Costs of Congestion in Canada's Urban Areas: Methodological Considerations

David Kriger, P.Eng., MCIP, iTRANS Consulting Inc., Ottawa

Fannie Joubert, Les conseillers ADEC inc., Montréal

Mark Baker, P.Eng., Delcan Corporation, Ottawa

Cristobal Miller, Transport Canada, Ottawa

Paper prepared for presentation
at the *Emerging Best Practices in Urban Transportation Planning (B)* Session
of the 2005 Annual Conference of the
Transportation Association of Canada
Calgary, Alberta

ABSTRACT

'Congestion' is commonly cited as a major urban ill by the public, politicians and the media. Urban transportation authorities aim continuously to manage (if not altogether eliminate) the problem through a variety of measures.

But what do we mean by congestion? How do we quantify it? What is its cost? Some Canadian urban areas have attempted to answer these questions. However, methods, data, approaches and assumptions have varied. A recent Transport Canada study has provided the first comprehensive and systematic analysis of congestion: The *Costs of Congestion in Canada's Transportation Sector* study developed congestion indicators for the nine largest urban areas in Canada (Québec City, Montréal, Ottawa-Gatineau, Toronto, Hamilton, Winnipeg, Calgary, Edmonton and Vancouver).

The indicators were based upon data that were derived from each urban area's travel demand forecasting models. Although the models all produce the same outputs (i.e., simulations of vehicle [and other] trips), there are structural and methodological differences among them. The resultant indicators thus are not easily compared; however, they do provide different perspectives on congestion in these urban areas, which contain over half the population of Canada.

Accordingly, the research required the development of common means to measure congestion and extract the requisite data from the nine models. Among other results, the research found differences ranging from how expressways and arterials are defined in model networks to the time periods and trip purposes that are considered in each model, to the ways in which speeds were calibrated. Some of the differences are subtle, while others are more obvious; but all have an impact on the practice and application of travel demand forecasting (as well as to the measurement of congestion).

This paper reviews these differences, as well as the points of commonality, and discusses their implications on the analysis of congestion. It explains the role of models in analyzing congestion, and provides a basis for urban authorities to conduct their own congestion analyses. The paper also provides some suggestions for further research.

1. INTRODUCTION AND PURPOSE

1.1 Introduction

What is *congestion*? How can it be measured? What is its monetary cost? The public, politicians and media all talk about *traffic congestion* as a serious urban problem, with the topic regularly being at or near the top of the lists of urban ills in opinion surveys. However, the ability to *measure* congestion – that is, to quantify and value it - is fundamental to being able to address it.

Some urban authorities have attempted to quantify congestion and its monetary value. Studies in Toronto (1987), Ottawa-Gatineau (1990) and Vancouver (and 1996 and 1999) examined congestion and quantified its impact on trucking costs (the Vancouver work also included the costs to auto drivers). However, the methods and – especially – the sources of data varied considerably among the studies, thereby limiting the ability to interpret and replicate the findings.

By far the most comprehensive analyses of urban congestion and its costs have been conducted by the ministère des Transports du Québec (MTQ) for the Montréal region. An annual congestion cost of $0.5 billion was estimated according to 1993 travel conditions, and subsequently was updated to $0.9 billion for 1998 ($0.8 billion for automobiles and $0.1 billion for trucks). These model-based studies provided a prototype for the current research, in that they defined congestion and methods and developed costs that reflected the primary impacts of congestion (time, operating costs, fuel costs, emissions).[1]

In recognition of these considerations, Transport Canada initiated the first comprehensive study of congestion in urban Canada, as part of its ongoing research in understanding the factors that influence sustainable transportation. Transport Canada's specific interest was in the greenhouse gas (GHG) emissions that are caused by vehicles operating under congested conditions. In its ongoing sustainability research, Transport Canada also recognizes that *pricing* has an important role in understanding traveller behaviour (that is, in the way that people make decisions about their travel). The understanding of congestion and its costs provides an essential basis for urban authorities to develop their own ways to address their transportation needs, sustainable transportation and climate change.

Among other objectives, the study – entitled the *Costs of Congestion in Canada's Transportation Sector* - included the development of:

- An analytical basis of definitions and measures.
- Methods for *consistent* measurement.
- Recommendations for methodological and data enhancements to improve *comparability* over time.

The study applied the resultant measures and indicators to Canada's nine largest urban areas (from east to west, these are Québec City, Montréal, Ottawa-Gatineau, Toronto, Hamilton, Winnipeg, Calgary, Edmonton and Vancouver). The nine urban areas represented just over half (51%) of Canada's population in 2001. The quantification of the components of congestion then was translated to monetary values, in order to develop a total cost of congestion for each urban area.[2]

1.2 Purpose

Transport Canada's original intent was to use these indicators in order to be able to *compare* congestion among the nine urban areas, both as a means to increase the awareness of the public and politicians (at all levels of government) as to the importance

1 Gourvil, L. and F. Joubert. 2004. Évaluation de la congestion routière dans la region de Montréal. (*"Estimation of roadway congestion in Montreal."*) ADEC Consultants and the ministère des Transports du Québec. Montréal: Ministère des Transports du Québec.

2 The cost information has not yet been released by Transport Canada and, accordingly, it is not discussed further in this paper.

of congestion in sustainable transportation, and to provide improved analytical methods that urban authorities could apply in their own plans to address congestion.[3]

The study did succeed in generating considerable attention in the subject of congestion among authorities at all levels of government (including – notably – cities and provinces that are not represented among the nine largest urban areas); and the *methods* that have resulted from this research are already being applied in urban transportation planning.

However, the *comparability* of the indicators has proven to be limited, for the fundamental reason that the underlying data from each urban area (more precisely, the travel demand forecasting models from which these data are drawn) differ in composition, structure, currency and coverage.

This is in contrast to the situation in the United States, whose *Urban Mobility* indicators (which were an important basis for this research) are based upon a common set of data for all 85 urban areas that are included in that study; the commonality of these data enables the comparison.[4] On the other hand, the 'top-down' data in the US provide a more limited perspective than the 'bottom-up' model-based data in the nine Canadian urban areas (and, in fact, have been criticized as masking many of the nuances and dynamics that are fundamental to understanding the phenomenon of congestion).

The differences in data and models constituted a recurring theme throughout the study. On the other hand, *identifying* the underlying differences and possible ways to *address* them, has proven to be an important outcome of this study; both for possible future initiatives in this topic and in helping urban and provincial transportation planning authorities consider their own modelling capabilities and needs.

The purpose of this paper, then, is to discuss these methodological differences from the perspective of congestion analysis. The paper is organized as follows: Section 2 provides a context for the discussion by reviewing briefly how congestion is analyzed, and the role of models in this analysis. Section 3 presents an overview of the nine urban models and the data upon which they are based, how they compare with each other and the implications of the various differences in the models and data differences on congestion analysis. Section 4 concludes the paper with a presentation of possible future directions for further analysis and research.

It is important to note that this paper is neither a critique nor a judgment of the state of the practice in urban transportation modeling in Canada, especially given that each of the models was developed for a particular purpose. Rather, the paper examines the current and potential adaptation of these types of models to the specific analysis of congestion.

---

3 Transport Canada has stated unequivocally that the results of this research will *not* be used to rank potential funding contributions for transportation investments.

4 However, differences in the way these data are sampled (e.g., the spacing between count locations on a particular facility), collected (e.g., some urban areas use automatic counting equipment, while others do not) and treated (e.g., different aggregation levels) imply some inconsistency even among these 'common' data. Recent research has attempted to replicate the 'bottom-up' approach used in this study (for example, by the Texas DOT).

2. CONTEXT

2.1 Perspective

It is important to note at the outset that the analysis of congestion is considered to have two perspectives: the *engineering* approach, which focuses on the direct and physical characteristics of congestion; and the *economic* approach, which considers congestion in a market context of travel supply and demand and in terms of its broader societal impacts. The reconciliation of the two approaches can be problematic, given that *urban* transportation planning *practice* (the source of the data and models upon which the study was based) focuses on the engineering approach. The consideration of the broader economic approach is not as common; or at least is not treated as explicitly as the engineering approach in urban transportation planning practice. It was agreed at the outset that the study would be based upon the engineering approach; however, both approaches have their merits and both were recognized in the study.

2.2 Definition of Congestion

The most common theoretical definitions of congestion in the literature make reference to *vehicle flow* (actual throughput, measured as vehicles per hour) and *road capacity* (available capacity for throughput, measured as vehicles per hour); i.e., congestion is 'too many vehicles attempting to use the same road space at the same time.' This provides the basic reference to the engineering analysis of congestion; that is, explaining congestion according to the physical characteristics of the road.

However, congestion can also be considered in economic terms, as a level of traffic that is greater than an optimal level, at which the full costs to maintain the level of service are paid. Implicitly, congestion is the result of travellers' decisions that are made without consideration of the total costs to society. This treatment broadens the consideration of the subject, since it expresses congestion in terms of its inconvenience to people, economic activity and society as a whole.

As noted, this study focused upon the engineering definition, although it clearly recognized the importance of the economic definition of congestion. Accordingly, in consultation with urban and provincial authorities, the study adopted the following general definition:

> "The inconvenience and increased costs that travellers impose on each other while using their vehicles, attempting to use the road network at the same time, because of the relationship that exists between traffic density and speed (with due consideration of capacity)."

This definition links economic considerations (inconvenience and costs) with engineering considerations (actual and available throughput). It also broadens the engineering consideration from a count of vehicles in a given time to recognition that the speed at which these vehicles move and their density (number of vehicles per unit distance) also must be considered.

2.3 Measurement of Congestion

Generally, congestion happens at a certain level of traffic on a given road in relation to a fixed capacity or according to certain characteristics of vehicle flow. In order to quantify congestion, one must determine the "threshold" at which congestion becomes apparent and is deemed unacceptable. It is only against this reference that the socio-economic costs can be accumulated. The evaluation of congestion depends greatly on this threshold and can be very sensitive to it.

Congestion is a function of a reduction in speeds (i.e., which is the direct cause of loss of time [delays]) and leads to increased vehicle operating costs and increased air pollution and GHG emissions. Therefore, the setting of a threshold that is directly related to travel speeds is most appropriate. A speed-based threshold thus appears to account for more of the impacts of congestion than would a threshold that is based on capacity. Because it is concerned with a reduction in speeds, it circumvents the problems that are associated with the use of free-flow conditions. Therefore, the threshold was based upon a percentage of the free-flow speed.

In other words, it is important to note that although free-flow conditions can be fixed, the percentage of free-flow speed that represents the threshold varies according to local conditions (quantitative) and perceptions (qualitative). The process of selecting the values of the threshold is a function of three related tasks:

- <u>Review of observed travel time – traffic flow conditions</u>. Typically, observations exist only for a small number of road and highway sections in a given urban area. These provide quantifiable reference points (that is, they define a range) for identifying appropriate threshold values for that urban area.

- <u>Perceptions of local 'users' (i.e., travellers)</u>. Because the observations reflect a range, it is necessary to take into account local perceptions of the point(s) – that is, the traffic conditions and speeds – at which travellers 'perceive' congestion to begin: identical conditions that are viewed to be non-congested in one city may be perceived as highly congested in another city, or even in different parts of the same city. Often, the professional judgment of transportation planning professionals commonly must be used, given the general unavailability of documentation on local perceptions.

- <u>Extrapolation of these threshold values to reflect the entire system</u>. This entails the testing of different threshold values in the urban area's travel demand forecasting model, which typically is the only systematic, quantitative means for extrapolating the limited observations from specific road sections across an entire network. The model results are then compared with the aforementioned observed thresholds (quantitative comparison) and the perceptions of local travellers or professional judgment (qualitative comparison), in order to identify appropriate threshold values for that urban area.

On this basis, a range of threshold values of 40% to 60% was adopted for different types of facilities (expressways and arterials: by definition, congestion is considered to occur only on these higher-order facilities[5]). It is important to note that some urban areas consider thresholds of 70% and 80% to be more reflective of local perceptions of congestion. These higher thresholds also were closer to the traditional level of service boundaries that are used to identify the need for new capacity in many long range transportation master plans (e.g., service levels D, E and sometimes F).[6]

With the aforementioned measure of congestion in place, indicators of congestion could now be derived. After a review of several candidate indicators from the United States and Europe, the following five indicators were selected from the candidate indicators, based upon relevance to the Canadian situation, data availability, data quality and 'replicability:'

- Travel delay (extra time spent in congestion).
- Wasted fuel (due to slower speeds).
- Roadway congestion (relative importance of links with high volumes).
- Travel rate (additional time required; related to travel delay but expressed as a ratio).
- Transit congestion (travel delay accrued by bus and rail transit that operates in mixed traffic). The need for a transit indicator was identified by stakeholders.

2.4 Role of Models in Congestion Analysis

Ideally, the thresholds should be based upon extensive observed data for each urban area. However, as discussed in Section 2.3, the required data (specifically, measurements of traffic volumes and speeds, and the variation by hour of day, day and year) were available only for limited sections of roads over limited time periods in each urban area. Generally, these available data were too sparse to serve as the basis for a meaningful extrapolation.

As well, the available observations cannot be easily compared among urban areas: the quality, coverage, frequency, currency, collection methods and even the type of basic traffic and speed information is known to vary among urban areas and provinces. This is in contrast with the situation in the United States, in which the Federal Highway Administration's *Highway Performance Monitoring System* (HPMS) database provides a nation-wide, uniform (if incomplete) set of observations for urban areas (and anecdotal evidence suggests that there remain problems of consistency and applicability at individual urban areas, even with this broad, uniform coverage – also, the aggregated nature of the data limits the ability to capture the dynamics of congestion).[7]

---

[5] That is, congestion certainly is known to exist on collectors and local streets; however, for the purposes of a region-wide analysis, it is considered to be somewhat localized in impact. Moreover, the aforementioned MTQ study found that the percentage of congestion on these facilities is relatively small, with 95% of the congestion in Montréal occurring on the expressways and arterials.

[6] It is essential to note that the service levels used for these transportation plans are a function of volume and capacity; whereas the threshold used to define congestion is a function of speed. Thus, the two measures cannot be equated directly.

[7] According to the FHWA's website, "the HPMS is a national level highway information system that includes data on the extent, condition, performance, use, and operating characteristics of the Nation's highways." Stratified samples are used to collect the data, which include "limited"

Moreover, the stakeholder consultation meetings revealed that detailed observed data were available only for expressways in Toronto and Montréal (and only for some sections of these expressways); and no consistent data were available for arterials. In addition, the definition of what constituted an 'arterial' or an 'expressway' varied among urban areas.

Urban areas generally have found that a more cost-efficient way to profile urban travel is to combine selective data collection with the development of a travel demand forecasting (transportation) model. These models simulate travel demand, typically for autos and public transport (and, in some urban areas, also for pedestrian trips, cycling trips or trucks), across an entire urban region.

Accordingly, outputs from the transportation (travel demand forecasting) models in each urban area were requested as a key source of data. This required the development, in consultation with the urban authorities, of common categorizations of how the nine urban models define expressways, arterials and other road links, and how they treated the modelling of travel demand. As well, there was a need to reconcile the speeds computed in the models with actual posted speeds and/or observed free-flow speeds. Through this analysis, it became apparent that there were many differences in data and modelling methods, and that there were gaps in the available data. Nonetheless, given these constraints and with the extensive cooperation of the urban and provincial authorities, a consistent set of measures was developed for this research.

In sum, it is important to note that neither the available data nor the transportation models provide the complete picture of travel. However, for the purposes of this research, the available transportation models provide the best platform upon which to develop the measures and indicators. These models and the available data are reviewed in the next section.

3. ROLE OF TRANSPORTATION MODELS IN ESTIMATING CONGESTION

3.1 Categorization of the Available Models

Table 1 summarizes the different relevant features of the nine urban models and associated data. The nine models can be grouped into three main classes according to their structure. This categorization was important in explaining the differences and in defining common bases for the derivation of the required data:

information on travel (notably, daily traffic counts on sampled sections of expressways, arterials and other types of roads). The HPMS data also can be combined with data from the US Census and other sources. (See http://www.fhwa.dot.gov/policy/ohpi/hpms.)

Table 1. Key Attributes of Models and Reference Data

Organization	Base year	Time(s) of day modelled	Model formulation	Transit modal split formulation	Modes	Treatment of peak spreading	Reference data	Comments
Ministère des Transports du Québec (la région de la Capitale nationale)	2001	AM peak period (peak hour for assignment)	Trip assignment only – OD survey trip tables used instead of demand models	Not applicable	Auto driver Auto passenger	Yes - by varying peak hour factor according to observation	2001 OD survey Screenline and cordon counts	The presence of truck trips is taken into account on principal links Transit trips are simulated in MADITUC
Ministère des Transports du Québec (la Grande région de Montréal)	1998	AM peak period (peak hour for assignment)	Trip assignment only – OD survey trip tables used instead of demand models	Not applicable - but some modal split is done externally	Auto driver Auto passenger Light truck Heavy truck	Can vary peak hour factor (% of peak period) according to observation	1998 OD survey Screenline and cordon counts (1998 and earlier) Real-time speed-delay surveys (some expressways and arterials)	Transit trips are simulated in MADITUC New OD survey conducted in 2003; model to be updated
TRANS Committee (Ottawa-Gatineau)	1995	PM peak period (peak hour for assignment)	Four-step demand model	Logit modal split and diversion curves	Auto driver Auto passenger Transit	2½ hour peak period demand model is factored to yield peak hour auto and transit trips Factors are held constant but could be varied	1995 OD survey Annual screenline and cordon counts	New OD survey scheduled for 2005; model recalibration to follow AM and PM peak period models calibrated in 1992, using 1986 OD (first bi-period model in Canada)
City of Toronto	2001	AM peak period (peak hour for assignment)	Four-step demand model	Disaggregate nested logit mode choice model, accounting for mixed modes (ie, park-and-ride)	Auto driver Auto passenger Transit / mixed modes Pedestrians	3 hour peak period demand model is factored to yield peak hour auto and transit trips	Transportation Tomorrow Survey (2001 and earlier) Annual screenline and cordon counts Real-time speed-delay surveys (Highway401 FTMS)	"GTAModel" allows for some activity-based modelling Modelling of home-work trips accounts for employment categories
City of Hamilton	2001	AM peak period (peak hour for assignment), with some limited analyses in the PM peak hour	Four-step demand model (Fratar [growth factor] used for trip distribution)	Logit modal split	Auto driver Auto passenger Transit	3 hour peak period demand model is factored to yield peak hour auto and transit trips	Transportation Tomorrow Survey (2001 and earlier) Annual screenline and cordon counts Speed-delays later	'Two" peaks: internal and to GTA and elsewhere in southern Ontario Model recalibrated 2003

Table 1. Key Attributes of Models and Reference Data

Organization	Base year	Time(s) of day modelled	Model formulation	Transit modal split formulation	Modes	Treatment of peak spreading	Reference data	Comments
City of Winnipeg	1986	AM peak hour	Four-step demand model	Logit modal split	Auto persons (factored to separate drivers and passengers) Transit	N/A	1992 OD survey Screenline and cordon counts Speed-delay studies on several roads	City considers model to be out-of-date Existing formulation being transferred to TransCAD
City of Edmonton	1994	Five 1-hour periods consisting of AM / PM peaks and shoulders, plus off-peak daytime hour	Nested logit with choice behaviour for trip generation, distribution, time of day and mode	Incorporated within generation, distribution, time of day and modal choice 'chain'	Auto driver Auto passenger Transit Pedestrians Cyclists Commercial vehicles	Model allocates trips between the peak hour and peak shoulder	1994 OD survey Annual screenline and cordon counts	Pioneered simultaneous choice behaviour models in Canada (rather than traditional sequential choices)
City of Calgary	2001	Five 1-hour periods consisting of AM / PM peaks and shoulders, plus off-peak daytime hour	Nested logit with choice behaviour for trip generation, distribution, time of day and mode	Incorporated within generation, distribution, time of day and modal choice 'chain' Accounts for P+R	Auto driver HOV2, HOV3 Transit / P+R Pedestrians Cyclists School buses Commercial vehicles, incl. taxis	Model allocates trips between the peak hour and peak shoulder	2001 OD surveys, including commodity flow survey Annual screenline and cordon counts	Based upon Edmonton formulation
TransLink (Greater Vancouver)	1996 (1999 for trucks)	AM peak period (peak hour for assignment), with PM peak period / hour model based on AM model	Four-step demand model Truck demand modelled separately (generation, distribution)	Logit modal split Accounts for mixed modes (P+R)	Auto driver HOV2 / HOV3 Transit / P+R Pedestrians Cyclists Light trucks Heavy trucks	2 hour peak period demand model is factored to yield peak hour auto and transit trips Factors are held constant but could be varied	1996 OD survey / 1999 truck OD survey 1996 and 1999 screenline and cordon counts	Separate light- and heavy-truck model is integrated within main passenger model

- Five of the models (Ottawa-Gatineau, Toronto, Hamilton, Winnipeg and Vancouver) follow the traditional four-step paradigm: trip generation (which estimates the magnitude of trips starting and ending at traffic 'zones'); trip distribution (distributes these trips among the zones as a function of accessibility and of the magnitude of population, jobs, etc., at each zone); modal split (allocates the trips by mode as a function of modal travel time and costs); and, trip assignment (loads the trips onto the specific road network and transit routes). This paradigm is the most commonly used transportation model structure around the world.

- The Montréal and Québec City models simulate only trip assignment. Instead of demand models (trip generation, trip distribution and modal split) within the travel demand forecasting framework, the MTQ projects the base origin-destination survey matrices using detailed demographic projections and taking into account employment growth and improvements to the transportation network. The same method is applied to both urban areas.

- The Edmonton and Calgary models combine the generation, distribution, trip start time (i.e., a model of when the trip actually takes place) and modal split into a series of *simultaneous* decisions: in other words, whereas the traditional four-step model treats these as sequential decisions (and assumes that the start times of these trips are fixed to the peak hour), the two models recognize that the decisions are all related. One result is that the two models can simulate varied trip start times, i.e. modelling a shift between the peak hour and the shoulders of the peak (if congestion so dictates): Most of the other models account for the peak hour as a factored proportion of the peak period.

3.2 Review of the Available Models and Data

An important goal of the study was to *apply* the recommended measure and indices to the nine largest urban areas in Canada. However, based upon the information summarized in Table 1, and notwithstanding several areas of commonality, it became clear that several differences in model structure and, especially, gaps in data required modifications to how the indices actually were developed.

The points of commonality and differences in the models and data are summarized below. **It again is important to note that it was not the purpose of either the research or this paper to criticize or judge the models in any way.**

- All nine urban areas used the EMME/2 software, although the Winnipeg model subsequently was being transferred to the TransCAD software. (The choice of software is not necessarily the determinant of a model's structure; however, common software simplifies the exchange and processing of information and, especially, the coding of networks.)

- Each model simulated, at a minimum, auto vehicle trips during either the a.m. or the p.m. peak hour (or, in the case of Edmonton and Calgary, the peak *half*-hour ['crown']; which was simulated for both the a.m. and p.m. peaks and from which the data could be extrapolated to represent peak *hour* conditions). The

Vancouver and Hamilton models both had some abilities to model the second (p.m.) peak. The Edmonton and Calgary models also simulated the shoulders of the two peak periods, as well as an off-peak hour.

- The development of some of the indicators required a comparison between congested (peak) and free-flow conditions. Given their bases in the peak hours, the models all were able to simulate congested travel conditions.

  However, the treatment of free-flow conditions proved to be more challenging. Essentially, free flow represents conditions of unconstrained travel – for example, as might occur in the middle of the night. However, generally data that describe either the magnitude or the distribution (origins-destinations) of these trips do not exist. Accordingly, a common simplified treatment is assign a unit matrix to the road network, where the unit trips (i.e., 1 vehicle-trip between each origin-destination pair) are intended to represent minimal travel on the road network.

  However, in some circumstances, even this small value was higher than the actual demand (as reflected by the [few] screenline traffic counts that existed for the times of day at which free-flow occurred). The traffic patterns varied considerably from those of other times of day. The unit trips tended to be assigned to the highest-capacity facilities. Finally, the use of a unit matrix proved problematic in urban areas that had a large number of zones, since the greater the number of zones, the greater the magnitude of trips on the system.

  Notwithstanding these limitations, for the purposes of this analysis the urban areas developed free flow assignments using unit matrices, the values of which ranged between 0.01 and 1. (The City of Edmonton, in fact, tested several values for the unit matrix, ultimately developing the free-flow link travel times by assigning a null (0-filled) matrix and applying these to the congested vehicle flow on each link.) However, it was recognized that more data would be required in order to more accurately depict free flow conditions (specifically, new traffic counts and origin-destination data).

- All models, except those for Montréal and Québec City, simulated transit trips. Transit in Montréal and Québec City was simulated in a separate model (MADITUC). In addition to simulating the auto and transit modes, the Montréal, Calgary, Edmonton and Vancouver models also simulated truck trips (albeit in different ways). Walking and cycling was simulated, again in different ways and at different levels of detail, in the Toronto, Edmonton, Calgary and Vancouver models. The Toronto, Edmonton, Calgary and Vancouver models also simulated mixed modes; notably, park-and-ride (transit / auto). The Toronto model distinguished between transit and commuter rail (GO Rail).

- All models simulated several trip purposes; notably, the home-work commute but also the home-school commute. These represented the common, non-discretionary peak trips (i.e., these were trips that must be made regularly and commonly between the same origin and destination). Discretionary trips also were modelled, albeit to different levels of categorization (e.g., shopping, business related to work, serve passenger [pick-up and drop-off], recreational, etc.).

- Each model was based upon travel (origin-destination) surveys, which quantify urban travel behaviour. However, these surveys – i.e., the base years of the models – varied in age. Calgary, Toronto and Hamilton had or were developing recent models, based upon 2001 data; but Winnipeg's model was based on data from 1986 and 1992. The urban areas also had screenline and cordon counts (which count vehicles and their occupants at strategic points in the transportation network, distinguishing the counts by type of vehicle). Some urban areas also had speed-delay studies (which quantify the time it takes to travel on a particular route, including stops at intersections or due to congestion, at different times of the day [i.e., under different levels of congestion]). Finally, as noted real-time speed / volume data were available only in Toronto and Montréal, and only for limited sections of the respective expressway systems.

- There were differences in the areas covered by the models. The need to provide a consistent definition of what is meant by an 'urban area' arises because of the need, in turn, to differentiate in the model results between urban / suburban roads (where congestion is presumed to occur) and rural roads (where congestion presumably is minimal). These areas can be defined by political boundaries, the actual developed urban area, and/or reasonable commuting distances. For example, Calgary's model included the commutershed well beyond the city boundaries; whereas a large area of rural and farm lands is included within the Ottawa-Gatineau city limits (and the model).

- Differences in classification of the models' road networks proved important, because of the need to extract data from expressways and arterials in order to develop the indicators. These differences reflected in part different legal functional classifications and definitions, and in part different model coding philosophies (e.g., whether the ramps on a full interchange were coded explicitly, or as turning movement on a simple intersection). This also was reflected in how the models' volume-delay functions were structured.

  Some urban areas 'tagged' each link by type (for example, the Toronto and Ottawa-Gatineau models), such that each link was explicitly categorized as an expressway, arterial, etc. In contrast, the Vancouver model used a more holistic definition according to lane capacity and posted speed (but independent of functional classifications). Another example was that of Edmonton, where although several sections of arterials are of sufficient length and have the geometric and operational characteristics that are commonly associated with expressways, they must be considered as arterials, and were designated as arterials in the model, because they are controlled and accessed via at-grade intersections (rather than by grade-separated interchanges).

  A related issue concerned the definition of arterials and lower-order roads: The definition of arterials varied considerably among urban areas. There also was the need to differentiate arterials from other roads, such as collectors and local roads.

- Typically, the travel demand forecasting models were calibrated to observed link volumes and characteristics from each urban area's most recent travel (origin-destination) survey. The calibrations depicted a traffic equilibrium on the transportation networks; that is, under times of peak loading they depicted peak traffic volumes and (average) speeds. However, the treatment of speed in calibration tended to be more generalized: it is common to look only at the *reasonableness* of *average* speeds across the system or on sub-sections of the system or on specific corridors. Accordingly, some urban areas expressed concern that the use of link-level speeds could introduce inaccuracies and biases in the results, because the model speeds were not calibrated to this level of detail.

The treatment of speed is an important consideration for congestion analysis, given the basis of the threshold in speed. In particular, there was a general lack of real-time speed-volume data. These were needed to quantify the values chosen for the thresholds. As noted, only Toronto and Montréal had these data; and only for certain sections of their expressway systems. However, some cities had speed profiles (i.e., travel time surveys and traffic counts; e.g. Winnipeg) which, although these were not real-time, could be used as references. (Similarly, as noted, no local data existed regarding the community's perceptions of congestion; accordingly, judgment and comparisons with the findings of other cities were used.)

- Lack of common classifications for traffic counts. A standard, nation-wide definition for vehicle classification counts does not exist in Canada, although there are similarities in the data.

4. CONCLUSIONS: IMPLICATIONS FOR CONGESTION MODELLING AND RESEARCH

The research described in this paper was the first comprehensive attempt to review, define, quantify and assess a cost to urban traffic congestion in Canada's largest urban areas. The topic of congestion, and Transport Canada's research initiatives in particular, has attracted considerable attention among urban and provincial transportation planning authorities.

Differences among the travel demand forecasting models upon which the analysis was based, and a lack of certain types of data, are critical considerations in congestion research. The 'bottom-up' approach used in this study allowed a greater in-depth analysis than would be possible, for example, with the 'top-down' approach used for the *Urban Mobility Report* in the United States (which, as noted, uses the same source of data for all of the urban areas it considers). Neither approach provides a complete perspective on congestion. However, this paper's illustration of the many differences among urban travel models, and in the data upon which they are based, underscores the complexity of congestion analysis and, in turn, the difficulty of trying to describe the phenomenon in simple terms.

It could be argued that a multi-city comparison of models is of limited use to a given urban or provincial authority, which must design its models and data to address its own specific local issues and mandates. However, congestion is one of those 'local issues' – and a commonly-cited one at that: this study provides a means of quantifying and measuring the phenomenon, using an urban area's existing model and data.

The paper, and the study upon which it is based, also provides a basis for potential further research in (at least) three topics that are becoming increasingly important to urban transportation plans:

- <u>Consideration of non-recurrent congestion</u>. The research upon which this paper is based examined "recurrent" congestion (that is, congestion that occurs from the regular, daily build-up of traffic and which, therefore, is somewhat predictable). However, stakeholders in the consultation meetings identified "non-recurrent congestion" also as being important - that is, random delays caused by accidents, inclement weather, natural disasters, truck spills, stalled vehicles, etc. The evidence suggests that the resultant variability in trip times (i.e., reliability) can be more problematic to travellers than is recurrent congestion: the impact is felt by transit providers (a key tenet of whose market is their requirement for schedule adherence) and goods transporters (who must pass on the costs of delay to consumers of the goods they carry).

  However, methods to address non-recurrent congestion are only recently emerging. There is also a considerable lack of data – notably, the aforementioned lack of real-time speed-volume data, but also more precise information regarding how these data change by type of incident, duration, queues, etc., on a region-wide scale. The emerging use of network micro-simulation models and the growing availability of precise, low-cost data collection technologies (such as GPS) suggest much potential for addressing this topic.

- <u>Time-of-day modelling</u>. The treatment of time-of-day choice for trip-making, in addition to modal choice and route choice modelling, is an emerging factor in model development. This is manifest in part by the simulation of peak-spreading (that is, how drivers behave in the face of maximum traffic volumes at the 'crown' of the peak, by advancing or delaying the start time of their trip to precede or follow the crown): this provides a more precise representation of peak period travel. The subject also is important in addressing non-recurrent congestion, because it can simulate (in some now-emerging treatments) how travellers choose their departure time in order to ensure that they arrive at a specified time (or window) at their destination.

  Time-of-day modelling also is expressed through the simulation of travel in different time 'slices' at different times of the day: this allows for a more accurate representation of daily vehicle activity (vehicle-hours and vehicle-kilometres travelled), as opposed to expanding the peak period values. In turn, the multi-'time slice' treatment is important because the daytime off-peak (the time between the a.m. and p.m. commuter peak periods) has seen the fastest growth in travel, to the point that some facilities in some North American cities now operate at a day-long peak period.[8] As well, it can allow for a more accurate treatment of traffic conditions under free-flow.

  In Canada to date, only the Edmonton and Calgary models have incorporated peak-spreading and different time slices, although other cities have recognized

[8] Highway 401 at the Keele Street interchange in Toronto is commonly cited as an example of a facility that operates at virtually a day-long peak period.

both topics as future modelling issues. Edmonton and Calgary model the peak 30 minutes (that is, the highest half-hour of congestion ['crown']) within the two-hour peak period.

It is noted that peak-spreading function tends to minimize the occurrence of hourly link volumes that are greatly in excess of hourly capacity (compared with models that lack this feature, which do not dissipate [spread] the peak hour demand) and that, accordingly, some bias could be introduced when comparing forecasted congestion levels. This also is recognized as a constraint, although one that is more applicable to forecasts of peak spreading (since all model calibrations are intended to replicate actual current conditions). However, it would be appropriate to examine the implications of peak-spreading and time-of-day modelling in forecasts of congestion, as part of possible future research.

- <u>Value of time</u>. Although not discussed previously in this paper, means of incorporating the 'value' to travellers of congestion-related delay are fundamental to the analysis of congestion and its costs. However, there is a general lack of available, up-to-date data: Value of time data have been collected in a few Canadian cities and in some provinces, but only for the purpose of assessing the potential demand for specific new roads and highways under tolling. Accordingly, given their specificity, it is difficult to extrapolate these data to represent both an entire urban area and 'everyday' travel conditions. Moreover, "willingness-to-pay" (that is, the additional premium, for example, that toll road drivers might place on having their 'tolls' collected electronically [i.e., without having to stop]) and variable pricing open up further analytical dimensions that are only recently being considered in Canada. In the event, much of the existing value of time data in Canada is proprietary (because the data are used to predict toll revenues), is dated or reflects inter-urban (rather than urban) values.

5. ACKNOWLEDGEMENTS

The work described in this paper is based upon the study *Costs of Congestion in Canada's Transportation Sector*, prepared for Transport Canada in March 2005 by Delcan Corporation, Les conseillers ADEC Inc. and iTRANS Consulting Inc. Cristobal Miller served as Transport Canada's project manager, and David Kriger of iTRANS served as the consultant project manager. Other contributors to the research whose roles must be acknowledged are: André Bourbeau of Transport Canada, Mark Baker of Delcan Corporation, and Fannie Joubert and Gilles Joubert of Les conseillers ADEC inc.

Appreciation also is expressed to the urban and provincial authorities who kindly provided data, reports, information, comments and advice for the aforementioned study.

The views expressed in this paper are those of the authors, and do not necessarily reflect those of Transport Canada, its consultants or of the aforementioned urban / provincial authorities. No funding commitments by any agency are implied by the research topics that are suggested in Section 4.

6. REFERENCES

(1) Gourvil, L. and F. Joubert. 2004. Évaluation de la congestion routière dans la region de Montréal. (*"Estimation of roadway congestion in Montreal."*) ADEC Consultants and the ministère des Transports du Québec. Montréal: Ministère des Transports du Québec.