Relevance and Reliability of Area-Wide Congestion Performance Measures in Road Networks

Carlos A. Morán T.

Doctoral Dissertation
Royal Institute of Technology
Stockholm, Sweden 2010
© Carlos Moran 2011

Relevance and Reliability of Area-Wide Congestion Performance Measures in Road Networks

Royal Institute of Technology –KTH
School of Architecture & Built Environment
Transportation and Logistics Division

TRITA-TEC-PHD 11-002
ISSN 1653-4468
ISBN 978-91-85539-60-4
Abstract

For operational and planning purposes it is important to observe, predict and monitor the traffic performance of congested urban road links and networks. This monitoring effort describes the traffic conditions in road networks using congestion performance measures. The objective of this research is to analyse and evaluate methods for measuring congestion and congestion performance measures for monitoring purposes. For this objective, a selection of the required aspects of the performance measures in the literature is considered. The aspects to be considered can be classified into two categories: A first group relates to the statistical aspects of these indicators, i.e. reliability. The second relates to their ability to capture the impacts of congestion, i.e. relevance. The reliability and relevance of the congestion performance measures are evaluated. A recommendation of the most suitable indicator is provided at the end of the study.
Acknowledgements

The work presented in this thesis has been carried out at the Division of Transportation and Logistics, Royal Institute of Technology, Stockholm, Sweden.

I am deeply grateful to Professor Harilaos Koutsopoulos, who supervised the later part of this thesis, for patiently teaching me how to conduct research. His constant support and his insight, enthusiasm are sincerely appreciated. I would also like to thank Karl-Lennart Bång who supervised the earlier part of this thesis for all the fruitful comments and discussions as well as helpful advising. I would like also to mention my gratitude to Karin Brundell-Frej for her encouragement, guidance and insight and her advice.

I would like mention that this research would not have been possible without the founding and cooperation with the Swedish Road Administration (swe: Trafikverket), Stockholm City Traffic Office (swe: Stockholmsstads Trafikkontoret) and Centre for Operations and Maintenance (swe: Centrum för Drift och Underhåll - CDU).

The pursuit of this work would not have been possible without the encouragement of my family, my former and current colleagues at Traffic and Logistics, and former teachers in Chile. Refreshing laughs and conversations with colleagues have been most supportive in my everyday research work.

Stockholm, September 2011

Carlos Morán T.
# Contents

Notation ........................................................................................................................................ 11
1 Background ................................................................................................................................ 13
  1.1 Introduction .............................................................................................................................. 13
  1.2 Research Motivation ................................................................................................................ 15
  1.3 Knowledge gap and research questions ................................................................................... 16
  1.4 Thesis structure ........................................................................................................................ 18
2 Literature study .......................................................................................................................... 19
  2.1 Definition of congestion – all transport modes ........................................................................ 19
  2.2 Approaches for Defining Congestion in road networks For Monitoring Purposes ................ 20
  2.3 Inventory of congestion performance measures for road networks ....................................... 22
  2.4 Approaches for modelling the effects of congestion ............................................................... 28
  2.5 Conclusions of the literature research ...................................................................................... 31
  3.1 Background .............................................................................................................................. 33
  3.2 Problem formulation ................................................................................................................ 33
  3.3 Theoretical assumptions ......................................................................................................... 33
  3.4 Empirical work ....................................................................................................................... 34
  3.5 Results .................................................................................................................................... 34
  4.1 Background .............................................................................................................................. 35
  4.2 Problem formulation ................................................................................................................ 35
  4.3 Theoretical assumptions ......................................................................................................... 35
  4.4 Empirical work ....................................................................................................................... 35
  4.5 Results .................................................................................................................................... 36
  5.1 Background .............................................................................................................................. 37
  5.2 Problem formulation ................................................................................................................ 37
  5.3 Theoretical assumptions ......................................................................................................... 37
  5.4 Empirical work ....................................................................................................................... 37
  5.5 Results .................................................................................................................................... 38
  6.1 Background .............................................................................................................................. 39
  6.2 Problem formulation ................................................................................................................ 39
  6.3 Theoretical assumption .......................................................................................................... 39
  6.4 Empirical work ....................................................................................................................... 39
  6.5 Results .................................................................................................................................... 40
  7.1 Background .............................................................................................................................. 41
  7.2 Problem formulation ................................................................................................................ 41
  7.3 Theoretical assumption .......................................................................................................... 41
List of Figures

Figure 1: Elements of a performance based planning process (Cambridge Systematics, 2000) ......................... 14
Figure 2: Topics treated in each paper and overview of thesis ........................................................................ 18
Figure 3: Inventory of congestion performance measures categories ............................................................. 23
Figure 4: Congestion increases with traffic levels. Source (TfL, 2003b) .......................................................... 26

List of Tables

Table 1: research motivation: aspects and research questions ........................................................................ 16
NOTATION

CBD : Central Business District
CPM : Congestion Performance Measure
\( f_i \) : Vehicle flow on link \( i \), and
GIS : Geographical Information System
\( i \) : Link in the set of links \( N \)
\( l_i \) : Length of the link \( i \).
LOS: : Level of Service
MFD : Macroscopic Fundamental Diagram
MOE : Measure of Effectiveness
\( N \) : Set of links considered in the network
SRA : Swedish Road Administration
\( S_i \) : Speed observed on link \( i \)
\( S_{network} \) : Network Speed
TfL : Transport for London
\( tt_i \) : Travel time on link \( i \)
TTI : Travel Time Index
TR : Travel Rate
VKT : Vehicles Kilometres travelled.
1 BACKGROUND

1.1 INTRODUCTION

Urban settlements have grown in size and this tendency does not seem to be stopping. In these big agglomerations of people, the demand for travelling concentrates in certain periods. These periods are, in some cases, longer than just a peak hour. It is during these periods when congestion occurs. Mobility and accessibility are reduced and many other problems take place that not only reduce the quality of life but also the prospects for economic development (Weisbrod et al., 2001). One of the aspects that has received increased awareness is the increase in pollutant emission caused by traffic congestion. Regional and local governments have endeavoured to diminish all these congestion impacts in the congested areas and to relieve the suffering of the traveller through a wide range of actions, investments and policies. These efforts, like all the efforts made by politicians, are evaluated using measures of effectiveness.

1.1.1 PERFORMANCE MEASURES AS A KEY FOR DECISION MAKING

Turner et al. (1996) provide recommendations for the selection of these measures of effectiveness in a framework of major investments in transportation. For example, matching the measures of effectiveness with the goals and objectives of the study, quantifying impacts when possible, not using subjective judgement and identifying the error levels of calculation in relation to the measured level.

The selection of the measures of effectiveness has to be done carefully, because they will direct funds of future investments. Cambridge Systematics (2000) introduces these measures of effectiveness as performance measures. This study provides a framework of how these performance measures fit in the overall planning process of the transport agencies for general transportation policies as shown in Figure 1. The diagram shows the core and key role of the performance measures in the performance-based planning process. This framework can be applied for example for congestion reduction, safety improvements, traffic calming, etc.

In this general decision making context, Sinha (2007) provides a list of properties that a performance measure should have:

- Appropriateness: It should be an adequate reflection of at least one goal or objective of the system action.
- Measurability: It should be easy to measure in an objective manner as well as to estimate with available resources and tools. Measurable results should be within an acceptable degree of accuracy and reliability.
- Dimensionality: It should fit the dimensions of the evaluated problem. The performance measure should be comparable across time periods and geographic regions.
- Realistic: It should be possible to estimate without excessive effort, cost or time.
- Defensible: It should be clear and concise so it can be communicated effectively to certain intended audiences. This is often possible when the performance measure is clear and simple in its definition.
• Forecastable: For planning purposes, it should be possible to reliably estimate its values based on current forecasting tools.

![FIGURE 1: ELEMENTS OF A PERFORMANCE BASED PLANNING PROCESS (CAMBRIDGE SYSTEMATICS, 2000)](image)

1.1.2 PERFORMANCE MEASURES IN THE CONTEXT OF ROAD CONGESTION IN URBAN NETWORKS

These properties, provided in the general context of infrastructure planning have been specified for a road transportation context (SRA, 1999):

- First, a good congestion performance measure (CPM) should be statistically efficient and reliable; specifically related to the fact there always will be a certain distortion or statistical bias caused by the sampling method\(^1\).
- Second, a good congestion performance measure should have a monotone relationship with the defined goal or impact caused from congestion. Furthermore, it has to be as highly correlated as possible with the impacts caused by congestion.
- And third, a good congestion performance measure should be understandable and preferred by the intended audience.

---

\(^1\) For monitoring purposes it is not possible to collect data from the whole population but only from a limited sample.
1.2 RESEARCH MOTIVATION

The objective of this research is to analyse and evaluate methods for measuring congestion and congestion performance measures for monitoring purposes. This analysis presented in this thesis covers a variety of aspects.

Evaluating “appropriateness” requires defining goals or objectives for a policy or measure. If the goal is to reduce congestion, then defining congestion is required. The literature and previous monitoring efforts do not agree on a definition of congestion. This is the first aspect to be addressed in the intended analysis.

Previous monitoring efforts have used the output of long term planning models. As such, the indicators are static and insensitive to the dynamics of congestion. A consistent analysis requires taking into account the time dependent characteristics of congestion. This is the second aspect to be addressed in the intended analysis.

The “measurability” aspect requires the estimations to be within an acceptable degree of accuracy and reliability (Sinha 2007). In the same way, it is necessary to identify the error levels of calculation in relation to the measured level (Turner et al 1996). These makes sense with the “statistically efficient” requirements that a good congestion performance measure should fulfil (SRA 1999). All these aspects can be summarized as the congestion performance measures reliability. This is the third aspect to be considered.

When the policy goal is to reduce delays, a good performance measure should be a good reflection of the variation in the impacts (Sinha 2007) or to be as highly correlated as possible with the delay impacts. This is the fourth aspect needing analysis.

Another shortcoming is that the CPMs are based on a single deterministic level, when comparing congestion to a reference or non-congested level (Merritt and Bang, 2000, Schrank and Lomax, 2005, TfL, 2008, City of Stockholm, 2006). This approach ignores how motorists perceive congestion and does not take into account their preferences. This is the fifth aspect to be addressed in the present study.

Recent increased awareness has targeted emissions and sustainability as a desired goal on the politicians’ agenda. Previous efforts have been done at the intersection level (Choi and Frey, 2010), at the metropolitan level (Bigazzi and Bertini, 2009a), or at a more highly aggregated level (Ramjerdi et al., 2008), but not at the network level. This evaluation of the ability of these congestion performance measures to capture pollution impacts of congestion needs attention. This is the sixth aspect to be considered.

The fourth, fifth and sixth aspects refer to the ability of indicators to capture the impacts of congestion and can then be then summarized as the congestion performance measures’ relevance.

Sinha (2007) and SRA (1999) refer to the defensible property of the performance measures and consideration of the intended audience. In many cases, the intended audience will be non-expert audiences like politicians. In these cases, it will be necessary to summarize all the aspects that a good congestion performance measure should fulfil. This is the seventh & final aspect considered in the analysis in the present study.
1.3 **Knowledge Gap and Research Questions**

Several studies and reviews about congestion indicators and inventories can be found in the literature (Lomax et al., 1997, Schrank and Lomax, 2005, Bremmer et al., 2004) but analysis of the aspects mentioned above that make “a good congestion performance measure” has not received any attention in the literature. The following table summarizes some of the aspects that have not been discussed in earlier studies and corresponding research questions that arise.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Research questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. Congestion performance measures should be an adequate reflection of congestion.</td>
<td>What are the definitions of congestion for monitoring congestion in urban road areas? What are the suitable congestion performance measures for the definition approaches?</td>
</tr>
<tr>
<td>ii. Indicators use output data from long term planning models and thus are not sensitive to the dynamics of congestion.</td>
<td>How can the congestion indicators be adapted or further developed to consider the time dependent aspects of congestion?</td>
</tr>
<tr>
<td>iii. Identifying the error levels in relation to the measured level.</td>
<td>What are the error levels for the estimated measured levels? How can the reliability aspects of the estimation be analysed? What indicators perform better and are hence more reliable?</td>
</tr>
<tr>
<td>iv. During congested periods mobility and accessibility are reduced. Congestion performance measures should be an adequate reflection of delay impacts.</td>
<td>How can congestion performance measure capture the impacts caused by congestion in terms of delay?</td>
</tr>
<tr>
<td>v. Current congestion performance measures consider only a single deterministic reference level. This hinders the possibilities of capturing the variability across users.</td>
<td>How can the variability across users’ preferences be incorporated in area-wide congestion indicators?</td>
</tr>
</tbody>
</table>
vi. There is an increase in pollutant emissions during congestion as well as an increased awareness of environmental impacts caused by traffic congestion. Congestion performance measures should be an adequate reflection of environmental impacts.

Do congestion performance measures capture the impacts caused by congestion in terms of emissions? What are the factors that affect the relevance aspects of congestion performance measures?

vii. A list of properties that an ideal congestion performance measure has been identified, but no guidelines are found in the literature of how to compound them in practice.

How do the different aspects that good congestion performance measures should fulfill act together?

These aspects are important and necessary for performance measures to be useful and efficient tools for policy makers and traffic engineers. The need for addressing the deficiencies and gaps in the area of congestion performance measures inspires this research. The research aims to provide better understanding about congestion performance measures for monitoring the road system, taking into consideration the time dependent aspects that congestion has in dense urban areas.

In conjunction with the Stockholm congestion charging trial, a great amount of traffic data has been collected for estimating the traffic conditions in different areas of the city and for different traffic facilities. This provides a unique opportunity for evaluating congestion and analysing congestion performance measures in terms of reliability and relevance.

The goal of this research is to address the issues discussed in Table 1. This can be illustrated as shown in Figure 2 below.

The first two research questions refer to the topic of defining congestion and how to incorporate the time dependent aspects in the analysis. One or both of these aspects are treated across all the papers.

Research question iii refers to the Reliability aspects of congestion performance measures and is treated in Papers A and B. They provide a better understanding about the statistical properties of congestion performance measures in respect to the error level with relation to the measured level.

Research questions iv, v, & vi refer to the Relevance aspects of congestion performance measures and can be found in Papers D, C & E. They analyse how congestion performances measures capture the impacts caused by congestion in terms of delays and emissions, the factors that influence their capabilities of describing these impacts.
1.4 Thesis Structure

The structure of the thesis is as follows. Chapter 1 introduces the research, motivation and objectives. Chapter 2 reviews the literature. Chapter 3 to 7 briefly describe the Papers A to E. Each of these sections is subdivided into: background, problem formulation, theoretical assumptions, empirical work, and results. Chapter 8 presents the conclusions of the study. Chapter 9 presents the future research recommendations. The final section includes the references.
2 LITERATURE STUDY

The literature study is divided in five sections:

- One crucial aspect for congestion performance measures is to match the objective or goal with what it is intended. The literature study analyses definitions of congestion in its first section in the general context of transport studies.
- The second section describes in more detail the definitions of congestion in the context of traffic monitoring road networks. It also presents approaches for defining congestion.
- In the third section, a brief inventory of selected congestion performance measures is presented.\(^2\)
- Figure 1 also shows that performance measures require targeted “data”. This “data” is utilized by “analytical models”. In the case of road networks, data is usually not available for the whole network, nor for all vehicles in the network, and it corresponds to a small sample of the total population. Thus, a process for estimating this data is required, i.e. estimating the traffic descriptive parameters for the rest of the network or for estimating other traffic descriptive parameters that are not widely observed, e.g. average queues, delay, etc. This procedure is also called the process of modelling the effects of congestion in traffic performance. The fourth section deal with this aspect.
- The last section summarises the conclusions from the literature research.

2.1 DEFINITION OF CONGESTION – ALL TRANSPORT MODES

The word congestion has its roots in the Latin “congest” that refers to overfilling and overcrowding. “Congest” is the past participle form for the verb “congerere” that refers to “carry together or collect” (“con”: together, with; “gerere”: to carry). Currently this term is more generally used for describing phenomena in the areas of telecommunications, medicine, electric power transmission, and traffic.

The meaning of the word congestion in transportation studies has not been unique. Differences are due the objective, conditions, scope and assumptions of the studies. However, congestion has always been perceived as an “undesired” phenomenon that unavoidably occurs in road traffic networks. The causes and the effects of congestion have been the basis for defining congestion in the past. However, definitions based on the effects have the advantage that they can be understood by non-experts such as policy makers and road users (Lomax et al., 1997).

2.1.1 FACTORS AFFECTING THE MEASURED LEVEL OF CONGESTION

Perceptions of the transport system users of different levels of congestion may vary depending on a series of factors and situations that have been categorized according to geographic scope, locus\(^3\), transportation mode, road type, time of day/week, planning context and level of detail (Lomax et al., 1997)

\(^2\)An extended inventory of congestion performance measures is available in Moran (2008).
2.1.2 Recurrence of Congestion

Lomax et al (2007) refers to recurrent congestion as an event that occurs regularly or every day at peak hours causing undesired travel time delays. These congestion levels are similar during similar periods. On the other hand, the literature refers to non-recurrent congestion as unexpected increases in travel time, mobility reduction and queue building caused by random incidents.

The differentiation between recurrent and non-recurrent congestion lacks importance when the objective is to study the performance of the road network and how robust it is against random incidents. Differentiating between recurrent and non-recurrent congestion will provide information on how the traffic control centres perform their functions and duties. The present study will not differentiate between these two types of congestion.

2.2 Approaches for Defining Congestion in Road Networks for Monitoring Purposes

There are many definitions of congestion. They can be grouped in two approaches:

- Bottleneck Approach
- Travel-Time Approach

2.2.1 Bottleneck Approach

The Highway Capacity Manual (HCM2000) does not directly define congestion. It defines instead “Congested flow - A traffic condition caused by a downstream bottleneck” (TRB, 2000). This definition is applied primarily to specific road links. If the traffic demand exceeds bottleneck capacity it will result in an accumulating queue that can cause upstream blockage influencing other parts of the network. This definition of congestion relates to the number of queuing cars on the studied link. The amount of queuing is also a good estimator of the effects of congestion for the occupants and non-motorized travellers in the studied zone.

2.2.2 Travel-Time Approach

Studies focused on economic effects have shown that travel time measures offer the best means for estimating the economic impacts of congestion (Weisbrod et al., 2001). Congestion in this context can be defined as “Time spent over what would be experienced under ‘un-congested’ or ‘free flowing’ conditions.”

The definition above is used in the London Congestion Charging Scheme (TfL, 2003a). Congestion is considered to be the difference between the measured or experienced travel rate (exper.) and the travel rate measured during nighttime or uncongested conditions (uncong.), i.e. the Excess Delay (ExD). The Swedish Road

3 Specific type of zone
Administration considers congestion as the “reduction in travel speed as time lost caused by a high saturation level” (SRA, 1999).

2.2.3 Discussion about Approaches of Defining Congestion

The discussion about the approaches of defining congestion focuses on three issues:

- Measurability
- The threshold for triggering the existence of congestion
- Variability in the reference level or basis for comparison of congestion

**Measurability**

These approaches to defining congestion present advantages and shortcomings. The bottleneck approach relates demand exceeding capacity in one or more points of the network. Since demand cannot be always observed in the field, the tracked measure is related to the presence of queues. The definition does not require a reference value for uninterrupted flow installations. When applied to urban zones, this definition has the disadvantage that it needs knowledge of where queues are detected. Queues can be observed partly at intersections and partly on road links.

- For queues at intersections, the interesting measure will be the number of cars in the queue\(^4\). This measure has the disadvantage that queues might exist even in low traffic levels.
- Measuring queues on link has its own complications. Accumulations of moving vehicles (e.g. roads that present high density of traffic or moving jams) lead to a vague threshold for the existence of congestion. Traffic density\(^5\) becomes a more pertinent parameter in this case. However, estimating density for monitoring purposes is not currently a feasible task.

The travel-time-based approach is not strongly affected by the duality of the nature of possible queues nor by random queues in intersections. The focus of this approach is on travel time (or journey speed) between two points. Other aspects, for example, density or driving behaviour, have certain effects but their values are ignored. Thus, the travel-time-based definition better suits routes comprised of road links with different traffic characteristics. This approach strongly depends on the definition of the reference level. Thus, managing measures that consider a reference level that matches the optimum level of utilization of the infrastructure (if can be defined a priori) will then be optimum. On the other hand, considering other values for reference level (e.g. posted speed) will benchmark current conditions with non-optimum levels.

**The threshold for triggering the existence of congestion**

\(^4\) It can alternatively be considered the number of passengers affected, length of the queue or any other quantity that refers to a number of observed units.

\(^5\) It can alternatively be considered vehicles-equivalent/segment, flow/capacity or any other proper quota.
The different approaches are not necessarily consistent and may trigger the existence of congestion at different traffic levels. An example of this occurs when the situation changes from no-flow to total gridlock. In the beginning, both approaches will report no congestion when the network is empty. Gradual increases in traffic will not result in perceptible effects. At this point, the drivers on the road will not perceive any changes in the travel impedance until the upper boundary of the uncongested situation is reached. Further increments in traffic will cause significant interference between drivers, but no queues. At this point, the bottleneck definition does not indicate the existence of congestion. However, this level will trigger the appearance of congestion according to the travel-time-based definition. Platoons, moving jams, and situations where desired speed cannot be maintained (but traffic flow is still below breakdown levels) may increase the difference between the approaches.

**Variability in the Reference Level or Basis For Comparison of Congestion**

Systems are called congested when demand for a service exceeds capacity. This fact has become apparent for both approaches.

The bottleneck approach requires, at some point, to identify bottlenecks. Identifying a bottleneck occurs when \textit{flow} exceeds \textit{capacity}. \textit{Flow} has been widely considered to randomly vary. Elefteriadou et al (2006) point out that field measurements show that capacity is not an upper bound for Level of Service E (maximum flow occurs after the occurrence of congestion i.e. breakdown according to the HCM). Their study concludes that \textit{capacity} is a random variable. This identification of bottlenecks requires the study of the difference between two random variables. This fact is also important in dense urban areas where queues might appear even in low traffic levels at signalized intersections.

This difficulty also exists for the travel-time approach. Lomax (1997) recognizes the need of differentiating between times of the day etc. He points out the need to address different perceptions and opinions that different drivers have about what is acceptable and unacceptable congestion.

2.3 **Inventory of Congestion Performance Measures for Road Networks**

Increases in travel time due to congestion are not the same for all drivers. The increases in travel time can be described using for example, average value, the worst traffic conditions or the spread in the observed values. Each one of these descriptors relates to a certain aggregation method. The focus of the present study is on congestion performance measures for monitoring urban road networks. It is then more suitable to adopt indicators that relate to the “average value” of the conditions in the system.

Figure 3 shows that congestion performance measures, as related to car traffic, can be categorized as:

- Category 1: Central Value indicators.
- Category 2: Dispersion indicators.

Category 1 describes the average value of certain parameters. This group might be subdivided in two groups depending on the input data needed for calculation.
Subgroup 1: Design defined data.

Subgroup 2: Field observable data.

Subgroup 1 corresponds to congestion performance measures that are derived from the planning and design process required for road infrastructure. Planning and design usually consider certain objectives and aims to satisfy certain standards (for example, “the minimum road capacity for a road should be higher than 1700 vehicles per hour per lane travelling at a speed of 70 km/h”). The parameters are usually defined by previous analyses and studies, such as national road plans or demand forecasting studies. This subgroup of congestion performance measures uses input variables that are usually not observable in the field such as, for example, traffic demand.

Subgroup 2 corresponds to congestion performance measures that relate to maintenance and operation of road infrastructure. Performance measures in this subgroup use data from simulation models or collected on field. These indicators are more suitable for monitoring purposes and are hence the focus of the present study.

Category 2 corresponds to congestion performance measures that aim to capture variability in traffic conditions. They describe the dispersion around the average value of parameters such as travel time, journey speed or other traffic descriptive parameters, indirectly neglecting the central value. These indicators require more detailed data. Lately, focus on these indicators has increased due to the growing public interest in the reliability of travel times.

The following section briefly describes the design defined data central value indicators. Later, the Field observable data/Central value indicators are presented. Empirical estimation and efforts for quantifying each of the described measures are respectively included in every section. A more detailed inventory of congestion performance measures can be found in Morán (2008)
2.3.1 DESIGN DEFINED DATA/CENTRAL VALUE INDICATORS

A good example of the indicators in this category is the Level of Service (LOS). LOS is presented in the Highway Capacity Manual. The objective of the HCM is to provide a consistent system of methods for the evaluation of the quality of service of highways and street facilities. The HCM does not set policies regarding a desirable or appropriate quality of service for various facilities or traffic systems. Its objectives include providing a logical set of methods for assessing transportation facilities, assuring that practitioners have access to the latest research results, and presenting sample problems.

HCM presents LOS in an easy-to-understand methodology of analysis and performance measures for single homogenous road segments. LOS describes conditions on road links and there is no direct methodology for aggregation. LOS has been criticised by analysts and experts in the area, but it is still in use due to the easy-to-communicate properties.

2.3.2 FIELD OBSERVABLE DATA/CENTRAL VALUE INDICATORS

The literature presents a series of indicators that use a stepwise scale. For a certain value of a traffic parameter (speed, flow, etc.) a discrete value of congestion or colour is assigned. This continues the idea of level of service described above. Given that these indicators are qualitative and not quantitative, they fail to fulfil important aspects of a good congestion performance measure. Thus, these congestion performance measures will not be described in the following study.

The field observable data measures relate, in general, to aggregate values of field observed data or parameters, e.g. average travel time for a certain road segment. Almost all these congestion performance measures comprise an intrinsic comparison of an observed or measured situation with a reference level. This reference level is usually defined arbitrarily. The reference congestion level might consider a low traffic level or a level that yields an optimum use of the infrastructure.

The congestion performance measures listed below consider the average of the traffic descriptive parameter used as input (i.e. speed, flow, travel time). Non-linear functions and any correlation between variables requires distributional information in order to avoid bias when estimating the mean value of the functions (Zhao and Kockelman, 2002).

2.3.2.1 EXCESS DELAY – ExD

Transport authorities and researchers have been carrying out comprehensive surveys and travel time studies in the London area since the late sixties. Different types of surveys have been used to measure car traffic evolution. These studies have detected a significant increase in travel demand, specifically for road users. A demand management strategy implemented to deal with congestion in London was the congestion pricing. Congestion pricing has existed since February 2003. It is an extra fee for car users in central London. The result of these measures showed improvement in the traffic situation in the city centre (i.e. charged zone) but an increase in travel times in the streets immediately outside the boundary of the zone.

24
The first yearly report (TfL, 2003a) describes the key measures of congestion and the associated methodology. The network speed is defined as the ratio between the total distance travelled on the network and the total travel time in the network as shown in Eq. 1. The network speed ($S_{\text{network}}$) is usually expressed in km per hour.

$$S_{\text{network}} = \frac{\sum_{i \in N} l_i \cdot f_i}{\sum_{i \in N} t_{t_i} \cdot f_i}$$  \hspace{1cm} \text{EQ. 1}

where $N$ is the set of links considered in the network, $i$ is one of these links, $t_{t_i}$ is the travel time on link $i$, $f_i$ is the vehicle flow on link $i$, and $l_i$ is the length of the link $i$. Studies of the value of the network speed in limited conditions have recognized the numerator as the traffic throughput. When the speed tends to zero, the numerator tends to capacity and remains finite and non-zero, but the denominator tends to “infinity” (Evans, 1992b). The reciprocal of the network speed, i.e. the travel rate, is preferred. The operational definition for travel rate ($TR$) then becomes:

$$TR = \frac{1}{S_{\text{network}}} = \frac{\sum_{i \in N} t_{t_i} \cdot f_i}{\sum_{i \in N} l_i \cdot f_i}$$  \hspace{1cm} \text{EQ. 2}

The definition of congestion for Transport for London (TfL) is the “difference between the average network travel rate and the uncongested network travel rate in minutes per kilometre; i.e. the delay, ‘lost travel time’ or ‘excess travel rate’”. This indicator is similar to the “urban congestion indicator” used in Australia\(^6\). The Excess delay ($ExD$) is then formulated as shown in Eq. 3:

$$ExD = TR_{\text{observed}} - TR_{\text{uncongested}}$$  \hspace{1cm} \text{EQ. 3}

The uncongested network travel rate is estimated using data collected when the traffic flows are very light. In urban areas the typical value is 1.5 minutes per kilometre.

Figure 4 shows the general relationship between the experienced travel conditions and traffic levels. The experienced conditions consider two sections: “uncongested travel rate” (in light grey) and “average congestion” (in dark grey). The TfL-congestion definition in Eq. 3 corresponds to the dark grey. The Excess delay can be then defined as shown in Eq. 4:

$$ExD = \left(\frac{\sum_{i \in N} t_{t_i} \cdot f_i}{\sum_{i \in N} l_i \cdot f_i}\right)_{\text{experienced}} - \left(\frac{\sum_{i \in N} t_{t_i} \cdot f_i}{\sum_{i \in N} l_i \cdot f_i}\right)_{\text{uncongested}}$$  \hspace{1cm} \text{EQ. 4}

where $N$ is the set of links considered in the network, $i$ is one of these links, $t_{t_i}$ is the travel time on link $i$, $f_i$ is the vehicle flow on link $i$, and $l_i$ is the length of the link $i$.

\(^6\) http://www.algin.net/austroads/ (accessed 11\(^{th}\) February 2009)
The estimation of the Congestion Performance Measure at TfL was outsourced to the traffic consultant MVA\(^7\). The following section is a summary of the approach (Taylor, 2005).

The calculation of the congestion performance measure requires flow and travel time information. Six time periods were considered (before morning peak, morning peak, after morning peak, before afternoon peak, afternoon peak, and after afternoon peak). The flow values were previously obtained from simulation models. Flow counting information was gathered in correspondence with the evaluation of the scheme, but the coverage of this counting does not allow the estimation of the flow for the whole evaluated network. The objective of the counting is to detect increments in the flow and is not to be used for congestion estimation.

The travel time is collected using a survey car that moves throughout the year. The initial plan considered reporting congestion levels every two months. For every report, each link was surveyed at least twice for every time period. An automatic plate recognition system (ANPR) was also used to extract travel times. The travel time estimation from this system was not reliable. The main cause was vehicles that stopped between two points.

2.3.2.2 TRAVEL TIME INDEX – TTI

The Texas Transportation Institute conducted the Urban Mobility Study (Schrank and Lomax, 2005) based on data over the previous 20 years. The current measurements cover around 50 cities in the US-country. (The total number of cars is calculated using detectors. This data is transferred to the Federal Highway Administration that maintains the highway performance monitoring system (HPMS). The Travel Time Index (TTI) aims at measuring congestion levels, delays, and the resulting social cost.

\[ TTI = \frac{tt_{\text{peak}}}{tt_{\text{freeflow}}} \]  

EQ. 5

Where \(tt_{\text{peak}}\) corresponds to the peak travel time, and \(tt_{\text{freeflow}}\) to the free flow travel time.

\(^7\) http://www.mvaconsultancy.com
Peak travel time is estimated according to the Travel Time Data Collection Handbook or based on the speeds measured directly on the field or using empirical/model-based flow-road design relationships (Turner et al., 1998). Free flow travel time is estimated from empirical data of related studies, signalised speed or signal timing parameters on interrupted-flow roadways. This measurement does not deal with travel time variability.

The proposed aggregation of this measure uses Vehicle-kilometres travelled (VKT) as shown in Eq. 6:

\[
TTI_{\text{Area}} = \frac{\sum_{i \in N} VKT_i \cdot TTI_i}{\sum_{i \in N} VKT_i}
\]

where \(TTI_i\) is the travel time index for link \(i\). Vehicle-kilometres travelled for a road link \(i\), i.e. \(VKT_i\) is defined as \(VKT_i = \sum_{l \in N} f_i \cdot l_i\) where \(N\) is the set of the links in the analysed network, \(f_i\) is the flow on link \(i\), and \(l_i\) is the length of link \(i\).

2.3.2.3 Relative Speed Reduction – RSR

The Swedish Road Administration (SRA, 1999) has used “Relative speed reduction (RSR)” for describing the level of congestion and traffic performance in urban road networks. The objectives of the report were:

- To create a methodology and to identify a congestion performance measure for describing and analysing road congestion, considering as a basis the existing models and the need for strategic transport planning.
- Create a GIS database for estimating conditions and analysing deficiencies in Sweden’s largest cities.

\(RSR\) is best suited for road links without major intersections. It is only sensitive to speed variations.

The relative speed reduction was originally defined at the link level and is formulated as shown in Eq. 7:

\[
RSR = \frac{s_{\text{freeflow}} - s_{\text{obs}}}{s_{\text{freeflow}}}
\]

Where \(s_{\text{freeflow}}\) is the free flow speed (measured or posted) and \(s_{\text{obs}}\) corresponds to the observed or measured speed during the peak traffic period. The free flow speed is used instead of recommended speed because it reflects the possible speed better than the recommended one. This performance measure presents difficulties in its application in downtown areas or zones with a significant number of intersections. It is sensitive to speed variations, but does not depend on the flow directly. It does not deal with travel time variability. The aggregation at the network level is not obvious.

2.3.2.4 Mean Journey Speed – MJS

---

8 Vehicle-kilometres travelled for a road link \(i\) is defined as \(VKT_i = \sum_{l \in N} f_i \cdot l_i\) where \(N\) is the set of the links in the analysed network, \(f_i\) is the flow on link \(i\), and \(l_i\) is the length of link \(i\).
The Swedish government has set as a goal that speed should not decrease. The SRA has therefore decided to use “Mean Journey Speed” measured during the morning peak hour as its guiding congestion performance measure. The SRA intends to use this measure to describe the traffic performance on arterial streets, considering only the peak traffic direction.

2.3.2.5 PERCENTAGE EXTENSION IN TRAVEL-TIME – PET

The Environmental Office in Stockholm – MAK (Swedish: Miljöavgiftskansliet) was in charge of the evaluation of the congestion charging trial in Stockholm. The proposed congestion performance measure was the average increase in travel times in comparison with travel times when no congestion exists (MAK, 2006). They considered the non-congested travel time to be the travel time during nighttime from 02:00 to 04:00.

PET is formulated at the link level as shown in Eq. 8:

\[ PET_i = \frac{t_{ti}}{t_{ti}^0} - 1 \]

EQ. 8

Where \( t_{ti} \) is the observed travel time on link \( i \) and \( t_{ti}^0 \) is travel time during nighttime. A zero value means free flow; while 100% means that travel time is twice the free flow travel time. The aggregation method is not operationally defined.

2.3.2.6 QUEUES

The bottleneck approach is related to the demand exceeding capacity at a given location (TRB, 2000). However, demand cannot be empirically observed when it exceeds capacity. A field observable traffic measure is queue or dense flow. Earlier studies have considered queues. A shortcoming with this approach is the definition of the queuing threshold because it depends on the traffic facility under analysis (i.e. highways might show moving jams while intersections show totally stopped vehicles). Also, the estimation of queues presents data collection difficulties and no agreement exists regarding the aggregation methodology (Morán and Bang, 2010). Empirical studies have used the two-fluid modelling approach for estimating area-wide parameters (Herman and Prigogine, 1979). These efforts have used empirical data collected and have provided high explanatory value (Amini et al., 1998). However, the assumptions of two-fluid modelling and the required data collection restrict the applicability. There is no monitoring system currently using this modelling approach.

The comparison of congestion performance measures from the bottleneck and travel-time approaches using empirical data has been done but only applied to a single road segment (Bang, 2006). For area-wide analysis, no analysis has been carried out comparing both approaches, neither empirically, nor using microsimulation.

2.4 APPROACHES FOR MODELLING THE EFFECTS OF CONGESTION

Modelling congestion aims at estimating the necessary parameters that are the input for the MOEs, i.e. the congestion performance measures as shown in Figure 1. The literature presents several methods for estimating...
the effects of congestion. These methods have used traffic relationships, queuing theory, as well as computer tools. These efforts can be grouped in four categories.

2.4.1 **Deterministic Queuing Approaches**

These are the earliest developed approaches and they mainly focus on the analysis of single links. Morán and Bång (2006) refer to a model for modelling the congestion phenomena. Unfortunately, its strong assumptions of deterministic arrivals and deterministic discharge rates make it unrealistic for most cases in urban road networks.

2.4.2 **Stochastic Queuing Approaches**

The analysis mentioned above can also be carried out by considering stochastic behaviour. At a link level, the impact of stochastic arrival and discharge rates can be estimated using M/M/1 or M/G/1 queuing theory. Queuing networks can then be used for modelling networks (Jackson, 1957). Queuing network models utilize a performance index that allows comparing the overall performance of the network. Nevertheless, the main obstacle for the use of this modelling method in real applications is the estimation of the demand as it cannot be observed in the field when congestion occurs. Furthermore, queuing networks, under general regime service conditions, also require the estimation of the residual service times in all the servers of the queuing network, which in turn requires the assumption of a Poisson distribution for the arrivals (Asmussen, 1993).

2.4.3 **Speed-Flow Relationships**

The speed-flow relationship was originally devised at a link level. These relationships are functions that describe the behaviour of two variables for a cross section. They represent the fundamental relationship between flow-speed-density and their estimation is based on empirical studies. They have been widely used to estimate the time savings or speed improvements on highway investments. The estimation of the effects of policies on congestion have been successful but limited to expressway conditions (Li, 2002). Later studies have attempted to describe the traffic conditions for area networks using the Macroscopic Fundamental Diagram (MFD). Daganzo & Geroliminis (2008) have considered a reformulation of the traffic parameters originally defined at link level in order to describe conditions at an area level. Through this, it has been possible to identify a relationship between flow (for an area), speed (for an area) and traffic density (also for an area) similar to the ones existing at link level.

A shortcoming of network analysis is that relationships at the link level do not consider stops at the intersections. Also, the curves are based on the assumption that queue levels at the beginning of the study period are zero. The study period needs have homogeneous flow levels. Cities where the road networks are limited by geography or highly populated cities are examples where long AM-peaks and PM-peaks can be observed. The dynamics of road traffic under these long congestion periods is not captured by the speed-flow relationships. Furthermore, speed-flow link-level relationships need knowledge of geometric data and cross section of the road, as well as, that these geometric characteristics are homogeneous along the units of analysis.
If the variability of street geometric designs is also considered, then the method becomes unfeasible. Additionally, the use of the MFD does not allow a direct comparison across different road networks.

Using speed-flow relationships for translating estimated values into economic terms using demand-supply analysis, for searching for optimum levels, or for benefit-cost analyses has been controversial. The literature reveals a fruitful discussion (mainly related to the congestion charging subjects) consisting in a large number of comments and rejoinders that evidences these conflicts [(Else, 1981) & (Else, 1982) versus (Nash, 1982), (De Meza and Gould, 1987) versus (Evans, 1992a); (Evans, 1992b) & (Evans, 1993) versus (Hills, 1993) (May et al., 2000),(May et al., 2001) versus (Hills, 2001); (Ohta, 2001a) & (Ohta, 2001b) versus (Verhoef, 2001)]. Among the questions that have found differing answers are: Does equilibrium exist between demand and supply? Is it unique? Is it stable? How can a social optimum be reached? How are the first and second order optimal conditions fulfilled by the potential equilibriums? Which units should be considered for measuring travel demand?


2.4.4 Simulation models

Simulation models are popular tools for the analysis of transport networks. They facilitate the evaluation of alternatives before implementation. Simulation models have a number of advantages over analytical models. They can carry out larger numbers of operations and high-speed analysis and, thus, can consider for example interactions between supply and demand at a level that was not feasible before. However they have some disadvantages. Some of the disadvantages are generic to simulation models. One of them is the relatively large computational cost compared to analytical solutions; another is the need for calibration to conditions specific to the traffic system to which they are applied. Barceló (2010) proposes three categories of simulation models grouped according their level of detail:

- Macroscopic models are based on fluid dynamics and they consider flow as a homogeneous fluid that propagates through the network. They are the most efficient when considering large-scale problems but they are unable to capture behavioural elements such as departure time choice or interaction between vehicles in the same road segment.
- Mesoscopic models use an increased level of detail considering each single vehicle (or packets of vehicles). They consider some behavioural aspects, like reaction to information, but interaction between vehicles is not specifically modelled.

---

9 A more detailed description of these discrepancies can be found in Moran (2008)
Microscopic models represent each individual vehicle and the relationship with other vehicles and road objects in their immediate surroundings. Thus, they can provide a good picture of safety aspects. However, they require large efforts for calibration and validation to cover even small areas.

2.5 CONCLUSIONS OF THE LITERATURE RESEARCH

Congestion performance measures need to fulfil several criteria in order to be useful tools in the decision-making framework. The literature research referring to this has been presented and represents the basis for the identification of the knowledge gaps in this area. These gaps can be summarized as follows:

2.5.1 SUITABLE INDICATORS FOR MONITORING PURPOSES

The problem of defining congestion has been presented and the factors that affect the definition of congestion reviewed. Two approaches for defining congestion in road networks have been considered: bottleneck approach and travel-time approach. The characteristics, advantages and disadvantages of each approach have been reviewed. A simple comparison has been made that shows how considering different approaches will indicate the existence of congestion in one case, but not in the other. Analyses of how different approaches and their corresponding suitable indicators differ are not registered in the literature and thus are needed.

This gap relates to research question i as shown in Figure 2 and is considered across all the papers.

2.5.2 CONSISTENT AGGREGATION FROM A SYSTEM PERSPECTIVE FOR A TIME DEPENDENT ANALYSIS

Earlier efforts for congestion monitoring have considered the output data from planning models that are not sensitive to the dynamics of congestion. Congestion performance measures that take into account the time-dependent aspects of congestion are required. An inventory of congestion performance measures for monitoring purposes is also presented. Some of them have been originally developed at the network level, others have been developed at the link level with associated aggregation method, while others are just presented at the link level but are used to describe a situation of a road network. A consistent adaptation of the formulation of these indicators has not been found in the literature.

This gap relates to research question ii as shown in Figure 2 and it is considered in papers B-E.

2.5.3 RELIABILITY ASPECTS OF CONGESTION PERFORMANCE MEASURES

Some literature exists related to travel time variability, however neither the variability nor the reliability of the central value estimators has been studied earlier. This might have been for two reasons: i) Limited amount of available data for estimations, and ii) earlier efforts to monitor congestion attempt to estimate indicators using existing data, instead of designing data collections efforts for the estimation of a certain measure. The reliability aspects in terms of the error levels for the estimated measured levels have not been studied. A methodology for this analysis is also missing. An analysis of which estimators are more effective in terms of sample size has not been carried out. These studies are necessary for estimating significant differences in traffic conditions requiring smaller sampling efforts.
This gap focuses mainly on research question \( iii \) as shown in Figure 2 and is addressed in Papers A and B.

2.5.4  **Variability across users and reference level**

Some of the congestion performance measures explicitly consider a reference level. This reference level has been defined in different ways usually related to free flow traffic conditions. Also, this reference level considers a single deterministic value that does not consider the variation in the preferences of the users. A network-level congestion performance measure that takes into account the users’ point of view of congestion has not been found in the literature.

This aspect focuses mainly on research question \( v \) as shown in Figure 2 and is addressed in Paper C.

2.5.5  **Relevance aspects of congestion performance measures**

Earlier studies estimating delay as a function of the flow level have been at a link level. Aggregation for road networks has considered only a static approach. Regarding emission, several efforts have recently been carried out for different levels of aggregation, i.e. link, intersections, metropolitan areas, etc (Bigazzi and Bertini, 2009b, Choi and Frey, 2010). However, the capability of the area-wide indicators to capture the impacts of congestion on pollutant emissions has not been studied.

This gap focuses mainly on research questions \( iv \) & \( vi \) as shown in Figure 2 and it addressed in Paper D and E.

2.5.6  **Congestion performance measure as useful tool for decision making**

The literature on performance-based planning presents several properties that a congestion performance measure should fulfil. These general recommendations need to be applied to the specific case of congestion. These properties have not been evaluated and these analyses are required. Special care has to be focused on the trade-offs between improvement in one property and worsening in others.

This gap relates to research question \( vii \) as shown in Figure 2 and is addressed in Papers D and E.
3.1 BACKGROUND

For operational and monitoring purposes it is important to be able to observe traffic performance of congested networks. Traffic planners are generally restricted to use performance measures obtained from macroscopic transport models. However, due to the dynamics of characteristics of urban traffic and congestion, such measures are insufficient for monitoring areas with severe congestion problems and they do not support selection of short-term solutions.

3.2 PROBLEM FORMULATION

The objective of the paper is to review and evaluate existing congestion performance measures suggested in the literature considering data collection efforts in Stockholm. The main focus is on the statistical aspects of the congestion performances measures.

3.3 THEORETICAL ASSUMPTIONS

Previous studies recognize that it is essential that performance measures be consistent with the goals and objectives of the process for which they are being used (Lomax et al., 1997). The study starts by considering different approaches for defining congestion, and the different indicators which will be more suitable for each definition.

Reliability is considered as the ability of an indicator to reproduce a similar value, if measured again. The reliability of an average trend indicator depends on the variance in the sample. If the variance (i.e. standard deviation) is smaller, then the average trend indicator will be more reliable. Thus, the variance becomes a good basis for comparison between congestion indicators.

- A first issue is the comparison of variances between indicators expressed in different units. For example, the deviation of the Excess Delay\(^{10}\) indicator is expressed in minutes-per-kilometre and the deviation of the Travel Time Index indicator is expressed as a percentage. The basis for comparison chosen is the coefficient of variation (i.e. ratio of the standard deviation and average value).
- A second issue is the estimation of the variance value for each congestion indicator. The considered data collection methods are not automatic and they often provide small sample sizes on each surveyed link and, thus, the estimation of the variance of the congestion performance measures cannot be done with usual methods. However, the variance of the travel time, speed, and flow is more readily available. The

---

\(^{10}\) When the paper was submitted to the conference, the “Excess Delay” indicator was called “congestion” as by TfL.
variance of the area-wide congestion performance measures is then estimated using the variance of the traffic parameters at the link level and the way these parameters are included in the formulation of the indicators. This method not only allows better estimation of the variance of the congestion indicators, but also takes into consideration the existing correlations between traffic parameters (for example how the travel time can be correlated with flow and between links).

3.4 Empirical Work

The study includes a literature review of the existing congestion definitions and congestion performance measures, a descriptive analysis of the traffic descriptive parameters and finally the estimation of the variance of the traffic descriptive parameters. Later, the variance of the traffic descriptive parameters is used for the estimation of the variance of the congestion performance measures.

3.5 Results

A comparison of the data needs and descriptive capabilities of the indicator is presented. Furthermore, an index of the statistical efficiency and reliability of different indicators is proposed and empirically estimated for a reduced sample of data. Later, a graphical comparison is carried out for visual identification of the indicators that perform better. The Excess Delay, Travel Time Index and Relative Speed Reduction indicators showed the highest reliability. The study recognizes the need to differentiate between the type of users and their purpose.

4.1 BACKGROUND
Congestion indicators available in the literature are mainly devised for road links. For these link level indicators, aggregation methods to a network level are seldom included in the literature. Paper A identified the need to consider the intended audience when calculating an indicator, proposed an index of statistical efficiency and reliability, and highlighted the need for considering the time dependent aspects of the dynamics of congestion.

4.2 PROBLEM FORMULATION
The objective of the study is to further develop the link-based congestion performance measures to make them suitable for an area-wide analysis. For these indicators, Paper B aims to identify and estimate the indicators’ reliability and accuracy based on large-scale traffic field data collection.

4.3 THEORETICAL ASSUMPTIONS
Paper B makes the same theoretical assumptions as in Paper A. However, in order to compare the index of the statistical efficiency and reliability among different congestion indicators, a statistical test is carried out. The statistical test recognizes the level of measurement of this index of reliability and the assumption of the standard errors, and then uses a Wilcoxon signed ranked test for finding statistically significant differences.

4.4 EMPIRICAL WORK
The whole data collection for the Stockholm Congestion Charging Trials (SCCT) before and after study is used for estimating the observations of Coefficients of Variation that are used as the index of reliability for comparison between the indicators.

First, the quality of the empirical data is analysed and a more robust methodology for estimating unbiased traffic parameters is presented that allows capturing time dependency aspects of congestion. Second, the standard error (i.e. variance) of the traffic parameters is estimated, which is later used as an input for the empirical estimation of the variance of the congestion performance measures.

Third, the final sample of empirical observations of Coefficients of Variation is obtained considering various approaches for defining road networks (i.e. variations in definition of time intervals, geographical location and road hierarchy). Various approaches for defining road networks were considered because the literature did not present an agreement on the best approach for defining road networks for monitoring purposes.

Finally, a statistical test is applied to the final sample of Coefficient of Variation in order to identify systematic differences between the indicators.
4.5 RESULTS

From the analysis of the data, strong shortcomings were found in the method for defining the reference level of congestion.

The coefficients of variation are tested pairwise. This allows constructing reliability rankings. One of the calculated rankings used a Congestion Performance Measure aggregated by vehicle-kilometres, the other by road segment length. This was done in consideration of the results in Paper A related to the target audience of the congestion indicators. The obtained rankings did not differ in the overall order, but differed slightly in the statistical significance in some cases. Among the considered indicators, the results showed that Travel Time Index had the highest reliability and Relative Speed Reduction the lowest. In the middle range of Congestion Performance Measures, Percentage Extra Travel-time showed better reliability than Excess Delay although this was not always significant.

5.1 **BACKGROUND**

Typically, congestion indicators for the travel-time approach compare traffic conditions under congestion to some reference level. The reference level is usually based on the free flow speed (TRB, 2000), nighttime traffic levels (TfL, 2003a, Morán and Bang, 2010), posted speed limits, etc. These approaches of defining the reference level, however, ignore the fact that drivers, under free flow conditions, would travel at various speeds, reflecting their own preferences (desired speed). It is therefore of interest to evaluate the impact of using different reference levels for the calculation of congestion indicators.

5.2 **PROBLEM FORMULATION**

The objective of this paper is to propose a methodology for taking into consideration, in the definition of the reference level for calculating congestion indicators, the preferences of the individual commuters and, hence, develop congestion indicators from the system users’ point of view.

5.3 **THEORETICAL ASSUMPTIONS**

Each congestion indicator, for example the Travel Time Index, depends on observed values \(x_{obs}\) and reference values \(x_{ref}\), that is \(TTI = f(x_{obs}, x_{ref})\).

Current methodologies consider as a reference value the posted speed for all the users in the network. In this case the indicator can be expressed as \(TTI_{POST} = f(x_{obs}, x_{POST})\). Paper C considers that the reference value (for example desired speed) has a known statistical distribution, \(f_{s}(s_{ref})\), among the users of the system and, thus, the indicator can be calculated considering the expected value \(TTI_{STOCH} = E[TTI] = \int_{0}^{\infty} TTI(\text{obs}, S_{ref})f_{s}(s_{ref})ds\).

5.4 **EMPIRICAL WORK**

The study consists of two parts: an analytical study and a numerical case study.

The analytical study considers the formulation of the indicators and analytically estimates the impact of considering the assumption that the reference level for drivers follows a distribution instead of being fixed for all. The results support the hypothesis that the formulation of the indicators has an impact when considering the different reference level methods.

Later, a case study using microsimulation is presented that numerically estimates the magnitude of the differences when using alternative reference levels.
5.5 Results

The analysis shows that some area-wide indicators are biased and tend to overestimate or underestimate congestion, depending on the desired speed distribution.

A case study, using microsimulation of a small congested network, was used to illustrate the impact of using a more appropriate definition for congestion indicators. The results show that the errors ranged from 1% to 4% depending on congestion levels and that the error was larger when using the posted speed than when using the mean desired speed. It is also observed that some indicators present a constant error across different levels of congestion (e.g. TTI) while others present decreasing errors with increased congestion levels.

6.1 BACKGROUND

Congestion indicators for traffic monitoring have used the output from long-term planning models which makes them insensitive to the dynamic and time-dependent aspects of congestion. Furthermore, despite the improvements and advances in recent years in the field of data collection methods and microsimulation, the indicators are still using data aggregated in time at the same level as the long-term planning tools. Also, the indicators were devised to measure the impact in travel time (i.e. delay). Recent increased awareness of the environmental impacts of traffic requires the evaluation of the ability of these indicators to capture other impacts of congestion, such as pollution.

6.2 PROBLEM FORMULATION

The ability of area-wide indicators to monitor congestion impacts has neither been studied in terms of available enhanced data capabilities nor from the environmental aspects. The objective of the study is to compare the abilities of congestion indicators to capture the impacts of traffic congestion on an area-wide level. For this purpose, existing congestion indicators are further developed in order to make them suitable for a dynamic analysis.

6.3 THEORETICAL ASSUMPTION

It is assumed that the quality of the current instantaneous emission models allows for comparison among the indicators.

6.4 EMPIRICAL WORK

The analyses existing in the literature are further developed, consistent with the time-dependent aspects of congestion, as well as the network-level aggregation. For example, an aggregation considering vehicle-kilometres travelled (suitable for administrators that know both flow and length of the links) is proposed the Relative Speed Reduction indicator. In this way the indicator becomes as shown in Eq. 9

$$RSR_{WA} = \frac{\sum_{i} V KT_i \cdot RSR_i}{\sum_{i} V KT_i} \quad \text{EQ. 9}$$

Simulations are carried out in VISSIM using an already calibrated simulation platform for small subnetworks in Stockholm with morning traffic conditions. The subnetworks were chosen because they are good examples of
dense urban centres with mixed traffic conditions. The pollutants considered were CO2, CO, NOX, HC, and Fuel Consumption.

A correlation analysis between the indicators and the environmental impacts was carried out. The socio-economic impacts are expressed in “total area-wide values divided by the vehicles-kilometres travelled” during the period of analysis to allow for a better comparison of the two areas that differ in size and traffic levels.

6.5 Results

Using both visual and statistical analysis, the indicators that, in general, performed better across different areas are \( RSR_{WA} \) for Emissions and \( ExD \) for Delay. The descriptive capabilities of these two indicators proved to be more robust to variations in percentage of heavy vehicles and other network parameters. The descriptive capabilities of this indicator will make these indicators to be robust for comparison between different networks as well as the transferability of the results.

Results from previous papers (i.e. paper A and paper B) regarding the statistical confidence of the indicators were confirmed by the present study. However, comparing indicators across several areas, the present study identified that the statistical confidence (i.e. reliability) and socio-economic relevance can be properties that act in an opposed manner.

When considering single network analysis and NOX emissions, none of the indicators performed satisfactorily.

7.1 Background

Previous papers on relevance (i.e. paper D) used a simulation platform of two areas of dense urban areas to analyse the capabilities of the indicators of capturing the impacts of congestion. “Groupings” in data for each of the simulated networks were identified. The factors that could have caused these “groupings” were the differences between the considered sites. These differences are: topology of the network, car fleet composition, and percentage of heavy vehicles.

7.2 Problem Formulation

The present study is devoted to the analysis of the factors that impact the relevance of the Congestion performance measure, that is, the factors that impact their ability to capture the impacts of congestion. A special focus is on pollutant emissions.

7.3 Theoretical Assumption

It is assumed that the quality of the current instantaneous emission models allows for comparison among the indicators.

7.4 Empirical Work

The factors that are considered are: car fleet composition and percentage of heavy vehicles. For the case of car fleet composition, the classes of vehicle considered were European Gasoil cars 1 to 4 (i.e. G1, G2, G3, G4) and European diesel cars (D4). For each of the simulated areas, four scenarios were considered: 1/4, 1/2, 3/4 and the original percentage of heavy vehicles.

7.5 Results

The study focused first on the car fleet composition and examined its impact on the relevance of the indicators. The results showed that even when the emissions varied, the indicators that were poor in relevance continued to be poor. However, attention has to be considered when applying the CPMs in areas or cities where the car fleet varies considerably (for example a city with only gasoil vehicles and cities with a certain percentage of diesel vehicles). The study focused later on the percentage of heavy vehicles. It was observed that, for the range of variation considered, this parameter did not affect the relevance of the CPMs.
8 Conclusion

During the last decades, transport agencies have endeavoured to cope with congestion problems with a wide range of actions and policies. Each transport agency has applied different congestion performance measures at a local level. Although monitoring programs have been devised to follow the development of congestion, data collection efforts at the network level have been limited. In the same way, methods using macroscopic models are not sensitive to the dynamics of congestion. This threatens consistent performance-based transport planning. The gaps in the knowledge in the area were earlier summarized in Figure 2 “Topics treated in each paper and overview of thesis” shown also below.

<table>
<thead>
<tr>
<th>Research questions</th>
<th>PAPERS</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. What are the definitions of congestion for monitoring congestion in urban road areas? What are the suitable congestion performance measures for the definition approaches?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ii. How can the congestion indicators be adapted or further developed to consider the time dependent aspects of congestion?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>iii. What are the error levels for the estimated measured levels? How can the reliability aspects of the estimation be analysed? What indicators perform better and are hence more reliable?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>iv. How can congestion performance measure capture the impacts caused by congestion in terms of delay?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v. How can the variability across users’ preferences be incorporated in area-wide congestion indicators?</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>vi. Do congestion performance measures capture the impacts caused by congestion in terms of emissions? What are the factors that affect the relevance aspects of congestion performance measures?</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>vii. How the different aspects (that good congestion performance measures should fulfill) act together?</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

A methodology that allowed a time dependent analysis (research question i & ii) was considered. This thesis contributed in two main areas: Reliability (research question iii) and Relevance (research question iv, v & vi). The overall recommendations are presented at the end of the present chapter (research question vii).

8.1 Reliability

Reliability refers to the ability of an indicator to provide statistical significance with a smaller data collection effort. This research contributed by providing a framework for analysing the reliability of congestion...
performance measures. This analysis used the data collected in correspondence with Stockholm congestion charging trials.

The reliability section identifies from a select group of existing indicators (and their further development to capture traffic dynamics), the most reliable indicator for monitoring congestion in urban road networks. The main difficulty of this section was to carry out a comparative analysis of indicators that are expressed in different units. For this purpose, the operational formulation of the measures, the standard error of the input parameters and the coefficient of variation were considered. In this way, it was possible to identify the indicator that has the smallest error levels in relation to its estimated level.

8.2 Relevance

Relevance is considered as the ability of an indicator to capture both delays and environmental impacts caused by congestion. This research contributed by providing better understanding of how congestion performance measures capture the impacts caused by congestion (e.g. delay and pollution). This analysis used a microsimulated platform and instantaneous emission models in dense urban areas.

Based also on the analysis of the approaches of defining congestion, the issue of defining a reference level as the comparison ground of congestion was identified. Considering the variation of the reference level across different road users provided estimations that are more relevant to the impacts caused by congestion. For this purpose, the analysis considered the further development of the indicators in order to include a statistical distribution for the reference level (instead of a single deterministic value). An estimation of how this impacts each congestion performance measure’s estimation in relationship to delay was presented. This allowed indicators to better consider the preferences of individual drivers.

Also, using the simulated data, it was possible to identify the indicators that best captured the impacts of congestion regarding pollutant emissions. In the same way, the factors that affected the abilities of capturing congestion impacts were analysed. The percentage of heavy vehicles was a factor that did not affect of the indicators’ abilities, i.e. relevant indicators remaining relevant and poor relevant indicators remained as poor relevant indicators. Variations in the car fleet show to impact in the descriptive capabilities. As a result, it is not recommended to compare areas where car fleets are significantly different.

8.3 Overall Recommendations

Several aspects to take into account when evaluating congestion performance measures have been described, for example: time-dependent aspects of congestion, preferences of the users, ability to capture environmental impacts, statistical efficiency of the indicators in terms of required sample size, correlations between the input parameters required for the calculation of the congestion performance measures, etc. Given the diversity of the aspects analysed, it is likely that one indicator will best perform in one aspect, and another indicator in another aspect. In order to provide a recommendation about the most suitable indicator, it will be necessary to know some of these aspects.
This study aimed to analyse the trade-offs between the desirable aspects that a good congestion indicator should fulfil. This research showed that there was evidence that relevance and reliability act in an opposite direction. This means that “reliable” indicators like $R_{SL}$ and $TTI_{WA}$, that show statistical significance even with small sample sizes, can be poorly correlated with some of the congestion impacts. On the other hand, “relevant” indicators like $R_{SRWA}$ and $ExD$, that capture the impacts of congestion, might require larger sample sizes to have statistical significance.

However, in order to provide an overall general recommendation that considers a balance between all evaluation dimensions, a decision framework is necessary. Resnik (1987) refers to the Maximin rule in the framework for a decision under ignorance. This rule will compare the minimum utilities (i.e. disadvantages) provided by the selection of each indicator and will choose the indicator whose minimum is the maximum for all the minimums. Applying this framework requires answering the following questions:

a. What are the disadvantages of using a very reliable indicator, if it has low relevance (i.e. a measure that is estimated with statistical confidence requiring small sample sizes, however, it poorly correlate with the impacts of congestion)?

b. What are the disadvantages of using a very relevant indicator, if it has low reliability (i.e. a measure that varies as well as the impact varies, however, it requires large data collection efforts to show a statistically significant difference)?

Given that the purpose of the congestion performance measures for monitoring purpose is to provide information to decision makers or road users, then the disadvantages of a. are larger than vice versa. This means that the minimum minimum-utility occurs when choosing a reliable indicator with low relevance. In the same way, the maximum minimum-utility occurs when choosing a relevant indicator with low reliability. Furthermore, the Maximin rule is applied to the preferences or disadvantages at a single moment in time. Considering the latest improvements in data infrastructure and developments in traffic data collection methods, the availability and quality of data has dramatically increased in recent years and this trend is expected to continue. Thus, it is expected that the disadvantages of considering a less reliable indicator will diminish over time.

For readers interested in having a unique answer to a problem that has no unique answer, it can be pointed out that the Relative Speed reduction weighted using vehicles-kilometres (i.e. $R_{WA}$) is the preferred indicator.
9 Future research

The long-term goal of this research is to contribute to the improvement of performance-based planning. For this purpose, two important aspects (i.e. reliability and relevance), that had not been previously analysed, could be examined in detail thanks to the exceptional data collection in conjunction with the Stockholm Congestion Charging Trials. The emergence of GPS and other ubiquitous technologies is improving the possibilities of collecting congestion data in dense urban areas as well as for empirical estimations of the congestion situation.

Another issue of increasing interest is the variability in travel time. Considering also the improvements in data collection technology and availability of data, it could become feasible to incorporate the variability around the central value as one of the aspects to evaluate the congestion performance measures considered in this study.

The present study considered the impact of space correlations among the traffic descriptive parameters and considered links for the reliability study. Another interesting dimension to consider is the time dimension. In this way, time correlation analysis can improve the reliability analysis of the congestion indicators. Unfortunately the quality of the SCCT data did not allow this in great extent. Future studies should aim to further analyse these correlations. The new data collection techniques mentioned above can help in this regard. Based on improved empirical data, it will be useful to identify under which conditions the effects in the time dimensions are significant and should be considered.

The sustainability aspects covered in the present study were connected to the emissions caused by vehicles that suffer congestion. A holistic view of the transportation system should also consider the impacts that congestion has on other road users. For example, studying how residents perceive congestion in the neighbourhoods. It might be expected that in this situation, queue indicators will be a better measure of the congestion impacts than travel-time-based indicators. Another example can be to consider other sustainability aspects, i.e. including the efficiency that other transport modes have together with cars. This will provide a better descriptor of the performance of the transport system. Further research should be focused on efforts to integrate these aspects.

It was earlier argued that congestion performance measures have to be defensible. This becomes important when communicating the concept to non-experts. The acceptance of a congestion performance measure by politicians, for example, is important. Furthermore, studies on the information in the form of congestion performance measures that travellers can use need to be undertaken.
10 REFERENCES


Phone: +81-3-3501-7761

Fax: +81-3-3501-7762

http://www.jste.or.jp/.


ISSN: 0097-8515.


