Annex 9: Measuring Congestion, Reliability Costs and Selection of Calculation Method Direct Costs

Congestion indicators

There are two general approaches for measuring congestion; an operational approach that has had the favour of those responsible for constructing and managing road networks and an economic-based approach that has generally been used to prioritise public expenditures for transport. The former is typically concerned with observable features of roadway performance (speed, flow, density, queue length and duration), whereas the latter has typically focused on extrapolating physical measures into monetary values that can then serve to guide policy through cost-benefit analysis.

In the former context, engineers have sought to deliver *technically "optimal" roadway performance* whereas economists have attempted to determine *economically "optimal" levels of congestion*. A review of national and regional practice among Working Group countries highlighted that the former approach – measuring physical and technical system performance – seems to be the overwhelmingly dominant approach.

Indicators that refer to time, service level or delay typically incorporate some arbitrary definition of the reference travel speed (e.g. free-flow as determined by design, legal operating speeds, or an arbitrary percentage of the free flow travel speed) that make no reference to what *users* may consider an economically optimal speed. Of course these indicators can be used as inputs to generalised cost calculations to derive economically optimal traffic levels. The use of such economically optimal traffic levels was surveyed as part of this study but most respondents confirmed that physical indicators and link flow maximisation were the main features of congestion measurement used in their experience. Furthermore, it seems that relatively few jurisdictions seem to track or otherwise monitor the *variability* of traffic performance via reliability indicators.

The manner in which these indicators are actually derived can be broken down into three broad approaches; those derived from point-related measurements (vehicle count, flow), temporal/speed indicators extrapolated or derived from the former (link travel time and delay) or spatial indicators (density, queue length, congested lane kilometres, etc). There is some evidence (see box), that point-related measurements of travel time (delay, speed, travel time and Level of Service) dominate the measurement of congestion. There also seems to be mixed views on the accuracy of these indicators, alone, to deliver an accurate understanding of congestion on the roadway network. The following table inventories a broad set of congestion indicators:

Table: Congestion Indicators: Inventory

Indicator	Description	Notes
1. Speed Based Indicator		
Average Traffic Speed	Average speed of vehicle trips for network	Does not adequately capture congestion effects
Peak Hour traffic speed	Average speeds of vehicle trips during peak hours	Can serve as a benchmark for reliability measures based on actual average or median speeds
2. Temporal/Delay-based indicators		
Annual Hours Of Delay	Hours of extra travel time due to congestion	All delay-based indicators depend on a baseline value for calculating the start of "delayed" travel – when this baseline is free-flow speed, the term "delay" becomes misleading since it is not at all clear that travellers on the network would ever be able to achieve delay-free speeds at peak hours.
Annual Delay Per Capita	Hours of extra travel time divided by area population	
Annual Delay Per Road User	Annual Delay Per Road User	
Average Commute Travel Time	Average commute trip time	
Estimated Travel Time	Estimated travel time on a roadway link (used in conjunction with variable message signs)	
Congested Time	Estimate of how long congested "rush hour" conditions exist	
Delay per road kilometre	Difference between reference travel time and congested travel time per network kilometre	
Travel Time In Congestion Index	Percentage of peak-period vehicle or person travel that occurs under congested conditions	The use of the travel time index and the travel time rate also depend on the identification of a baseline value for signalling the start of congested conditions – when this value is based on free flow speeds, the same reservation as noted for other "delay"-type indicators holds
Travel Time Index	The ratio of peak period to free-flow travel times, considering both reoccurring and incident delays (e.g., traffic crashes).	
Travel time Rate	The ratio of peak period to free-flow travel times, considering only reoccurring delays (normal congestion delays).	
3. Spatial Indicators		
Congested Lane Miles/kms	The number of peak-period lane miles/kms that have congested travel	Spatial indicators also depend on threshold values. These may be based on the median/average speeds typically achieved or on free-flow speeds (see note above).
Congested Road Miles/kms	Portion of roadway miles/kms that are congested during peak periods	
Network Connectivity Index	An index that accounts for the number of nodes and interchanges within a roadway network	This is an indicator of the potential for congestion to arise, whether or not this potential is realised depends on a number of other factors

4. Service level/capacity indicator		
Roadway Level Of Service (LOS)	Intensity of congestion delays on a particular roadway or at an intersection, rated from A (uncongested) to F (extremely congested).	These indicators have had the favour of roadway managers. They typically reference the design capacity of a roadway and are typically implicitly used to maximise throughput up to the design capacity of the roadway link in question.
Roadway Saturation Index	Ration of observed flow to design capacity of roadway	
5. Reliability Indicators		
Buffer time Index	See planning time index below	These indicators try to capture how road users typically make trip decisions on congested networks – they explicitly take into account the importance to many users of making trips "on time" rather than simply making trips at a high rate of speed.
Congestion Variability Index	An index relating the variability of travel speeds on the network	
Planning time index	An index that accounts for a time buffer that allows an on-time arrival for 95% of trips on a network	
Mean vs. variance travel times	Measure of the standard deviation of travel times on a link or on the network for a given period	
Distribution of travel times: Percentile - mean	Measure of the difference between the 80th or 90th percentile of the travel time distribution and the median or 50th percentile	
6. Economic cost/efficiency indicators		
Annual Congestion Costs	Hours of extra travel time (generated by travel below reference speed) multiplied by a travel time value, plus the value of additional fuel consumption. This is a monetised congestion cost.	As noted above, the selection of free-flow speeds when trying to account for "congestion costs" is highly problematic.
Current marginal external congestion costs	The additional external costs (not borne by users) of every additional vehicle/use entering the network	
Total deadweight loss	The sum total of the overall losses (costs benefits) incurred for a given level of use/traffic	
Average deadweight loss per vehicle/km	The dead weight loss divided by the number of vehicles/km giving rise to that loss.	
7. Other indicators		
Congestion Burden	The exposure of a population to congested road conditions (accounts for availability and use of alternatives)	
Excess Fuel Consumption	Total additional fuel consumption due to congestion.	Again, determining the point of reference for "additional" fuel consumption can be problematic if based on free-flow speeds
Excess Fuel Consumption Per Capita	Additional fuel consumption divided by area population	

Source: VTPI(2005) and COMPETE, (2006).

Among the multitude of available indicators, one can discern *three broad families of primary indicators and performance measurements* that could usefully transmit a more accurate picture of congestion and its burden. These primary indicators of congestion could be used to track both system performance as well as to derive the economic impacts of congestion. These indicators relate to system performance in relation to:

1. *Travel time* (and thus the average speeds experienced on the roadway at peak hours).
2. Travel quality (and primarily to trip reliability and predictability).
3. The exposure of urban peak-hour travellers to roadway congestion (e.g. roadway users travelling on congested roads vs. all urban travellers in peak hours*)*.

Travel Time Indicators

Of this indicator "families", the first is most developed and most widespread. These measures of system performance can either only reference average speeds or can go one step further and try to relate these average speeds to some benchmark figure – typically free-flow speeds. Free flow speeds are those speeds which drivers self-select on what are commonly considered "empty" roads – e.g. roads with so few vehicles on them that vehicles do not impede each others progress.

These speeds tend to gravitate around the maximum legal speeds posted for each road type although on urban roads these speeds might be less due to stops at intersections. The use of free-flow speeds as a benchmark is understandable since it can be seen as a replicable and readily useable "objective" figure. Problems arise, however, when the difference between free flow speeds and experienced speeds are labelled "delay". Delay is the technically correct term for the difference but is often semantically misconstrued as referring to an attainable target for peak hour travel – e.g. "zero-delay". It should be noted that, most dynamic cities cannot afford to deliver free-flow speeds at peak hours, nor would they want to live with a road network that could deliver these speeds at peak hours. Discourse based on delay as measured in reference to free-flow travel times can thus be biased towards an unattainable and likely undesirable congestion management goal (e.g. zero delay). In this case, the use of an outcome neutral indicators or indices is preferable.

Another way of side-stepping the issue of involuntarily biasing congestion management policies towards the delivery of free-flow travel speeds is to select speeds other than free-flow to serve as the benchmark value. These may be the legally posted speed or some manifestation of "normal" or "expected" travel speeds on the particular type of road. Such selected benchmarks can more realistically convey the deviation from expected travel speeds but make it difficult to compare different regions. Canada provides one example of how free-flow indicators can be nuanced for policy purposes.

The average travel speed for a given road link and for a given representative period may seem to be a natural candidate for such a benchmark – but average speeds can hide the impact of extraordinary non-recurring congestion-causing events. Therefore, it may make more sense to select the *median* speed as opposed to the *mean average* speed as a reasonable proxy for use as an "expected" or "normal" travel time performance benchmark.

Travel Quality / Travel Time Variability Indicators

By travel quality, we mean principally those elements that contribute to *smooth and predictable* travel conditions. Travel-time *variability* and its converse, travel time *predictability*, are at the heart of this family of indicators. These metrics are important in order to provide system managers, roadway users and policy-makers with a realistic assessment of how well traffic and congestion management policies are delivering consistent and therefore "plan-able" travel times. Success in delivering such dependable travel conditions is important since these can greatly contribute to reduced traveller stress – even in light of relatively slow *average* travel speeds.

While travel times can vary according to departure time for different vehicles travelling at roughly the same time period on the same road (*vehicle-to-vehicle variability*), most travel-time variability indicators seek, rather, to capture the change in travel times for vehicles travelling during the same time periods on the same roads *but on different days*. While habitual roadway users are likely to make a rough heuristic determination as to how much time they should account for in order to make a reasonable percentage of their trips on time based on their past experiences, this is not true for the large minority of those who have not sufficient experience to make such judgements. Research into roadway traffic composition has underscored that up to 20-40% of roadway users during peak hours are not habitual travellers for any given road.

Congestion indicators that effectively relate the variability of travel times can allow non-habitual users to make realistic assessments of their travel time requirements (and the "time buffers" necessary to allow for in order to have a good chance of arriving on time) as well as provide more realistic assessments of habitual users' travel and buffer time requirements.

Travel time variability can refer to the difference in travel time between *different* vehicles undertaking the *same trip with the same departure and arrival times*, the difference between the same trip undertaken at *different departure and arrival times* and/or the difference between the same trips undertaken at the same time on *different days*.
In this context, the measure of variability is related to the frequency distribution, and the standard deviation, of the same trips started at the same time, but on different days – that is the day-to-day variation in travel times. The distribution provides insight into what is hidden by average speed data – namely, if the average is composed of more uniform and predictable trips or by highly diverse and unpredictable trips.

Indicators of Exposure to Roadway Congestion

The final leg of congestion indicator families relates to how *roadway* congestion impacts *total* transport system performance. This is an underdeveloped indicator "family" and would ideally seek to provide a relative measure of how many urban travellers are affected by congestion. As noted earlier, roadway congestion is a still a temporal phenomenon and thus, ideally, policy-makers might seek to understand what percentage of travel takes place in congested conditions. For commuting travel, it seems obvious that the bulk of road travel will take place during the morning and evening peak periods. Less obvious, however, is the importance of peak period travel for other travel purposes. For instance, the United States Federal Highway Administration reports that "most motor carriers work aggressively to schedule and route their truck moves outside of peak periods and around known bottlenecks. Truck volumes typically peak during the midday, especially on urban Interstate highways, and are relatively high in the early morning and at night compared to automobile volumes". Thus, it might be surmised that truck travel is relatively less exposed than commuter traffic to congested travel conditions.

Another form of exposure to congestion relates to the number of travellers caught on congested roads versus the total number of travellers during daily peak periods. This type of indicator would seem to be potentially useful as it could guide policy interventions seeking to improve total transport system performance. Obviously, the policy importance of roadway congestion in a city where 98% of peak hour travel takes place upon the roads is different from that of a city where only 60% of peak hour travel takes place upon the road. For such an indicator to be helpful it must also seek to capture the relative quality of road vs. public transport. This necessarily would have to seek to compare travel times and travel predictability among the different modes – along with some measure of road vs. public transport accessibility to desired destinations. Much travel by public transport (non-separated tramway and bus travel) employs the same congested roads as cars and is therefore exposed to the same congestion as the latter.

While elements of such a composite indicator exist within various road and public transport administrations, operational holistic indicators of traffic congestion exposure across modes are still to be developed. This is an area where further innovation and research is required.

Commonly used performance measure(s) that reflects congestion levels on roads

In this section, a set of commonly used performance measures that reflect congestion levels on roads are briefly explained and their drawbacks are distinguished.

Roadway congestion index

This index allows for comparison across metropolitan areas by measuring the full range of system performance by focusing on the physical capacity of the roadway in terms of vehicles. The index measures congestion by focusing on daily vehicle miles traveled on both freeway and arterial roads.

Drawback(s): None

Travel rate index

This index computes the "amount of additional time that is required to make a trip because of congested conditions on the roadway." It examines how fast a trip can occur during the peak period by focusing on time rather than speed. It uses both freeway and arterial road travel rates.

Drawback(s): Measure can be difficult for public to understand

Travel time index

This index compares peak period travel and free flow travel while accounting for both recurring and incident conditions. It determines how long it takes to travel during a peak hour and uses both freeway and arterial travel rates.

Drawback(s): Requires separation of recurring and incident delay. Measure can be difficult for public to understand

Travel delay
Travel delay is the extra amount of time spent traveling because of congested conditions. The TTI study divided travel delay into two categories: recurring and incident.
Drawback(s): Requires separation of recurring and incident delay

Travel rate
Travel rate, expressed in minutes per mile, is how quickly a vehicle travels over a certain segment of roadway. It can be used for specific segments of roadway or averaged for an entire facility. Estimates of travel rate can be compared to a target value that represents unacceptable levels of congestion.
Remark: Included in many other calculations

Delay rate
The delay rate is "the rate of time loss for vehicles operating in congested conditions on a roadway segment or during a trip." This quantity can estimate system performance and compare actual and expected performance.
Remark: Included in many other calculations

Total delay
Total delay is the sum of time lost on a segment of roadway for all vehicles. This measure can show how improvements affect a transportation system, such as the effects on the entire transportation system of major improvements on one particular corridor.
Drawback(s): None.

Relative delay rate
The relative delay rate can be used to compare mobility levels on roadways or between different modes of transportation. This measure compares system operations to a standard or target. It can also be used to compare different parts of the transportation system and reflect differences in operation between transit and roadway modes.
Drawback(s): Measure may be difficult for public to understand because result is a number with no units.

Delay ratio
The delay ratio can be used to compare mobility levels on roadways or among different modes of transportation. It identifies the significance of the mobility problem in relation to actual conditions.
Drawback(s): Measure may be difficult for public to understand because result is a number with no units.

Congested travel
This measure concerns the amount and extent of congestion on roadways. Congested travel is a measure of the amount of travel that occurs during congestion in terms of vehicle-miles.
Drawback(s): Formula requires length of congested roadway segment

Congested roadway
This measure concerns the amount and extent of congestion that occurs on roadways. It describes the degree of congestion on the roadway.
Drawback(s): Formula requires length of congested roadway segment

Accessibility

Accessibility is a measure of the time to complete travel objectives at a particular location. Travel objectives are defined as trips to employment, shopping, home, or other destinations of interest. This measure is the sum of objective fulfillment opportunities where travel time is less than or equal to acceptable travel time. This measure can be used with any mode of transportation but is most often used when assessing the quality of transit services.

Drawback(s): Requires information on trip objective. Most often used with transit services.

Speed reduction index

This measure "represents the ratio of the decline in speeds from free flow conditions." It provides a way to compare the amount of congestion on different transportation facilities by using a continuous scale to differentiate between different levels of congestion. The index can be applied to entire routes, entire urban areas, or individual freeway segments for off-peak and peak conditions.

Drawback(s): Measure may be difficult for public to understand because result is a number with no units. Result is relative to free flow speed, which is difficult for motorists to comprehend.

Congestion severity index

This index is "a measure of freeway delay per million miles of travel." This measure estimates congestion using both freeway and arterial road delay and vehicle miles traveled.

Drawback(s): None

Lane-mile duration index

This index is a measure of recurring freeway congestion. This index measures congestion by summing the product of congested lane miles or kms and congestion duration for segments of roadway.

Drawback(s): Results would be poor since not all freeway segments in area collect traffic data.

Level of service (LOS)

LOS differs by facility type and is defined by characteristics such as vehicle density and volume to capacity ratio. Congested conditions often fall into a LOS F range, where demand exceeds capacity of the roadway. Volume to capacity ratios could be compared to LOS to reach conclusions about congested conditions; however, there is no distinction between different levels of congestion once congested conditions are reached.

Drawback(s): Is difficult to distinguish between levels of congestion once congested conditions are reached.

Queues

Queues or traffic back-ups best represent the public's view of congestion. Queues can be measured using aerial photography, which can often determine performance measures such as LOS and queued volume.

Drawback(s): Difficult to estimate queues using available traffic data. Can be measured by use of aerial photography but is costly and site specific.

Travel Time Reliability

In transport planning, reliability performance is generally expressed by the probability of realizing trips within a certain travel time. As travel times depend on many factors, the travel times in a given network have some randomness arising mainly from interaction between users and available network capacity as well as variations in road capacity due tom external factors.

There are numerous indicators used to express the reliability system performance. While the reliability of system performance of public transport is often expressed by the punctuality of arrivals and/ or departures at stops and stations, reliability of system performance in private transport is measured by a wide variety of temporal indicators.

Commonly used travel time reliability indictors

In this section, a range of suggested indicators are presented, taking into account some considerations regarding the situation for which they can best fit.

The Standard deviation

In situations where there is a need to look at the variability in travel times around an average value and it is expected that this variability is not much influenced by (a limited number of) extreme delays, the travel time distribution will be not very much skewed. In these cases, statistical range indicators can be considered useful. The standard deviation of travel times can be used to describe the extent of travel time dispersion. A further consideration to use the standard deviation as a reliability indicator is due to pragmatic reasons and applicability in the cost-benefit analyses.

The 95- percentile value

To overcome the eventual problem of not giving much specific attention to possible extreme, the 95-percentile value of the distribution can be used or added to the analyses; this indicator is very appropriate to focus on the width of the travel time distribution and can be very useful to analyze the development of high travel time values. However, as long as this indicator is not combined with information on average expected travel times or delays, the indicator does not directly represent reliability.

The Buffer time

The use of so-called "buffer time" related indicators is becoming more and more common. The buffer time can be explained as the extra percentage of travel time due to travel time variability on a trip that a traveler may take into account in order to have a high probability of arriving on time. Examples of buffer time related indicators are the Buffer Index and the Planning Time Index, used in the US Federal Highway Administration's Urban Congestion Reports, aimed at monitoring traffic congestion and travel reliability on a national scale.

Buffer Index

The buffer index represents the reliability of travel rates associated with single vehicles. This measure may be beneficial to the public because it tells them how congestion will affect them as individuals.
The buffer index shows the effect of congestion on the reliability of travel rates along the roadway. The extra percentage of travel time a traveler should allow in order to be on time 95 percent of the time is represented by the buffer index as follows:

$$Buffer\ Time\ Index = \frac{TT95 - \overline{T}}{\overline{T}}$$

Where:

TT95: 95^{th} percentile of travel time
$\overline{T}$: Average travel time

Planning Time Index
The Planning Time Index represents the extra time most travelers should add to a free flow travel time so as to be fairly confident of arriving at the destination by a certain time. The measure differs from the Buffer Index in that it includes recurring delay as well as nonrecurring delay. For example, a planning time index of 1.60 means that travelers plan for an additional 60% travel time above the free-flow travel time to ensure on time arrival most (95%) of the time.

$$Planning\ Time\ Index = \frac{TT95}{TTfreeflow}$$

To summarize, the consultant believes buffer time related indicators such as the Buffer Time Index and Planning Time Index are appropriate monitors to describe and communicate travel time reliability to planners as well as network users. Other more simple measures such as travel time percentiles, median travel times and the standard deviation of travel time may also serve as appropriate indicators, but they should be used with caution, as relevant characteristics of the travel time distributions could be easily overlooked. For instance, using the standard deviation of travel time as a utility component in route choice may results in biased outcomes.

Selection of Performance Measures for GCMA

To define a measure or measures of performance that reflect congestion levels along specific corridors, two performance measures are selected that focus on :

- travel time delay
- travel time reliability

Basically, the data needed to support these measures are available in Smart Traffic Centers (Texas Transportation Institute), therefore allowing most agencies to perform the necessary calculations to validate the measures' accuracy.

Total delay measure

Total delay was chosen as a performance measure because it relates to delay and the data needed to calculate this measure are readily available. The travel times can be derived by using speed and the length of a route. This measure will help transportation professionals determine the delay for all vehicles traveling over a segment of roadway during a specific time period and thus to assess the severity of the congestion. Total delay could also allow transportation professionals to estimate how improvements within a transportation system affect a particular corridor or the entire system.

Total delay may be useful to traffic managers because it represents delay for all vehicles. Time lost for all vehicles is more important for roads that have higher volumes because higher volumes mean that more travelers are affected by the time lost, which can mean more community money is wasted. A comparison of delay among different segments of roadway is also possible when using total delay. Total delay shows the effect of congestion in terms of the amount of lost travel time. The sum of time lost on a segment of roadway due to congestion for all vehicles is represented by total delay as follows:

Total delay (PCU-hr) = [Actual travel time (hr) – Acceptable travel time (hr)] x Traffic Volume (PCU)

The Buffer Index as the travel time reliability measure

A recent US national Cooperative Highway research Program report concludes that the Buffer Index appears to relate particularly well to the way in which travelers make their decisions (NCHRP 2008). The Buffer Index is useful in the user's assessment of how much extra time has to be allowed for uncertainty in travel conditions. It hence answers simple questions such as "how much time do I need to allow?" "When should I leave?"

In other words, the buffer index is chosen for the project as a performance measure because it relates to the reliability of an individual vehicle trip, and also is useful to both the public and transportation professionals. The travel rates used in this calculation can be derived from average speed readings and the length of a route. This measure will help transportation professionals determine the impact of congestion on one vehicle traveling on a segment of roadway during a specific time period. The buffer index could also be useful in alerting motorists of the anticipated changes in travel time on particular segments of roadway so trips could be planned accordingly.

In addition to the Buffer Time Index, the Planning Time Index represents the total travel time that should be planned when an adequate buffer time is included. In the NCHRP report both these indicators are advised as cost effective measures to monitor travel time variation and reliability.

Identification of Congested Locations

The goal of road administrators are focused on assessments and management of the road systems in urban areas in ways that *maximised* the ability of existing infrastructure to handle current and expected future traffic demand and *minimised* traffic delays and the associated personal, business and resource impacts including personal and productive time lost, fuel wasted and air quality degradation. Identification of existing and potential future congestion locations is the first step toward transport system management.

The approach typically involves:

- Measurement of traffic speeds and flows.
- Estimates of maximum achievable speeds and flows during uninterrupted traffic flow conditions (but taking into account speed limits and intersection capacity).
- Assessments of actual speeds and flows in relation to maximum achievable speeds and flows. These are often defined in terms of percent below posted speed (or below off-peak speeds at prevailing flows), roadway volume/capacity ratio, speed-flow charts and intersection levels of service.
- Identification of congested locations throughout the network based on overall Levels of Service (LOS) or another form of categorization.