The purpose of this note is to suggest ways of thinking about and studying congestion within the broader context of transportation system performance and planning. It draws on a review, reported elsewhere, of the technical literature on congestion-related issues.

In outline, this note first defines congestion, showing that congestion is just one dimension of transportation system performance. Mobility, accessibility, reliability, and safety, among others, are also performance dimensions that need to be considered in transportation planning. There are inevitable tradeoffs between these that decision-makers must reconcile: for example, traffic calming measures intentionally reduce automobile speeds in the interests of safety and livability. A study that focuses too narrowly on only one or a few of these dimensions runs the risk of not generating the results needed to make informed decisions about these tradeoffs.

Next the note discusses various ways to measure congestion and, to a lesser extent, other performance dimensions. This is less straightforward than it might seem, and the choice of measures can influence the way in which congestion is understood and addressed. A congestion study needs to carefully select the measures to use based on its objectives, scope and the type of recommended actions it is considering.

Current understanding of the factors that influence congestion is next presented, distinguishing between those that influence overall travel levels, network characteristics and travel choice. A congestion study would normally need to consider the full range of these factors in order to identify the contribution of each to congestion levels in the study area. Statistical analysis or transportation modeling would typically be used to quantify these effects. A study that limits itself to only one or a few of these factors might ascribe changes in congestion levels to only those factors, possibly missing other contributing factors that are equally or more important.

Lastly the note outlines the steps that would normally be taken to fully study congestion-related issues in a complex urban area such as New York City. A study of this type can be quite a substantial undertaking, potentially involving public outreach, extensive data collection, statistical analysis or model development, and application of the understanding developed in this way to evaluate possible transportation system improvements across the range of performance dimensions. Although the specific schedule and resource requirements depend on the details of the study, accomplishment of all the required steps could potentially take a year or more.

Our hope is that this note will contribute to a meaningful public discussion of congestion-related issues in the City.
Defining Congestion

Although anyone who has lived in a populated area is surely familiar with the concept of congestion, precise definitions of this phenomenon can be difficult to pin down. “Too much traffic” and “loss of speed” are common descriptions, while more formal definitions include “a situation in which demand for road space exceeds capacity” [1] or “the incremental delay caused by vehicle interactions” [2]. In general, a good definition of congestion acknowledges both the supply of road capacity as well as the demand for road travel. More contemporary definitions do not focus exclusively on vehicles, recognizing in addition the people trying to travel and their perception and valuation of travel options [1].

A shortcoming of many congestion definitions is their emphasis on mobility: how easily people can move about. But individuals care about far more than this when they travel. People do not typically travel for the sake of traveling; they travel to access activities and opportunities [2, 3]. Certainly, improved vehicle speeds can improve accessibility, but so also can proximity and density, and even delivery services and telecommunications. People want a reliable trip, and may not care if this predictability comes with a somewhat longer travel time [1, 4]. People want a comfortable trip, a convenient trip, an affordable trip [5, 6]. More broadly, society wants travel to be safe, equitable and environmentally friendly [2].

Balancing these different performance dimensions requires making tradeoffs, both in definitions and in later analysis and mitigation efforts. While it may be tempting to narrowly define congestion in terms of traffic speeds and traffic volumes, this neglects the other transportation performance dimensions. In New York City, the Department of Transportation has been working toward safety goals through the Vision Zero program, and towards improved accessibility to all areas of the City through increased emphasis on non-automobile modes [7, 8]. The lower speed limits, traffic calming measures and bike and bus lanes seen throughout the City demonstrate commitment to these goals, and certainly provide many benefits to citizens that a strict view of road speeds would obscure. As a city thinks about congestion, it must recognize these different dimensions and establish a definition of congestion that is appropriate for its concerns and goals.

To establish a definition that is appropriate for its own needs, a city can begin with the two most common perspectives on congestion:
• **Engineering perspective**: focuses on the physical usage of roads by measuring traffic flows, levels of service, or speeds. This can be useful to identify system bottlenecks and improve speeds, and is the approach taken by many local or regional governments [9, 10, 11, 12], as well as by the well-known Texas Transportation Institute’s Urban Mobility Report [13].

• **Economic perspective**: recognizes that the costs of providing congestion-free roads typically exceed society’s willingness to pay for them, so some amount of congestion appropriately balances social costs and benefits. Economic congestion studies [14, 15, 16, 17, 18] often report the total costs of congestion, accounting for factors such as the delay time and additional user costs (such as for fuel) resulting from driving in congested conditions.

Both of these perspectives may however be myopic by not recognizing the other dimensions of transportation system performance. Economic approaches are generally preferred [2], because they can be extended to include other dimensions. An approach that recognizes that there is some acceptable level of congestion, that considers multiple modes, and that includes factors like emissions and safety in total travel costs, is a good starting point for a congestion study.

The definition should also decide what “types” of congestion it wants to include. Some congestion, like that from peak hour commuting, is recurrent so regular travelers know to plan for it. Other congestion delays are caused by random events like vehicle crashes or poor weather. When there are so many vehicles on the road that they begin to impede each other’s movements, congestion may be physically evident, but if traffic conditions simply do not meet expectations, drivers experience a relative congestion. In dense urban areas, some congestion may be tolerable—an understood side effect of city living—but these levels can turn excessive as the problem worsens. In defining congestion for a city-wide study, it is recommended to focus on excessive congestion and to distinguish between recurrent and non-recurrent congestion (which have different mitigation needs).

Finally, definitions of congestion are dependent on the scale of interest. A regional study needs to define congestion differently than one looking for the causes of particular bottlenecks [19, 20]. Congestion levels vary over the day, so concerns about traffic conditions in the peak-of-the-peak vs. over a longer period may also require different congestion concepts [20]. Some congestion definitions incorporate all modes of travel (cars, transit, bike, freight, etc.) but most acknowledge only cars. Especially in cities with extensive use of non-auto modes, a multimodal characterization can be useful to incorporate variability in modes’ speeds, stopping patterns and efficiency (including space requirements and occupancy) [2]. Defining congestion for a particular scale is especially dependent on context and motivation, but for a city like New York, acknowledging all modes, looking city-wide as well as at particular corridors of interest, and considering meaningful time periods within the entire day will provide a more holistic sense of congestion’s impacts on transportation system performance.
Measuring Congestion

Just as there are many ways of defining congestion, many metrics have been developed to measure and monitor it. Good congestion metrics should be simple and unambiguous; replicable and analytical; transferable across modes, time and space; continuous over a range of values; and able to describe and predict conditions [2, 21, 22]. Several overarching measurement categories are provided in the table below [1, 2, 21], with specific examples and comments citing some of their strengths and weaknesses.

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>Average or Nth percentile speeds in particular periods, locations and directions</td>
<td>Easy to measure, meaningful to user; Challenging to establish a point of comparison</td>
</tr>
<tr>
<td>Temporal/Delay</td>
<td>Hours of delay (total, per capita or distance), Duration and frequency of congested conditions, Average travel time, Travel Time Index (TTI)</td>
<td>Consistent with typical perception of congestion; Challenging to establish a point of reference for “delay”; Travel time metrics ignore distance traveled</td>
</tr>
<tr>
<td>Spatial</td>
<td>Number or portion of congested roadways, Network Connectivity Index</td>
<td>Challenging to establish a point of reference for “congested”</td>
</tr>
<tr>
<td>Service Level/Capacity</td>
<td>Level of Service, Roadway Saturation Index</td>
<td>Easy to measure, common and well understood; Too location-specific and traffic-centric, generally not multimodal</td>
</tr>
<tr>
<td>Reliability</td>
<td>95th vs. 50th percentile travel time (Planning Time Index), average levels vs. variability of travel time</td>
<td>Link congestion with reliability; Not sufficient as a stand-alone metric</td>
</tr>
<tr>
<td>Economic Cost/Efficiency</td>
<td>Cost of congestion, often including value of time lost, value of reliability, or additional fuel consumption</td>
<td>Potential to be multimodal and link congestion to numerous other dimensions; Challenging to gather sufficient data and establish a point of comparison</td>
</tr>
<tr>
<td>Energy/Emissions</td>
<td>Additional fuel consumption (from stops/starts or changes in speed)</td>
<td>Link congestion with environmental concerns; Not sufficient as a stand-alone metric, challenging to establish a point of reference</td>
</tr>
<tr>
<td>Exposure</td>
<td>Portion of travelers caught in congestion</td>
<td>Potential to be multimodal and capture temporal distribution of congestion; Underdeveloped category</td>
</tr>
</tbody>
</table>

All of these measures can provide useful ways of quantifying and understanding particular aspects of the congestion issues in an area. However, many of them are very traffic-centric. The following measures, at a minimum, are recommended for a comprehensive and multimodal assessment [2]:
• **Congestion Cost per Capita**: this indicator can potentially include the costs of direct and indirect impacts of congestion across multiple modes, while relating these costs to the size of the affected population;

• **Hours of Delay per Capita**: this indicator can incorporate multiple modes and the role of population while being consistent with the usual understanding of congestion’s effects; and

• **Travel Time per Capita**: this indicator can incorporate multiple modes as well as the effects of density and land use on trip making.

Again, a comprehensive transportation system assessment generally needs a variety of measures to capture all relevant aspects of congestion, as well as other performance dimensions. In fact, these other dimensions – like accessibility, safety and service quality – have their own sets of indicators that can be even more complex to quantify. For example, accurate accessibility metrics require not only temporally and spatially detailed data, but also information about people’s abilities and destination preferences. Very aggregate accessibility measures (like the number of jobs that can be reached within a given travel time from a location) can be developed from available data, but may lack the detail needed to understand the impacts of accessibility on travelers in different situations. Similarly, although basic statistics about travel safety are available, discussions of how to incorporate this dimension into travel costs can become heated. There is also less clarity about how travelers themselves perceive these factors and incorporate them in decisions.

Data collection efforts to establish values for these metrics can be significant [19, 21], particularly when temporally and spatially detailed data are required. This level of resolution has high value, though, because aggregate congestion metrics (such as city-wide average speed over the entire day) can mask specific local details about changes to traffic conditions and the reasons for them.

Local or state governments can be a source for traffic volumes and speeds, frequently collected with road sensors. Similar data can also be purchased from private companies, which collect and combine data from a variety of sources. Taximeters can make taxis a convenient source of speed data. Government-collected travel, land use and socio-economic data, or data from transit agencies about their performance, are sources that can be used to measure other performance dimensions. Some of this data is collected by the Census and American Community Survey (e.g. journey to work data for commute trip characteristics). Special purpose data collection efforts, such as travel surveys, may be needed to derive information on particular issues, for example to understand people’s value of time or reliability in the local context or to relate travel speed to fuel use and emissions. Municipal open data initiatives, like that of New York City [23], are increasingly effective repositories of useful data.
3 Understanding Congestion

Having discussed above the definitions and measures of congestion, this section now looks at some of the key factors that can have an impact on congestion levels, as revealed through research carried out around the world. These factors – not all of which may apply in New York City – can broadly be divided into:

- Factors that influence travel levels;
- Factors that influence system characteristics; and
- Factors that influence the choice of mode for travel.

Factors that influence travel levels

- **Economic activity.** As the US economy rebounds from the 2008 recession – having regained nearly 9 million jobs lost during the downturn – traffic congestion across major cities has returned to pre-recession levels [13]. Moreover, vehicle-miles traveled (VMT) across the nation, which were decreasing during the recession years (2008-2010) have been increasing since 2011 [24]. Increasing employment levels increase the demand for travel across all modes [25] for both obligatory (e.g. commuting to work) and discretionary (e.g. leisure) trips. Freight traffic also increases with a rebounding economy [26]. The overall increase in travel from an improved economy can manifest itself in increased congestion levels in urban areas and on intercity routes.

- **Population and income.** The population of an area affects local travel levels. Real (i.e. inflation-adjusted) income has historically also exhibited a strong correlation with the number of vehicles owned and total vehicle miles traveled. Studies have shown that when real income goes up by 10%, vehicle ownership increases by nearly 4% in the short term (up to one year) and by more than 10% in the long term. The corresponding volume of traffic grows by 2% and 5% in the short and long term respectively [27]. Similarly, according to another study, a 50% increase in income per adult leads to a 15% and 23% increase in VMT in the short and long term respectively [25]. The difference between short- and long-term changes is stark because commuters (who contribute a large fraction of VMT) cannot easily change their work or residence locations [25].

- **Land use and spatial structure.** Land use factors such as development density and mix, regional accessibility, population centrality, jobs-housing balance and others collectively have
an impact on the volume and extent of travel. For instance, location of development in close
proximity to an urban center reduces per capita travel mileage [28] [29].

- **Tourism.** Iconic cities like New York City are popular tourist destinations attracting a large
volume of tourist flow annually. The number of annual visitors to New York City has increased
from 36.2 million in 2000 to 54.3 million in 2013 [30]. This fast-growing tourism volume can
place heavy demands on infrastructure such as roads and public transportation systems [31].

- **Fuel costs.** People and companies generally consider cost when making travel decisions and
fuel costs are among the most palpable of travel costs. Fuel costs are directly related to fuel
prices and inversely related to vehicle fuel efficiency. VMT increases when fuel costs per mile
decrease and vice-versa [25] [32] [33]. This effect of fuel cost on VMT is more pronounced in
times of high fuel price volatility and media coverage [32].

### Factors that influence transportation system characteristics

- **Traffic calming measures (e.g. Vision Zero).** Traffic
  calming measures such as speed limit reductions,
  establishment of slow zones, creation of speed bumps
  and increased speed enforcement are often introduced
  with the primary goal of improving traffic safety and
  livability [7] [8]. These measures can provide important
  transportation and social benefits, but may potentially
  reduce vehicle speeds.

- **Adding and removing capacity.** Occasionally, capacity
  is added to road and transit networks, as for example the recent 7 Subway Extension in New
  York City. More commonly, existing systems are transformed. Urban streets and street
designs are evolving to include improved facilities for cyclists and pedestrians in many cities
[34]. Provision of exclusive lanes (e.g. bike or bus lanes) improves safety and accessibility for
non-auto users but may reduce the width or number of travel lanes for motorized vehicles
[35]. On highways, provision of separate toll or high-occupancy vehicle lanes may reduce the
number of lanes available for general vehicular traffic. These measures are nonetheless
pursued because of an emphasis on broader transportation system performance.

- **Vehicle and parking restrictions.** Many urban drivers spend time looking for parking. It has
  been reported that this parking search traffic can constitute up to 30% of the traffic in some
urban areas [36, 37, 38], though this figure has high spatial and temporal variability. In
addition to overall demand for parking, its impact on traffic levels is affected by the amount of
and prices for curbside parking, as well as those for off-street parking garages and lots.

- **Freight traffic and parking.** Illegal parking of delivery vehicles is estimated to be the third
  leading cause of non-recurrent urban congestion, behind vehicle crashes and construction
works. A 2005 study estimated the total delay caused by illegally-parked delivery vehicles in
large U.S. urban areas at around 500,000 million vehicle-hours or $10 billion in lost time
annually [39]. This problem is exacerbated by increased economic activity and e-commerce
levels, both of which tend to increase freight traffic.

- **Disruptions.** Specific events can disrupt transportation system operations, leading to poor
  service, loss of capacity and delays.
a. Vehicle crashes are the single most important cause of non-recurrent traffic congestion [40], producing a temporary loss of capacity and long delays. Measures to improve traffic safety can thus also reduce non-recurrent congestion levels. Studies have shown the introduction of digitally-dispatched for-hire vehicle (FHV) services can reduce alcohol-related traffic deaths [41].

b. Slowdown or closure of transit services due to inclement weather, transit strikes or maintenance and repair works can shift travel to other modes, exacerbating delays on them. A study of the 2003 Los Angeles transit workers’ strike found that average highway delay increased by 47% when transit services ceased [42]. As Superstorm Sandy demonstrated, extreme weather events can lead to long-term disruption of transportation services and travel conditions.

c. Special events in the city can lead to temporary lane or road closures and detours. These physical changes can be compounded by traveler confusion, with people trying to navigate uncertain conditions.

- **Construction/road works.** Work zones account for about a quarter of the delay attributed to non-recurrent congestion [40] and lead to reduced speeds, detours and diversions. Although transportation and public works departments make considerable efforts to minimize the impact of construction/road works on traffic, these works are still the second most significant source of non-recurrent congestion in urban areas [40].

**Factors that influence travel choices**

A traveler’s choice from among the available modes of transportation is influenced by each mode’s service characteristics including cost, travel time, schedule convenience, access/egress convenience, reliability, comfort, safety and others. Changes in any of these factors can affect mode choices and potentially traffic levels. Research into these influences has traditionally focused on factors such as travel time and cost. Understanding the demand for newer modes such as digitally-dispatched FHVs requires better understanding of traveler assessment of factors such as schedule and access/egress convenience.

In some cases, a traveler might have a choice of when to travel, and this choice is both influenced by and influences traffic conditions at different times of the day. Transportation policies such as peak/off-peak pricing differentials (e.g. Port Authority bridge tolls) can also affect the choice.

The relationship between congestion and the factors identified above can be quantified with the help of statistical methods and transportation network models. Changes over time in a city’s congestion levels will generally be due to a combination of many of the above factors. The interrelationships between the factors adds to the complexity of the analysis and makes it challenging to disentangle the independent contributions of each of them. In any case, a narrow focus on one specific factor, such as FHVs, will likely lead to inaccurate conclusions about the true underlying causes of any congestion changes.

Similar statistical and modeling methods can be used to understand the relationship between these factors and other performance dimensions such as accessibility, reliability and safety. A holistic congestion study will want to include evaluation of all performance dimensions and develop an understanding of connections and trade-offs between them.
4 Studying Congestion

The preceding discussion has presented state-of-art best practice approaches to understanding congestion, including in the broader context of transportation system performance. In practical terms, a study that applies these approaches would entail a number of inter-related steps.

**Define study objectives.** Congestion is just one area of transportation system performance, along with other areas such as mobility, accessibility, reliability, safety and others. A congestion study should consider the full range of performance areas of concern, as any proposed improvements will inevitably involve tradeoffs among different performance areas, and a study that focuses narrowly on one or a few areas will not generate the information needed to understand impacts in the other areas and to make such tradeoff decisions. Selecting suitable measures to quantify congestion and the other performance areas is an equally important early decision for the study. Inputs from stakeholders, obtained through a robust communications and outreach program, can be a valuable source of ideas for these definitional steps. It is likely that a New York City study would want to consider, in addition to congestion itself, a number of other performance areas: accessibility performance, with measures such as access to jobs or to underserved areas; mobility performance, including measures that capture shifts away from single-occupant vehicles to active and alternative modes; safety performance, with measures of conflict reduction; and others.

**Assemble needed data.** Many different types of data are required for a congestion study, both to establish baseline and historical travel conditions in the study area, as well as to provide inputs to statistical analyses and transportation model development. Some data types are routinely collected by government agencies and are readily available. Other data types are less commonly available and a study must make special arrangements to collect these. Private for-profit vendors (e.g. of anonymous cell phone movement data or of aggregations of overlapping traffic data sources) increasingly provide such data. Quality assurance and cross-checks are also required to ensure that this data are suitable for any subsequent applications. It is also frequently the case that obtaining an in-depth understanding of traveler behavior for a study requires the design, fielding and analysis of study-specific traveler surveys. This can be a significant undertaking. The duration and cost of a congestion study’s overall data collection effort depend closely on the specific nature and scale of its data collection activities, but in an area such as New York City, a data collection effort lasting several months and costing in the hundreds of thousands of dollars would not be out of reasonable range.
Analyze the assembled data. Collected data can be analyzed to shed light on particular issues of concern or, as described above, through statistical methods or transportation modeling approaches to develop a more comprehensive and holistic understanding of the relationships between congestion (and the other selected measures of transportation system performance) and possible explanatory factors. In the current New York City context, for example, a statistical analysis might investigate quantitative relationships over time between average speeds and a variety of possible socio-economic and transportation system factors that are postulated to affect speed. Similarly, a transportation model might, for example, attempt to simulate the daily travel patterns of FHVs and their interactions with general traffic. Prior to use, any developed statistical or transportation models should be subjected to a rigorous process of validation, involving a variety of checks and verifications; and also to extensive sensitivity and scenario testing, to establish ranges of applicability of the model results.

Apply the understanding. With the understanding of the transportation system developed from the preceding steps, it then becomes possible to assess possible actions to address the identified issues. The specific nature of the actions to be considered will depend on the scope of the study. In all cases, the assessment of possible actions needs to consider their effects across the range of selected performance areas and measures, and any decision should consider the inevitable tradeoffs between these, as revealed through the congestion study. New York City has a wide range of possible actions to address congestion-related issues, ranging from major infrastructure investments, to tolling and congestion pricing options, transportation demand management initiatives and development of new travel options such as digitally-dispatched FHVs.

A study of this sort, conducted at the scale of the New York City area or even just a single borough, is a considerable undertaking, and could potentially take a year or more to accomplish all the required steps. The specific duration and other resource requirements of the study will depend closely on the scope, the need for and nature of new data collection, the suitability and applicability of already-available results, and other details.
Thinking about Congestion | Report

Works Cited


4) York City, 2013.


