CONGESTION AND ITS EXTENT

Robert L. Bertini, Portland State University

ABSTRACT

The objective of this paper is to discuss current definitions of metropolitan traffic congestion and ways it is currently measured. In addition, the accuracy and reliability of these measures will be described along with a review of how congestion has been changing over the past several decades. First, the results of a survey among transportation professionals are summarized to assist in framing the issue. Most respondents linked the measurement of congestion to the increased travel time that occurs during peak periods. Roughly half of those surveyed felt that congestion measures are at least somewhat accurate, and about 80% of those surveyed feel that congestion has worsened over the past 20 years. The paper includes a literature review of current trends in congestion definition and measurement and presents some basic theory about how traffic parameters should be measured over time and space. A brief description of possible congestion measures over corridors and entire door-to-door trips is provided. Additional analysis of recent congestion measures for entire metropolitan area is provided, using Portland, Oregon and Minneapolis, Minnesota as case examples. Some discussion of the stability of daily travel budgets and alternative viewpoints about congestion are provided along with some conclusions and perspectives for future research.

INTRODUCTION

"You're not stuck in a traffic jam, you are the jam." - German public transport campaign (Kay, 1997)

Congestion—both in perception and in reality—impacts the movement of people and freight in our society and is deeply tied to our history of high levels of accessibility and mobility. Along spatial and temporal dimensions, some say that traffic congestion has been around since ancient Rome (Downs, 2004), that it wastes time and energy, causes pollution and
stress, decreases productivity and imposes costs on society equal to 2-3% of our GDP (Cervero, 1998). From a technological perspective, it has been noted that an automobile is “a conveyance which is capable of moving a mile a minute, yet the average speed of traffic in the large cities is of the order of 11 mph.” (Traffic in Towns, 1963). For 2002, it was estimated that congestion “wasted” $63.2 billion in eighty-five metropolitan areas during 2002 because of extra time lost and fuel consumed, or $829 per person. (Texas Transportation Institute, 2004) Some refer to these kinds of estimates as misleading since the prospect of eliminating all congestion during peak periods is “only a myth; congestion could never be eliminated completely.” (Downs, 2004). While some research emphasizes that “rush hour is longer than an hour in the morning and an hour in the evening and few people are ‘rushing’ anywhere,” others say that “gridlock is not going to happen because people change what they do long before it happens.” (Garrison and Ward, 2000) Some view congestion as a “problem” that individual drivers are subject to, while others emphasize that the users of transportation networks “not only experience congestion, they create it.” In fact, it has been shown that most people make travel decisions based on an expectation of experiencing a certain amount of congestion; while “few consider the costs their trips impose on others by adding to congestion.” (Mohring, 1999) The objective of this paper is to discuss current definitions of metropolitan traffic congestion and ways it is currently measured. In addition, the accuracy and reliability of these measures will be described along with a review of how congestion has been changing over the past several decades.

FRAMING THE ISSUE

In order to assist in framing the issues to be addressed in this paper, an unscientific survey about metropolitan area congestion was distributed by email to more than 3,500 transportation professionals and academics, with a total of 480 responses. Four qualitative questions were asked:

- How do you define congestion in metropolitan areas?
- How is congestion in metropolitan areas measured?
- How accurate or reliable are traffic congestion measurements?
- How has metropolitan traffic congestion been changing over the past two decades?

Respondents were also provided with an opportunity to comment on congestion in general.

Definition of Congestion

In attempting to define congestion, a total of 557 responses were provided since many responses included separate definitions for freeways and signalized intersections. As shown in Figure 1, survey respondents mentioned time, speed, volume, level of service (LOS) and traffic signal cycle failure (meaning that one has to wait through more than one cycle to clear the queue) as the primary definitions of congestion. Respondents who used the term “LOS” were not more specific; typical LOS measures include volume/capacity, density, delay, number of stops, among others. The majority of the responses included a “time” component—travel time, speed, cycle failure and LOS are all related to the fact that users
experience additional travel time due to congestion. It is clear from these responses, that some definitions of congestion rely on point measures (e.g., volume and time mean speed) and some rely on spatial measures (travel time, density and space mean speed).

The definitions of congestion that are related to point measures include vehicle count (expressed as flows) and time mean speed extracted from point detectors (often extrapolated over a segment to estimate link travel time). This point-based travel time estimate can be compared to a free flow travel time for a particular link, where the difference between actual and free flow travel times is the delay. Multiplying the number of vehicles to pass a point during a given time interval by the length of an associated segment will reveal the vehicle-miles traveled (VMT). Multiplying the number of vehicles passing the point during a certain time interval by the travel time for the associated link will reveal the vehicle-hours traveled (VHT). Definitions of congestion that are related to spatial measures include density, queue length and actual segment travel time (such as that recorded by a probe vehicle). Delay can

---

*Figure 1. Congestion Survey Results*
be calculated as actual travel time over a segment minus the free flow travel time for that segment.

Several survey responses included important comments. It was pointed out that “if we want to reduce congestion we need to be able to define it and quantify it.” Some were willing to define congestion as “anything below the posted speed limit,” or below some “speed threshold (e.g., <35 mph).” Others noted that “congestion is relative,” “a perception” and, “I know it when I see it.” One response pointed out that congestion was not so much of a concern anymore since “fortunately we have had an economic collapse.” Other surveys have found similar results. In the U.K. respondents defined congestion as stop-start conditions (38%); a traffic jam with complete stops of 5+ min at a time (24%); having to travel at less than the speed limit (19%); and moving very slowly at less than 10 mph (17%). (U.K. Department for Transport, 2001) A conference of European Transport Ministers concluded that there is no widely accepted definition of road congestion, but that an appropriate definition might be: “the impedance vehicles impose on each other, due to the speed-flow relationship, in conditions where the use of a transport system approaches its capacity.” (European Conference of Ministers of Transport, 1998) Finally, a survey conducted by the National Associations Working Group for ITS included a response describing the difficulty in defining congestion: “you know it when you see it—and the severity of the problem should be judged by the commonly accepted community standards.” (Orski, 2002)

**Measurement of Congestion**

There were 682 responses to the question asking how congestion is measured (multiple response). As shown in Figure 1, most responses were related to time: delay, speed, travel time and LOS, all of which include the notion that actual travel time can be a primary measure of congestion. Other measures included volume/capacity (point measure) and queue length and density (spatial measures). The small number of responses in the “Other” category included such measures as number of stops and travel time reliability. One responded summed up the issue by stating “it is never truly measured.” The literature includes a wide array of possible congestion measures including: volume/capacity (disregards duration), VMT, VMT/lane mile, hours of delay per 1000 VMT, speed, occupancy, travel time, delay, LOS and reliability. In the U.K. survey several helpful measures of congestion were identified: delay (51%); risk of delay (20%); average speed (18%); and amount of time stationary or less than 10 mph (11%).

**Accuracy and Reliability of Congestion Measurements**

Next, respondents were asked, “how accurate or reliable are traffic congestion measurements?” There were 525 responses to this question, indicating that respondents had mixed feelings about the accuracy/reliability of congestion measurements. As shown in Figure 1, about half of the responses indicated that the measurements are accurate or somewhat accurate, while the other half indicated that they are inaccurate or variable. Many
comments indicated that congestion measurements are often based on very small sample sizes, that they are relative and variable and that they should be presented with confidence intervals rather than purely deterministic values. One comment stated that congestion measurements are “reasonably accurate despite the fact that they measure the wrong things,” while another stated that “congestion perception is a personal thing based on personal experiences and anecdotes.” Finally, one comment indicated that “the result is really just a snapshot in time.”

**Changes in Congestion**

The final survey question aimed to assess the extent to which urban traffic congestion has changed over the past two decades; most respondents indicated that congestion has worsened. Some respondents indicated that more transportation options (such as transit and intelligent transportation systems) are now available and some indicated that congestion has gotten better. The response is not surprising since the U.S. population increased by 24% to 282 million from 1980-2000, the number of highway vehicles has increased 39% to 225 million and the passenger car VMT has increased 44% during the same period. Respondents commented that the impact of congestion has increased in both spatial and temporal dimensions, including the spread of peak periods. Others pointed out that change has been relative, depending on the area and on the user. For example, “western cities are experiencing increasing congestion, as both population and per capita miles of travel increase. Some rust belt cities, on the other hand, are experiencing decreasing population and thus are experiencing decreases in congestion.” Some respondents indicated that drivers have been conditioned to tolerate more congestion, using an analogy about how to boil a frog: “… start out by putting him cold water, gradually raise the temperature, and he won’t figure it out (and thus escape) so he stays ’til he's cooked! We're all gradually getting cooked!”

Several responses pointed out that the congestion will always be a by-product of a healthy, vibrant urban area. In this context, it was pointed out that traditional traffic/transportation engineering antidotes to congestion have been reactive in nature, and that roadways are not improved until there is a problem. Another comment stated that “current conceptions of congestion have more to do with preserving the world as it was, rather than preparing us for the world as it will be,” and another stated that there is “too much focus on congestion, there should be more attention to accessibility. What can people get to in a reasonable period of time (20-30 minutes)?” Several respondents indicated that there has been a transformation from the mentality that you can “build your way out of congestion” to the point where other options such as HOV lanes and reversible lanes” are available.

**Other Survey Comments**

Survey respondents had the opportunity to offer comments and suggestions. These included sometime conflicting suggestions to consider alternative transportation strategies such as: congestion pricing, more transit, demand management, regional/statewide approaches, more
funding, higher gas taxes, more frequent systematic assessment of traffic operations, driver and citizen information and education, probe vehicles, multi-jurisdictional traffic signal operations, incident reduction, better land use policies, consideration of user vs. system goals, consideration of impacts on health and air quality, focus on moving people during the peak hour, more operations improvements, ITS improvements, more understandable measures, consideration of how trucks affect capacity and simply building more lanes. On the other hand, other respondents suggested: consider equity impacts of congestion pricing, don’t build more lanes, don’t waste money on ITS, don’t waste money on transit and “don’t spend too much time on it.”

LITERATURE REVIEW

The Federal Highway Administration (FHWA) defines traffic congestion as: “the level at which transportation system performance is no longer acceptable due to traffic interference.” Because there is a relative sense to the word “congestion,” the FHWA continues their definition by stating that “the level of system performance may vary by type of transportation facility, geographic location (metropolitan area or sub-area, rural area), and/or time of day,” in addition to other variations by event or season. (Lomax, Turner and Shunk, 1997) The definition of congestion is imprecise and is made more difficult since people have different perceptions and expectations of how the system should perform based on whether they are in rural or urban areas, in peak/off peak, and as a result of the history of an area.

Congestion can vary since demand (day of week, time of day, season, recreational, special events, evacuations, special events) and capacity (incidents, work zones, weather) are changing. Most researchers agree that recurrent congestion (due to demand exceeding capacity (40%) and poor signal timing (5%)) makes up about half of the total delay experienced by motorists, while nonrecurrent congestion (due to work zones (10%), incidents (30%) and weather (15%)) makes up the other half. It has been shown that four components interact in a congested system (Lomax, Turner and Shunk, 1997):

- Duration: amount of time congestion affects the travel system.
- Extent: number of people or vehicles affected by congestion, and geographic distribution of congestion.
- Intensity: severity of congestion.
- Reliability: variation of the other three elements.

Because of the issue of user expectation, one proposal is to define “unacceptable congestion” as the travel time in excess of an agreed-upon norm, which might vary by type of transportation facility, travel mode, geographic location, and time of day. (Lomax, Turner and Shunk, 1997) “A key aspect of a congestion management strategy is identifying the level of ‘acceptable’ congestion and developing plans and programs to achieve that target.” (Lomax et al., 2001)
The FHWA has initiated a Mobility Monitoring Program based on measured travel time in which they are trying to answer a mobility question: “how easy is it to move around?” and a reliability question: “how much does the ease of movement vary?” The primary measures include (Turner, et al., 2002; Lomax, Turner and Margiotta, 2001; Jung, et al., 2004):

- **Travel time index**: ratio of travel conditions in the peak period to free-flow conditions, indicating how much longer a trip will take during a peak time (a travel time index of 1.3 indicates that the trip will take 30 percent longer). The calculation of this index assumes a capacity of 13,000 daily VMT/lane-mile on freeways and 5,000 daily VMT/lane-mile on principal arterials and compares measured VMT to these assumed capacity values.

- **Average duration of congested travel per day** (hours): “How long does the peak period last?” Trips are considered across the roadway network at five-minute intervals throughout the day. At any time interval, a trip is considered congested if its duration exceeds 130% of the free-flow duration. When more than 20% of all trips in a network are congested in any five-minute time interval, the entire network is considered congested for that interval. The total number of hours in which the network is designated as congested is reported in this measure.

- **Buffer index**: this measure expresses the amount of extra time needed to be on-time 95 percent of the time (late one day per month). Travelers could multiply their average trip time by the buffer index, and then add that buffer time to their trip to ensure they will be on-time for 95 percent of all trips. An advantage of expressing the reliability (or lack thereof) in this way is that a percent value is distance and time neutral.

A recent synthesis examined more than 70 possible performance measures for monitoring highway segments and systems (NCHRP, 2003). From users’ perspectives, key measures for reporting the quantity of travel included: person-miles traveled, truck-miles traveled, VMT, persons moved, trucks moved and vehicles moved. In terms of the quality of travel, key measures included: average speed weighted by person-miles traveled, average door-to-door travel time, travel time predictability, travel time reliability (percent of trips that arrive in acceptable time), average delay (total, recurring and incident-based) and LOS.

A review of the literature reveals that transportation agencies have adopted particular definitions of congestion for their purposes. INCOG, the regional council of governments in Tulsa, Oklahoma defines congestion as “travel time or delay in excess of that normally incurred under light or free-flow travel conditions.” (INCOG, 2001) INCOG applies their policy of identifying recurring congestion and documenting its magnitude. To do so, traffic counts are compared to capacity and then the ratio of volume/capacity is expressed as a level of service. Tulsa uses traffic counts (and traffic volume forecasts) as an initial screen to locate congested routes and future problems. In Rhode Island, the state DOT recognizes that “congestion can mean a lot of different things to different people.” As a result, the state attempts to use objective congestion performance measures such as percent travel under posted speed and volume/capacity ratios. In Cape Cod, Massachusetts, a traffic congestion indicator is used to track average annual daily bridge crossings over the Sagamore and Bourne...
bridges. (Cape Cod Center for Sustainability, 2003) This very simple measure was chosen for this island community since it is appropriate, easy to measure, and since historic data are available to monitor long-term trends. In the State of Oregon, the 1991 Transportation Planning Rule (TPR) uses VMT as a primary metric, with a goal of reducing VMT by 20% per capita in metropolitan areas by 2025.

In Minnesota, freeway congestion is defined as traffic flowing below 45 mph for any length of time in any direction, between 6:00 a.m. and 9:00 a.m. or 2:00 p.m. and 7:00 p.m. on weekdays. Michigan defines freeway congestion in terms of LOS F, when the volume/capacity ratio is greater than or equal to one. Since the function of the transportation system is to provide transport of people and goods, and its benefits are a function of the number of trips served, in California “congestion” is defined as the state when traffic flow and the number of trips are reduced. The California Department of Transportation (Caltrans) defines congestion as occurring on a freeway when the average speed drops below 35 mph for 15 minutes or more on a typical weekday. There is currently a proposal to change the definition of congestion to be measured as the time spent driving below 60 mph, based on analysis of 3363 loop detectors at 1324 locations as part of the California Performance Measurement System (PeMS) database (Varaiya, 2002). The State of Washington DOT aims to provide congestion information (in plain English) that uses real time measurements, reports on recurrent congestion (due to inadequate capacity) separately from nonrecurrent congestion (due to incidents). This includes the measurement of volumes, speeds, congestion frequency, and geographical extent of congestion, travel time and reliability. The Washington DOT also focuses on travel time reliability and predictability by presenting a “worst case” travel time for a set of corridors such that commuters can expect to be on time for work 19 out of 20 working days a month (95 percent of trips), if they allow for the calculated travel time. (Washington State Department of Transportation, 2004)

**SOME BASIC THEORY**

Many common traffic measurements are derived from the basic traffic flow parameters—flow, density and speed. This section describes how these fundamental measures can be applied at the level of the roadway segment, a corridor and over an entire door-to-door trip.

**Segment Level**

Figure 2 illustrates some basic points about traffic flow. In Figure 2(a), a set of vehicle trajectories on a time-space plane is shown in the context of a roadside observer (or detector) at location x. (Daganzo, 1997) During time interval t, an observer would count 7 vehicles passing point x. Flow, a point measure, is defined as the number of vehicles that pass a point during a particular time interval; in this case 7/t, usually expressed in vehicles/hour. Under certain circumstances the “capacity” of the highway at point x might be estimated, and the actual measured volume could be compared to that theoretical capacity value in the form of a volume/capacity ratio. If a roadway section is operating with a volume/capacity ratio >1, it is
usually said that the facility is providing poor LOS to travelers. Speed could also be measured at point \( x \), for example by a radar gun. If the arithmetic average of the speeds measured at a point is taken over a measurement interval \( t \), this is called the time mean speed.

\[
\begin{align*}
v_f &= \text{Free Flow Speed} \\
v_i &= \text{Actual Travel Time} \\
v_v &= \text{Delay} \\
v_e &= \text{Extrapolated Travel Time} \\
v_= &= \text{Free Flow Travel Time} \\
\end{align*}
\]

Figure 2. Segment Level Measures of Congestion

Figure 2(b), which also shows a set of vehicle trajectories on a time-space plane, illustrates that some key traffic flow parameters are measured over space. For example at time \( j \), the number of vehicles on the segment \( d \) at that instant would be counted as six vehicles. The
density at time $j$ is the number of vehicles on the section at that time divided by the section length, in this case $6/d$, usually expressed in vehicles/mile. The actual travel times of vehicles can also be recorded over space; in this case for vehicle $i$, its travel time is shown as $v_i$. The free flow travel time for segment $d$ might be assumed to be $v_f$. Therefore, for vehicle $i$ on this roadway segment the delay is defined as $v_i - v_f$.

Depending on what data collection system is available, sometimes a point measure, such as speed, can be applied over a roadway segment. As shown in Figure 2(c), if a roadway is equipped with measurement sensors, a sensor’s area of influence can be assumed to be the distance between the upstream and downstream midpoints between each detector pair. In Figure 2(c) this would be equal to $0.5s_1 + 0.5s_2$. For federal reporting purposes, as part of the Highway Performance Monitoring System (HPMS) limited point-level count and speed measurements are taken on a sampling of urban roadway locations for one 48-hour period every three years and extrapolated over the entire roadway network. Figure 2(a) also shows how a point speed measured at location $x$ can be extrapolated to determine segment travel time, $v_x$ for vehicle $i$. This can be used to estimate the delay for vehicle $i$, $v_i - v_f$.

Figure 2(d) illustrates the basic relation between fundamental traffic flow variables on a density-flow plane. (Coifman and Mallika, 2004) The relation is approximated as a triangle, where zero flow occurs when there are no vehicles on the facility—density and flow are both zero. Zero flow also occurs when density increases to a level such that all vehicles must stop—the speed and flow are zero. (FHWA, 2003) Figure 2(d) illustrates four distinct traffic states: 1, 2, 3 and C. Traffic states 1, 3 and C fall on the unqueued branch of the flow-density relation, while state 2 falls on the queued branch. Figure 2(e) illustrates several issues related to the measurement of traffic parameters along a highway segment where an incident of some kind occurred at a bottleneck location $b$ at time $t_1$. (Coifman and Mallika, 2004) Prior to time $t_1$, traffic along this hypothetical road segment was in state 1. At $t_1$ traffic transitioned abruptly to state 2, with lower flow, lower speed and higher density. A shock passed backwards. The traffic state just downstream would be characterized by state 3. At some time $t_2$, the bottleneck was deactivated, so a backward moving recovery wave passed upstream until it intersected with the initial shock at time $t_3$. During the recovery, traffic flowed at state C until it returned to state 1. It is very important to understand how, when and where bottlenecks occur on a highway. For example, in Figure 2(e) if a detector were located at location $a$, it never would have “seen” queued traffic and furthermore if a vehicle entering the road section at time $t_f$ had received traffic speed information recorded at locations $a$, $b$ or $c$ prior to $t_f$ that motorist would have had no way of predicting the actual delay that was later experienced. Also, if a detector were located at point $c$, traffic speeds would remain high throughout the entire time period shown, falsely characterizing the segment conditions as unqueued. Thus, depending on where traffic conditions are monitored, it is possible to misreport actual conditions; this would adversely affect congestion measures for a segment.
Corridor Level

It is possible to compute congestion-related measures over a larger freeway corridor where more detection locations are available. Travel time can be calculated from real-time or archived freeway sensor data. For example, Figure 3(a) shows travel time versus time for one day on northbound Interstate 5 in Portland, Oregon. This was performed over this 22-mile corridor using data from inductive loop detectors at 25 locations that recorded count, occupancy and speed at 20-sec intervals. The figure also illustrates the cumulative travel time and free-flow travel time (dashed line) throughout the day. As the cumulative line deviates from the cumulative free-flow travel time the travel time increases can be clearly observed. At
Access to Destinations

7:05 the travel time increased from 23 min. to 28 min. Similarly, at 19:42 the travel time decreased from 49 min. to 24 min. The free-flow travel time on this day was approximately 24 min.

One of the costs of congestion is delay, defined as the excess time required to traverse a section of roadway compared to the free flow travel time. As shown in Figure 3(b), the average delay was calculated for northbound Interstate 5 over five weekdays. Delay was estimated based on the difference between actual travel time and the free-flow travel time on the freeway segments. Total delay for each detector station, defined as the sum of all delay at that station throughout the day, is shown on a three-dimensional plot in Figure 3(c) for the southbound direction. For locations that indicate higher delays, as an example, a DOT can focus its incident response efforts to reduce further delays. From this plot one can see several spikes of delay that occurred at key bottlenecks along the corridor.

Figure 3(d) shows a speed plot for northbound Interstate 5 on one day, where the color variation represents the average speeds measured at 20-second intervals at six detector stations. In addition, 20 express bus trajectories recorded from an automatic vehicle location (AVL) system have been superimposed over the speed plot, indicating that the loop detectors can provide a good indication of mean travel time for a corridor. The slopes of the trajectories changed at nearly the same locations where the freeway speed declined (darker red color). This method was used to show how accurately the speed is reported by the loop detectors. Statistical analysis was used to validate that there was no evidence of difference between the means at the 95% level of confidence.

### Table 1. Percentage of U.S. Lane Miles and Vehicle Miles Traveled by Facility Type, 2000

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Total Lane Miles</th>
<th>% Lane Miles</th>
<th>Total VMT (Million)</th>
<th>% VMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freeway</td>
<td>250,873</td>
<td>3.1</td>
<td>845,711</td>
<td>30.7</td>
</tr>
<tr>
<td>Arterial</td>
<td>952,218</td>
<td>11.6</td>
<td>1,150,012</td>
<td>41.6</td>
</tr>
<tr>
<td>Other</td>
<td>7,036,534</td>
<td>85.4</td>
<td>771,645</td>
<td>27.8</td>
</tr>
<tr>
<td>Total</td>
<td>8,239,625</td>
<td>100.1</td>
<td>2,767,368</td>
<td>100.1</td>
</tr>
</tbody>
</table>

### Consideration of Total Trip

As shown in Table 1, freeways comprise about 3% of the lane miles in the U.S., but they carry more than 30% of the traffic. (FHWA, 2002) Most congestion measures are relevant for a particular link, with a focus on freeways since that is where the most traffic is located and where sensors are in place. Some researchers point out that what is relevant for the traveler is the entire door-to-door trip (Taylor, 2002). For example, Figure 4 is based on Taylor (2002) and illustrates a hypothetical vehicle trajectory (solid line) on a time-space plane. For this trip, the traveler walks to her car, travels on a local street, collector and arterial, followed by a freeway segment, an arterial, a parking lot, and finally walks from her car to her workplace. This trip took 36.1 min and traversed 10.6 miles. As shown on the x- and y-axes of Figure 4,
the congested freeway component of the trip (at 25 mph) accounted for 57% of the distance and 40% of the total travel time. On this trip 60% of the travel time occurred off of the freeway. If we focus on the freeway segment and imagine a solution that would return freeway speeds to free-flow conditions (60 mph), the trip time would be reduced to 27.7 min, as shown by the dashed line. As shown on the x-axis, this would reduce the freeway segment’s share to 22% of the travel time; now 78% of the trip time would occur off the freeway.

![Diagram showing door-to-door trip times](image)

**Figure 4. Door-to-Door Trip Times**

**METROPOLITAN LEVEL MOBILITY MEASURES**

When thinking about ways to measure congestion at the metropolitan scale, it is important to remember that our current perceptions are strongly influenced by what happened during the 1960s and 1970s in the U.S. This period (within the memory of many of today’s drivers) was one of relatively low congestion since the Interstate system construction era provided much greater expansion in travel capacity than the growth in travel during the same period. (Lomax, Turner and Shunk, 1997) The result was that in many large urban areas traffic congestion actually decreased. This recent experience frames the debate in that some would like to try to return mobility levels to those earlier conditions.
The Texas Transportation Institute prepares an annual national study comparing congestion and other performance metrics from 85 metropolitan areas. (Texas Transportation Institute, 2004) Using data from the 2002 version of this report, a recent study was conducted to begin tracking transportation performance in Portland, Oregon (Portland State University, 2004; Gregor, 2004) for the past 20 years. For example, Figure 5(a) shows trends in the proportional change in VMT, population and size (sq. mi.) of the Portland-Vancouver urbanized area, along with total annual travel time in peak periods since 1982. With growth in population, land area and the Oregon economy, VMT has increased. But as the urban area did not see increases in the ratio of size/population, travel time remained nearly constant. Figure 5(b) shows that daily VMT on Portland area freeways more than doubled between
Congestion and Its Extent

1982 and 2002, and has also doubled on arterials. Lane miles on arterials have been added at a rate greater than the increase in VMT. However, lane miles on freeways have increased by only 25 percent over the past 20 years. The gap between VMT and lane miles on freeways may explain the declining speeds on Portland freeways. Figures 5(c) and 5(d) show similar data for the Minneapolis-St. Paul, Minnesota urbanized area.

![Figure 6.](image)

As part of the Portland performance measures analysis, the Portland-Vancouver urbanized area was compared to 26 other urban areas with populations between 1-3 million. (Portland State University, 2004) As shown in Figure 6(a), when graphically comparing Portland to...
other urban areas from the Urban Mobility Report, the colored lines are for six western peer cities: Phoenix, Sacramento, San Diego, San Jose and Seattle, plus Portland. The grey lines are for the remaining cities the 1-3 million population category, and the dashed black line represents the average value measured across all 27 Large cities. Figure 6(a) shows that annual congestion delay for peak period travelers in Portland increased from 7 hours per year in 1982 to 46 hours per year in 2002, and has been close to the mean value for similarly sized cities. It had been below the average before 1992, and exceeded the average after that. Shorter-than-average travel distance coupled with lower-than-average travel speed has leveled off the delay actually experienced by travelers. Figure 6(b) shows that Portland annual travel time per peak period traveler has remained below average. Despite increases in delay, travel time has not changed noticeably in the Portland-Vancouver urbanized area. Again, shorter-than-average travel distance has eased the impact of congestion on travel time.

![Travel Time Index, 1982-2002](a)

![Travel Time Index, 1982-2002](b)

![Travel Time and Population 2002](c)

![Travel Time and Travel Time Index 2002](d)

*Figure 7. Comparing Urban Area Travel Time Index, Population and Travel Times, 1982-2002*
Figure 6(c) shows 20-year trends on delay with a focus on Minneapolis and 8 other cities with populations between 2-3 million: Atlanta, Denver, Phoenix, Seattle, Tampa, Baltimore, San Diego and St. Louis. The average trend line is shown for all 27 cities with populations between 1-3 million. As shown, Minneapolis began somewhat below average in the delay measure but now falls above average. The trend in mean travel time is similar, yet the more recent data illustrate that the travel time for Minneapolis is about average among all cities in the 1-3 million population range.

Using comparisons as in Figure 6, Figures 7(a) and 7(b) illustrate the Travel Time Index (TTI) trends between 1982-2002 for Portland and Minneapolis, respectively. The TTI is an estimate of how much longer it takes on average to travel on the major road system during peak times vs. off-peak times considering the effects of everyday recurring congestion and the effects of congestion due to incidents. The TTI is the ratio of travel time in the peak period to the travel time at free-flow conditions. A value of 1.35 indicates a 20-minute free-flow trip takes 27 minutes in the peak. These figures show that the TTI for both cities is comparable with other peer cities in this category. Figure 7(c) shows a scatter plot of population vs. peak period travel for the 27 cities with populations between 1-3 million. Portland’s population is 13th out of the 27 large cities (25th out of all 85 cities), and the amount of travel per peak period traveler is 19th out of the 27 large cities. The population of Minneapolis is 5th out of the 27 large cities, yet the annual hours of delay per peak period traveler is only the 16th highest. Figure 7(d) shows that the annual amount of travel per peak period traveler in Portland is among the 9 lowest when compared to other large cities, while the Travel Time Index for Portland is among the top 6 out of the 27 large cities. In the case of Minneapolis, the Travel Time Index is 11th among the 27 large cities.

GOING BEYOND CONGESTION MEASURES

The results just described are aggregate in the sense that they are derived from very general HPMS measurements that were designed for one purpose and used for another. It is thus important to consider going beyond these measurements when attempting to grasp the issues related to congestion.

Travel Time Budget

It has been mentioned above that the journey to work time has remained stable in recent years despite large increases in VMT. Some authors have stated that “average journey to work time changed little as suburban road and highway expansion has accommodated growing number of trips.” (Garrison and Ward, 2000) It is said that typical commutes are actually becoming shorter. (Garreau, 1991) When looking back at development patterns of even ancient cities, the average travel time to work locations has been relatively stable for a few thousand years, and that people have maintained a total daily travel time budget of roughly one hour over the past 5,000 years. (Lomax, et al., 2001; Ausubel and Marchetti, 2001; and Shafer and Victor, 1997) This point has been linked to an assessment of the size of cities as transportation
technology has evolved. (Crawford, 2000) For example, if we consider that the maximum accepted (average) commute time is about 45 minutes, it has been reported that the city of Istanbul was approximately six miles in diameter in the sixteenth century. In that case and if people walked at about four miles per hour it would have taken 45 minutes to walk from the city’s periphery to the city center. (Garreau, 1991)

Several studies have attempted to document these phenomena more specifically. In one study it was shown that the average weekly commuting time for males and females in the U.S.S.R. between 1910-1990 ranged between 4.2 and 5.8 hours week, which would have translated to 50-70 minutes per day with a five day work week. (Grübler, 1990) Another study analyzed aggregate survey data from 1958 and 1970 in Washington, D.C. and Minneapolis-St. Paul, and found that in both cities and at both times, travelers averaged approximately 1.1 hours of travel per day. (Ryan and Zahavi, 1980)

Finally, a comprehensive analysis of daily travel between 1965-1995 in various U.S. and international cities, it has been noted that it appears that people travel about one hour per day. As technology has evolved, the distance traveled per day in the U.S. has increased approximately 2.7% per year from 4 km (walking) in 1880 to approximately 80 km per day in 1990. (Ausubel, Marchetti and Meyer, 1998) As an extract from this work, Figure 8 shows representative data for studies of the U.S. and a dozen other countries since 1965. Despite one obvious outlier from a California study, the figure does indicate that much of the available data centers on the 60 minute range.
Other Viewpoints

A number of authors have been presenting slightly different views about traffic congestion. Some have noted that successful cities are places where economic transactions are promoted and social interactions occur, and that traffic congestion occurs where “lots of people pursue these ends simultaneously in limited space.” (Taylor, 2002; Downs, 2004) It has also been stated that congestion is not necessarily all bad, since it can be a sign that “a community has a healthy growing economy and has refrained from over-investing in roads.” (Cervero, 1998) Similarly, it has been noted that unpopular places rarely experience congestion (Garrison and Ward, 2000) and that declining cities have actually experienced reductions in congestion. (Taylor, 2002)

Given the limitations of metro level congestion indices, some alternative techniques have been proposed. For example, a congestion burden index (CBI) was proposed to account for the presence of commute options. (Surface Transportation Policy Project, 2001) The CBI is the travel rate index multiplied by the proportion of commuters who are subject to congestion by driving to work. For example, the 1999 Portland travel rate index was 1.36 (rank 8), and the transit share was 0.14. So the CBI was $1.36 \times (1 - 0.14) = 1.16$ (rank 14). As another indicator that the provision of transportation choices in an urban area is helpful, the transportation choice ratio was also proposed (Surface Transportation Policy Project, 2001), which is calculated by dividing the hourly miles of transit service per capita by the lane miles of interstates, freeways, expressways and principal arterials for each metro area. It has also been recognized that there is an interaction between personal lifestyles and traffic congestion. Some have noted that during peak periods, only one-third to one-half of all trips are work trips (Lomax, et al., 2001).

Knowing that congestion is often poorly measured, there are few standard indices and it is difficult to compare congestion across metro areas and years, a capacity adequacy (CA) has been proposed. (Boarnet, Kim and Parkany, 1999) The CA system establishes six capacity levels for highway classifications between principal arterials in rural areas to major urban expressways, based on peak hour traffic flow rather than daily VMT. The CA is calculated as $100 \times \frac{\text{capacity}}{\text{volume during present design hour}}$. In this equation, the capacity is estimated and design hour volume is based on the 30th highest hour (rural) or 200th highest hour (urban). This analysis was performed at a county level for California counties; the CA for each highway was weighted by ADT and summed for the entire county.

Current Research

Current research reveals some fundamental changes in how vehicles use the highway system. For example, very high sustained freeway flows have been measured, more than 20% greater than was once considered to be a theoretical maximum (Lomax, et al., 2001; Cassidy and Bertini, 1999). This means that drivers are accepting very short headways, such that one vehicle’s hesitation can cause other vehicles to brake suddenly (Lomax, et al., 2001). In addition, it has been shown that under some circumstances, freeway flows drop when
Congestion forms. On Canadian and German freeways, this drop is in the range of 5-10% (Cassidy and Bertini, 1999; Bertini, et al., 2004), while in Los Angeles, there are reports that the uncongested flows of 2,000 to 2,500 vehicles per hour per lane (vphpl) to about 1,400 to 1,600 vphpl. (Garrison and Ward, 2000). Finally, earlier studies have mentioned the need to protect transit vehicles from congestion (Traffic in Towns, 1963). Recent developments in bus rapid transit and transit signal priority are taking advantage of opportunities to exploit some gains in improving person travel through congested corridors. (Byrne, et al., 2004) There is a need to continue these and other research programs in order to improve our understanding of how the transportation system operates at both microscopic and macroscopic levels.

CONCLUSIONS

Congestion on the nation’s highway system continues to increase, both in reality and in people’s perceptions. Increasingly, congestion can no longer be addressed by simply expanding capacity. Limitations in transportation finance do not allow all needs to be met, and in many areas, public acceptance of highway capacity expansions is limited because of the impact on the environment. One implication of this situation is that since congestion cannot be eliminated the standard methods for measuring and reporting system performance in those terms are no longer very useful. We can no longer simply evaluate the effects of road widening projects on vehicles using limited, aggregate measures such as VMT, the volume/capacity ratio and LOS, nor is it helpful to apply arbitrary speed or volume thresholds across all facility types. These limited measures are usually derived from simple, limited data (e.g., average volumes, number of lanes) extrapolated over large segments of the network and do not consider the impacts on different types of users. This limits the specificity of performance reporting to large areas and generalized effects. Given new developments that allow for more robust data collection and demands for reporting actual system performance, we can no longer rely on the old way of system performance measurement.

Improvements or changes to the transportation system will impact different users differently—and the magnitude of that impact depends on the type of travel (e.g., freight, commute, recreation) and when their travel needs occur. Therefore, we now need to develop the ability to assess how different system users and society in general are affected by congestion and how that would change with different congestion mitigation actions. For example, reducing congestion on a highway serving a retail center might not be as beneficial as reducing congestion on a freight route because shoppers may be less sensitive to congestion delay than manufacturers and shippers, especially where just-in-time delivery is an important business practice.

In order to reliably estimate how congestion affects different travelers we need three things. First we have to know who is on the congested highway links and how and why they’re traveling. Second we need to understand the trip characteristics that are important to travelers (e.g., travel time, reliability). Third, we need data that can be used to estimate these important...
trip characteristics. For example, if truck movements to and from a high tech manufacturing area are occurring on a congested highway segment, and if travel time reliability is an important travel characteristic, then we must be able to collect performance data that can be used to estimate travel reliability. Future efforts to define and measure traffic congestion should include these important principles.

ACKNOWLEDGEMENTS

The author appreciates the dedicated support and input of Brian Gregor, Oregon Department of Transportation. Tim Lomax of the Texas Transportation Institute generously supplied the advance 2004 Urban Mobility Report (2002 data). Sonoko Endo conducted the comparative analysis of Urban Mobility Data. Chris Monsere, Jennifer Dill and Jacob Baglien also assisted with data analysis. Matt Lasky, Steve Hansen, Alex White, Aaron Breakstone, Erin Qureshi and Abram VanElswyk assisted with the literature review. Thanks to Prof. Joe Sussman for the quote. This research has been supported by the Oregon Department of Transportation and the Portland State University Center for Transportation Studies. Any views presented here, or any errors or omissions are solely the responsibility of the author.

REFERENCES


http://www.odot.state.or.us/tddtproved/papers/cms/CongestionOverview021704.PDF


http://www.nawgits.com/itsforum/apco/index.cgi?noframes;read=1348


