidelity with which is can represent dynamic traffic conditions at an operational level of detail.

1.6 Dynamic Modeling

The INTEGRATION model can consider virtually continuous time varying traffic demands, routings, link capacities and traffic controls, without the need to pre-define an explicit common time-slice duration. This implies that the model is not restricted to hold departure rates, signal timings, incident severities, or even traffic routings at a constant setting for any particular period of time. Consequently, instead of treating each of the above model attributes as a sequence of steady-state conditions, as needs to be done in most rate-based models, all of these attributes can be changed on virtually a continuous basis over time.

The microscopic approach also permits considerable flexibility in terms of representing spatial variations in traffic conditions. For example, while most rate-based models consider traffic conditions to be uniform along a given link, INTEGRATION permits the density of traffic to vary continuously along the length of a link. Such dynamic density variation, along for example, an arterial link, permits the representation of platoons departing from traffic signals and the associated propagation of shockwaves both in an upstream and downstream direction.

Finally, it is important to note that, while the model is primarily microscopic, these microscopic rules have been carefully calibrated in order to still capture most of the target macroscopic traffic features that most traffic engineers are more familiar with, such as link speed-flow relationships, multi-path equilibrium traffic assignment, and uniform, random and oversaturation delay, as well as weaving and ramp capacities. The main challenge in the design of INTEGRATION has therefore been to ensure that these important macroscopic features automatically become emergent behavior that arises from the more fundamental microscopic model rules that are needed to represent the system dynamics using a single integrated approach.

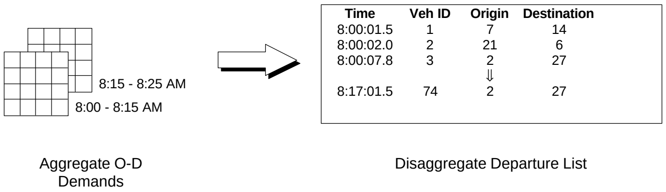
2. TRAFFIC FLOW FUNDAMENTALS

The manner in which INTEGRATION represents traffic flows can be best presented by discussing how a typical vehicle initiates its trip, selects its speed, changes lanes, transitions from link to link, and also selects its route.

2.1 Initiation of Vehicle Trips

Prior to initiating the actual simulation logic, the individual vehicles that are to be loaded onto the network need to be generated. As most available O-D (Origin-Destination) information is macroscopic in nature, INTEGRATION permits the traffic demand to also be specified as a time series histogram of O-D departure rates for each possible O-D pair within the entire network. Each histogram cell within this time series can vary in duration from 1 second to 24 hours, and the duration of each cell is independent from one O-D pair to the next, or from one time period to the next. When the same O-D is repeated within the departure list for an overlapping time window, the resulting vehicle departures are considered to be cumulative.

The actual generation of individual vehicles occurs in such a fashion as to satisfy the time-varying macroscopic rates that were specified by the modeler within the model's input data files, as illustrated in Figure 1. It can be noted that the model simply disaggregates an externally specified time varying O-D demand matrix into a series of individual vehicle departures prior to the start of the simulation. For example, if the aggregate O-D input data requests departures at a uniform rate of 600 veh/hr between 8:00 and 8:15 AM, a total of 150 vehicles will be generated at headways of 6 seconds.



Time	Veh ID	Origin	Destination
8:00:01.5	1	7	14
8:00:02.0	2	21	6
8:00:07.8	3	2	27
		⇓	
8:17:01.5	74	2	27

Figure 1: Conversion of aggregate O-D traffic demands into disaggregate departure list

It should be noted that, as the externally specified demand file is disaggregated, each of the individual vehicle departures is tagged with its desired departure time, trip origin and trip destination, as well as a unique vehicle number. This unique vehicle number can be utilized to trace a specific vehicle towards its destination. It can also be utilized to verify that subsequent turning movements of vehicles at, for example, network diverges are assigned in accordance to the actual vehicle destinations, rather than some arbitrary turning movement probabilities, as is the case in many microscopic models that are not assignment based.

2.2 Determination of Vehicle Speed

When the simulation clock reaches a particular vehicle's scheduled departure time, that vehicle is entered into the network at its origin zone. From this point the vehicle will begin to proceed in a link-by-link fashion towards its final destination. Upon entering this first link, the vehicle will select the particular lane in which to enter. This is usually the lane with the greatest available distance headway.

Once the vehicle has selected which lane to enter, the vehicle computes its desired speed on the basis of the distance headway between it and the vehicle immediately downstream of it but within the same lane. This computation is based on a link specific microscopic car following relationship that is calibrated macroscopically to yield the appropriate target aggregate speed-flow attributes for that particular link (5,6). Having computed the vehicle's speed, the vehicle's position is adjusted to reflect the distance that it travels during each subsequent deci-second. The updated positions, that are derived during one given deci-second, then become the basis upon which the new headways and speeds will be computed during the next deci-second.

The macroscopic calibration, of the microscopic car-following relationship, ensures that vehicles will traverse that particular link in a manner that is consistent with that link's desired free-speed, speed-at-capacity, capacity and jam density. Figure 2 illustrates the direct correspondence between the more familiar macroscopic speed-flow and speed-density relationships, and the less familiar car-following relationship that is plotted in terms of speed-headway. This correspondence is illustrated for three different traffic conditions, which are identified as points *a* , *b* and *c*.

It can be noted from the speed-flow relationship that point *a* represents uncongested conditions, point *b* represents capacity flow and point *c* represents congested conditions. However, the attributes of points *a*, *b* and *c* are more difficult to discern from the speed-density and speed-headway relationships, which simply represent mathematical transformations of the same relationship.

Qualitatively, it can be noted from the speed-headway relationship that vehicles will only attain their desired free speeds when the headway in front of them is very

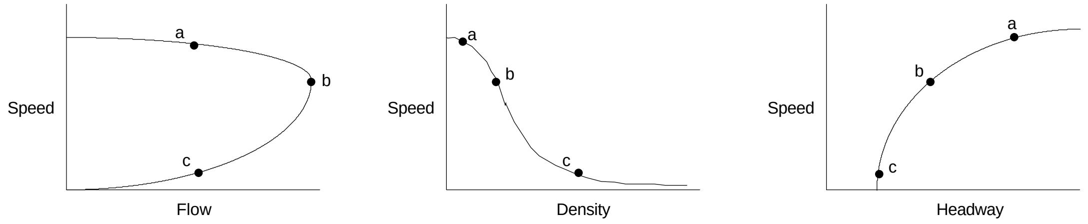


Figure 2: Determination of microscopic speed from corresponding macroscopic relationships

large. In contrast, when the distance headway becomes sufficiently small, as to approach the link's jam density headway, the vehicle will decelerate until it eventually comes to a complete stop.

A natural by product of the above car following logic, is that INTEGRATION represents all queues as horizontal rather than vertical entities. The representation of horizontal queues ensures that queues properly spill back upstream in space and time, either along a given link, or potentially across multiple upstream links. Furthermore, the representation of horizontal queues also ensures that the number vehicles in the queue will be greater than the net difference between the arrival and departure rate, as the tail of the queue grows upstream towards the on-coming traffic. Finally, the use of the above speed-headway relationship enables horizontal queues to consider variable density, depending upon the associated speeds of vehicles within the queue.

2.3 Lane Changing Logic

When a vehicle travels down a particular link, it either may make discretionary lane changes, mandatory lane changes, or both, as illustrated in Figure 3. Discretionary lane changes are a function of the prevailing traffic conditions, while mandatory lane changes are usually a function of the prevailing network geometry.

In order to determine if a discretionary lane change should be made, each vehicle computes three speed alternatives at deci-second increments. The first alternative represents the potential speed at which the vehicle could continue to travel in the current lane, while the second and third choices represent the potential speeds a vehicle could travel in the lanes immediately to the left and to the right of the vehicle's current lane. These speed calculations are made on the basis of the available headway in each lane and a pre-specified bias, for a vehicle to remain in the lane in which it is already traveling and to move to the shoulder lane when possible.

The vehicle will then elect to try to change into that lane which will permit it to travel at the highest of these three potential speeds. For example, in Figure 3 vehicle *D* may elect to leave the shoulder lane for the center lane in order to increase its headway and therefore the speed at which it can comfortably travel. Such lane changing, while discretionary, is still subject to the availability of an adequate gap in the lane to which the vehicle wishes to move.

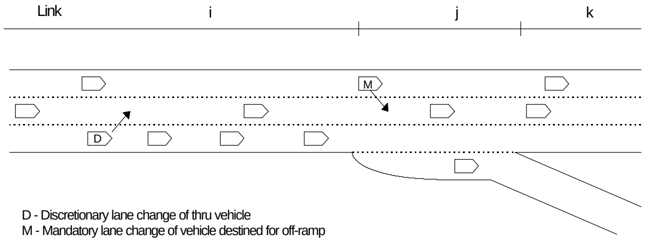


Figure 3: Discretionary and mandatory lane changes

While discretionary lane changes are made by a vehicle in order to maximize its speed, mandatory lane changes arise primarily from a need for a vehicle to maintain lane continuity at the end of each link. For example, in Figure 3 vehicle *M* would ideally desire to remain in the median lane, in order to maintain a higher speed. However, since this vehicle must access the off-ramp, it must first enter the deceleration lane prior to exiting link *j*.

In general, lane continuity requires that eventually every vehicle must be in one of the lanes that is directly connected to the relevant downstream link onto which the vehicle anticipates turning. A unique feature of INTEGRATION's lane changing model is that the lane continuity at any diverge or merge is computed internal to the model, saving the model user the extensive amount of hand coding that would be necessary in representing explicitly link continuity in networks with several thousands of links.

Once a lane changing maneuver has been initiated, a subsequent lane change is not be permitted for a pre-specified minimum amount of time. In the first instance, this minimum ensures that lane changes involve a finite length of time to materialize. It also ensures that two consecutive lane changes cannot be executed one immediately after the other. Furthermore, while the actual lane changing maneuver is in progress, the vehicle is modeled as if it partially restricts the headway in both the lane it is moving from, and the lane it is changing into. This concurrent presence in two lanes will result in an effective capacity reduction beyond that which would be observed if the vehicle had not made any lane change. The relationship of this impact to the speed and capacity of weaving sections is discussed later.

2.4 Link-to-Link Lane Transitions

Upon approaching the end of a link, the above mandatory lane changing logic will ensure that vehicles will automatically migrate into those lanes that provide direct access to the next desired downstream link. When the end of the first link is actually reached, the vehicle is automatically considered for entry onto the next downstream link.

The entry onto this downstream link is subject to the availability of an adequate minimum distance headway that is required in order to absorb the new vehicle without violating the downstream link's jam density. In addition, any available headway beyond this minimum is also utilized to set the link entry speed of the vehicle in question. If the maximum headway in the downstream link is insufficient to accommodate the vehicle in question, the vehicle will be held back on its original link until an acceptable headway becomes available. Consequently, congestion in one downstream link can constrain the outflow rate of one or more upstream links, such that queues can also spill back across multiple links.

Any available downstream capacity is also implicitly allocated proportionally to the number of inbound lanes to the merge. For example, if at a diverge all lanes have a saturation flow rate of 2000 veh/hr/lane, and two 2-lane sections merge into a single 3-lane section, the combined inflow from the two inbound links will be limited to 6000 veh/hr when the downstream link is not congested. However, if an incident were to have reduced the capacity of the 3-lane section to only, say 4000 veh/hr, the two inbound approaches would then only have a reduced combined outflow capacity of 4000 veh/hr available to them as well.

The exit privileges of a particular link may also be constrained by a conflicting opposing flow. In this case, the opposed vehicles would need to delay their entry into their next downstream link until a sufficient gap appeared in the opposing traffic stream. On a single lane approach, such gap seeking would also delay any subsequent vehicles that share the use of the same lane, even if subsequent vehicles are not opposed. However, on a multi-lane approach, unopposed vehicles maybe able to utilize the residual capacity in the remaining lanes.

On the basis of the above logic, vehicles proceed toward their destination in a link-by-link fashion, where their speeds, as well as longitudinal and lateral positions, are updated each deci-second until the vehicle's final link is reached. When the vehicle reaches the end of this final link, the vehicle is removed from the simulation, any trip statistics are tabulated, and any temporary variables assigned to that vehicle are released.

2.5 Route Selection and Traffic Assignment

The selection of the next link to be taken is determined by the model's internal routing logic (7,8). There exist several different variations to the model's basic assignment technique, and hence the details of these are beyond the scope of this simulation oriented paper.

In general, there exist many different ways within the model in which the next downstream link can be determined. Both of these techniques can be static and deterministic, or stochastic and dynamic. However, regardless of the technique that is utilized to determine these routings, all of these routings are eventually conveyed to the simulated vehicle using a look-up table format, as shown in Figure 4. This routing look-up table format provides, for each vehicle class, an indication of the next link to be taken towards a particular destination. The look-up table is also indexed based on the current link that is being traversed, as indicated next.

Upon the completion of any link, a vehicle queries the relevant look-up table, based on the current link that is being traversed. This permits the vehicle to determine which link it should utilize next to reach its ultimate destination in the most efficient manner. When this next link is completed, in turn, the process is repeated until eventually a link is reached whose downstream node is the vehicle's ultimate trip destination. In order to provide for multipath traffic assignments a set of multiple trees is utilized concurrently during a given time period, while different sets of trees may be utilized to represent time-varying multipath routings.

The key simulation feature to be noted within this traffic assignment process, is that turning movements (and therefore all mandatory lane changes) are based vehicle-specific path based turning movements, rather than more arbitrary turning movement percentages.

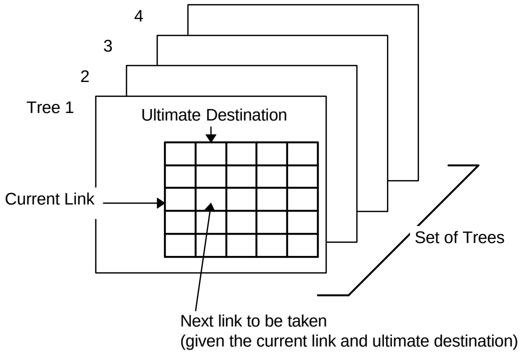


Figure 4: Illustration of the routing tree table concept

2.6 Link Use and Turning Movement Restrictions

One of the features, which allows the model to better represent the operational quirks of many actual networks, is the restriction of the use of either specific links, link lanes, and/or specific turning movements.

Restrictions on the use of links can be implemented for a specific subset of vehicle types. It is therefore possible to represent either the restricted availability of a certain link to only HOV vehicles, or the availability of a certain toll booth to a vehicle that possesses a specific toll collection technology (9). Alternatively, this feature can also be utilized to model the impact of a truck subnetwork within a more general overall road network.

It is also possible to restrict certain lanes to specific vehicle types. This feature may be used to model, for example, an HOV lane that is exclusive to one vehicle type. Alternatively, a given vehicle may be constrained to utilize only a given lane, such as for example, a truck lane, by restricting this vehicle from utilizing all other lanes. In either case, these restrictions are sufficiently flexible to permit a vehicle, turning onto or off of the link, to pass through these restricted lanes in order to complete their desired turning movement

A third type of restriction that is possible is that vehicles can be confined to only make certain turning movements from certain lanes. This ability permits the modeling of exclusive versus shared lanes. It is critical to properly model the impact of advanced/leading phases and/or estimating the number of vehicles that maybe able to make a right-turn-on-red before a through vehicle blocks the lane.

The final type of restriction can be applied to a specific turning movement. It is typically utilized to represent banned turning movements at intersections for certain periods of time. However, the same feature can also be utilized to represent time dependent access restrictions to the use of a particular reversible lane or special on-ramp.

2.7 Simulation of Incidents and Diversions

The continuous nature of the simulation and assignment model permits incidents to start at any time (to within one minute) and to be of any duration. The can also be of any severity, blocking from 0 to 100% of the available capacity. In addition, any specific group of lanes can be blocked at any point along the link, and the blockage can be of any length. Incidents may also be modeled concurrently at different locations, or different incidents may be modeled at the same location at different instances of time. The net effect of the incident is that it reduces the saturation flow and/or the maximum speed of each targeted lane on the given link.

INTEGRATION's routing logic does not at present explicitly respond to the occurrence of an incident. Instead, it responds to any delay that arises from the flow or speed restrictions associated with the incident. This indirect response has the effect that diversion does not occur until the delay experienced by vehicles becomes sufficiently large as to make an alternative route more desirable. Similarly, the model may sustain diversions, even after the actual blockage at an incident site has already been cleared, but when some residual queues remain to produce on-going delays.

3. FREEWAY SECTIONS

The INTEGRATION logic, while facility independent, is designed to deal with a number of situations which are usually perceived to be unique to freeway sections, such as merges, diverges and weaving sections. It should be noted, however, that many of these elements also appear on surface streets.

3.1 Modeling Merges

In general, when two traffic streams merge, all available merge capacity is allocated using entitlements that are in proportion to the non-queued capacities of the two merging links. However, since one of the merge lanes may not be able to fully utilize its entitlement, or as the total merge capacity may be further reduced by queues spilling back into the merge, the actual merge capacity always needs to be allocated dynamically.

At an on-ramp merge, therefore, queues may form downstream of the ramp, upstream of the ramp on the freeway, on the actual on-ramp, or on both, depending upon the prevailing demands. For example, when an acceleration lane is present, following the ramp merge, the queue will automatically be modeled as occurring immediately upstream of the lane drop. This is shown in the upper portion of Figure 5.

When the queue grows to then fill the entire merge area, the queue may then spill back onto the on-ramp or onto the upstream section of the mainline freeway. The exact location will then depend upon the split in the vehicle arrival rates. However, if there is no acceleration lane provided, the queue will form upstream of the on-ramp merge, as indicated in the lower portion of Figure 5. The split of the queues on the on-ramp is again a function of the relative vehicle arrival rates on the main-line and the on-ramp.

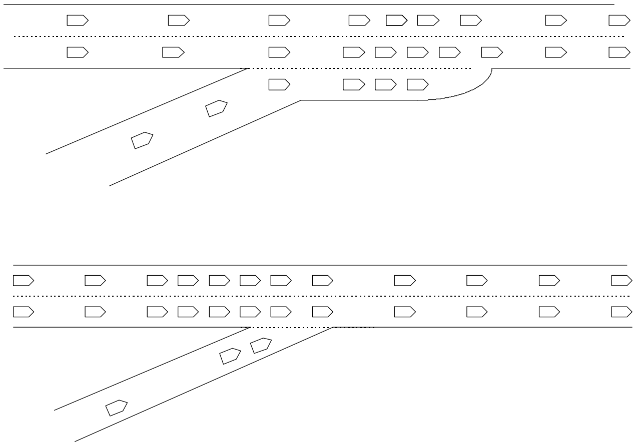


Figure 5: Congestion formation at a merge.

Once the above merge flow rates are determined, INTEGRATION computes the appropriate shockwaves upstream of the merge on either the mainline or the on-ramp. Furthermore, the absence of an explicit time slice in the model's analysis permits the formation of such queues to be analyzed over both very short and relatively long time intervals. For example, the formation of merge queues over time periods from, say 15 minutes to several hours, can be modeled if typical peak period demand overloads are to be considered. However, if upstream of the particular on-ramp, a traffic signal is present, the model can also consider short term merge over-saturation for 30 - 60 seconds at a time, each time the upstream traffic signal discharges its queues in a cyclical fashion.

It is important to note that the allocation of queues to different upstream arms at a merge is critical to estimating the relative travel times on each of these links. Errors in estimating these relative travel times not only affect the resulting MOEs in isolation, but also have a significant impact on any dynamic traffic assignments or diversions that consider these MOEs within their routing objective function.

3.2 Modeling Diverges

At a diverge, queues may form when one of the discharge arms receives more demand than there is available capacity. Such limitations on off-ramp inflow capacity may result from poor off-ramp geometry or from limitations on the net capacity of a traffic signal that is located at the downstream end of the off-ramp. Alternatively, the mainline section may spillback at a diverge due to downstream mainline congestion, even though there is sufficient capacity available both upstream of the diverge and on the off-ramp.

When one of the diverge arms, say an off-ramp, possesses insufficient capacity, as illustrated in Figure 6, vehicles destined for this bottleneck will queue on the upstream section of the diverge. Consequently, this queue may eventually constrain the flow of through vehicles, where these through vehicles may not even be utilizing the off-ramp in question. The INTEGRATION model computes the resulting queue spill-back as a function of the prevailing off-ramp oversaturation, as well as the extent to which vehicles utilizing the off-ramp tend to congregate in the lane feeding the off-ramp. Research to-date has shown that assumptions, with respect to the extent to which vehicles voluntarily queue in only the shoulder lanes, have a pronounced impact on the ultimate mainline flow delay.

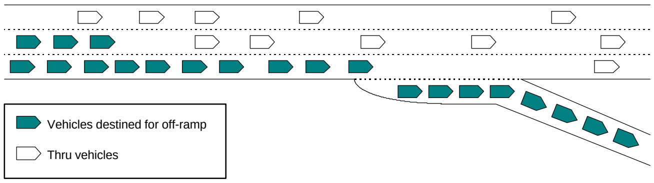


Figure 6: Queue formation due to an off-ramp.

An important impact of such queue spill back at diverges is that the link as a whole no longer complies to standard FIFO (First-In-First-Out) queuing logic. Instead, considerable differences in link travel times may arise, depending upon the destination links of the vehicles. Such differences complicate any routing logic, which may need to consider that different link users will arrive at the next downstream link following a different lag time, even though they may have entered the link concurrently. In addition, while the two traffic streams may not share the same travel times, they do interact considerably. This interaction is due to the fact that the split in off-ramp versus through vehicles will significantly alter the travel times experienced by both groups, even if the total arrival flow on the link remains relatively constant.

3.3 Modeling Weaving

Within INTEGRATION the final impact of a weaving section is a direct function of the prevailing car-following and lane changing behavior (10,11). The associated basic logic is illustrated in Figure 7.

Specifically, it can be noted that vehicles, who are engaged in a lane-changing maneuver, occupy space in both the lane they are leaving from and in the lane to which they are changing into. The fact that a single lane changing vehicle consumes capacity in both lanes makes a single weaving vehicle temporarily have the impact of essentially two vehicles. This effect, which is proportional to the duration of the lane change and the fraction of vehicles making lane change maneuvers, results in a dynamic calculation of weaving capacity in the model.

The reduced effective capacity of weaving sections, in which a large number of vehicles are making lane changes, not only is more pronounced at the onset of queuing, but also reduces the prevailing speeds in the weaving section prior to the onset of queuing. Total throughput capacity is usually also reduced by the fact that the availability of lane changing gaps is not uniform but random. This randomness has a further speed and capacity reduction impact on weaving flows, beyond that which would be considered by a deterministic impacts of the above weaving logic.

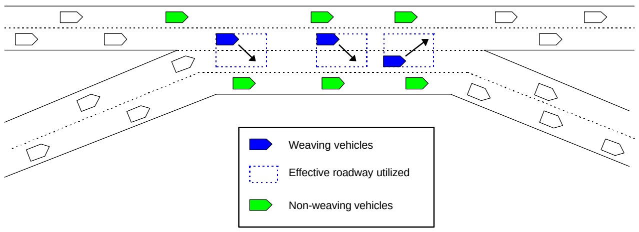


Figure 7: Capacity impact of weaving vehicles.

The final weaving logic is sensitive to the type of weave, as different weave types require different numbers of lane changes per vehicle. The model is also sensitive to the length of the weave, as a longer weaving section permits the impact of the lane changes to be spread out over a longer length of road. It is important to note that weaving logic and impacts are emergent features of the default model logic, and therefore do not require the modeler to tag specific sections as being weaving sections. Therefore, any area, in which a large number of mandatory lane changes are necessary, will automatically experience weaving impacts and the capacity reduction will dynamically depend upon the mix of weaving versus non-weaving flows. This effect is therefore also automatically captured on arterials, where a rapid succession of alternate turning movements may create implicit weaving section bottlenecks.

4. TRAFFIC SIGNALS

The second main roadway element, within most urban areas, is the presence of signalized intersections. While the highway sections between signalized intersections often operate in a fashion analogous to lower speed freeways, the behavior of traffic at signals is quite unique, as indicated below.

4.1 Modeling of Signal Cycles

Within INTEGRATION, a signalized link is identical in virtually all respects to a freeway link, except that the exit privileges to this link may periodically be suspended and the free-speed and saturation flow rates usually take on slightly lower values (12).

The suspension of exit privileges is set to occur when the traffic light indicates an effective red. When the light is red, vehicles must still obey the link’s car-following logic, except that a red traffic signal is considered to act as an additional vehicle that is positioned at the end of each lane on the link. This virtual vehicle therefore creates a reduction in the vehicle’s perceived headway, and causes subsequent vehicles that approach a red signal to slow down as their headway to this virtual object decreases. Eventually, the virtual object causes the first vehicle to come to a complete stop upstream of the stop line. Subsequent vehicles then automatically queue upstream of the first vehicle in a horizontal queue, where the minimum spacing of vehicles in this horizontal queue is governed by the jam density that has been coded for this link.

As shockwave theory applies to both freeways and arterials, the rate at which the tail of queue moves upstream along the link can be determined from a standard hydrodynamic analysis, as the ratio of the “arrival rate at the tail of the queue”, divided by the “net difference between the density of the queued vehicles and the density of the arriving traffic”. The dynamic nature of the model’s car-following logic also permits the rate, at which this queue grows, to vary dynamically when the arrival rate varies as a function of time during the cycle.

4.2 Shockwaves at Traffic Signals

When the effective green indication commences, the virtual object that represents the signal at the stop line of each lane is removed and the first vehicle in queue faces an un-interrupted headway along the downstream link. The initial acceleration of the first vehicles in the queue as well as the subsequent impact of the model’s car-following logic on any additional vehicles causes two shockwaves to form concurrently, as shown in Figure 8.

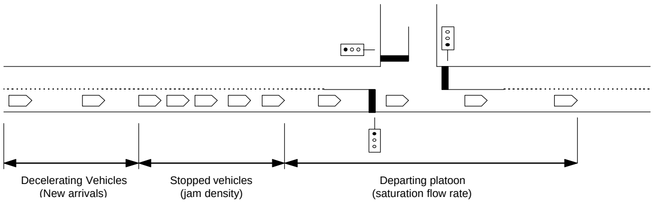


Figure 8: Traffic flow dynamics at a traffic signal

The first shockwave moves downstream from the traffic signal stop-line and consists of the front of the surge of traffic that crosses the stop line at saturation flow. The second shockwave moves upstream from the traffic signal stop-line, as queued vehicles start to accelerate when the vehicles ahead of them accelerate up to the speed-at-capacity. This backward moving shockwave, therefore, consists of the dividing line between those vehicles that are still stationery and those vehicles that already have begun to accelerate to a speed corresponding to the saturation flow rate. The speed of this second shockwave is again a function in the model of the speed flow characteristics of the link.

4.3 Uniform, Random and Oversaturation Delay

If the link is under-saturated, the queues that form during the red indication will be served, as indicated during cycle 1 in Figure 9.

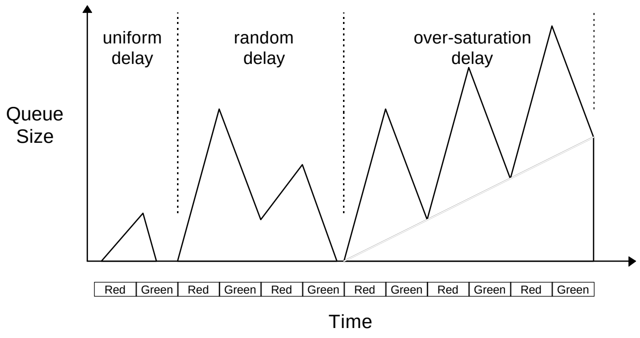


Figure 9: Uniform-random-over-saturation delay.

In this situation, the discharge at saturation flow rate will cease prior to the end of the effective green, and any subsequent vehicles will pass the stop-line at their arrival rate without being required to stop. However, if the signal is over-saturated, vehicles will continue to move at saturation flow rate until the end of the effective green. This will let a residual queue of unserved demand remain to be served in the subsequent cycle, as indicated during cycles 4, 5 and 6 in Figure 9.

Cycles 2 and 3 exhibit an intermediate state in which oversaturation occurs during one cycle due to randomness in arrival rates, but where during a second cycle this oversaturation queue can be completely cleared. One of the main complexities, of estimating the above delays within analytical intersection delay models, is circumvented in INTEGRATION through its use of a microscopic simulation approach. This approach makes the above uniform, random and oversaturation delays become emergent behavior.

4.4 Gap Acceptance Modeling at Traffic Signals

One of the most complex modeling tasks, in estimating the capacity of both isolated and coordinated traffic signals, is the treatment of permissive left turns and/or right turns on red. Within INTEGRATION, a microscopic gap acceptance model is utilized to reflect the impact of opposing flows on either one of these opposed movements (13). This opposition is automatically customized by the model at each intersection by means of built in logic that specifies which opposing movements are in conflict with each movement on the link of interest. This internal logic also determines which of the turning movements are opposed within a shared lane. Given the above data, the model automatically provides opposition to left turners, when the opposing flow link discharges concurrently. However, it also allows the opposed link's discharge rate to revert back to the unopposed saturation flow rate when the opposed link is given a protected phase.

The incorporation of this logic within INTEGRATION permits the model to evaluate the impact of protected versus permissive left turns, the impact of leading versus lagging greens, and an assessment of the impact of the duration of the left turn phases in great detail. In addition, the gap acceptance logic can work concurrently with the queue spill-back model to determine when or if vehicles in a left turn bay may spill-back into the through lanes, or conversely when the through lanes may spill-back to cut off entry into the left turn bays. The combination of lane striping, to change certain lanes from being exclusive to being shared, and the selective opposition of vehicles, depending upon the direction in which they are turning, permits the implicit computation within the model of shared lane saturation flow rates.

4.5 Stop and Yield Signs

Exactly the same mechanism, which is described above for the simulation of left turns, can also be utilized to model the impact of stop or yield signs. In this case, different critical gap sizes are identified and several links may concurrently oppose a given turning movement. The simulation logic within INTEGRATION is set up to automatically model the hierarchy in gap acceptance priority of one movement over a lower priority movement. The above hierarchy not only permits the modeling of all-way stop sign controlled intersections, but also allows for a consideration of impedance when multiple traffic streams seek gaps in the same opposing traffic flow.

4.6 Signal Coordination

The INTEGRATION model is also capable of evaluating the impact of alternative signal effects. Specifically, the discharge pattern downstream of any upstream intersection is preserved (subject to dispersion resulting from variability in vehicle speeds, and the inflow and outflow of vehicles at mid-block intersections that are not signalized) as it travels down the link to become the inflow pattern of the next traffic signal.

INTEGRATION is not constrained, however, to operating all traffic signals on a common cycle length. Consequently, it is possible to simulate explicitly the impact of a lack coordination at the boundary of 2 coordinated subnetworks and/or to evaluate the impact of removing a single traffic signal out of a coordinated network in order to explore the relative benefits of placing one intersection under critical intersection control.

5. MEASURES OF EFFECTIVENESS

It is implicit, in the earlier discussion of the use of speed-flow and car-following relationships, that the INTEGRATION model does not have built into it an explicit link travel time function in a fashion similar to most macroscopic or planning oriented traffic assignment models. Instead, link travel time emerges as the weighted sum of the speeds that are encountered by each vehicle as it traverses each link segment. This distinction introduces both a level of complexity and accuracy not present in most of these other models.

Specifically, the dynamic temporal and spatial interactions of shockwaves, which form upstream of a traffic signal, or along a freeway link that is congested, are such that the final link travel time is neither a simple function of the inflow nor the outflow of the link. Instead, the travel time is a complex product of the traffic flow dynamics along the entire link, and the temporal interactions of this flow with the signal timings and flow oppositions at the end of these links.

The strength of a microscopic approach is that, beyond the basic car-following/ lane-changing/ gap-acceptance logic, there is no need for any further analytical expressions to estimate either uniform, over-saturation, coordination, random, left-turn or queue spill-back delay. While such complexity precludes the simplicity of a functional relationship, such as for example the Bureau of Public Roads relationship, it also permits 2 distinct travel times to be properly considered for the same flow level, depending on whether forced or free-flow conditions prevail, and can deal much more readily with the concurrent presence of multiple vehicle/driver types on the same link.

5.1 Estimation of Link Travel Time and Stops

The model basically determines the link travel time for any given vehicle by providing that vehicle with a *time card* upon its entry to any link. Subsequently, this *time card* is retrieved when the vehicle leaves the link. The difference between these entry and exit times provides a direct measure of the link travel time experience of each vehicle. Furthermore, each time a vehicle decelerates, the drop in speed is also recorded as a partial stop on the above *time card* and eventually provides again a very accurate explicit estimate of the total number of stops that were encountered along that particular link.

It is noteworthy that INTEGRATION will often report that a vehicle has experienced more than one complete stop along a given link. Multiple stops arise in this case from the fact that a vehicle may have to stop several times before ultimately reaching the link's stop line. This finding, while seldom recorded by or permitted within macroscopic models, is a common observation within actual field data for links on which considerable over-saturation queues occur.

5.2 Estimation of Fuel Consumption

The INTEGRATION model computes the speed of vehicles each deci-second. This permits the steady state fuel consumption rate for each vehicle to also be computed each second on the basis of its current speed. In addition, by tracking the change in speed, from one time instant to the next, it is also possible to determine the amount of additional fuel that is likely to have been consumed by the vehicle due to any acceleration and deceleration cycles.

The default coefficients, that are utilized to estimate the above steady speed and acceleration oriented fuel consumption, are usually derived internal to the model. The default vehicle is a 1992 Oldsmobile Toronado (14,15), but the derivation of these coefficients for any other vehicle can be performed on the basis of the published EPA city and highway mileage ratings. These base fuel consumption rates are modified in view of the prevailing ambient temperature. Additional fuel consumption penalties are typically also assigned while a vehicle's engine is warming up during the first part of its trip.

The above fuel consumption analysis features are built into the model and are executed every second for every vehicle in the network. They are also applied in a fashion that is consistent across all facility types, operating regimes, and control strategies. This consistent internal use of the same general fuel consumption model permits a very objective assessment of the fuel consumption implications across a wide range of potential traffic or demand management strategies.

5.3 Vehicle Emissions of HC, CO and NO_x

A series of compatible vehicle emissions models are coupled to the above fuel consumption model. These models, which estimate hydrocarbon, carbon monoxide and nitrous oxide emissions, also operate on a second by second basis (15). They are sensitive to the speed of the vehicle, the ambient temperature and the extent to which the vehicle's catalytic converter has already been warmed up during an earlier portion of the trip.

Applications of these models have shown that the emission of these three compounds is often related to vehicle travel time, distance, speed and fuel consumption in a highly nonlinear fashion. Consequently, traffic management strategies, which may have a significant positive impact on one measure, are not always guaranteed to have an impact of either the same magnitude or sign on any of the other measures. As a result, the types of analyses that can be performed extend far beyond the capabilities of EPAs MOBILE5 model (16), which considers a single fixed speed profile for any given average speed and considers primarily the number of vehicle miles traveled as the main predictor variable.

The execution of the INTEGRATION model, for the EPA city and highway speed profiles, has yielded emission estimates consistent with those estimated by MOBILE5. In contrast, the analyses of other speed profiles, which still yield the same average speed, have also been shown to often yield very different emission quantities.

5.4 Aggregation of Statistics by Link and O-D Pair

The same *time card* concept, that is used for recording a vehicle’s travel time and number of stops on a particular link, is also utilized to track the fuel consumption and cumulative emissions for each vehicle on each link. Furthermore, internal to the model, these statistics are further aggregated, both for all links traversed by a particular vehicle, and for all the vehicles that have traversed a particular link. The former statistics can be aggregated at the O-D level by time period or vehicle type, or they can be aggregated by time period for each link or by cell within a latitude/longitude grid. When emission data are tracked by latitude and longitude as a time series, these data can in turn be provided as input to an external air quality emission model of the atmospheric conditions for an entire urban area.

In addition to tracking the number of lane changes occurring within the network and counting the number of vehicle passes, the model also provides an estimate of cumulative accident risk. This accident risk is estimated on a second by second basis by cross-multiplying the distance driven by a particular vehicle against the accident rate per unit distance for that link and vehicle type. The latter unit distance accident risk can be facility type dependent, reflect the impact of the presence of congestion, and may also reflect the use by a particular vehicle of a given ATIS technology. The use of the model in this capacity also permits the estimation of accident risk reduction as a function of the level and quality of ATIS deployment.

5.5 Loop Detectors and Vehicle Probes

INTEGRATION’s microscopic approach also permits rather realistic representations of the expected surveillance data that would be obtained from loop detectors and/or vehicle probes.

Specifically, when a vehicle traverses any location that is considered to be a loop detector site, the model records a vehicle count, the vehicle’s speed, and the estimated vehicle detector occupancy. These data are then accumulated into a 20, 30 or 60 second reading, which can be provided as an additional model output. The ability to locate vehicle detectors anywhere on a link, and/or to locate multiple vehicle detectors on a given link, permits considerable flexibility in the collection of additional model statistics. In addition, it permits the evaluation of alternative surveillance levels in support of real-time control and/or various incident detection schemes. Detectors can also be coded as trapping only a certain vehicle type in order to represent, for example, a special bus or truck sensor.

Certain vehicles can also be flagged as being vehicle probes (18). In the most basic case, a separate record is generated when a probe vehicle starts or ends a trip. In a more intermediate level of analysis, a separate record can be generated each time a probe vehicle completes a link, while in its most detailed form, a separate record can be recorded for each probe vehicle every second. The latter more detailed statistics are most useful in tracking the speed profiles of vehicles along a given link.

6. CONCLUSIONS

The INTEGRATION model is somewhat unique in that it has tried to combine various microscopic details of car-following, lane changing and gap acceptance behavior with such macroscopic features as traffic assignment, coordination delay and speed-flow relationships. These features are available on networks ranging from a single intersection or weaving section up to several thousands of links. During the past decade, ITS has created a need for models that can concurrently consider these diverse dynamic network attributes, while advances in computer memory and computational speed have made the application of this type of microscopic model feasible.

The objective of this paper has been to describe some of the main traffic flow features of the model and to discuss some of the rational for including them. It is anticipated that these descriptions will make it easier for potential model users to both appreciate the opportunities and limitations of the model prior to considering its use. This overview should also have provided a context within which complementary papers that deal with specific model elements, such as weaving, HOV lanes or gap acceptance can be viewed.

7. REFERENCES

1. *Van Aerde, M., (1985), Modeling of Traffic Flows, Assignment and Queuing in Integrated Freeway/Traffic Signal Networks, Ph.D. Thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Canada.*

2. *Van Aerde, M., and Yagar, S., (1988),* Dynamic Integrated Freeway/Traffic Signal Networks: A Routing-Based

Modeling Approach, *Transportation Research Record A, Volume 22A, Number 6, pp. 445-453.*

3. Van Aerde, M., and Yagar, S., (1988), Dynamic Integrated Freeway/Traffic Signal Networks: Problems and Proposed Solutions, *Transportation Research Record A, Volume 22A, Number 6, pp. 435-443.*

4. Van Aerde, M., and Yagar, S., (1990), Combining Traffic Management and Driver Information in Integrated Traffic Networks, *Third International Conference on Road Traffic Control, IEEE Conference Publication Number 320, London, England.*

5. Van Aerde, M., (1995), A Single Regime Speed-Flow-Density Relationship for Congested and Uncongested Highways, *Presented at Transportation Research Board 74th Annual Meeting [950802], Washington, D.C.*

6. Van Aerde, M., and Rakha, H., (1995), Multivariate Calibration of Single Regime Speed-Flow-Density Relationships, *VNIS/Pacific Rim Conference, Seattle, WA.*

7. Rilett, L., and Van Aerde, M., (1991), Modeling Distributed Real-Time Route Guidance Strategies in a Traffic Network that Exhibits the Braess Paradox, *VNIS Conference Proceedings, Dearborn, MI.*

8. Rilett, L., and Van Aerde, M., (1991), Routing Based on Anticipated Travel Times, *Application of Advanced Technologies in Transportation Engineering, Proceedings of the Second International Conference, ASCE, pp.183-187.*

9. Robinson, M., and Van Aerde, M., (1995), Examining the Delay, Safety Impacts and Environmental Impacts of Toll Plaza Configurations, *VNIS/Pacific Rim Conference Proceedings, Seattle, WA.*

10. Van Aerde, M., Baker, M., and Stewart, J., (1996), Weaving Capacity Sensitivity Analysis using INTEGRATION, Transportation Research Board 75th Annual Meeting, Washington, D.C.

11. Stewart, J., Baker, M., and Van Aerde, M., (1996), Analysis of Weaving Section Designs using INTEGRATION, *Transportation Research Board 75th Annual Meeting, Washington, D.C.*

12. Rakha, H., and Van Aerde, M., (1996), A Comparison of the Simulation Modules of the TRANSYT and INTEGRATION Models, *Transportation Research Board 75th Annual Meeting, Washington, D.C.*

13. Velan, S., and Van Aerde, M., (1996), Relative Effects of Opposing Flow and Gap Acceptance on Approach Capacity at Uncontrolled Intersections, *Transportation Research Board 75th Annual Meeting, Washington, D.C.*

14. Van Aerde, M., and Baker, M., (1993), Modeling Fuel Consumption and Vehicle Emissions for the TravTek System, *VNIS Conference Proceedings, Ottawa, Canada.*

15. Baker, M., and Van Aerde, M., (1995), Microscopic Simulation of EPA Fuel and Emission Rates for Conducting IVHS Assessments, *ITS America Conference,Washington, D.C.*

16. US EPA, (1993), User's Guide to MOBILE5A: Mobile Source Emissions Factor Model.

17. Hellinga, B., Baker, M., and Van Aerde, M.,(1995), Linking ATIS/ATMS and Environmental Plume Dispersion Models, *VNIS/Pacific Rim Conference, Seattle, WA.*

18. Van Aerde, M., Hellinga, B., Yu, L., and Rakha, H., (1993), Vehicle Probes as Real-Time Sources of Dynamic O-D and Travel Time Data, *Advanced Traffic Management Systems Conference, St. Petersburg, FL.*