The costs of congestion reappraised
February 2013

Ian Wallis and David Lupton
Ian Wallis Associates Ltd

NZ Transport Agency research report 489

ISBN 978-0-478-39451-1 (electronic)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Wallis, IP and DR Lupton (2013) Costs of congestion reappraised. *NZ Transport Agency research report 489*. 65pp.

Ian Wallis Associates Ltd, L5, 2 Woodward Street, Wellington 6011
Telephone 64 4 472 2354, ian@ianwallis.org

Keywords: (costs) of congestion, Auckland, New Zealand, speed-flow functions, (traffic) congestion

An important note for the reader

Erratum

24 May 2013

Page 57 – Table A.1: Summary of STCC estimates of Auckland recurrent congestion costs

Interpeak – Total annual costs ($ million) figure corrected to read 381.2

Acknowledgements

We acknowledge with gratitude the support of the NZ Transport Agency (NZTA), the funder of this project.

In addition, we would particularly like to thank the following organisations and individuals who made substantial contributions to shaping the project and ensuring that its outputs will be of maximum value to end users:

- Members of the project steering group, nominated to monitor project progress and to provide advice to the consultants as appropriate: Sandy Fong (coordinator), Ernie Albuquerque (NZTA), Joanne Leung (MoT), Ingrid van Elst (NZ Treasury), John Davies, Nik Vorster (Auckland Transport)
- Jojo Valero of Auckland Council, who undertook runs of the Auckland Transport Model (ART3) to the specification of the consultants
- Ron Allan who provided additional advice on the theoretical aspects
- Peer reviewers, who provided independent appraisal of the draft report and whose comments have been taken into account in preparing this final research report: Ken Gwilliam (Visiting Professor, Institute for Transport Studies, University of Leeds) and Eric Verhoef (Professor of Spatial Economics, VU University of Amsterdam).

Abbreviations and acronyms

ARC	Auckland Regional Council (subsumed into Auckland Council 1 November 2010)
ART3	Auckland Regional Transport Model
ADT	average daily traffic
BTCE	Bureau of Transport and Communications Economics (Canberra)
BTRE	Bureau of Transport and Regional Economics (Canberra)
CAS	Crash Analysis System (NZTA)
DfT	Department for Transport (UK)
DWL	deadweight loss
EEM	*Economic evaluation manual* (NZTA 2010)
EMME/2	transport model suite (tradename)
FTE	full-time equivalent (employees)
GDP	gross domestic product
GHG	greenhouse gas (emissions)
HPMS	Highway Performance Monitoring System
LoS	level of service
LRMC	long-run marginal costs
NZTA	New Zealand Transport Agency
pa	per annum
pcu	passenger car unit
SRMC	short-run marginal costs
STCC	*Surface transport costs and charges study* (Booz Allen Hamilton 2004)
TTI	Texas Transportation Institute
UMR	Urban mobility report (produced annually by the Texas Transportation Institute)
v/c	volume/capacity
VKT	vehicle kilometres travelled
VOC	vehicle operating costs

Contents

Executive summary

Objective

The case for providing additional road capacity and public transport alternatives in major cities and metropolitan areas is based in part on the high perceived costs to society of congested roads. In the case of Auckland, some estimates put road congestion costs over $1 billion a year. The objective of this research project was to develop an improved approach to assessing the costs of urban traffic congestion and to make corresponding estimates of the 'costs of congestion' in Auckland.

This objective was addressed by reviewing previous studies and theory and devising a measure of the cost of congestion that is meaningful and quantifiable, and which could be applied in the context of planning (comparing transport strategies) or monitoring (measuring achievements on the ground).

Concepts and definitions – congestion and the costs of congestion

Road vehicles interact with one another. Adding an extra vehicle to a traffic stream has a tiny but quantifiable effect on other vehicles, which are slowed by a small amount. When traffic is sparse the interaction is minimal and the effect of the extra vehicle is to increase the total traffic flow. As traffic flow increases, travel speed reduces. The flow cannot keep increasing as speed keeps falling because at zero speed there is zero flow. Hence there must be a point, call it 'capacity', at which the flow rate is a maximum. At capacity, the marginal vehicle slows the traffic stream by an amount that leaves the total flow rate unchanged. If more vehicles try to use the road than it has capacity to carry, the total flow actually reduces, queues start forming either on the road – traffic slows to a crawl – or off the road at motorway on-ramps – or at home as people stagger (ie defer or advance) their start times.

At what stage should we call this 'congestion'? Three possible ways of defining *congestion* are identified in the literature:

1 Economists define congestion as the presence of interactions between vehicles on the road. In effect all roads that carry significant traffic volumes are congested.

2 Users perceive roads to be congested when speeds fall below an acceptable level, which may differ by location and over time. We call this 'perceived congestion'.

3 Engineers classify a road as congested when more vehicles are attempting to use the road than it has capacity to carry.

We believe the engineering definition provides a clearly defined measure that best encapsulates what people understand by congestion. Our proposed definition of **congestion** is thus:

Congestion occurs when the demand for the road exceeds its capacity.

Commonly used measures of the *costs of congestion* reflect these definitions. The most commonly quoted measure compares the actual travel time with the free-flow time. This is consistent with the economists' definition of congestion even though attempting to provide free-flow conditions would be uneconomical, if not impossible. The concept most favoured in the economic literature is the deadweight loss (DWL) from congestion. It is also based on the economists' definition of congestion, but compares the actual travel cost with the travel cost that would result if road users were charged the marginal social cost of their travel. As normally calculated, the DWL assumes the road is operating at less than capacity and thus that an optimal toll results in a loss of benefit for those who are 'tolled off'.

Calculating the cost of congestion by comparing observed speeds with an arbitrary benchmark speed is consistent with perceived congestion. It has the advantage that the benchmark reflects people's perceptions and that elimination of congestion could be achievable, but its failing as an objective measure is the arbitrariness of the benchmark itself.

If we adopt the engineering definition of congestion, the **cost of congestion** is logically based on the difference between observed conditions and the network operating at capacity. Our proposed definition is:

> **The cost of congestion is the difference between the observed cost of travel and the cost of travel when the road is operating at capacity**.

Observed delays are only part of the true cost of congestion. People change their schedules or mode of transport to avoid delays and make extra time allowances in case of unexpected delays. We therefore include schedule delay costs, reliability costs and other applicable social and environmental costs in estimating the total cost of congestion.

The proposed measure of the cost of congestion meets the objectives of the study. It has practical meaning and elimination of congestion is achievable: if roads are managed so they operate close to capacity, the cost of congestion will be zero. This measure can also be supported on efficiency grounds: it compares the current cost with the cost if the network were operating at maximum efficiency.

Past applications

Most attempts to calculate the cost of congestion have been based on the 'free-flow' definition. While everyone agrees that free-flow is impractical as an objective – it requires roads to be virtually empty, which would be impossibly wasteful and expensive – it has the advantage of being easily measurable. Previous estimates made for Auckland – which gave rise to the widely quoted $1 billion a year figure – involved comparisons with free-flow conditions.

The Texas Transportation Institute publishes estimates of the costs of congestion based on the free-flow approach for over 100 larger American cities each year, and the data can be used to compare cities and study trends over time. Their free-flow estimates indicate that costs generally increase with the size of the city, but in almost all cases are less than $US1000 per head per year. It appears that cities with comprehensive rail systems have lower costs per head, although this could be attributable to other factors.

Estimates have been made for some Canadian cities based on the perceived congestion approach. The costs of congestion vary depending on the average speed that is chosen as 'acceptable'.

Where the DWL has been calculated (Paris, London and the main Australian cities) the estimated cost is significantly less than the free flow estimate: by a factor of 10 for London and Paris and by a factor of two for the Australian cities.

Estimating the costs of congestion for Auckland

The proposed method

Applying the proposed measure of the cost of congestion requires the ability to estimate travel times and costs in the 'congested' and 'uncongested' states. For our Auckland analyses, the Auckland Regional Transport model (ART3) was used to provide this data. ART3 represents the traffic in Auckland using an EMME/2 transport network that was calibrated to closely replicate travel patterns observed in a major transport survey undertaken in 2006. All estimates are based on the 2006 network and 2006 levels of demand. Prices have, however, been adjusted to 2010.

The Auckland models were used to estimate three sets of travel times, as follows:

- The existing or 'congested' situation was represented by a morning peak 'base year' run of ART3.
- For 'free-flow', all links were set to their free-flow speed, which is based on the link type.
- The 'uncongested' travel times were calculated by setting the volumes on the links to the lesser of (i) the observed link volume and (ii) the capacity of the link, then running the ART3 model to calculate the resulting travel times.

Key network-wide statistics for each case are shown in table ES 1. Note that the total number of trips is the same in each of the three cases (454,000 trips) and the vehicle-kilometres are almost the same (6.4–6.6 million)[1]. Travel times and speeds reflect the three situations above.

Table ES 1 ART3 modelling results (2006 AM peak) 'free flow', 'congested' and 'uncongested' cases

Measure	Free flow	Congested (existing)	Uncongested (capacity)
Total travel time (min)	5,517,975	9,029,505	8,366,584
Average speed (km/h)	69.9	44.4	45.7
Average trip time (min)	12.2	19.9	18.4

Travel time delay costs only

Comparing the travel times only, and assuming an average value of time (including an allowance for unreliability) of $26.20 per vehicle hour based on the NZTA Project Evaluation Manual gave an annual congestion cost of $766 million if the current (ie congested) time is compared with 'free flow' and $145 million if compared with flow at capacity (ie 'uncongested') flow.[2]

Schedule delay costs

The schedule delay cost of those who stagger (ie, defer or advance) their trip times were also estimated. The approach used was based on the observed delay cost for those who travel at the peak and placing that in the context of the total duration of the commuter period, including the 'shoulders' of the peak. By looking at the distribution of travel times observed on the Auckland motorways, we estimated that the schedule delay cost adds between 65% and 70% to the observed travel time delay costs. This proportion is consistent with international studies.

Other congestion-related costs

In addition to the above cost components, estimates were made of any changes in vehicle operating costs, environmental costs and crash costs to the extent that these differ with the level of congestion.

Including all congestion cost components, we concluded that the costs of congestion in Auckland are approximately **$1250 million per year** when compared with free-flow conditions, or **$250 million per year** when compared with the network operating at capacity.

Other measures of congestion

Although we do not recommend them as measures, we were able to make some estimates of the cost of congestion under the perceived congestion and DWL approaches. These fall between the free-flow estimate and the proposed capacity-based estimate. They are presented in more detail in the main report.

[1] They differ slightly because routes through the network are selected based on travel time.

[2] Except where noted, all monetary figures are given in 2010 NZ$.

Short-run and long-run costs

The provision of infrastructure is economically optimal when the short run marginal cost (SRMC), which is the cost imposed by the marginal user without adjusting capacity, is equal the long run marginal cost (LRMC), which is the cost of adjusting capacity for the marginal user. We were able to use the property of networks operating at capacity to calculate the SRMC. We estimated that the SRMC for the Auckland network in peak periods was $7.86 for an average peak trip.

The LRMC can be estimated from the costs of constructing or expanding the motorway system. Based on recent Auckland projects, we estimate that the LRMC averages about $8.35 per peak trip. Currently then, the capacity of the Auckland network is thus not far from being optimal.

Note that this analysis is at an abstract network-wide averaged level. A real network contains many links each with different demands, and demand for individual links may exceed capacity even when the overall capacity of the network is adequate. Actual investment decisions must be based on cost-benefit analysis for individual projects, which can take these and other factors into account.

Applicability to other cities

The approach to estimating the costs of congestion developed and applied to Auckland in this project should be readily applicable in other New Zealand cities, and other cities internationally, that have transport models or road traffic assignment models.

If the costs of congestion are calculated relative to free-flow conditions, then all larger New Zealand cities where there is a noticeable slowing of traffic in the peak period will have a calculable cost of congestion. It is likely to be less per head of population than the Auckland estimate. However the costs of congestion measured relative to the network operating at capacity may be too small to calculate outside Auckland and Wellington, as the number of road links where demand exceeds capacity is believed to be small.

Abstract

The purpose of this research was to develop improved approaches to assessing the costs of urban traffic congestion and to make corresponding estimates of the costs of congestion in Auckland (New Zealand).

Various definitions of congestion were reviewed and it was found that the concept of congestion is surprisingly ill-defined. A definition commonly used by economists treats all interactions between vehicles as congestion, while a common engineering definition is based on levels of service and recognises congestion only when the road is operating near or in excess of capacity. A definition of congestion based on the road capacity (ie the maximum sustainable flow) was adopted. The costs of congestion on this basis are derived from the difference between the observed travel times and estimated travel times when the road is operating at capacity.

Estimates were made of the annual costs of congestion in Auckland, based on this definition and also relative to free-flow travel conditions. These estimates covered: the travel time and reliability differences for travel in peak periods; vehicle operating cost, environmental cost and crash cost changes associated with the differences in travel speeds; and schedule delay costs associated with travellers who adjust their time of travel to avoid the congested peak periods.

1 Introduction

1.1 The project

The research project outlined in this report was undertaken as part of the NZ Transport Agency (NZTA) research programme by consultants Ian Wallis Associates Ltd (Ian Wallis and David Lupton) in 2010–11.

Its overall objective was:

> *to develop improved approaches to assessing the costs of urban traffic congestion and to make corresponding estimates of the 'costs of congestion' in Auckland.*

1.2 Project background and context

Transport system (principally road traffic) congestion is perceived by Auckland residents and businesses as the single largest problem with living and doing business in Auckland, according to a number of authoritative market surveys. The 'costs of congestion' are often a headline-grabbing item in public debates on the congestion issue, and are often used by those advocating the provision of additional road capacity and/or some form of 'congestion charging'.

For the last 15 years, the costs of congestion in Auckland have been widely quoted as around $1 billion per annum. This figure is based primarily on a 1997 (Ernst & Young) study, which drew on analyses of the situation in 1991, with the figures adjusted and broadly updated for traffic growth and inflation.

The Ernst & Young estimates are thus based on analyses relating to the situation more than 20 years ago. This passage of time is sufficient, on its own, to warrant re-examination and updating of the original work. In addition, several other factors would support review of the original study and its interpretation:

- The methodology estimated the costs of congestion related to a free-flow situation throughout the day, although in practice no city would attempt to achieve such an outcome. Thus, in practice any reduction in the costs of congestion in Auckland resulting from economically-worthwhile improvements to the transport system would be very much less than the $1 billion per annum (pa) figure.
- Further, some commentators assume that, if Auckland's transport system were improved, then the reduction in the costs of congestion would translate directly into productivity gains and a corresponding increase in gross domestic product (GDP). We consider this assumption is invalid.
- Apart from these issues of interpretation, and the fact that the costs of congestion estimate is based on 1991 data, the methodology used in the Ernst & Young report appears to contain some significant flaws.

Given this context, this research project was commissioned to address the objective set out above.

1.3 Overview of research approach

Road vehicles interact with one another. Adding an extra vehicle to a traffic stream has a tiny but quantifiable effect on other vehicles, which are slowed by a small amount. When traffic is sparse the interaction is minimal: road use is essentially non-rival, having many of the characteristics of a public good[3]. As traffic flow increases, travel speed diminishes, initially slowly. As traffic becomes heavier, speed falls at an increasing rate. The flow cannot keep increasing as speed keeps falling because at zero speed there is zero flow. Hence there must be a point, call it 'capacity', at which the flow rate is a maximum. At capacity, the marginal vehicle slows the traffic stream by an amount that leaves the flow rate unchanged; use by one more vehicle exactly displaces another, giving road use the characteristics of a private good.

If more vehicles try to use the road than it has capacity to carry, queues start forming either on the road – traffic slows to a crawl – or off the road at motorway on-ramps or even at home as people deliberately stagger (ie defer or advance) their start times.

At what stage should we call this 'congestion'? Surveys undertaken by the Department for Transport in Great Britain[4] revealed a wide spectrum of views. The two extremes were:

> *Traffic is congested if there are so many vehicles that each one travels slower than it would do if the other vehicles weren't there.*
>
> *Traffic is congested if there are so many vehicles that they are brought to a standstill or can only crawl along.*

One aim of the project was to devise a measure of congestion cost that is meaningful and quantifiable, and reflects the impact of interventions, which may be policy changes or investments. The measure should be applicable in the context of planning (comparing transport strategies) or monitoring (measuring achievements on the ground). The interventions themselves were not considered in the context of this research: what is important is that any proposed measure of congestion costs is responsive to a wide range of interventions.

The measure of congestion should relate to both the short and long term, as the time-scale will then fit with the distinction made between the short and long term in transport. The short term is about managing current available capacity: the measure should be responsive to traffic management, public transport services and fares, and infrastructure pricing. The long term is about making investment decisions about additional capacity and analysing and evaluating possible impacts from such decisions on the level of congestion. The long-term issue that needs to be addressed is whether the network is optimal and the extent to which the cost of congestion is a result of the network rather than how it is managed.

1.4 Report structure

The rest of this report is structured as follows:

Chapter 2 examines concepts and definitions of congestion and the costs of congestion, leading to the preferred definitions for use in the project.

Chapter 3 reviews estimates, and the basis for these estimates, of the costs of congestion made for various cities internationally.

Chapter 4 applies our preferred methodology to derive estimates of the costs of congestion in Auckland, based first on the current network. It also addresses the long-term question on whether the Auckland network is the optimal size.

---

[3] Public goods are goods that one individual can consume without reducing their availability for others (ie consumption is non-rival) and where the cost of preventing individuals from consuming the goods is high.

[4] As quoted in Grant-Muller and Laird (2007).

Chapter 5 draws out the report's conclusions and implications and comments on the scope for refinement of the project methodology and its possible extension to other cities.

Chapter 6 lists the main references used in the project.

Appendix A reviews previous estimates of the costs of congestion in Auckland.

Appendix B derives unit economic benefit parameter values required as inputs to the estimation of the economic 'costs of congestion'.

Appendix C reviews recent road projects in Auckland and draws conclusions on the long-run marginal costs of expanding capacity.

2 Congestion concepts

2.1 What is congestion?

Congestion occurs as a result of interaction between vehicles on the road. The effect of that interaction is expressed in the fundamental law of traffic – the speed-flow curve (TRB 2011). This was described by Goodwin (1997) in his inaugural lecture at University College, London in 1997 as follows:

> *… the fundamental defining relationship of our field, the speed-flow curve … shows that the more traffic uses a road, the slower it goes, the effect becoming more and more severe as the traffic flow approaches the maximum capacity of the network, until finally overload is so extreme that all vehicles are unable to move. We may define congestion as the impedance vehicles impose on each other due to this relationship. It helps us to understand, that the underlying cause of congestion is not roadworks or taxis or accidents: it is trying to operate with traffic flows too close to the capacity of the network, when any of these transient incidents will have a disproportionate effect.*

This quote from Goodwin illustrates well the imprecision of the common definitions of congestion. While defining congestion as 'the impedance vehicles impose on one another' (implying congestion is anything less than free flow) he goes on to say that congestion occurs from 'trying to operate with traffic too close to capacity' (ie only when the effect is significant). This imprecision pervades the literature.

2.2 Types of congestion

To clarify the question 'What is congestion?' or at least to understand what answers others have reached, an extensive literature review was undertaken. The starting point for the modern study of congestion is William Vickrey, who in his 1969 paper (Vickrey 1969), identified six causes of congestion:

1. Simple interaction on homogeneous roads: where two vehicles travelling close together delay one another
2. Multiple interaction on homogeneous roads: where several vehicles interact
3. Bottlenecks: where several vehicles are trying to pass through narrowed lanes
4. 'Trigger neck' congestion: when an initial narrowing generates a line of vehicles interfering with a flow of vehicles not seeking to follow the jammed itinerary
5. Network control congestion: where traffic controls programmed for peak-hour traffic inevitably delay off-peak hour traffic
6. Congestion due to network morphology, or polymodal polymorphous congestion: where traffic congestion reflects the state of traffic on all itineraries and for all modes. The cost of intervention for a given segment of roadway increases through possible interventions on other segments of the road, due to the effect of triggered congestion.

Most of the subsequent literature concentrates on the second and third causes – that is congestion arising from interactions between multiple vehicles on a homogeneous road section or congestion due to bottlenecks. These are the types of delay modelled in the Auckland regional transport models. Trigger neck congestion is an important component of detailed intersection design models but is rarely considered explicitly in regional network models or theoretical studies.

The other causes of congestion (1, 5, 6) are seldom considered in detail. This is perhaps because they are considered inevitable and generally not susceptible to policy. Type 1 is, however, included when delay is calculated by comparison with free-flow conditions.

A Federal Highway Administration report (FHA 2005) identifies seven root causes of congestion: physical bottleneck, traffic incidents, work zones, weather, traffic control devices, special events and fluctuations in normal traffic. The report puts emphasis not only on travel time but also travel time reliability and techniques are developed to measure the cost of travel time reliability.

Grant-Muller and Laird (2007) review a number of definitions of congestion. Following the UK Department for Transport (DfT 2003) they distinguish between recurrent, non-recurrent and pre-congestion. They consider these in terms of travellers' expectations, predictability of occurrence and analytical measures. They stress the importance of the user's perceptions and note that these are affected by the driver's age, driving experience, journey purpose, the historical state of the network and other factors. Drivers in historically congested cities have a greater tolerance for traffic delays than those in areas with a recent history of relatively free-flow conditions.

We are concerned primarily with recurrent congestion. The Auckland transport models used in the case study (chapter 4) are calibrated to replicate a typical morning peak period without any allowance for unexpected events.

Grant-Muller and Laird identify three formalised definitions of recurrent congestion used by transport agencies as follows:

> *The situation when the hourly traffic demand exceeds the maximum sustainable hourly throughput of the link.* (Highways Agency et al 1997)
>
> *An interurban or peri-urban link is defined as being congested when the point average speed taken over 3 minutes is below 50% of the speed limit.* (DfT 2000)
>
> *An urban link (with a signalised exit) is defined as congested when traffic cannot exit the link within one cycle. An urban link with an unsignalised exit is defined as congested when traffic cannot exit the link within a time equivalent to one signal cycle (the cycle time equivalent was calculated by estimating what the cycle time would be if the link exit was signalised).* (DETR 2000)

They also quote two user-defined measures that are based on research by the UK DfT (2000):

> *Traffic is congested if there are so many vehicles that each one travels slower than it would do if the other vehicles weren't there.*
>
> *Traffic is congested if there are so many vehicles that they are brought to a standstill or can only crawl along.*

In other studies 'just over half of all respondents defined congestion on a motorway by a traffic jam with complete stops of five minutes. Forty-five percent considered a motorway to be congested if they had to travel at less than 20mph. Less than 20% considered the motorway to be congested if they had to travel at around 50mph'.

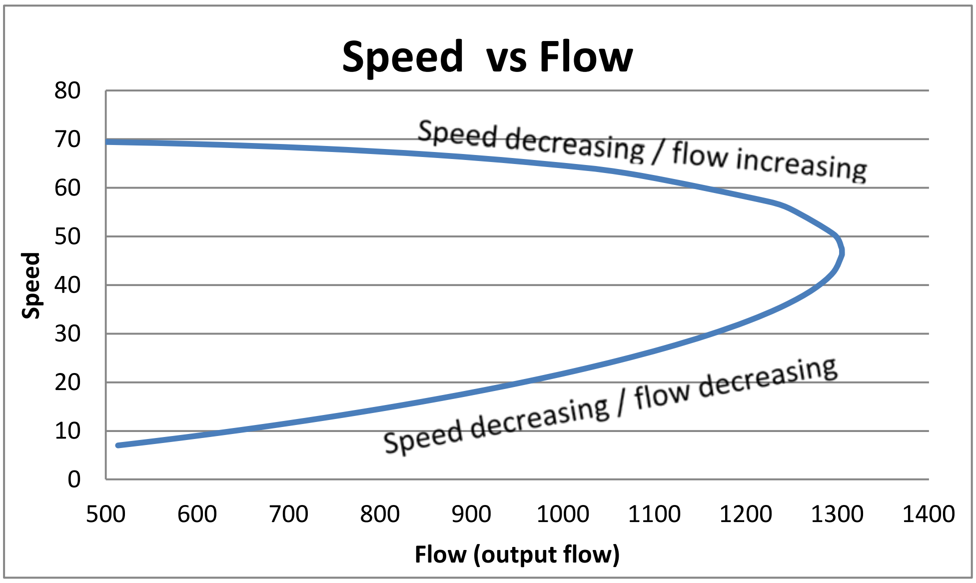
Together these definitions span the spectrum: the first of the user-defined definitions treats congestion as any deviation from free flow[5]; the transport agency definitions generally relate congestion to demand

[5] In most congestion studies (including the Auckland models) a nominal maximum speed (eg the legal speed limit) is used as the free-flow speed.

exceeding the road capacity; while the second and some of the third user definitions only count as congestion situations approaching deadlock where traffic comes to a standstill.

A further characterisation of congestion is provided by Deryche and Crews (1991). Noting that as the density of vehicles increases the flow at first increases and then decreases, they define type I congestion as the situation where speed is decreasing and traffic volume is increasing, and type II congestion as the situation when both speed and volume are decreasing.

Figure 2.1 Speed vs flow curve



Deryche and Crews state that while type I congestion can be controlled by means of regulation, taxes and balanced tolls, type II congestion calls for supplementary infrastructure (widening or doubling traffic lanes, building new roads, and so on)[6]. They also observe that most authors only deal with type I congestion.

According to Lindsey and Verhoef (1999), economists refer to the upper branch of the speed-flow curve as congested, and to the lower (type 2) branch as hyper-congested. In the engineering literature the upper branch is variously referred to as uncongested, unrestricted or free flow, and the lower branch as congested, restricted or queued. Lindsey and Verhoef consider the term queued is apposite for the hyper-congested branch in that queuing usually occurs in this state, but say congestion also occurs on the upper branch whenever speed is below the free-flow speed. For this reason, they use the economics terminology.

Truong and Hensher (2003) develop an alternative measure for the concept of traffic demand (and supply), and of ‘congestion’ in terms of traffic density rather than in terms of traffic flow. They use this to define a capacity utilisation level G_i and a utilisation index G_i/G equal to the ratio of actual traffic flow to the potential maximum flow. Truong and Hensher point out that when the traffic is below a minimum level, the road acts as a public good, whereas when it is at maximum capacity it behaves like a pure private good[7].

Verhoef (1997) points out the shortcomings of a static model. He distinguishes two families of dynamic approach: (i) the ‘bottleneck approach’, originally developed by Vickrey (1969) in which commuters using a road network all wish to be at work at the same time, but due to a limited capacity somewhere in the network (the bottleneck), this is physically impossible; and (ii) a second approach, originally proposed by

[6] A view not shared by the current authors.

[7] This point was also made by Hau (1992).

Henderson (1974) which uses the concept of 'flow congestion', and is capable of dealing with situations in which speeds decrease to still positive values due to congestion. He says that although the bottleneck approach may offer a realistic representation of congestion for some situations (eg at bridges or tunnels), it is not necessarily the best modelling approach in all cases, in particular because it is rather rigid in assuming that drivers either drive at the free-flow speed, or have a zero speed when waiting in the queue.

There are many reasons for the wide diversity in definitions of congestion. One reason is simply a matter of expectation. As Grant-Muller and Laird (2007) say, the perception of congestion depends on the driver and the historical state of the network. For this reason, measures based on user perceptions are not considered suitable. The more objective measures used by economists and engineers are preferred. There is one factor that appears to distinguish the economists from the engineers. Economists equate congestion with the existence of externalities, and therefore they see congestion from the moment vehicles start to interact.

2.3 Measures of congestion

In the literature there are two distinct approaches to measuring congestion: engineering approaches utilised by traffic engineers and transport modellers and economics-based approaches. The former approaches reflect an interest in the physical capacity of the network and the number of vehicles that can be accommodated. The latter approaches are concerned primarily with delays imposed by road users on each other.

2.3.1 Engineering definitions

Engineers identify a range of levels of service (LoS) to qualitatively describe the operating conditions of a roadway based on factors such as speed, travel time, manoeuvrability, delay and safety. The level of service of a facility is designated A to F, with A representing the best operating conditions and F the worst. There is a 'maximum flow' for each level, which is the maximum traffic that the road can accommodate and still provide that service level. Each level of service corresponds to a range of volume to capacity ratios.

Table 2.1 Level of service definitions

LoS	Description	Volume/capacity
A	Free-flow conditions with unimpeded manoeuvrability. Stopped delay at signalised intersection is minimal.	0.00 to 0.60
B	Reasonably unimpeded operations with slightly restricted manoeuvrability. Stopped delays are not bothersome.	0.61 to 0.70
C	Stable operations with somewhat more restrictions in making mid-block lane changes than LoS B. Motorists will experience appreciable tension while driving.	0.71 to 0.80
D	Approaching unstable operations where small increases in volume produce substantial increases in delay and decreases in speed.	0.81 to 0.90
E	Operations with significant intersection approach delays and low average speeds.	0.91 to 1.00
F	Extremely low speeds caused by intersection congestion, high delay and adverse signal progression.	Greater than 1.0

Source: Transportation Research Board. Highway capacity manual (TRB 1994)

Although the volume/capacity (v/c) ratios quoted in table 2.1 imply the road is not at 'capacity' until LoS E, in practice the maximum flow is at LoS D, and it could be argued that the maximum sustainable flow is actually somewhere between LoS C and LoS D. Roads with LoS D, E and F could be considered to be congested. The design capacity from an engineering perspective is equivalent to LoS C or somewhere between C and D.

In practice v/c ratios greater than 1 are rarely if ever observed and the v/c ratio is actually lower when there is congestion. Thus the ratios in table 2.1 are more properly described as demand/capacity ratios.

2.3.2 Economic concept of congestion

Whereas engineers relate the traffic to the practical capacity of the road, economists concern themselves with the optimum level of traffic. The concept of an optimal level of traffic on a network is well established in economic theory. Each vehicle on the road potentially imposes a cost on every other vehicle – called an externality because it is not taken into account in the decisions of the individual road user. Economists agree the condition for an economic optimum is that the price paid by each user should equal the marginal externality. This 'internalises the externality' so that each driver (assumed to be the decision maker) faces the cost they impose on society.

It can be shown (eg Verhoef et al 2008) that, if the charge for using each road link and the capacity of each link can be set by a regulator (or policy agency), and the objective is to set them so the net benefit to society (ie the benefits to users minus the cost of road provision) are maximised, the policy rules that should be followed are:

- the charge should equal the externality imposed by each vehicle
- the value of the travel time saving resulting from a unit increase in capacity should just equal the cost of providing that unit increase in capacity.

These are often referred to as the infrastructure pricing rule and the investment rule respectively. They are rules for ensuring an economic optimum and apply more generally than just to roads. They are rules for addressing externalities, not specifically for addressing congestion. However as noted above, most economists equate 'externality' with 'congestion' or at least assume that correcting for externalities also addresses congestion. For example Button (2004) says 'Road pricing, sometimes called congestion pricing or value pricing, aims to apportion scarce road space by market pricing rather than queuing'. For a bottleneck, the optimum price just replaces the queue by a varying toll that causes arrivals to be staggered, so it could be said that the optimum price eliminates the congestion. However the optimum price does not eliminate link congestion: link congestion remains even when the economically optimum price – equal to the externality – is charged[8]. This confusion would be removed if we agree to adopt the engineering definition and only refer to the lower branch of the speed-flow curve (ie 'hyper-congestion') as 'congestion'. It can be shown (see box) that optimal pricing does indeed eliminate hyper-congestion. As will be seen in the next section, most measures of the cost of congestion use the economists' definition of congestion.

> The optimum toll must eliminate hyper-congestion: if the toll did not, it would be possible to increase the flow through the link by increasing the toll. This would increase real benefits, so the toll could not have been optimal.
>
> Hence we conclude that the optimum toll must result in demand less than or equal to the capacity.

[8] With a bottleneck, congestion takes the form of a queue of traffic waiting to enter the bottleneck. The optimum toll just eliminates this queue – and thus the congestion. With link congestion, congestion takes the form of reduced speeds due to interactions between vehicles. The optimum toll ensures that the marginal vehicle takes the costs of interactions into account but it does not eliminate them.

2.4 Costs of congestion – approaches

2.4.1 Comparison with free flow

The most common measure of the cost of congestion estimates the cost of travel relative to a free-flow base case. This is normally represented as the travel times involved in making the trip in question in the early hours of the morning[9]. This is consistent with a definition of congestion that includes any interaction between vehicles (the economists' definition). It can be seen from the Auckland motorway examples (refer chapter 4) that true free flow only occurs where traffic volumes are very light.

While some authors reject the use of free flow as the base for calculating the cost of congestion, a study by the Australian Bureau of Transport and Communications Economics (BTCE 2000) says:

> *The definition has the merit of having as its reference point a relatively well-defined state of zero congestion. However, it is important to note that there is no implication that zero congestion is a possible or desirable goal for policy. In this respect, the 'cost of congestion' must be carefully distinguished from the 'cost of doing nothing about congestion'—a quite different notion, which is discussed later …*

However it goes on to say:

> *A definition of the cost of congestion based on the difference between current and hypothetical uncongested conditions is easy to understand, and appears to be a natural measure of the scale of the problem. Unfortunately, from the point of view of policy, it is a dead end. Eliminating congestion is not possible, and the cost of congestion, estimated in this way, provides no pointers to an improved use of the road network.*

Prud'homme (1998) describes the free-flow comparison which takes the empty road and the empty road speed as the reference situation as the **naive approach**. Although widely used, he finds it unconvincing. Roads are not built to be empty, and an empty road is not optimal or even desirable.

Newbery (1988) also considers a definition of the cost of congestion based on the extra costs involved compared with a road with zero traffic (ie free flow) but dismisses this as unrealistic as it would be uneconomic to build roads to be totally uncongested. Newbery and Santos (1999) call the comparison with free-flow conditions a misconception, stating: 'Free-flow at all times on all roads would be phenomenally inefficient and therefore it is not a benchmark for comparison'.

Koopmans and Kroes (2003) disagree. They say 'overall true free-flow operation is not an efficient situation. But why this should preclude comparing present conditions to free-flow is another matter. Measuring the size of a problem is not identical to stating that this size should or could be zero. … As in medicine, diagnosis and therapy should not be mixed up'.

2.4.2 Comparison with a defined threshold

Another approach, which Prud'homme (1998) calls the **arbitrary approach**, takes for reference a speed defined as acceptable, such as 50km/h on a non-urban road. When the speed of traffic falls below this threshold, the situation is called congested. Congestion costs are then defined as the difference between the time actually spent and the time that it would take if traffic were flowing at the acceptable speed, multiplied by the volume of traffic. He says this approach is slightly less absurd than comparison with free

[9] Other studies define free flow as all travel being at the posted speed limit, but this is better considered as an example of an 'acceptable speed' approach – discussed below.

flow, but says it is nevertheless unacceptable because it is doubly arbitrary. It is arbitrary to compare time actually spent to the time that would be spent if speed were different (and call it time lost), and it is even more arbitrary when the reference speed is itself arbitrary.

Prud'homme defines the **engineering approach** as being where the reference situation is the speed associated with the maximum flow on the road considered. When actual speed falls below this speed, the road is said to be congested. Congestion costs are again equal to the difference between the time actually spent by vehicles and the time these vehicles would spend if they were running at the 'maximum flow' speed. He has the same criticisms as for his first two definitions. He says this approach ignores the demand for the road. The definition of congestion costs must take into consideration what people are ready to pay in order to use the road, as represented by the demand curve.

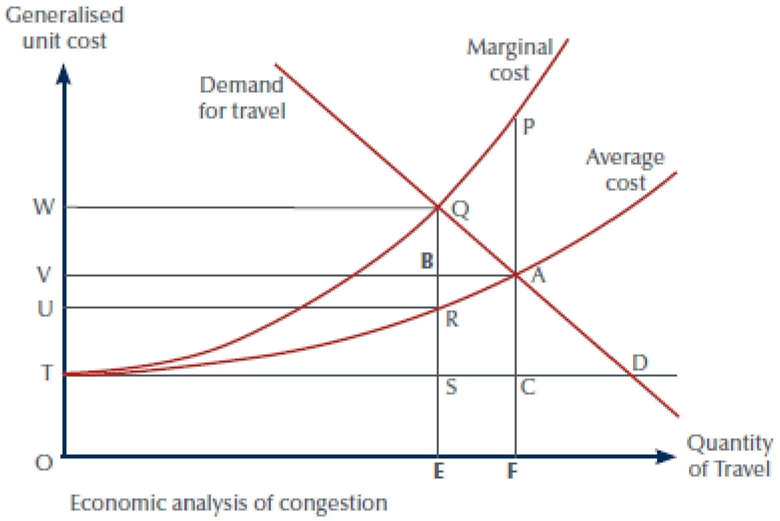
2.4.3 Approaches based on economic concepts

There are two approaches that Prud'homme calls 'economic' which take for reference the economically optimal situation. Consider figure 2.2, below. The 'equilibrium' condition is where the demand curve crosses the average cost curve, at A. At this point, the demand is OF and the generalised cost of travel is OV. The economically optimum traffic level is where the demand curve crosses the marginal social cost, at Q. The economic optimum quantity of travel is thus OE, which could be achieved by levying an internalising tax RQ. The difference between the effective level of road usage (F) and the optimal level of road usage (E) can be taken as a measure or an indicator of congestion: it is the 'excess' road usage, the amount by which road usage should be cut to take us to the economic optimum.

The proceeds of the tax required to achieve optimality is often considered as a measure of congestion costs. This is equal to the amount of the tax (RQ) times the optimal number of vehicles (E): that is UWQR. Prud'homme calls this **the half-economic** approach.

While Prud'homme says the reference situation here is meaningful – it is a desirable state of affairs – the amount of the tax is not a measure of cost. It is the proceeds of the tax required to reduce traffic to the optimal level, to the level that will maximise the economic surplus, and eliminate the congestion cost externality. To take the tax as a measure of cost is to confuse the means (the internalising tax) with the end (the elimination of the externality).

Figure 2.2 Economic concepts of congestion



Source: BTRE

Newbery and Santos (1988) also consider using the revenue that would be raised from an efficient congestion charge as the cost of congestion. However they reject it, saying this would be significant even if all roads were operating at or below capacity.

The approach Prud'homme describes as the **economic approach** takes the same reference situation as the previous approach, and produces the same indicator of congestion, the difference between E and F. But it defines congestion costs as the economic cost incurred by society when road usage is E instead of F. In terms of figure 2.2, it can be defined in two alternative, equal, ways. One is the area QPA. The other is the difference between the surplus associated with Q and the surplus associated with A which can be shown[10] to equal UVBR – BQA. Prud'homme claims that this is a true and realistic economic cost. It is the price we pay for not being at the optimum, for having F-E too many cars on the road. He defines this as the congestion cost.

In figure 2.2, the revenue that would be raised from an efficient charge would be WQRU, while the cost compared to zero traffic is VACT. Both of these are clearly significantly greater than the DWL.

Australia's Bureau of Transport and Regional Economics reaches the same conclusion as Prud'homme. It rejects free-flow as the reference point, saying (BTRE 2007) that '…essentially, there is no way to avoid all the total delay costs – since actual traffic volumes cannot travel at free-flow speeds at all times'. Instead BTRE measures the portion of total costs that could theoretically be saved if some traffic management strategy was capable of changing traffic conditions to the economically optimal level. It refers to the economic cost associated with such a change in traffic level as the deadweight loss (DWL) or 'the avoidable cost of congestion'[11]. Consider figure 2.2, where E is the efficient level of demand and F is the equilibrium level with congestion: the cost imposed by the additional traffic EF is the area EQPF while the benefit to that traffic is the area EQAF. The DWL is the difference, ie the area between the marginal cost curve and the demand curve represented by QPA. This is the same as Prud'homme's measure.

This is also Newbery's preferred measure. He defines it as the loss in social surplus associated with excessive road use. This is equal to the standard deadweight loss used by BTRE. Based on a worked example, Newbery estimates the deadweight loss is about one third of the revenue measure discussed above.

The BTRE points out:

> *A DWL value will still tend to be an upper bound for the actual social benefits achieved by any particular congestion reduction strategy – since it would take a perfectly variable management scheme, that targeted congestion by exact location and time of day (depending on the changing traffic levels on each of the network's road links), to obtain the economic optimum. However, it will be a substantially closer guide to actual obtainable benefits than a total delay cost estimate.*

There is a conceptual problem with the DWL as defined above. Note that in figure 2.2, the traffic volume is defined in terms of flow rather than demand for travel and the marginal cost is less than double the average cost. Hence the diagram as shown relates to the situation where the flow is still increasing as traffic densities increase – type I congestion as defined by Deryche and Crews (1991). The densities that lead to the lower tail on the speed/flow curve have not been reached. In this situation, using the engineering definition, there is no congestion. Figure 2.2 illustrates the deadweight loss from inefficient

[10] By using the property that the area below the marginal cost curve is the average cost, it can be shown that area QPA is equal to the area VBRU minus the area BQA. The interpretation of this is that QPA is the difference between the cost to society (the MC curve) and the benefit to the user (the demand curve) while VBRU is the benefit to those who continue to travel, and the area QBA is the loss of benefits to those who would be 'tolled off'.

[11] The term 'deadweight loss' is an analogy to the deadweight loss from taxation, a concept commonly used in economic literature. In fact the deadweight loss from the tax is QBA.

prices, not from congestion. If we had hyper-congestion, then imposing a toll would reduce the total demand but would increase the traffic flow. More vehicles would travel at their preferred travel time with an economically efficient toll than without it.

The conceptual problem arises because peak traffic is not really a 'steady state' flow phenomenon. Although it is sometimes useful to regard it as a flow, the total demand is in fact limited to a relatively short time span. If the road capacity is insufficient to meet the demand, or traffic is 'tolled off' over a particular time period

it can still be satisfied over an extended period. This time period would be shorter if the network operated at peak efficiency (ie at capacity) than if it were congested. For this reason, we propose to assume that all traffic is eventually satisfied, and that the effect of insufficient capacity – or inefficient pricing – is simply on the time taken for all vehicles to pass through the network or reach their destinations.

2.4.4 Relationship between engineering and economic measures

Prud'homme (1998) distinguishes between the economic and the engineering measures of the cost of congestion, preferring the former. Similarly, the BTRE study discussed above takes the economic optimum as the starting point for calculating the DWL. However, in practice identifying the economic optimum is not straight-forward: it appears to require full knowledge of the demand function and/or individual values of time as well as the cost function. BTRE calculates the costs of congestion relative to free flow and then uses the properties of the assumed demand and cost curves to estimate the proportion of the free-flow congestion cost that is 'avoidable congestion'. Other studies such as Fields et al (2009) use a volume: capacity ratio of 1.0 as the benchmark for avoidable congestion. This is also the measure used in the Singapore licensing scheme[12].

It can be seen (see box below) there is a close relationship between the economic and engineering definitions. Since the engineering definition that congestion occurs where the demand for travel exceeds the maximum capacity of the link(s) is relatively easy to apply, this can be used to provide a benchmark for measuring the economic cost of congestion.

> As discussed previously, the optimum toll cannot result in hyper-congestion: ie the equilibrium demand cannot exceed the road's capacity. Assume that we charge the optimum toll and that the equilibrium point is less than the capacity. Assume an upward sloping cost curve and that demand increases over time. As demand increases, the new equilibrium must have a higher toll but will also have a higher flow. But the flow cannot exceed the maximum flow. We conclude that the optimum toll results in an equilibrium demand that is equal to, or tends towards the capacity of the link as demand grows.

2.4.5 Peak spreading and other indirect costs

Koopmans and Kroes (2003) review a number of international studies and concludes 'The inclusion of costs of fuel, accidents and emissions does not seem to affect the estimates strongly, as the total of these costs, if included, is less than 20% of the time costs'. They also consider costs that result from changes in behaviour arising from congestion. Motorists can change their time of travel, their travel route, the mode of transport, their destination, or they can decide not to make the trip at all. In the longer run, people can

[12] In the Singapore licensing scheme, the toll which applies to vehicles entering the central district, is adjusted periodically to ensure the volume to capacity ratio of roads entering and within the CBD is less than 1.0 – ie that roads operate at capacity and hyper-congestion does not normally occur.

move house, firms can relocate, etc '...as these motorists (and people and firms) are not on the road (or not at their preferred time), we do not directly observe their costs'.

Koopman and Kroes consider some direct methods to estimate the costs of time losses and of other effects of congestion and find them wanting. They present an alternative way to estimate the cost of congestion using the monetary value of the consumer surplus. By quantifying the consumer surplus twice, once for free-flow traffic conditions and once for prevailing (actual) traffic conditions, the difference in consumer surplus value between those two conditions is determined. This difference, the loss in consumer surplus, is a measure of the 'cost' associated with the congested operation. Koopman and Kroes use the Dutch National Transport Model to determine values for the year 2000, together with the more traditional 'vehicle hours lost' values derived directly from the assignment method output (speed-flow effects only). This suggests the 'unobserved costs' are almost as high as the 'observed costs'.

2.5 Conclusion: proposed measures for congestion and congestion costs

2.5.1 Conclusions from the literature

To measure the costs of congestion, we first have to agree on a suitable definition of congestion. Three choices are identified in the literature: (i) most economists implicitly treat anything less than free flow as congested, so the economists' definition of congestion is the presence of interactions between vehicles on the road.-Practically all major roads are thus 'congested' which does not seem a very useful definition; (ii) users perceive roads to be congested when speeds fall below an acceptable level (which may differ by location and over time), we will call this perceived congestion; and (iii) engineers classify a road as congested when more vehicles are attempting to use the road than it has capacity to carry. The latter situation is defined by economists as hyper-congestion and corresponds to the lower branch of the speed-flow curve.

Commonly used measures for the cost of congestion reflect these definitions. Comparing the actual travel time with the free-flow time is consistent with the economists' definition even though attempting to provide free-flow conditions would be uneconomical, if not impossible, to achieve. The concept of the DWL from congestion is also based on the economists' definition of congestion, but attempts to identify the optimal level of congestion (as defined): it is the level of congestion that would result if the marginal road user was charged the social cost of their travel. The difference between the actual cost and the cost at the optimal level of congestion is called the avoidable cost or DWL from congestion. As normally calculated, the DWL assumes the road is operating in the upper branch of the speed flow curve and thus an optimal toll results in a loss of benefit for those who are 'tolled off'. While in theory the optimum can be determined as the point of intersection between the marginal social cost curve and the demand curve, calculating this point requires assumptions about the elasticity of demand and values of time that cannot be known with certainty.

Calculating the cost of congestion by comparison of observed speeds with an arbitrary benchmark speed is consistent with the user perceptions definition of congestion. It has the advantage that the benchmark can reflect people's perceptions and elimination of congestion could be achievable. Its failing as an objective measure is the arbitrariness of the benchmark itself.

While there is some debate, most theoretical papers reviewed favour the DWL measure based on the economists' definition of congestion. As will be seen in the next chapter, most practical studies calculate the cost of congestion as the difference between actual and free-flow time.

2.5.2 Proposed measures

We propose to adopt the engineering definition of congestion which is the same as the economists' definition of hyper-congestion. It can be stated as:

> *Congestion occurs when the demand for the road exceeds its capacity – ie the maximum sustainable flow.*

This definition is consistent with user perceptions, which generally do not recognise a road as congested until interactions between vehicles have a noticeable effect on speeds, but rather than being based on an arbitrary speed, it is based on the speed at the road's capacity. If we adopt this definition of congestion, the cost of congestion follows logically:

> *The cost of congestion is the difference between the observed travel time and the travel time when the road is operating at capacity – plus schedule delay costs, reliability costs and other applicable social and environmental costs.*

Since with this definition elimination of congestion is achievable, the cost of congestion measure has a practical meaning. It is a cost that can be avoided. It also can be justified on efficiency grounds: it is comparing the cost of the current network with the cost if the network was operating at maximum efficiency – ie with roads operating at their maximum capacity. Congestion can be seen as the situation where there is so much traffic that the network cannot operate efficiently.

There is a close relationship between the economic optimum – the level of demand that would result if the marginal road user was charged the social cost of their travel – and the capacity of the road (see box in section 2.4.4). This is another reason to prefer the proposed measure. It also means in practice, the proposed measure should give a similar result to the DWL for roads that are, in our terms, congested.

Because of its wider use as a measure and its use in previous Auckland studies, we will also calculate and quote the cost of congestion with respect to the free-flow situation.

A point made by the BTRE study is that even though the DWL is described as the avoidable cost of congestion, its dollar value would not necessarily be reflected in measures of national welfare such as GDP. This will also apply to our measures. While the timeliness and reliability of freight and service deliveries will impact on business productivity levels[13], a major proportion of the derived cost values refer to time savings of private individuals and private travel costs that play no part in the evaluation of GDP. Of course the benefits identified in conventional cost-benefit analysis for road construction have the same characteristics.

[13] These could, in theory, be calculated although the data available from the Auckland models with respect to freight flows and business trips is less reliable than the total trip data.

3 Congestion cost assessments: methods and estimates

This chapter reviews some previous estimates of the cost of congestion from various parts of the world. The figures come from different researchers using different definitions and are thus not strictly comparable. They nevertheless give a useful indication of the size of the problem and the implications of adopting different measures. See also Litman (2009) for a summary of various congestion cost estimates.

3.1 USA – Texas Transportation Institute

The Texas Transportation Institute (TTI) publishes an annual urban mobility report (UMR) (Schrank et al 2010) covering some 100 cities and municipalities in the USA. The procedures used in the report have been developed over a number of years and as a result of several research projects. The congestion estimates for all study years were recalculated in 2010 to provide a consistent data trend as the methodology had been amended.

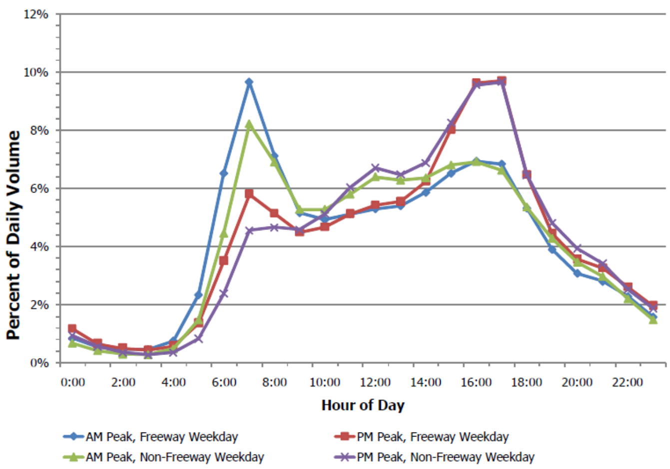
The TTI study measures congestion relative to a free-flow base case, represented by the travel times involved in making the trip in question in the early hours of the morning.

The UMR provides estimates of mobility at the area-wide level. The approach used describes congestion in consistent ways allowing for comparisons across urban areas or groups of urban areas. The calculation uses a dataset of traffic speeds from a private company (INRIX) that provides travel time information to a variety of customers; and volume and roadway inventory data from the Federal Highway Administration Highway Performance Monitoring System (HPMS).

The speed data is the annual average traffic speeds for each section of road for every hour of each day for a total of 168 day/time period cells (24 hours x 7 days). 'Real' rush hour speeds are used to estimate a range of congestion measures. Overnight speeds were used to identify the free-flow speeds that are used as a comparison. The HPMS files were used to provide vehicle and person volume estimates for the delay calculations. As the geographic referencing systems are different for the speed and volume datasets, a geographic matching process was performed to assign traffic speed data to each HPMS road section for the purposes of calculating the performance measures.

While there are some detailed traffic counts on major roads, the most widespread and consistent traffic counts available are average daily traffic (ADT) counts. The hourly traffic volumes for each section were estimated from these ADT counts using typical time-of-day traffic volume profiles developed from continuous count locations or other data sources. Figure 3.1 shows an example of the distribution profiles used.

Figure 3.1 Weekly traffic distribution profile for zero to low congestion



The calculation of congestion measures required establishing a congestion threshold, so that delay was accumulated for any time period once the speeds were lower than the congestion threshold. There has been considerable debate about the appropriate congestion thresholds, but for the purpose of the UMR methodology, the data was used to identify the speed at low volume conditions (for example, 10pm to 5am). This speed is relatively high, but varies according to the roadway design characteristics. An upper limit of 65mph was placed on the freeway free-flow speed to maintain a reasonable estimate of delay; no limit was placed on the arterial street free-flow speeds.

Most of the basic performance measures presented in the UMR were developed as part of the process of calculating travel delay – the amount of extra time spent travelling due to congestion. The INRIX speed data reflects the effects of both recurring (or usual) delay and incident delay (crashes, vehicle breakdowns, etc). The delay calculations are performed at the individual roadway section level for each hour of the week. Depending on the application, the delay can be aggregated into summaries such as weekday peak period, weekend, weekday off-peak period, etc.

The UMR calculates 14 different measures including travel speed, per vehicle and per person delays, wasted fuel, truck congestion costs, etc. The delay costs are all calculated with reference to the free flow travel speed.

$$\text{Daily Vehicle-Hours of Delay} = \left(\frac{\text{DailyVehicle-Miles of Travel}}{\text{Speed}} \right) - \left(\frac{\text{DailyVehicle-Miles of Travel}}{\text{Free-Flow Speed}} \right)$$

Estimated congestion costs for the larger American conurbations (over three million population) for 2009 are given in table 3.1. It can be seen that, with some notable exceptions, the estimated cost per person reduces with the size of the conurbation. It is also of note that, despite using free flow as the comparator, the costs per person are almost all under $1000 per year.

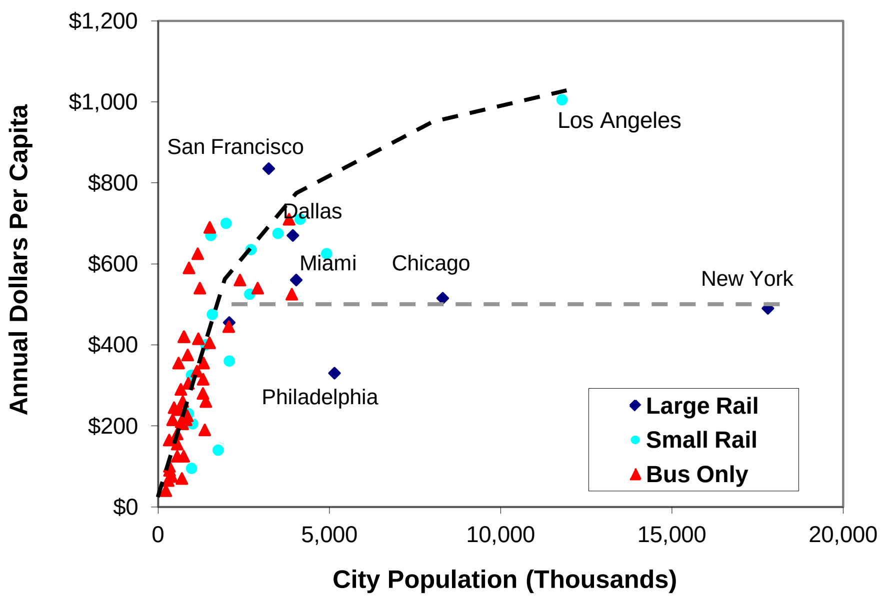
Litman (2011) plots TTI data from 2004 and notes it shows congestion costs increasing with city size, but not if cities have large, well-established rail transit systems (figure 3.2). He suggests this is the reason New York and Chicago have far less congestion than Los Angeles. Note that there appears to be a discrepancy between the Chicago figure in 2009 and the 2004 point plotted by Litman. This potentially weakens his conclusion. It is also possible that other factors such as road space per capita have a significant effect.

Table 3.1 Estimated annual congestion costs of large American conurbations, 2009

	Delay (000 hours)	Excess fuel (000 gal)	Truck congestion $million	Total congestion $m	Popu-lation (000)	Cost per person
Los Angeles-Long Beach-Santa Ana	514,955	406,587	3200	11,997	13,033	921
New York-Newark	454,443	348,326	3133	10,878	18,768	580
Chicago	372,755	276,883	3349	9476	8519	1112
Washington	180,976	148,212	945	4066	4454	913
Dallas-Fort Worth-Arlington	159,654	126,112	948	3649	5013	728
Houston	144,302	129,627	940	3403	3921	868
Philadelphia	136,429	106,000	967	3274	5337	613
Miami	140,972	109,281	883	3272	5350	612
San Francisco-Oakland	121,117	94,924	718	2791	4000	698
Atlanta	112,262	90,645	852	2727	4200	649
Boston	118,707	89,928	660	2691	4252	633
Phoenix	80,390	69,214	839	2161	3538	611
Seattle	86,549	68,703	659	2119	3187	665
Detroit	87,996	64,892	551	2032	3900	521
San Diego	71,034	60,057	450	1672	3048	549

Source: Urban mobility report (Texas Transport Institute 2010)

Figure 3.2 Congestion costs USA



Source: Litman (2011)

3.2 Paris

Section 2.4 described a number of measures of the cost of congestion. Prud'homme (1998) applied three of these to the Paris area – the economic congestion tax or DWL, the congestion tax and the comparison with free flow. His main findings are presented in table 3.2 for the year 1991. Previous estimates of the cost of congestion had put the cost at about 2% of the conurbation's GDP. Prud'homme therefore provides the GDP of Paris for comparison. While the congestion tax estimate (the tax that would be raised from an efficient congestion tax) and the free-flow based estimate are both in the range 1.5% – 2.0% of GDP, what he calls the economic congestion cost (ie the DWL) is an order of magnitude smaller.

Table 3.2 Estimates of congestion costs for Paris, 1991

	Billion francs	Percent of GDP	Per person francs (000)
Paris GDP	2,000	100	200
Economic congestion cost (DWL)	2.5	0.12	0.25
Congestion tax	33.2	1.66	3.3
Free-flow comparison	36.0	1.80	3.6

Source: Prud'homme (1998).

Note: 1000 French francs would be equivalent to about NZ$250 (US$200)

3.3 London

Prud'homme and Bocarejo (2005) use pre-charge and post-charge data for the London congestion charge scheme to estimate the DWL from congestion (which they call the economic cost of congestion) before and after the introduction of the congestion charging scheme. The costs (calculated in the same way as the Paris costs above) are shown to be a small 0.1% of the area GDP and are 90% eliminated by the charge. Charge proceeds are about three times the value of the congestion. The yearly amortisation and operation costs of the charge system appear to be significantly higher than the economic benefit produced by the system. The key statistics are summarised in table 3.3. Note that the revenue raised is a transfer cost and not considered a disbenefit, and that consequential benefits to public transport users are ignored.

Table 3.3 London congestion zone: key findings

	Pre charge	Post charge	Optimum charge
Road usage (1000 veh-km)	1390	1160	1055
Speeds (km/h)	14.3	16.3	18.5
Time for 1km (minutes)	4.2	3.6	3.2
Individual cost (euro/veh-km)	1.61	1.36	1.28
Social cost (euro/veh-km)	3.38	2.39	2.09
Charge (euro/veh-km)	-	0.56	0.81
Marginal congestion cost (euro/veh-km)	1.77	0.46	-
Congestion costs (1000 euros/day)	296	24	-
Benefits (1000 euros/day)	-	272	296
Charge proceeds (1000 euros/day)	-	650	854
Collection costs (1000 euros/day)	-	689	689
Benefits net of costs (1000 euros/day)	-	-417	-393

Source: Prud'homme and Bocarejo (2005)

Prud'homme and Bocarejo estimate the DWL from congestion in the charged zone amounted to about 296,000 euros per chargeable day in 2002, or about 75 million euros per year. This is only 0.03% of the GDP of London (255,000 million euros in 2001), and an estimated 0.11% of the GDP of the charging zone. This is very much in line with the findings of Prud'homme (1999) for the Paris area.

Using Prud'homme's figures and Transport for London's figure of 36km/h as the free-flow speed, one can estimate the cost of congestion in the congestion charging area relative to free flow to be 30 million euros per year. Because this applies to the congestion zone only, it is difficult to express it as a cost per person.

3.4 Canada –Transport Canada

Lindsey (2007) provides an example of a cost of congestion calculation using 'acceptable' levels of congestion as the comparison. He quotes statistics that had recently been compiled by Transport Canada quantifying the costs of travel delay, additional fuel consumption, and greenhouse gas emissions for the nine largest urban areas of Canada. Rather than taking free-flow conditions as the baseline, the study adopted a percentage of the speed limit as a threshold below which congestion could be considered 'unacceptable'. Since this threshold varies across municipalities and road networks, the study undertook calculations with thresholds of 50%, 60% and 70% of the speed limit. The costs of congestion for the nine urban areas estimated with the 60% threshold were about CAN$3 billion. Montreal and Toronto accounted for 70% of the total. In per capita terms, the annual cost ranged from CAN$17 per person for Hamilton to CAN$270 per person for Toronto.

3.5 Australia – BTRE

The Australian Bureau of Transport and Regional Economics (BTRE) uses an aggregate modelling approach to estimate both the free flow or total and the DWL or avoidable cost of congestion for the Australian state capital cities.

BTRE states that 'since total delay values are based on the value of the excess travel time compared with travel under completely free-flow conditions – an unrealisable situation for actual road networks – they are rather poor measures of the social gains that could be obtained through actual congestion'. They are quoted in their paper for comparison purposes only.

Calculation of the DWL requires assumptions about the formulation of the travel time function and the demand curve. An Akçelik curve is assumed for the travel time, which relates travel time to the demand for travel, and a constant elasticity function is assumed. BTRE states that the results are sensitive to the assumptions and that they should be treated as indicative.

The BTRE analysis gives an Australian total of $11.1 billion for total annual delay costs over the eight capitals for 2005 (excluding the cost elements for trip variability, vehicle operating cost and air pollution) – compared with around $5.6 billion for the preferable DWL valuation of delay. The distribution of these costs over the main metropolitan areas is shown in table 3.4. The DWL appears consistently to be about 50% of the total congestion figure (except in the smaller cities).

Table 3.4 Estimated costs of congestion 2005 (AUS$ billion pa)

Area	Total congestion (free flow)	Avoidable congestion (DWL)
Sydney	3.9	2.09
Melbourne	3.6	1.79
Brisbane	1.44	0.71
Adelaide	0.78	0.36
Perth	1.05	0.54
Hobart	0.80	0.03
Darwin	0.27	0.01
Canberra	0.18	0.07
Total	11.06	5.6

3.6 Other estimates

Lindsey (2007) quotes estimates of the annual costs of US$38 billion for the United Kingdom and US$800 million to $1 billion for Stockholm: these appear to be based on a comparison with free flow.

3.7 Auckland

Appendix A provides a summary and appraisal of previous New Zealand studies into the costs of congestion in Auckland. Only two independent estimates of Auckland's congestion costs were found to exist and these both use free flow as the reference for comparison. They are:

1 Work undertaken by Auckland Regional Council (ARC) in the mid-1990s, involving runs of the Auckland Regional Transport (ART) model for 1991, for the base (current) situation and a free-flow situation, with the resulting change in total vehicle hours being multiplied by standard values of travel time (from the NZTA (2010) *Economic evaluation manual* (EEM)) to derive an annual estimate of travel cost (time) savings from eliminating congestion. This work derived an estimate of costs (time) of congestion for Auckland of $570 million pa (mid-1990 prices).

   The ARC work was subsequently used as the primary basis of an augmented estimate of $755 million pa that was derived in the Ernst & Young (1997) study. That study appears to have provided the basis for many subsequent claims that the costs of congestion in Auckland are around $1 billion per year.

2 Analyses undertaken by/for the MoT *Surface transport costs and charges* (STCC) *study* (Booz Allen Hamilton 2004), using the ART model for 2001, in similar manner to the earlier ARC work above. This study estimated the costs of congestion for Auckland at $702 million pa (2002 prices).

   The STCC study also made estimates of annual costs of congestion for Wellington, Christchurch, other urban centres and for the inter-urban state highway network, as summarised in table 3.5.

It is evident that these two independent (but similarly based) analyses provide broadly similar estimates, after allowing for traffic growth trends and inflationary impacts between the 1991-based and 2001-based estimates. The STCC work is more recent and better documented, and is therefore regarded as the most authoritative source on the topic.

In one sense, these estimates could be said to understate the true costs of congestion, inasmuch as they make no allowance for the induced travel benefits that would result from the additional traffic 'induced' in a 'no congestion' situation; or any non-time economic benefits arising from the reduction in congestion, eg savings in trip time reliability and in-vehicle operating costs, and thus in fuel and GHG emissions. On the other hand, as noted elsewhere in this report, the estimates derived (relative to a free-flow situation) considerably overstate the savings in travel costs that could be achieved through optimised road pricing and/or optimised road investment policies as measured by, for example, the DWL.

Table 3.5 Surface transport costs and charges study – summary of congestion cost findings

Situation	Total cost – $mpa	Cost per person $pa
Main urban centres[(a)]		
Auckland	702	540
Wellington	101	210
Christchurch	77	204
Other urban centres with population >50,000	36	na
Inter-urban highways[(b)]		
Sections analysed[(b)]	28	
Other rural SH network[(b)]	5	
Incidents[(c)]	20	
Total	969	

Source: Booz Allen Hamilton (2004)

Notes:

[(a)] Figures for the main urban centres have been adjusted to allow for 'secondary' effects from changes in the level of congestion (eg peak spreading, modal change, trip suppression).

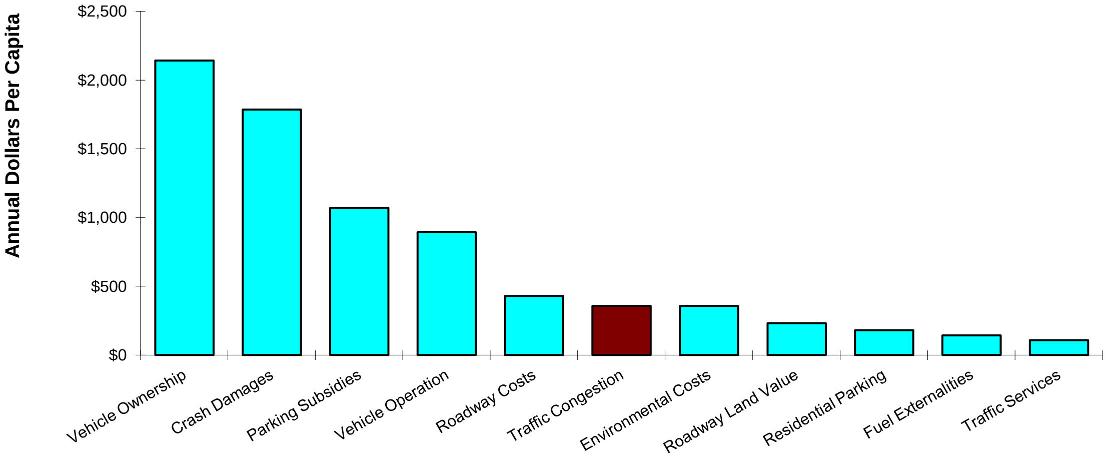
[(b)] Sections analysed were the busiest sections of the rural state highway (SH) network, accounting for 18% of route km and 40% of vehicle km travelled (VKT) on rural SH network.

[(c)] Indicative only.

3.8 Conclusions and implications

While there is a significant range in values for the cost of congestion estimates depending on the method of calculation and the size of the conurbation, these estimates, when expressed on a per capita basis, indicate that congestion is a moderate cost overall. Litman (2011) compares the costs calculated by the Texas Transportation Institute (TTI) for American cities with other costs of vehicle ownership and use. The basis for his calculations is not clear so the figures should be taken as indicative only. Nevertheless as shown by figure 3.3, even using the TTI figures, which are high compared to the economic measures of the cost of congestion, congestion is a moderate cost compared with other transport costs.

Figure 3.3 Costs ranked by magnitude (Litman 2011)



4 Assessment of costs of congestion in Auckland

4.1 The Auckland congestion problem

Auckland is a low density – highly motorised city. Since the main transit components of the De Leuw Cather report of 1965 were rejected, the primary emphasis of transport investment has been on developing the motorway network which, with the letting of tenders for the Waterview connection, is now considered to be nearing completion.

This emphasis on road-based solutions has led to dispersed development and a low public transport mode share. Until recently, public transport in Auckland has been seen as the mode of last resort for the transport disadvantaged rather than as a viable transport option. In the 2006 census, only 7.7% of journeys-to-work in Auckland were made by public transport, including only 1.2% by train. This situation is now changing with the construction of Britomart railway station, the Northern Busway and the planned electrification of the rail network. The proportion of 'choice' passengers on public transport nevertheless remains low by world standards for cities of comparable size.

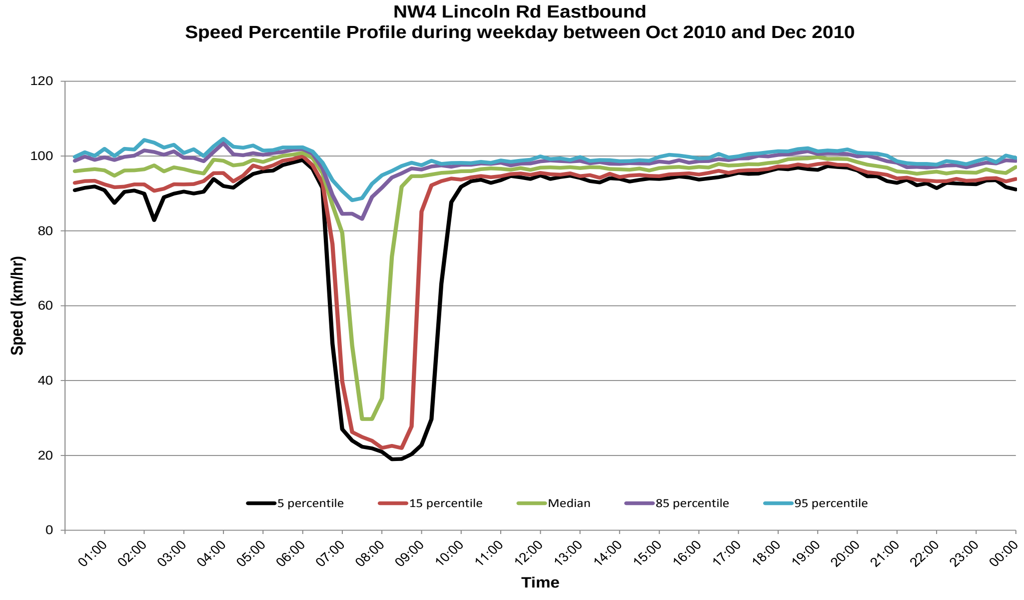
The result of this road-based strategy has been increasing pressure on the road network, particularly the Southern motorway, which was the first to be constructed. Long commute distances exacerbate the impact of reduced speeds on journey times, making congestion a hot topic amongst Aucklanders. This has created pressure for major expenditure on additional road capacity (as well as for public transport alternatives) despite the problem being relatively modest when compared with other large cities in the region.

4.2 Auckland traffic flows and congestion

The actual relationship between travel time, demand and capacity in Auckland can be studied using data collected from permanent traffic monitoring stations maintained by the NZTA at various points on the Auckland motorway system. These stations provide a rich source of data on traffic speeds and flows by time of day and over time. Figures 4.1 and 4.2 are example graphs showing the speed at points on the North Western motorway and the Southern motorway by time of day during the period October–December 2010. The figures show the 5th, 15th, 85th and 95th percentile speeds as well as the median speeds[14]. Figures 4.3 and 4.4 show the traffic volumes (flow) past these points over the same period. Both motorways have two lanes in each direction at these sites.

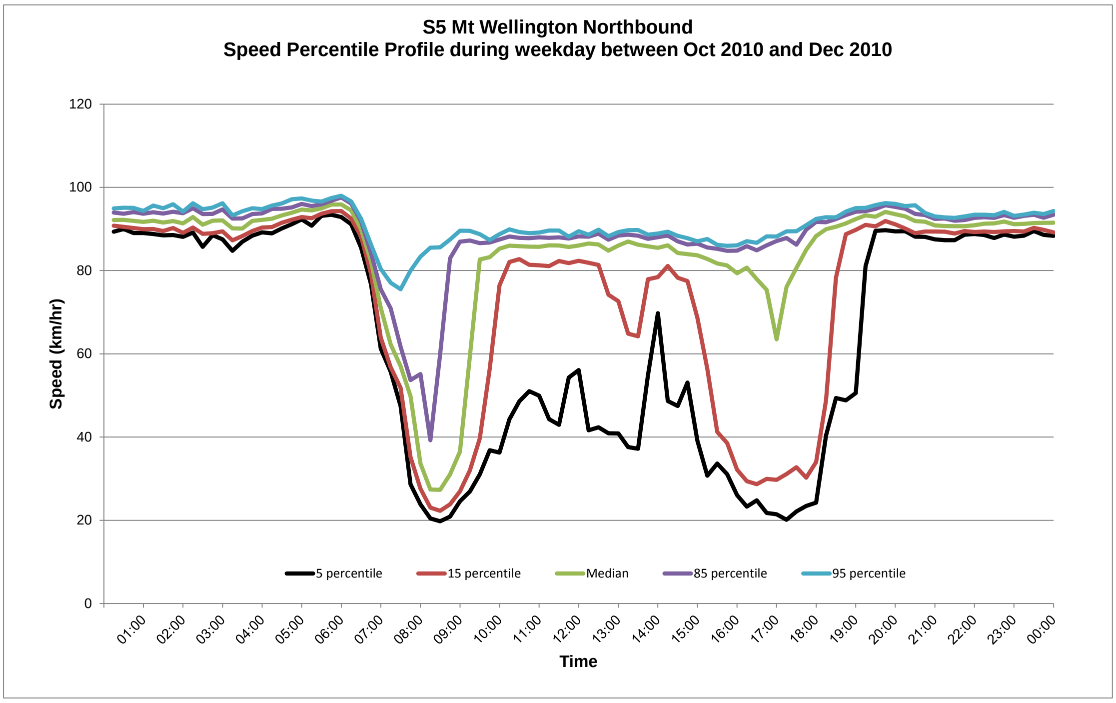
[14] The nth percentile speed is the speed at which the nth percentile vehicle travels, when vehicles are ranked in order of speed.

Figure 4.1 Speeds on the North Western motorway



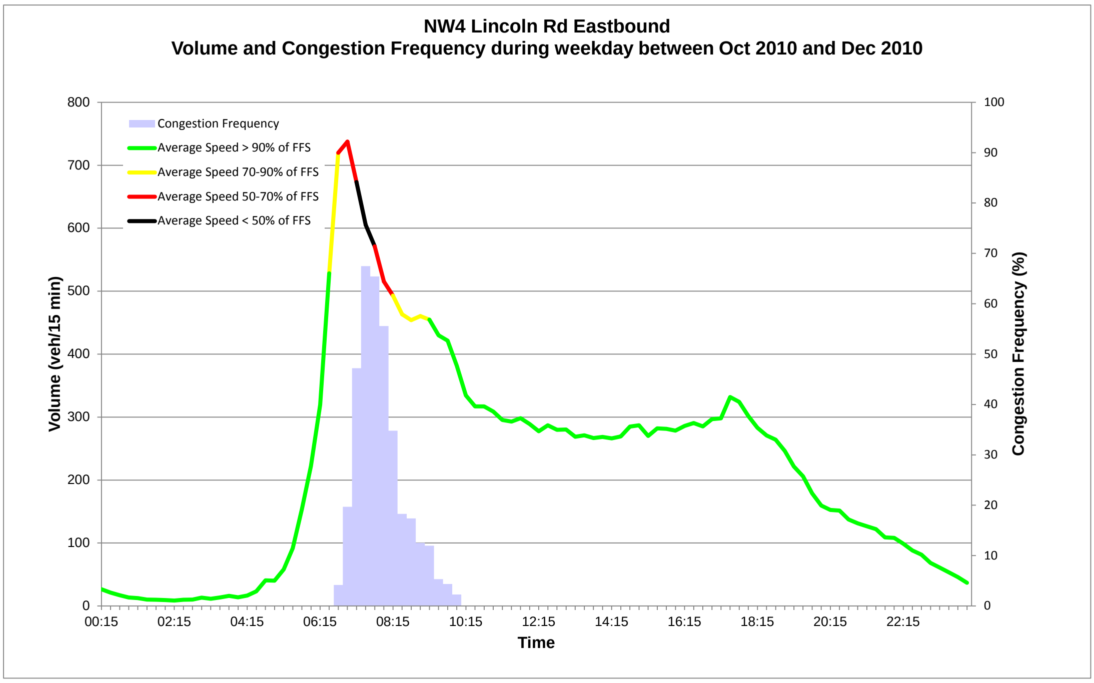
Source: NZ Transport Agency

Figure 4.2 Speeds on the Southern motorway



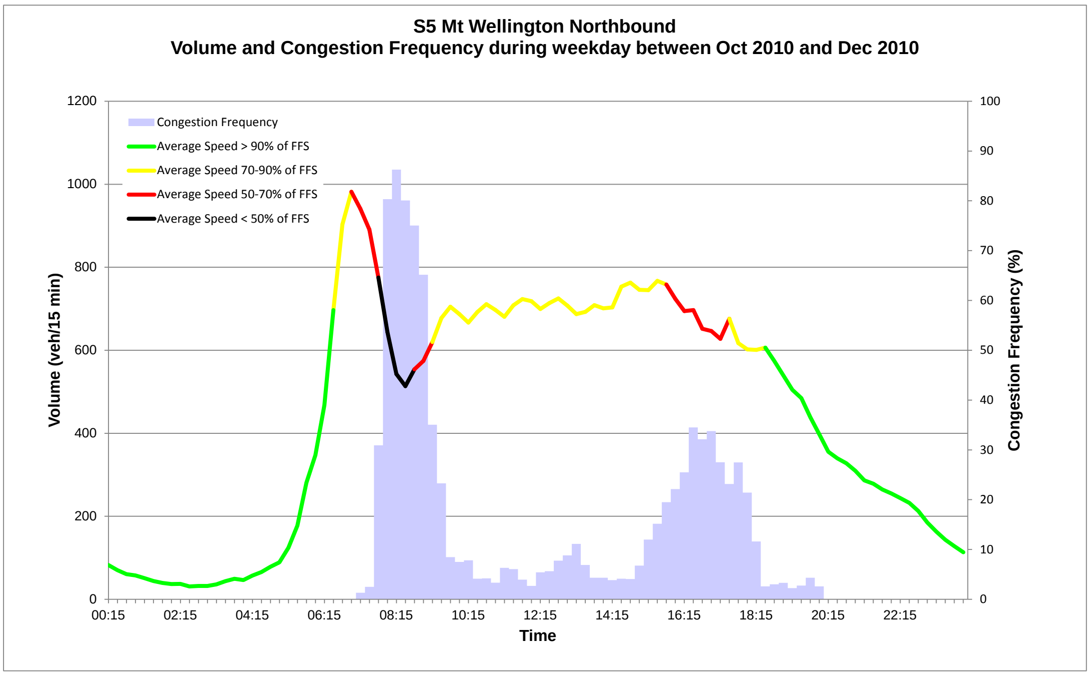
Source: NZ Transport Agency

Figure 4.3 Volumes and congestion levels on the North Western motorway



Source: NZ Transport Agency

Figure 4.4 Volumes and congestion levels on the Southern motorway

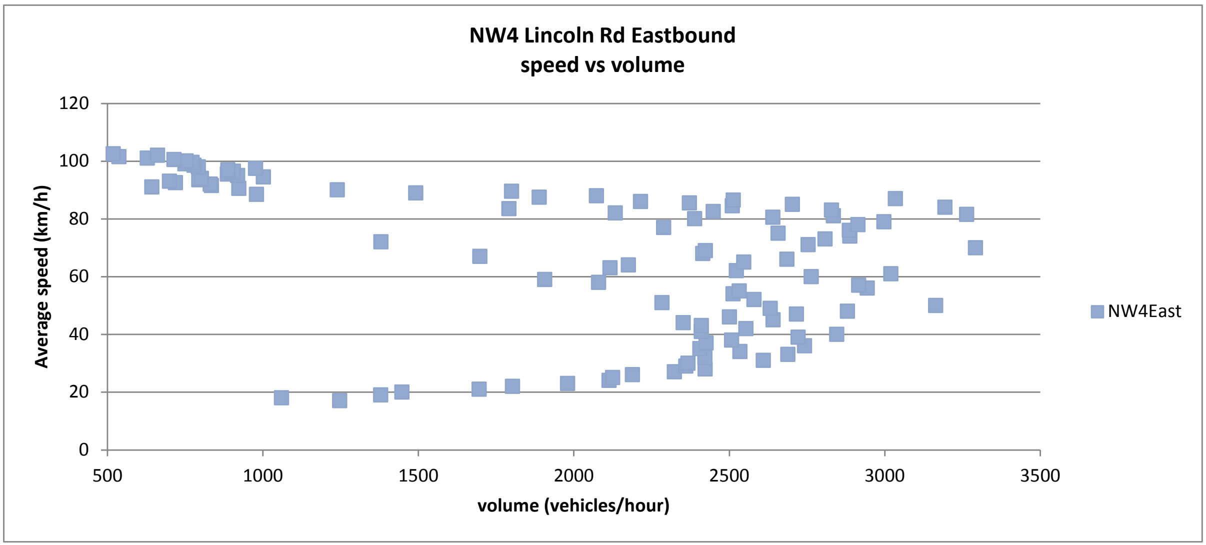


Source: NZ Transport Agency

The data shows clearly that traffic speeds and effective capacities are affected on a regular basis during the morning peak period on the North Western motorway eastbound and throughout the day on the Southern motorway northbound.

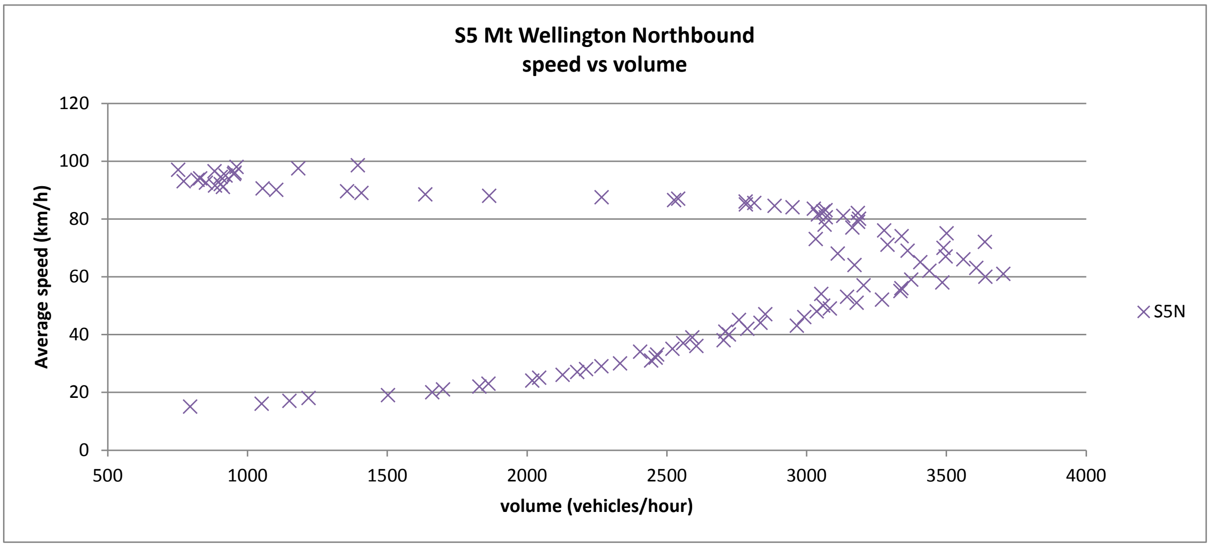
This data has been further analysed to show more clearly the relationship between average speed and traffic volume (flow). This is illustrated by figures 4.5 and 4.6.

Figure 4.5 Average speed vs flow – North Western motorway



Source: Consultant analysis/NZ Transport Agency

Figure 4.6 Average speed vs flow – Southern motorway



Source: Consultant analysis/NZ Transport Agency

These figures show a very clear relationship between the speeds and the traffic flow. A distinctive feature of the graphs is the low-speed section, where traffic speeds are low despite volumes being less than the maximum potential volumes. This is fully in accordance with international experience, the locus of the points is referred to as the 'fundamental law of road traffic' (TRB 2011), shown as figure 2.1. The lower sections of the graphs correspond to the low-speed sections on figures 4.1 and 4.2 and show the effect of congested conditions. The upper section of the graph is that referred to by economists as 'congestion' (but by engineers as 'normal' or 'uncongested'), while the lower section is referred to by economists as 'hyper-congestion' (and by engineers as 'congested' or 'queued').

At low demand, average speeds are close to the maximum permitted speed. As demand increases, the average speed falls very slowly initially and then at an increasing rate until the volume reaches about 3000 vehicles per hour (1500/hour/lane), at which point the flow starts to break down. Sometimes flows as high as 3700 are reached on the Southern motorway. However as more traffic attempts to use the motorway, speeds fall and with them the capacity of the motorway, so at higher attempted traffic levels, speeds and traffic volumes both fall. They continue to fall until the road is gridlocked and further traffic cannot physically join, or until demand eases. It can be seen that when the road performance is in the lower section, the effect of congestion is that the road operates inefficiently.

4.3 Approach to assessment

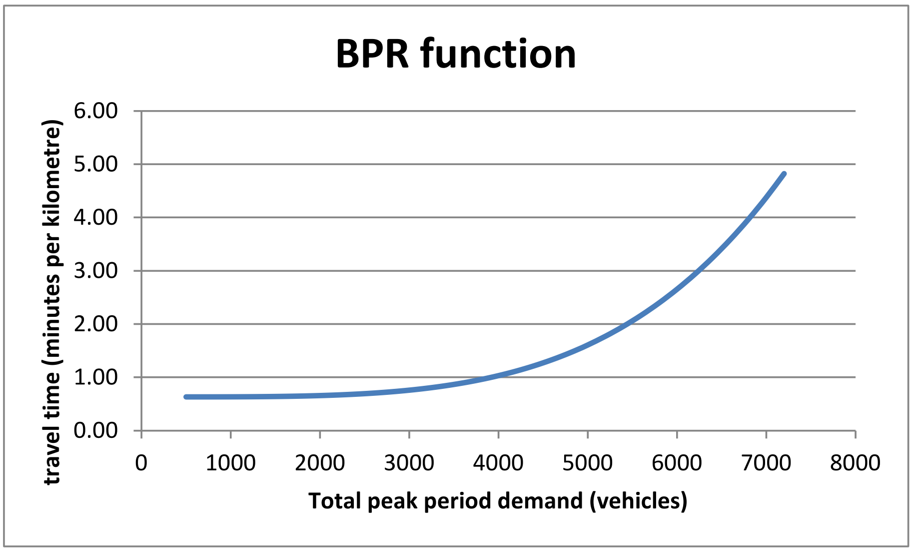
4.3.1 The speed-flow relationship

The volumes shown in figures 4.5 and 4.6 are the observed flows, not the underlying demand. We cannot determine the underlying demand directly from the observed flows. However we can use the relationship between demand and travel time developed in other studies. The most commonly assumed relationship between demand and travel time is the so called Bureau of Public Roads (BPR 1964) function[15].

The BPR function is of the form: $t = \alpha\, t_f * (1 + \beta(\frac{D}{K})^{\rho})$

where t_f is the free-flow time, D is demand over the peak period and α, β, ρ and K are constants. Figure 4.7 illustrates a BPR function with $\alpha = 1$, $\beta = 0.2, \rho = 4$ and K= 3,000. K is often referred to as the 'capacity' and the BPR function is interpreted as relating the travel time to the volume to capacity ratio D/K. However although this is an intuitively attractive way of thinking about the BPR curve, it is potentially confusing. The value used as K is not necessarily the maximum achievable flow. K could be set at 1.0 and the value of β adjusted accordingly.

Figure 4.7 Example of a BPR function



Source: Consultant analysis

How do we relate the demand to the flow? There is ambivalence in the literature, with some authors interpreting demand as the input flow (ie vehicles per hour attempting to travel). Verhoef in particular takes this view[16]. If we interpret demand as the total number of vehicles wanting to travel in the peak period and assume that everyone who wants to travel does attempt to travel, ie the number of vehicles

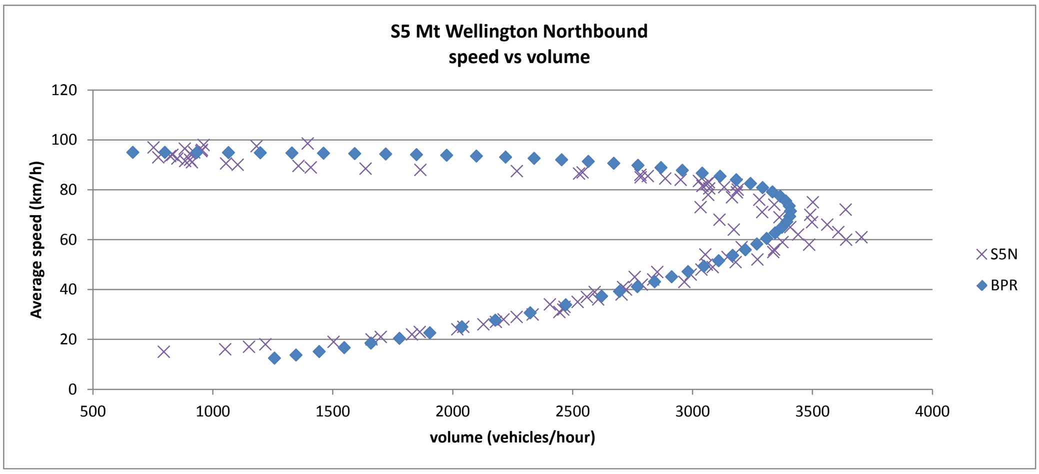
[15] The BPR function is a cost function relating travel cost (time) to the demand for travel.

[16] Private correspondence.

entering the road reflects the demand for travel, then the density will be proportional to the demand (the total number of vehicles) and the flow on a particular link (the number of vehicles per minute) will be proportional to the demand divided by the travel time for the link.

We might not expect this to be a particularly accurate representation of what happens – people react to congestion in many ways, such as by delaying their trips. To see how accurately the theory replicates the observed data, the above BPR function was converted to a speed/flow relationship by setting the speed to the inverse of the time per kilometre and the flow to the vehicles per minute (ie demand/travel time) and plotting these on the same graph as the observed speed vs flow data. This is shown as figure 4.8[17]. It can be seen that assuming a BPR function and assuming the flow is proportional to demand divided by the time per kilometre results in a theoretical function that closely tracks the observed data. This appears to validate our interpretation.

Figure 4.8 Comparison of BPR function and observed data – Southern motorway



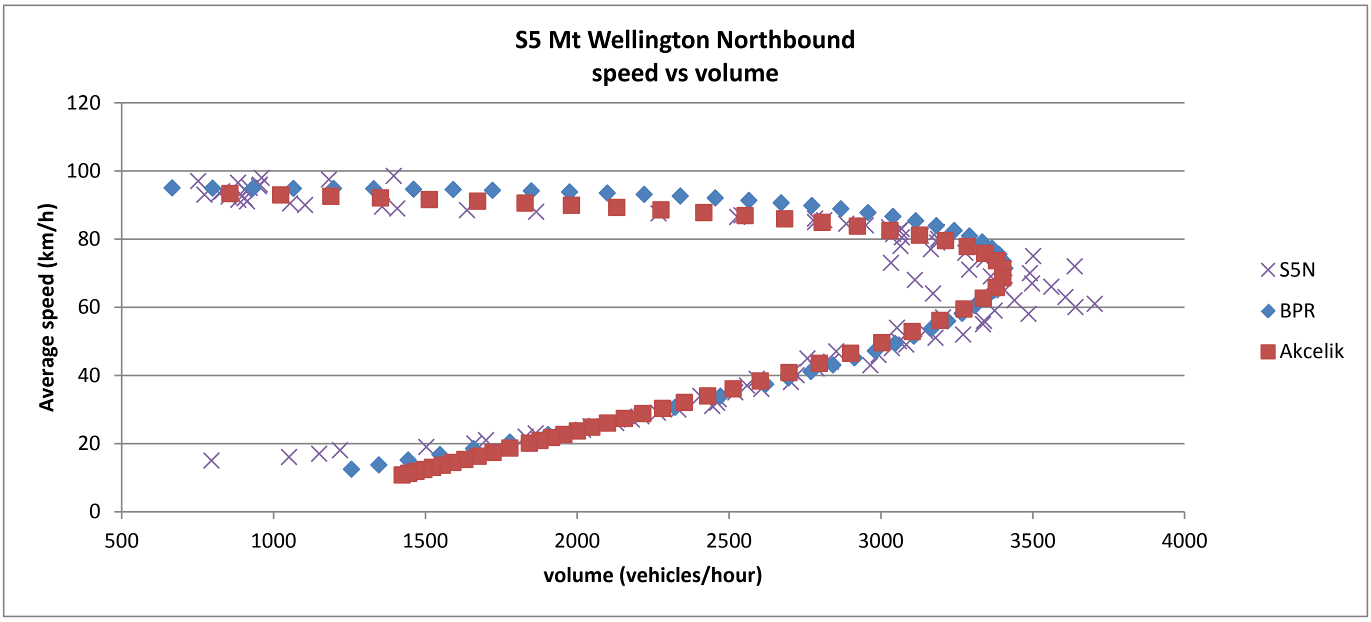
Source: Consultant analysis

4.3.2 Auckland Transport models

The Auckland Transport model suite uses Akçelik functions rather than the BPR functions to describe the relationship between demand and travel time. It can be shown that if we again calculate the flow as the demand divided by the travel time (ie assume the observed density is proportional to the demand) the coefficients of the Akçelik function can be chosen so that the derived speed vs flow function fits the observed data. This is shown in figure 4.9.

[17] Since the time is actually time per kilometre, the flow is (demand/travel time per km) * travel time at capacity flow.

Figure 4.9 Comparison of observed data, BPR and Akçelik functions



Source: Consultant analysis

The Auckland models use the Akçelik function because the formulation is based on queuing theory and is thus more 'explanatory' than the BPR function which was developed purely as the best fitting curve. We use the BPR function in the following discussion because of its analytical simplicity.

4.3.3 Implications for the costs of congestion

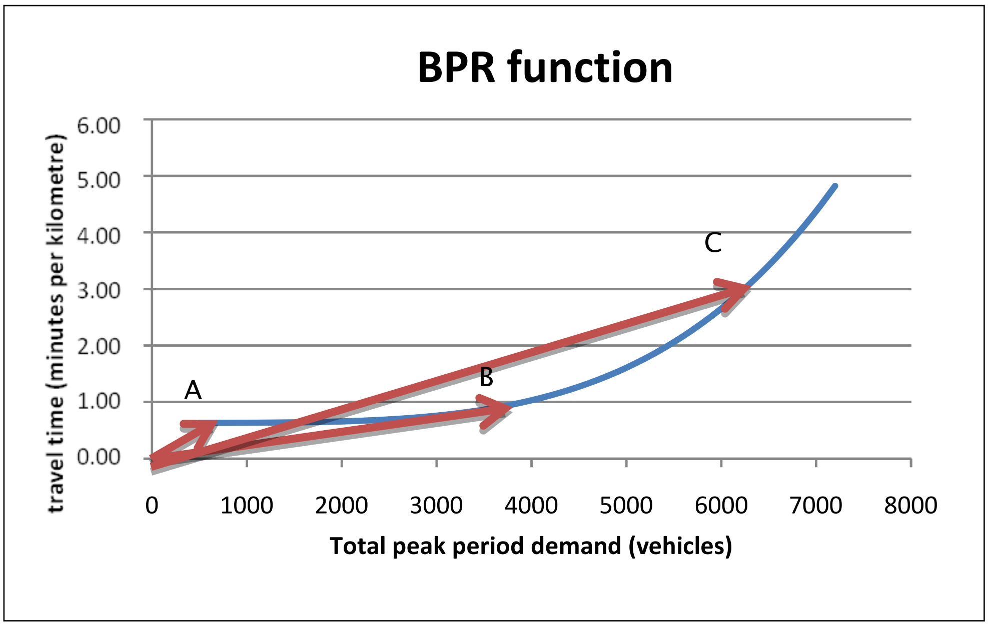
If we look at the BPR graph (figure 4.10) the flow is proportional to the inverse of the slope of a line joining the origin to the curve. The flow is a maximum when this line is just tangential to the BPR curve (OB). Before that, the flow is increasing with demand (OA). After that the flow is decreasing with demand (OC).

If we look at the shape of the speed vs flow curve, we can draw some tentative conclusions: First of all, free flow is only achieved at volumes well below the capacity of the road. It would be very expensive to build enough capacity to guarantee free flow. Yet speeds are not that much below free flow when the road is near capacity. Second, the two distinct halves of the curve have quite different properties. In the upper section (which corresponds to demand less than the road's capacity) speeds can only be increased if some people are displaced. We cannot make some people better off unless others are made worse off. Tolling a road in this situation is unlikely to be popular and may not be economically justified. In the lower section of the curve, increasing speed and increasing flow can go together. If we can find a way to move up the curve, people will travel faster and more will be able to travel; everyone will be better off. A toll that reduces demand in this situation would mean more people would travel faster. This confirms that the optimum toll must be a toll sufficient to keep demand no greater than capacity, and suggests the optimum toll would be that which keeps demand equal to the capacity of the road. This supports our earlier conclusion (section 2.5.2) that we should define congestion as the situation where demand exceeds capacity (ie we are in the lower branch of the speed flow curve) and the cost of congestion as the difference between the observed travel times and the times on a network when it is operating at capacity[18].

---

[18] Going back to the observed flows on the Auckland motorways, it can be seen that in practice there is considerable scatter about the point where the maximum flow occurs. Thus in practice, the reference point is the maximum sustainable flow, which is less than the maximum achievable flow.

Figure 4.10 Slope of locus: BPR function



Source: Consultant analysis

4.4 Costs of congestion: existing network

4.4.1 Time-saving costs (peak periods)

The proposed measure of the cost of congestion requires the ability to estimate travel times and costs in the congested and uncongested states.

For our Auckland analyses, the ART3 model has been used to provide data on existing traffic demand and travel times and to estimate the impact of a number of different assumptions on travel times. This has enabled us to estimate the costs of congestion. In these models, the relationship between speed and demand is represented by a series of curves developed by Akçelik which relate the travel time on a link to the observed free-flow time (eg at 4am), the road type and the demand. The properties of the Akçelik curves are similar to the BPR curve (figure 4.7) but are based on queuing theory, whereas the BPR function is simply a best fit to the observed data. In either case the travel time increases only slowly initially, but rises quickly as capacity is approached. It then continues to rise in response to further increases in peak demand. As shown in figure 4.9, both the BPR and the Akçelik curves replicate the observed data. The Auckland models were used to estimate three sets of travel times, as follows.

1 Congested times

The existing or congested situation was represented by a morning peak base year run of the ART3 model. The ART3 model represents the traffic in Auckland using an EMME/2 transport network that has been calibrated to closely replicate travel patterns observed in a major transport survey undertaken in 2006. The base year run represents the situation in 2006 with the existing network.

2 Free flow

To provide comparisons with earlier work, we undertook an ART3 run with all links set to their free-flow speed – note that this speed varies between links depending on their type.

3 Uncongested

The uncongested travel times – ie travel times assuming maximum flow, when links are operating at capacity[19] – were calculated by setting the volumes on the links in the EMME/2 network to the lesser of the observed link volume or the capacity of the link, and then running the model to calculate the travel times.

Key statistics for each case are shown in table 4.1.

Note that the total number of trips is the same in each of the three cases and the vehicle kilometres are almost the same. The travel times differ because in the free-flow case, the capacity of the network is assumed to be increased to enable faster speeds; in the uncongested case the demand during the peak two hours is limited to the capacity of the network and the remainder of the trips are assumed to be accommodated outside the peak two hours; while in the congested case, the peak two-hour demand exceeds the capacity of the network and queuing occurs so the time taken to meet the demand is more than two hours.

Table 4.1 ART3 modelling results (2006 AM peak) – 'free flow', 'congested' and 'uncongested' cases

Measure	Free flow	Congested (existing)	Uncongested (capacity)
Total vehicle trips	453,589	453,589	453,589
Total vehicle km	6,431,892	6,682,878	6,368,390
Total travel time (minutes)	5,517,975	9,029,505	8,366,584
Average speed (km/h)	69.9	44.4	45.7
Average trip time (minutes)	12.2	19.9	18.4

Source: Consultant estimates

To calculate the total costs of congestion, we need to multiply the difference in travel time by the number of travellers. The demand in the Akçelik function is expressed as a demand per peak two hours, but it is, in concept, a total demand rather than a rate[20]. If the demand exceeds the capacity of the network, implicit in the transport model is that the demand is satisfied but over a longer period. Thus if the volume to capacity ratio of a link averages 1.5 over the peak two hours, it implies that it will take at least three hours for the peak demand to dissipate if the link is operating at capacity[21]. If a toll reduces the peak demand to the capacity of the link, what happens to the rest of the vehicles? They can travel outside the peak period and the total traffic demand will still be met within the original time-frame. In fact because the link will operate more efficiently, the total time to serve the original traffic will be reduced. Thus we can use the total trip matrix in evaluating the costs of congestion.

The total travel time on the network reduced from 9.030 million minutes in the morning peak on the base (ie congested) network to 5.518 million minutes in the free-flow network. Using a standard value of travel time savings of $19.82 per vehicle hour in uncongested conditions, plus allowances for congestion and

[19] Assumed to be achieved through optimum pricing or other demand management. In effect we are using the engineering definition of congestion – in 'economic' terms this equates to the absence of hyper-congestion or treating an optimally priced network as the base.

[20] It is calculated using trip generation and distribution models that forecast the total demand for travel and assume that all peak travel demand can be met.

[21] Because an overloaded link operates at less than capacity, the effect of a v/c ratio of over 1.0 is to delay traffic by even more than the v/c ratio suggests.

unreliability (refer table B.1) based on the EEM, this gives an estimate of the 2006 costs of congestion in terms of time (and reliability) savings relative to free-flow conditions of some $1.53 million per morning peak period. Note that all these analyses are based on 2006 ART3 networks and matrices but have been adjusted to 2010 prices. While the traffic profile in the evening peak is different from the morning peak, and some congestion occurs outside peak times, we have adopted an annualisation factor of 500 to convert AM peak costs to annual costs, giving **annual time costs of congestion relative to free flow of $766 million.**

The travel time on the network in uncongested (optimised) conditions was 8.367 million minutes. This gives an estimate of the 2006 costs of congestion in terms of time savings **relative to optimised conditions of some $145 million per year**.

Note that these figures are based on 'normal' operating conditions on the road network (recurrent congestion) and make no allowance for the incidence of non-recurrent/irregular congestion (from crashes etc).

The data provided from the transport models is on a zone-to-zone basis. It is thus possible to look at individual or groups of zones to see where the costs of congestion are incurred. Analysis of the trip matrices shows that 50% of the trips account for 95% of the congestion cost. This suggests that properly targeted investment or demand management measures may be able to be used to address the problem.

4.4.1.1 Perceived congestion

Under the 'perceived congestion' approach, the cost of congestion is calculated by comparing the travel time with travel time based on a minimum acceptable speed. Although we do not recommend this as an appropriate measure, we were able to make some estimates of the cost of congestion under the perceived congestion approach. For the perceived cost of congestion, we said we had congestion whenever the overall speed between origin and destination fell below a specified minimum and counted the difference in time costs between observed time and the time at the minimum overall speed. Note that we only had data on the total travel time between origin and destination - if we were to refine this approach, we would set a different 'acceptable' speed for urban streets than motorways. 60km/h is higher than the speed limit for urban streets.

The equivalent travel time costs for the perceived congestion approach were estimated at:

Relative to 60km/h	**$703 million per year**
Relative to 50km/h	**$495 million per year**
Relative to 40km/h	**$262 million per year**

Note that the speeds shown here are not strictly comparable with the proposed approach which does distinguish between road types – the speed at capacity for urban streets is generally 30–35km/h while the speed at capacity for motorways is 75–80km/h.

4.4.1.2 Deadweight loss (DWL) from congestion

We do not recommend using the DWL approach and therefore did not collect the data required to calculate it. Calculation of the DWL requires estimation of the travel demand functions and the travel time functions. BTRE did this for the main Australian metropolitan areas – but at a very aggregate level. Their analysis indicated that the DWL was about 50% of the cost relative to free flow. This percentage applied across all 10 metropolitan areas and did not appear to be dependent on size.

Assuming a similar relationship would apply to Auckland suggests that the **DWL time cost** for Auckland would be around **$380 million per year**.

4.4.2 Schedule delay costs

The most significant indirect cost relating to congestion is the cost associated with avoiding the peak suffered by those who either travel before or after the peak travel period.

One way of estimating peak spreading costs – otherwise known as 'schedule delay costs' – is suggested by Verhoef (1997). Verhoef develops a conceptual model of peak travel assuming a one-link network of length L. If the value of time at home is α, the arrival time is t ($t_F \le t \le t_L$), free-flow speed is S*, the desired arrival time t*, the value of time at work for early arrival (ie t<t*) is β and the value of time at work for late arrival is γ, then for t* to be preferred, $\beta < \alpha < \gamma$ and the cost of travel c is given by:

$$c = (\alpha - \beta)\,(t^* - t_F) + \alpha L/S^* = (\gamma - \alpha)\,(t_L - t^*) + \alpha L/S$$

The first traveller ($t = t_F$) and the last traveller ($t = t_L$) experience free-flow speeds S*.

The implication of the assumptions is that the traveller cost c must be constant for the full peak period; otherwise people would change their time of travel.

To estimate the size of the schedule delay cost in practice would require knowledge about the reason for the timing of each trip. An upper bound for the cost as a proportion of the measured free-flow congestion cost can be estimated from the Auckland motorway data. Table 4.2 shows morning peak data for the Southern and North Western motorways.

Table 4.2 shows the observed maximum, peak average and minimum times to travel one kilometre on the Southern and North Western motorways in the morning peak. If we assume that the Akçelik functions used in the Auckland models have been calibrated to representative peak travel times, the calculated times will be close to the maximum 85 percentile times shown here. The calculated cost relative to free flow is the difference between the maximum time and the minimum time multiplied by the actual 'peak' travellers – assumed to be 7.15am to 9.15am. However if we assume that travellers outside the 7.15am to 9.15am window experience the same cost, but in the form of schedule delay, then the total cost could be as much as (max time – min time) * (travellers between 6.15am and 10.15am), and the schedule delay cost is the difference between these two. This is also shown in table 4.3.

It can be seen that the upper bound for the schedule delay cost is about 2/3 the calculated cost in the 7.15am to 9.15am period. Note that this is an upper bound to the cost: the contention that people travelling outside the 'peak of the peak' suffer schedule delay costs is based on the assumption that everyone has the same desired arrival time and values of time. In practice, the spread of travel times will in part be due to the spread of desired arrival times and destinations. The absolute and relative size of α, β and γ will also vary across the population.

We conclude that measuring the direct congestion costs of those who travel in the peak potentially overlooks the schedule delay costs of those who adjust their travel time to avoid the worst of the congestion delays. Based on the Auckland motorway data, an upper bound for the schedule delay cost is 65% – 70% of the calculated cost for the two-hour AM peak period. This is consistent with other studies (eg Koopmans and Kroes 2003) that suggest the schedule delay cost could be as much as the directly measured cost. The Koopmans figure includes people who switch to a less preferred mode in order to avoid congestion: this is likely to be relatively small for Auckland.

Table 4.2 Observed travel speeds and volumes – Southern and North Western motorways (85 percentile)

Time (starting)	Southern motorway			North Western motorway		
	Volume	Speed	Time (sec)	Volume	Speed	Time (sec)
6.15am	468	96.1	37.5	320	99	36
6.30am	697	91.8	39.2	528	95	38
6.45am	903	84.2	42.8	720	87	41
7.00am	982	75.5	47.7	728	79	45
7.15am	941	70.9	50.8	673	49	73
7.30am	891	61.7	58.4	605	30	121
7.45am	775	53,7	67.1	571	30	121
8.00am	644	55.1	65.3	515	35	102
8.15am	543	39.2	91.8	493	73	49
8.30am	514	59.9	60.1	463	92	39
8.45am	554	82.9	43.4	454	95	38
9.00am	574	86.9	41.4	460	95	38
9.15am	620	87.2	41.3	455	95	38
9.30am	677	86.6	41.6	430	95	38
9.45am	705	86.7	41.5	421	96	38
10.00am	687	87.5	41.2	382	96	38
10.15am	667	88.2	40.8	335	96	38
10.30am				317	96	37

Source: NZTA monitoring report

Table 4.3 Calculation of schedule delay cost

Measure	Southern motorway	North Western motorway	Totals
Max time (sec)	91.8	121.1	
Peak average (sec)	72.4	68.8	
Min time (sec)	37.5	36.2	
Calculated delay hours/day	82.0	99.9	181.9
Schedule delay hours/day	58.7	65.2	123.9

Source: Consultant estimates

4.4.3 Other components of congestion costs

Apart from travel time (including reliability) costs and schedule delay costs, other significant components of the costs of congestion are vehicle operating costs, environmental costs and crash costs. These cost components are estimated in appendix B and summarised as follows (refer table B.1):

- **Vehicle operating costs.** These are estimated at an average of $1.97/vehicle hour. This equates to $58 million pa relative to free-flow conditions, $11 million pa relative to optimised conditions.

- **Environmental (GHG) costs.** These are estimated at an average of $0.16/vehicle hour. These equate to $5 million pa relative to free-flow conditions, $0.9 million pa relative to optimised conditions.
- **Crash costs.** These **additional** costs when average speeds increase are estimated at $100 million pa relative to free-flow conditions, $10 million relative to optimised conditions. (However, note that these figures are indicative only.)

All the components of the costs of congestion that have been assessed are summarised in table 4.4, relative to both the free-flow conditions and the optimised conditions. It is seen that:

- **Relative to free-flow conditions**, the congestion costs associated with the base (2006) situation are **around $1250 million pa** (2010 prices). Excluding the schedule delay component (which has not been assessed in previous New Zealand studies, the costs would be around $730 million pa; this is (perhaps surprisingly) close to the best estimate hitherto available, from the STCC study, of about $700 million pa (2002 prices) – refer section A4.4.
- **Relative to optimised conditions**, the congestion costs associated with the base (2006) situation are **around $250 million pa** (2010 prices). In our view, this provides a more useful indicator of the annual economic costs that could potentially be saved by optimising (in economic terms) the level of traffic on the network.

The perceived and DWL estimates are included in table 4.4 for comparative purposes only. Their schedule delay and other costs are calculated on a pro-rata basis. The DWL is based on the BTRE estimate, which is a much higher proportion of the free-flow estimate than estimates by Prud'homme (1998) or Newbery (1999).

Table 4.4 Summary of costs of congestion components

Item	Annual cost estimates – $million pa (2010 prices)					
	Relative to free-flow	Relative to 60km/h minimum	Relative to 50km/h minimum	Relative to 40km/h minimum	Relative to 'un-congested'	Dead-weight loss (DWL)
Time costs (peak periods)	766	703	495	262	145	380
Schedule delay costs	522	479	337	179	99	260
Vehicle operating costs	58	53	37	20	11	29
Environmental costs	5	5	3	2	1	2
Crash costs	-100	-90	-56	-19	-10	-32
Totals	1251	1150	817	443	246	638

Source: Consultant analysis

4.5 Short-run vs long-run costs

Both the DWL measure and the proposed capacity-based measure of the costs of congestion calculate the costs relative to the operation of the existing network in a way that is economically optimal (in the case of the DWL) or efficient. This section looks to see whether the cost of congestion is higher than it should be because of inadequacies in the network itself and whether there is a prima facie case for further investment.

The standard rule to determine whether infrastructure investment is optimal (see for example Turvey 1968) is that the short-run marginal cost (SRMC), which is the cost imposed by the marginal user without adjusting capacity, must equal the long-run marginal cost (LRMC), which is the cost of adjusting capacity for the marginal user. It is optimal because if the SRMC was higher than the cost of adding capacity, adding more capacity would reduce the net cost (and vice versa).

There is a conceptual problem in applying this rule to roads on which traffic flows are peaked in nature. Since capacity is primarily required for the peak, we need to calculate the marginal costs for increases in the peak flows only – and calculate the capacity cost per unit of peak flow rather than for the all-day flow.

Note that the analysis is at an abstract level. A real network contains many links each with different demands. Practical considerations limit the design opportunities – investments are lumpy and often indivisible. Actual investment decisions must be based on cost–benefit analysis, which can take these and other factors into account.

4.5.1 Short-run marginal cost

The data from the uncongested model run described above can be used to estimate the short-run marginal cost on the existing network. It can be shown that when the road is operating at capacity, the SRMC is equal to the value of the average travel time (see box). The average journey time in the uncongested run – where times were set equal to the time 'at capacity' - was 18 minutes (table 4.1). Using this figure and the NZTA EEM value of time ($26.2/hour) implies that the short-run marginal cost for the existing network operating at capacity is $7.86.

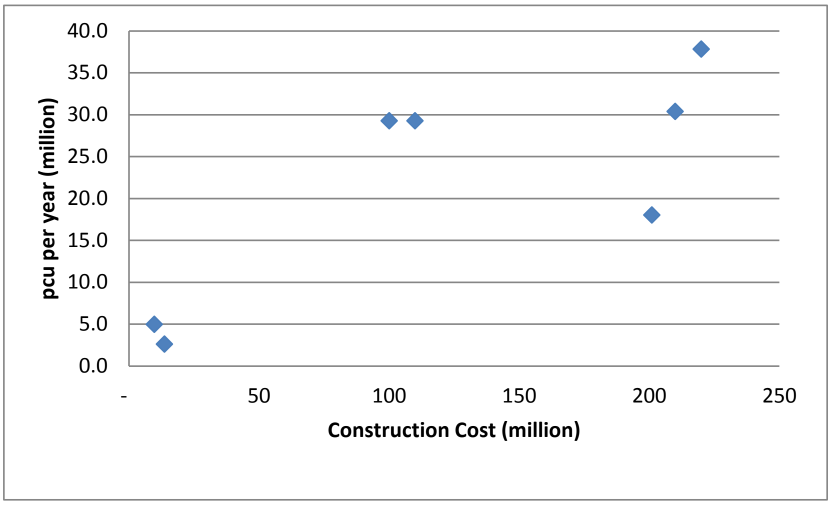
This method assumes that everyone has the same value of time. It should give a reasonable estimate of the SRMC provided the number of links where demand exceeds capacity – ie are congested – is small. If congestion is significant, only a subset of the population with higher values of time will be able to travel during the peak period and the method based on average values of time will underestimate the SRMC.

> Consider an increase in demand dN that results in an increase in travel time of dt. The effect of the increased demand is to increase the output flow to (N+dN)/(t+dt)*t. Output flow will be at a maximum if the increase dN does not result in any change in the output flow – ie when (N+dN)/(t+dt)*t = N
>
> ie when t =Ndt/dN.
>
> But SRMC = αN dt/dN
>
> Hence when the road is operating at capacity SRMC = αt.

4.5.2 Long-run marginal cost

We can make an approximate estimate of the LRMC from the cost of recent contracts for road widening and construction in Auckland. If we assume that the new construction is primarily to provide additional peak capacity, that the peak lasts 2.5 hours on average, and a lane carries 1800pcu per hour, we can calculate the construction cost per additional passenger car unit (pcu) capacity. Figure 4.11 shows the capital cost and the increase in peak capacity provided (pcu kilometres pa) for seven recent road capacity enhancements in Auckland (projects involving tunnelling were excluded). The average cost over the seven projects is $5.70 per pcu-km year. Details of the seven projects are provided in appendix C.

Figure 4.11 Construction cost per pcu-km – Auckland projects



If we assume an annual economic cost (depreciation plus opportunity cost) of 10% of the construction cost, this gives an average LRMC of 57 cents per pcu-km. From table 4.1 we can deduce that the average trip is 14.6 km, so for the average journey the LRMC and thus the optimum long run toll is 14.6 * 0.57 or about $8.35.

The SRMC and LRMC are quite close – certainly within the accuracy of this analysis – suggesting that the optimum network would have similar capacity to the current network.

The SRMC is based on the 2006 network and trip matrices, for which only a relatively small proportion of the links are operating at capacity. If traffic grows over time, this proportion will increase and with it the average travel time and SRMC.

5 Conclusions and implications

5.1 Assessment of the Auckland case study

As outlined in the previous chapter, we now have a possible method of calculating the costs of congestion. It involves comparing the travel time (and other economic impacts related to traffic speeds) in the existing or modelled situation with the travel time when all links are at or below capacity.

How does this stack up against the objectives we set? It is certainly measurable. Does it respond in a meaningful way to policy and investment initiatives?

- It is an achievable target: in theory the network could be managed to reduce the cost of congestion (as defined) to zero.
- Traffic demand management initiatives will reduce the total traffic demand, resulting in higher speeds. The measure will reflect the resulting reduced costs.
- Public transport improvements will similarly reduce traffic demand.
- Investment in new road space that results in time savings will result in an improvement in the measure. Any induced traffic will be credited with half the time savings. The cost of congestion measure will be equivalent to the normal measurement of consumer surplus used in economic evaluation.

While the objective of the analysis was to calculate the costs of congestion, the methodology identified also allows calculation of the optimum capacities on the network and the long-term equilibrium toll required on each link. This could be used to develop second-best toll schemes (such as cordon tolls), which could then be tested to see how they impact on the costs of congestion.

5.2 Potential methodology refinements

The methodology used to calculate both the free-flow measure and the proposed measure does have some unavoidable limitations.

- The data used is based on a 2006 network and travel survey. More recent data would improve the estimates.
- The method uses derived data rather than the original observations. While the ART3 model has been calibrated to replicate the observed travel speeds and times, this is never achieved precisely, and it is possible that in obtaining the best overall fit, speeds on the relatively few links that are hyper-congested may not be replicated as well as speeds on links that are operating normally. Given the importance of the hyper-congested links to this analysis, there may be value in additional work being undertaken with respect to these links.

The methodology used could be adapted to investigate further the concept of an optimal network, using an approach along the following lines. We know the optimal long-run level of the tolls from the LRMC analysis (section 4.5.2). We can determine what the optimal network looks like by modelling peak travel with the link tolls set at the LRMC values and the link times set equal to the times at capacity. The resulting link volumes will be the optimal capacity on the link. We can do this initially assuming that the network design stays unaltered and only the capacities are varied. Optimisation of the network design could then be undertaken in corridors where additional capacity is warranted by comparing the cost of widening the existing roads with the cost of provision of new facilities.

5.3 Potential application to other centres

The proposed method of calculating the cost of congestion uses data from the ART3 models. This is calibrated against observed data from 2006 Auckland travel surveys. The same approach should be readily applicable in other New Zealand cities that have transport models or road traffic assignment models.

If the cost of congestion is calculated relative to free-flow conditions, then all larger cities where there is a noticeable slowing of traffic in the peak period will have a calculable cost of congestion. It is likely to be less per head of population than the Auckland estimate.

However the costs of congestion relative to the capacity situation may be too small to calculate outside Auckland and Wellington, as the number of journeys made in what economists call hyper-congested conditions is believed to be very small. Only these journeys are included in our recommended measure of congestion costs.

6 References

Booz Allen Hamilton (2003) *Passenger transport evaluation and funding procedures (stage 2)*. Final report to Transfund NZ.

Booz Allen Hamilton (2004) *Surface transport costs and charges study: road congestion costs*. Working paper. Report to Ministry of Transport, Wellington.

Bureau of Public Roads (1964) *Traffic assignment manual*. Washington DC: US Department of Commerce, Urban Planning Division.

Bureau of Transport and Communications Economics (2000) *Urban* c*ongestion:* m*odelling traffic patterns, delays and optimal tolls*. Canberra.

Bureau of Transport and Regional Economics (2007) *Estimating urban traffic and congestion costs for Australian cities*. Working paper 71. Canberra.

Button, K (2004) *Road pricing*. Final report of ITS Center project. USA: DOT University Transportation Center, Center for ITS Implementation Research.

Department of the Environment, Transport and the Regions (DETR) (2000) *Tackling congestion and pollution: the Government's first repor*t. London.

Department for Transport (2000). *A measure of road traffic congestion in England – method and 2000 baseline figures*. London.

Department for Transport (2003) Congestion and accident risk. *Road safety research report no.44.*

Deryche, P-H and J Crews (1991) Urban concentration and road networks: two approaches to congestion. Part I. In *Flux 5*: 35–45.

Dowling, RB, R Singh and W Wei-Kuo Cheng (1998) Accuracy and performance of improved speed-flow curves. *Transportation research record 1646.*

Ernst & Young (1997) *Alternative transport infrastructure investments and economic development for the Auckland region*. Report for the Keep Auckland Moving Campaign (ARC and region's TLAs).

Federal Highway Administration (2005) *Traffic congestion and reliability: trends and advanced strategies for congestion.* Washington: Mitigation Cambridge Systematics with Texas Transportation Institute.

Fields, G, D Hartgen, A Moore and RW Poole Jr (2009) Relieving congestion by adding road capacity and tolling. *International Journal of Sustainable Transportation 3*, no.5: 360–72.

Goodwin, PB (1997) Solving congestion. *Inaugural lecture for the professorship of transport policy University College*, London, 23 October 1997.

Grant-Muller, S and J Laird (2007) *The cost of congestion*. Association for European Transport and Contributors.

Hau, TD (1992) *Economic fundamentals of road pricing – a diagrammatic analysis.* WPS 1070, Washington: The World Bank. pp4–5.

Hedges, A (2001) *Perceptions of congestion: report on qualitative research findings*. London: Department for Transport.

Henderson, JV (1974) Road congestion: a reconsideration of pricing theory. *Journal of Urban Economics 1*: 346–365.

Highways Agency, Scottish Office Development Department, The Welsh Office, The Department of the Environment for Northern Ireland (1997) *Design manual for roads and bridges, traffic flow ranges for use in the assessment of new rural roads.*

Koopmans, C and E Kroes (2003) Estimation of congestion costs in the Netherlands. In *Proceedings of the European Transport Conference*, October 2003.

Lindsey, CR (2007) Congestion relief: assessing the case for road tolls in Canada. *CD Howe Institute commentary no.248*.

Lindsey, CR and ET Verhoef (2000) Congestion modelling. Pp353–373 in: *Handbook of transport modelling, handbooks in transport 1*. DA Hensher and KJ Button (Eds). Amsterdam: Elsevier/Pergamon.

Litman, TA (2009) *Transportation cost and benefit analysis.* Victoria Transport Policy Institute.

Litman, TA (2011) *Smart traffic congestion reductions*. Victoria Transport Policy Institute.

Ministry of Transport (MoT) (1997) *Land transport pricing study – options for the future*. (Discussion document). Wellington: MoT.

Ministry of Transport (MoT) (2006) *Auckland road pricing evaluation study*. Final report, March 2006. Wellington: MoT.

Mogridge, MJH (1990) *Travel in towns: jam yesterday, jam today and jam tomorrow?* London: Macmillan.

Mohring, H and M Harwitz (1962) *Highway benefits: an analytical framework*. Evanston, IL: North-western University Press.

New Zealand Transport Agency (NZTA) (2010) Economic evaluation manual. Volume 1. *www.nzta.govt.nz/resources/economic-evaluation-manual/volume-1/docs/eem1-july-2010.pdf*

Newbery, D (1988) Road user charges in Britain. *Economic Journal 98*.

Newbery, D and G Santos (1999) *Quantifying the cost of congestion*. Cambridge, UK: Department of Applied Economics.

Pigou, AC (1920) *Wealth and welfare*. London: Macmillan.

Prud'homme, R (1998) Road congestion costs in the Paris area. *Paper prepared for the 8th World Conference on Transportation Research*. Antwerp, July 1998.

Prud'homme, R and JP Bocarejo (2005) The London congestion charge: a tentative economic appraisal. *PIARC Seminar on Road Pricing with Emphasis on Financing, Regulation and Equity*. Cancun, Mexico, April 2005.

Schrank, D, T Lomax and S Turner (2010) *TTI's 2010 urban mobility report.* Texas Transportation Institute.

Small, KA and ET Verhoef (2007) *The economics of urban transportation*. London: Routledge.

Transportation Research Board (TRB) (1994) *Highway capacity manual, special report 209*. Washington: TRB.

Transportation Research Circular (TRB) (2011) 75 years of the fundamental diagram of traffic flow. *Greenshields Symposium*. TRB, June 2011.

Truong, T and DA Hensher (2003) *Congestion charging and the optimal provision of public infrastructure: theory and evidence.* ITS WP-03-01.

Turvey, R (1968) *Optimal pricing and investment in electricity supply: an essay in applied welfare economics*. London: Allen and Unwin.

Verhoef, ET (1997) Time varying tolls in a dynamic model of road traffic congestion with elastic demand. *Tinbergen Institute discussion papers 97-028/3*. Amsterdam.

Verhoef, ET, A Koh and S Shepherd (2008) Pricing, capacity and long-run cost functions for first-best and second-best network problems. *Tinbergen Institute discussion papers TI 2008-056/3*. Amsterdam.

Vickrey, WS (1969) Congestion theory and transport investment. *American Economic Review 59*: 251–269.

Zhang, L, DM Levinson and SJ Zhu (2008) Agent-based model of price competition, capacity choice, and product differentiation on congested networks. *Journal of Transport Economics and Policy Volume 42*, part 3: 435–461.

Appendix A: Previous estimates of the costs of congestion in Auckland

A1 Introduction

This appendix summarises and reviews previous studies that have assessed the costs of congestion in Auckland.

It examines work from the following four studies:

- *Land transport pricing study* (MoT 1997)
- *Alternative transport infrastructure investments and economic development for the Auckland region* (Ernst & Young 1997).
- *Surface transport costs and charges* (STCC) *study* (Booz Allen Hamilton 2004)
- *Auckland road pricing evaluation study* (MoT 2006).

Only two independent estimates of Auckland's congestion costs were found to exist:

- Work undertaken by the Auckland Regional Council in the mid-1990s, involving runs of the Auckland Regional Transport (ART) model for 1991, for the base (current) situation and a free-flow situation, with the resulting change in total vehicle hours being multiplied by standard values of travel time (from the NZTA (2010) *Economic evaluation manual* (EEM) to derive an annual estimate of travel cost (time) savings from eliminating congestion. This work is quoted and/or applied in sections A2, A3 and A5.
- Analyses undertaken by/for the 2004 STCC study, using the ART model for 2001, in a similar manner to the earlier work above. This work is described in section A4 and also referred to in section A5 below.

These two independent (but similarly based) analyses provide broadly consistent estimates. The STCC work is more recent and better documented, and we would therefore regard it as the most authoritative source on the topic. It provides estimates of congestion costs relative to a free-flow situation.

A2 Land transport pricing study (MoT 1997)

This report notes that the 'Auckland Regional Council has estimated that traffic congestion costs its region $570 million per annum'.

No information is given on the basis of this estimate.

This estimate appears to be the same one quoted in the Ernst & Young report (see below).

A3 Alternative transport infrastructure investments. . . for the Auckland region (Ernst & Young 1997)

A3.1 Report assessment

In regard to the costs of congestion, this report states (p57):

> *Our study of the economic cost of delays (direct and indirect) to the manufacturing and distribution sectors in Auckland amount to some NZ$185 million per year. Utilising the Auckland region transport model, the total cost of delay is estimated to be some $570 million*

*per year. This latter model does not fully incorporate the economic costs calculated earlier. It is therefore possible to add the two costs to arrive at a **total cost of congestion for the Auckland region of $755 million**.*

The report provided the basis for the **$185 million pa** figure, as follows (pp17–20):

- Auckland region total sales in the wholesale trade, manufacturing, building/construction and business/financial service sectors were estimated at $43.2 billion in 1994 (source not given). These sectors also had 194,500 full-time equivalent (FTE) employees.
- The study survey of a small sample of businesses in these sectors estimated that local freight costs averaged 2.5% of total sales.
- On a pro-rata basis, local freight costs were therefore estimated at $1.1 billion pa and about 4900 FTE employees were involved in local freight functions.
- Because of peak period congestion, it was assumed that local freight distribution was confined to seven hours per day rather than the potential eight hours available in the absence of congestion; this was translated into a requirement for an additional 12.5% truck capacity relative to an uncongested situation.
- It stated that variable costs accounted for about 66% of total freight capacity costs, so that the 12.5% additional capacity required would translate to a 9.4% increase in total freight costs in the region.[22]
- Taking 9.4% of the $1.1 billion total local freight costs, it stated that ‘This equates to an annual congestion cost to firms in the manufacturing, distribution and service sectors in the Auckland region of $103 million (pa)’.
- The ‘indirect added value’ associated with these costs was then calculated:
  - The $103 million loss in business profits as a result of congestion would be reflected in lower returns to land, labour and capital.
  - Net income to households was estimated at 43% of all added value. Hence the loss in profits would reduce household spending by $44 million pa.
  - Each $1 million reduction in household spending would result in an estimated loss of 16 jobs and $0.5 million loss in added value. Hence the increased freight costs would (directly and indirectly) lead to a loss of 700 jobs and $125 million in regional value added.
  - An allowance was then made for the downstream impacts (‘induced added value’) of these losses: ‘These induced impacts could add another $60 million in regional added value and another 350 jobs’ (no further details of these estimates were given).
  - The assessment concluded that:

The results from this are quite revealing – if the wasted time and subsequent inefficiencies suffered in the delivery of goods and services through the congestion on our roads over the peak periods could be reduced by 1 hour of the working day, the Auckland economy would

[22] Assuming the 12.5% additional capacity requirement is correct, the 9.4% increase in costs appears to be a substantial over-statement. For instance, fuel and truck repairs/maintenance have been included in the 66% variable cost proportion, but are unlikely to vary significantly if the transport task is spread over a 7-hour or 8-hour day. Similarly, labour is unlikely to vary pro-rata.

> *benefit by some $185 million annually and 1,050 additional jobs would be created in the manufacturing, wholesale, construction and financial services sector.*

In regard to the basis for the **$570 million pa** figure, the report is very brief, saying only (p57):

> *Utilising the Auckland regional transport model, the total cost of delay is estimated to be some $570 million per year.*

No further details are given in the report. However, our understanding is that this estimate represents the difference in network travel time costs for peak period traffic volumes between actual peak period operation and free-flow (4am) conditions, based on 1991 ART model runs and applying EEM values of time (further details not known).

The report goes on to state:

> *This latter model does not fully incorporate the economic costs calculated earlier. It is therefore possible to add the two costs to arrive at a* ***total cost of congestion for the Auckland region of $755 million.***

A3.2 Some comments

We understand the above statement that the 'total cost of congestion for the Auckland region (is) $755 million' has been very widely quoted since 1997, and more recently has usually been rounded up to the statement that the total congestion costs in the Auckland region are around (at least) $1.0 billion per year.

However, the validity and value of this estimate seem dubious, from several perspectives:

- The major component of the cost estimate, ie the $570 million pa figure derived from the ART model, is not accompanied by any information on the basis of its derivation – so it is difficult to form any judgement as to its validity and merits.
- It appears highly likely that the estimate represents the difference between the travel (time) costs for peak period travellers between actual travel conditions and free-flow conditions. Since there is no economic case for expenditure on a road system to provide free-flow conditions for peak period traffic, the usefulness of such an estimate is dubious (it is liable to be misinterpreted).
- The secondary component of the cost estimate, ie the $185 million pa relating to delays affecting the Auckland manufacturing and distribution sectors, is explained in some detail. However, some aspects of the calculations behind this estimate seem very 'rubbery', in particular:
  - The assumption that congestion levels increase the freight sector capacity requirement by 12.5% – the real figure could be significantly higher or lower than this.
  - The assumption that this capacity requirement would translate into a freight sector cost increase of 9.4% – it seems probable the true figure would be considerably lower than this.
- Further, the assertion that the two components of the delay cost estimates can be added together is not supportable[23].

[23] The report (p19) estimates the direct time costs of delays, using 'standard' CBA values and assuming wasted driver time of one hour/day, at approximately $15 million pa. By contrast, it notes the report estimates of congestion costs to firms of $103 million pa, and thus concludes that the regional model analysis does not fully incorporate the congestion costs to firms. However, it appears that essentially the same cost items are being assessed in each case, but the differences in estimates are primarily the result of different assumptions in the two sets of calculation.

We thus conclude the costs of congestion estimates given in this report:

- are inadequately documented, in regard to their major component
- are of dubious validity, in respect to their minor component
- provide an estimate of the costs of congestion relative to free-flow conditions, which is of limited merits in policy terms
- are based on assessments relating to conditions some 20 years ago (eg ART model runs for the 1991 situation), and thus are increasingly outdated.

A4 Surface transport costs and charges study: road congestion costs (Booz Allen Hamilton 2004)

A4.1 Introduction

The *Surface transport costs and charges study: road congestion costs* (STCC) for MoT (Booz Allen Hamilton 2004) included an assessment of both the total and average costs and the marginal costs of congestion for the three main urban centres and also for the inter-urban state highway network. The following summarises the work undertaken and the results for estimation of the recurrent congestion costs for Auckland.

A4.2 Methodology

The methodology used the ART model to compare travel costs for both peak and off-peak periods, and hence on an annualised basis, between:

- the actual base situation (2001), and
- a situation with minimal congestion, represented by applying only 10% of the base traffic volumes.

Time costs only were estimated, including allowances for the incremental values of time applying in congested situations (refer EEM vol 1, appendix A4.3).

A4.3 Results and comments

The methodology and results are set out in detail in the STCC working paper. Table A.1 provides a summary of the modelling results and the derivation from these results of the annual 'costs of congestion' for Auckland.

Table A.1 Summary of STCC estimates of Auckland recurrent congestion costs (2001 conditions, ART model runs)

	Base case (2001)	No congestion case	Difference
AM peak – per peak period (2 hours)			
Trips	345,127	345,127	-
Vehicle hours	89,784	59,199	30,584
Time cost ($)	1,543,535	895,771	647,763
Ave time cost/veh hr ($)	17.19[b]	15.13[a]	21.19
Minutes/veh km	1.488	0.989	0.499
Annualisation factor			500
Total annual costs ($ million)			**320.4**
Total annual veh hrs (million)			**15.1**
Interpeak – per interpeak period (2 hours)			
Trips	322,528	322,528	
Vehicle hours	58,314	49,900	8414
Time cost ($)	1,125,122	895,740	229,382
Ave time cost/veh hr ($)	19.29[b]	17.95[a]	27.26
Minutes/veh km	1.230	1.058	0.172
Annualisation factor			1,697
Total annual costs ($ million)			381.2[24]
Total annual veh hrs (million)			13.8
Annual aggregate			
Total annual costs ($ million)			**701.6**
Total annual veh hrs (million)			**28.9**

Source: Booz Allen Hamilton (2004, table 2.2 and appendix B).

Notes:

[a] Consistent with base values of time for urban arterial routes (peak and inter-peak) in EEM (table A4.3, July 2002 prices).

[b] Allows for the proportion of travel taking place in 'congested' conditions, with base values of time being increased to include for congested values up to the maximum of $3.88 (AM peak) and $3.60 (inter-peak) per vehicle hour.

The following points should be noted:

- The total annual time costs of congestion were estimated at some $702 million pa, relative to a free-flow situation.
- This relates to 2001 traffic conditions and is expressed in July 2002 prices.
- The cost estimate is split broadly evenly between the weekday peak periods (four hours/day) and all other weekday and weekend periods (including the peak 'shoulders').
- This cost represents approximately 28.9 million vehicle hours pa of delays (equivalent to about 24 hours pa average per Auckland resident)[25].

[24] Erratum 24 May 2013 – figure corrected to read 381.2

[25] Based on a 2004 estimate of population in the Auckland urban area of 1.22 million.

We also note that this $702 million pa estimate may be compared with ARC's earlier estimate of the costs of congestion, as quoted in MoT (1997) and the Ernst & Young (1997) study, of $570 million (excluding the additional $185 million pa estimated by the latter for economic costs to the manufacturing and distribution sectors). ARC's earlier estimate was based on ART model runs for 1991, but it is not known whether these earlier estimates used the same methodology as the STCC analyses (no documentation is available for the 1991 analyses).

However, the two numbers compare reasonably closely. The STCC 2004 figure ($702 million) is some 23% higher than the 1991 figure of $570 million. Arguably a somewhat greater increase might have been expected over the 10-year period, given traffic growth trends and increases in value of time (with inflation) over this period[26].

A4.4 Conclusions

The STCC analyses of the costs of congestion for Auckland adopt a generally sound methodology and are documented in some detail. They are therefore preferred to the ARC 1991 analyses; although these have also used the ART model, the methodology and calculations have not been documented (or the documentation is not available), and they are 10 years older in terms of the base transport model used.

Thus the STCC finding that the costs of congestion in Auckland total some $700 million pa (2001 traffic conditions, 2002 prices) is supported as the best estimate available. However it needs to be kept in mind that:

- these costs are measured relative to free-flow conditions
- they allow for time savings only and not for any other adverse impacts associated with congestion.

A5 Auckland road pricing evaluation study (MoT 2006)

The *Auckland road pricing evaluation* (ARPE) *study* did not make any new estimates of the costs of congestion in Auckland, but did summarise and comment on the earlier studies, in particular the Ernst & Young (1997) study (including the ART model 1991 results) and the STCC study (with its ART 2001 model results).

The ARPE study notes in particular:

- Both studies estimated the total economic costs to society of congestion (relative to a free-flow situation), but the actual costs to the commercial sector are likely to be only a minority proportion of the total costs.
- None of the five road pricing schemes assessed in the ARPE study will eliminate congestion. Based on the STCC analysis, a reduction in economic costs in the range $80 million – $96 million pa as a result of such schemes was suggested as potentially achievable.

[26] We can find no reference to the price base for the 1992 estimates while the STCC estimates are based on 2002 prices.

Appendix B: Unit parameter values for estimation of costs of congestion

B1 Overview

This appendix derives unit economic benefit parameter values required as inputs to the estimation of the economic costs of congestion:

- The study's main traffic/economic analyses provide estimates of the annual change in vehicle hours of travel between the base (2006) estimates and the optimised congestion situation.
- This appendix derives factors to be applied to this change in vehicle hours to estimate the annual change in economic costs in $ terms.
- It also provides an indicative analysis of the proportion of the monetary value of time savings that relates to business and commercial travel, and hence is likely to impact on business/commercial sector productivity and potentially national gross domestic product (GDP).

Our derivation of unit benefit parameter values has drawn on the NZTA (2010) *Economic evaluation manual* (EEM) where this provides appropriate information. Where such information is not contained in the EEM, we have made use of other work undertaken for NZTA (and its predecessors), in particular analyses undertaken for the New Zealand Patronage Funding project (Booz Allen Hamilton 2003).

The following sections of this appendix address:

- values of time savings, including allowance for reduced traffic congestion and improved trip reliability
- vehicle operating cost savings
- safety (crash) costs/benefits
- environmental costs/benefits
- cost proportion relating to the business/commercial sector.

B2 Unit values of time savings

The EEM methodology estimates travel time benefits (from an infrastructure, pricing, etc scheme) as the sum of three components (EEM vol 1, appendix A4):

- base travel time benefits
- travel time benefits from reduced traffic congestion
- travel time benefits from improved trip reliability.

We cover each of these in turn.

B2.1 Base travel time values

EEM (vol 1, table A4.3) gives unit value of time per vehicle hour by road type and time period.

We note the following values given there for urban arterials (all in July 2002 prices):

- AM commuter peak $15.13

- daytime interpeak $17.95
- PM commuter peak $14.96
- weekday all periods $16.83
- weekend/holiday $14.09
- all periods $16.27.

Given that most congestion occurs in the AM and PM commuter peak periods (but a not insignificant amount outside these periods), we take the AM peak figure of $15.13/vehicle hour as a representative unit base estimate (at July 2002 prices)[27].

B2.2 Incremental travel time values for reduced congestion

EEM (vol 1, A4.4) states: 'For all bottleneck delay, the maximum increment for congestion from table A4.3 should be added to the base value of travel time'.

This maximum increment for congestion is given (table A4.3) as $3.88/veh hour for the AM commuter peak.

As most of the congestion reduction between our base case and optimised case is likely to relate to bottleneck delays, we adopt this unit value. (An alternative approach of estimating this incremental value based on the EEM formula for road links, based on their V/C ratios, is likely to result in a rather lower incremental value, maybe in the order of half this $3.88 figure.[28])

B2.3 Incremental travel time values for improved trip time reliability

EEM (vol 1, A4.5) provides a set of procedures for estimating the benefits from improvements in trip time reliability: these procedures relate reliability in large measure to the V/C ratios on the links and intersections traversed.

The procedures are relatively complex to apply, and would require the running of a detailed traffic model. Instead, we have taken a short-cut approach, based on experience from more detailed studies where the incremental value of reliability benefits has been derived as a proportion of the base travel time benefits. We were advised that, typically, the incremental reliability unit value is in the range 5% – 8% of the base travel time unit value[29]. We have therefore applied a proportion of 6.5% in this case.

B3 Impact on vehicle operating costs

Previous work has examined the relationship between changes in travel time costs and changes in vehicle operating costs (VOC) (both on a ¢/km basis) at different average speeds, based on EEM travel time and VOC unit values (Booz Allen Hamilton 2003, appendix F).

---

[27] In undertaking the project, we have reviewed the original basis for the derivation, by NZTA, of the figure of $15.13/vehicle hour. In doing so, we identified what appears to be a calculation error: the corrected value should be $14.60 in place of the $15.13 EEM figure (refer email correspondence Ian Wallis/Sandy Fong NZTA, 3 Nov 2011). However, we have used the $15.13 figure in our subsequent calculations.

[28] The STCC analysis of the costs of congestion in AKL, based on AM peak runs of the ART 2001 model and the Auckland Central Motorway Interchange (CMI) 2000 model, derived an average increment of $2.06/veh hour. However, this was derived from comparisons between the base situation and a free-flow situation, so a higher value would be expected when comparing with the optimised congestion situation.

[29] Advice (personal communication) from Andrew Murray (Beca) 15 April 2011.

Over the range of speeds examined in that work, it was found that the change in VOC was 5.5% – 6.0% of the change in travel time costs. On this basis, we assume that the VOC component of the cost of congestion is 6% additional to the TT component[30].

This 6% estimate applies to the base value plus incremental congestion value of time, at 2002 prices, consistent with the method used for the original Booz Allen Hamilton (2003) estimate.

B4 Impact on safety (crash) costs

Increases in congestion are likely to result in some reductions in crash costs, for a given traffic volume (VKT) on the network. UK research indicates that this effect is likely to be quite significant: in urban conditions, accident rates appear to fall quite sharply with increased levels of congestion, and the proportion of accidents that are fatal or serious also falls.' (Booz Allen Hamilton 2003, appendix C).

At this stage no detailed attempt has been made to estimate the magnitude of the crash benefits associated with the congested situation (relative to the optimum situation). However, a 'back of the envelope' assessment is as follows:

- Average social costs of road crashes in the Auckland urban area over the five-year (calendar) period 2006–2010 were $768 million pa ($2010) – based on Crash Analysis System (CAS) data analyses.
- The proportion of this total related to the peak periods (defined as 7am – 9am and 4pm – 6pm, seven days/week) was 27.3%, ie $209.7 million pa – based on further CAS analyses.
- Deducting the Saturday/Sunday peak periods (assuming the crash costs at these periods are two-thirds of the peak figure per weekday) gives a factor of 0.79, leaving a cost of $166 million pa for the weekday peak periods.
- We assume this cost would increase in direct proportion to average speed if congestion were reduced to its optimum level. The speed increase (table 4.1) is from 44.2km/h to 46.8km/h, ie 5.9% and hence the crash cost increase would be in the same proportion.
- This represents some **$10 million pa** on the base cost of $222 million pa, for the optimum level.
- For the free-flow level, the speed increase would be from 44.2km/h to 70.9km/h, ie 60.4%; the corresponding increase in crash costs would be $100 million pa.

This amount represents an indication of the annual safety **benefits** associated with the base level of congestion, relative to the optimum level.

B5 Impact on environmental costs

B5.1 Greenhouse gas emissions (GHG)

Increases in congestion will result in increases in fuel consumption and hence in GHG emissions. We provide an order-of-magnitude assessment of this effect, as follows:

- Carbon emissions are valued at approximately 4% of total VOC, at 2008 prices (EEM vol 1, A9.6).
- The fuel cost component accounts for c. 50% of total VOC (EEM 1, A9.7).

[30] A figure of 7% was used in the STCC study.

- At the margin (for changes in travel speed), we would expect that most of any VOC change relates to changes in fuel consumption. Hence, at the margin, changes in carbon emissions are likely to be valued at around 8% of any changes in VOC.

We therefore apply this 8% proportion to estimate the unit value of changes in carbon emissions from the unit changes in VOC estimated earlier.

B5.2 Local environmental impacts

At this stage, no attempt has been made to value the changes in any local environmental impacts associated with the changes in traffic speeds between the base situation and the optimum situation.

Such local impacts might include noise, air quality and water quality. Previous research (eg Booz Allen Hamilton 2003) suggests that any marginal changes in these costs through changes in average traffic speeds of the extent under consideration in this study would be very small (relative to the other cost items that have been valued).

B6 Proportion of costs relating to business/commercial sectors

This section provides an indicative assessment of the proportion of the monetary value of time savings that relates to business and commercial travel, and hence is likely to impact on business/commercial productivity and potentially national GDP.

The detailed NZTA spreadsheets used in estimating the EEM unit values of time were re-analysed to derive the component of the overall value of time for urban arterial travel, AM peak ($15.13/vehicle hour in $2002) that related to business/commercial travel, including an allowance for the costs associated with freight time (relating to inventories, etc). The analysis showed that freight/business time accounted for some 47% of the total value of time savings.

Since time (including reliability) savings account for the great majority of the congestion-related costs, we estimate for this indicative analysis that some 40% – 50% of our total congestion-related costs relate to the business/commercial sector. This range accommodates our expectation that this sector will account for somewhat less than 47% of the non-time related costs (ie vehicle operating, environmental and crash costs).

B7 Summary

Based on the analyses in the previous sections, table B1 provides our estimates of unit benefit values for those cost items on which values have been placed. It is seen that (in $/vehicle hour, at July 2010 prices):

- The overall value is $28.32/veh hour.
- The great majority ($26.19/veh hr, 92.5% of the total) of this value relates to the travel time item, including its congestion and reliability components.
- The base travel time value alone ($19.82/veh hr) accounts for 70% of the overall total.
- The figure of $28.32/vehicle hour is applied to the estimated total change in vehicle hours between the base (2006 actual) situation and the optimum situation to derive our estimate of the costs of congestion.

In addition, an order-of-magnitude estimate has been included for the increase in crash costs in moving to the optimum situation (ie a benefit of congestion). This is around $10 million pa for the optimum situation, about $100 million pa for the free-flow situation, as derived in section B4).

Overall, around 40% – 50% of the total congestion-related costs are estimated to relate to the business/ commercial sector, and hence these costs (or cost savings) have the potential to flow on to productivity and GDP impacts.

Table B.1 Summary of unit benefit values

Item	**Base date**	**Unit value @ base date ($/veh hr)**	**Factor to July 2010[(a)]**	**Unit value @ July 2010 ($/veh hr)**
Travel time savings				
Base value	July 2002	15.13	1.31	19.82
Congestion increment	July 2002	3.88	1.31	5.08
Reliability increment	July 2002	0.98[(b)]	1.31	1.29
Sub total	July 2002	19.99	1.31	26.19
Vehicle operating costs	July 2002	1.14[(c)]	1.73	1.97
Environmental costs				0.16[(d)]
Total[(d)]				28.32

Notes:

[(a)] Taken from EEM vol 1, appendix A12.3.

[(b)] Taken as 6.5% of base TT value (refer section B2.3).

[(c)] Taken as 6% of TT savings (base + congestion increment) value.

[(d)] Taken as 8% of the VOC change (refer section B5.1).

[(e)] Excluding crash costs (refer section B4).

Appendix C: Long-run marginal cost analyses – details of recent Auckland road capacity enhancement projects

Table C.1 shows the costs and estimated peak contribution in terms of pcu capacity of a number of recent motorway capacity improvements in Auckland.

The numbers of pcu are calculated assuming each lane provides for 1800 pcu per hour for 2.5 peak hours.

Table C.1 Costs for recent road enhancements in Auckland

Works	Upper Harbour – Greenhithe	Upper Harbour-Hobsonville	North Western widening	Waterview	Mt Roskill Extension	Manukau Extension	Northcote Sunnybrook auxiliary lane	Victoria Park Tunnel St Mary's Bay	Newmarket to Greenlane
Length (km)	6.5	4.5 (western ring) 3.0 (north western extension)	13	4.7	4.0	4.5	4.4	4.5	2.3
Lanes	4	4	1 additional +3 km west only	4 lanes provision for 6	4	4 lanes expandable to 6	1 (north)	Wellington Rd - Victoria Park 2 Victoria Park 3, St Mary's Bay 1	1 southbound
Interchanges	1 diamond, 2 half-diamond	2 complex large diamond 2 half-diamond	7	2 + diamond	2	3			
Overbridges					2	5			
Other features				2.8km in tunnel		Widening local roads	Widening overpasses		
Cost ($ million)	110	220	100+	1400	201	210	9.8	406	13.7
Completed	2007	2012	2011	2015				2012	
Peak pcu-km delivered	117,000	151,200	130,500	126,900	72,000	121,500	19,800	30,600	10,350
Annualised cost per pcu-km	$0.38	$0.58	$0.31	$4.41	$1.10	$0.69	$0.20	$5.31	$0.53