Travellers' benefits of reduced congestion

By Camilla Riff Brems, Ph.D., manager ModelCenter

Danish Transport Research Institute
Knuth-Winterfeldts Allé
Build. 116 V
DK-2800 Lyngby
Denmark

Email: cab@dtf.dk
Phone: +45 4525 6557
Fax: +45 4593

Travellers' benefits of reduced congestion

By Camilla Riff Brems, Danish Transport Research Institute

Congestion does not only lead to increased travel time but also to increased travel time variability. Even though travel time variability imposes significant costs on the travellers, effects of variability are still rarely included in appraisal of infrastructure projects. This paper suggests an approach, which based on distributions of travel time and the averseness of being late, measures the additional travel time related to travel time variability. Results from a pilot study indicate that inclusion of these effects in project appraisal will increase benefits of projects reducing congestion with up to 25% compared to the present approach.

Introduction

Less congestion and reduction of travel times are often the most important benefits of initiatives to improve the transport system capacity either through infrastructure enlargements or traffic guidance systems. Benefits of reduced travel time are usually modelled by average travel time based on speed-flow relationships for road segments. However, initiatives which can reduce congestion are also reducing travel time variability, and consequently, the present approach is underestimating the potential benefits of improving the transport system. A more correct estimate of benefits should include both average travel time and travel time variability.

As a consequence of travel time variability, travellers going to make a journey with a preferred arrival time tend to assign some additional time besides the expected travel time to complete the journey. This additional time may be quite significant, when the preferred arrival time is in some way fixed, e.g. by check-in at the airport. Slack time, safety margin, and hidden waiting time are some of the names for the additional time, which increases with both increasing variability and with travellers' increasing aversion of being late. The latter varies among travellers and for different purposes of the journey. In this way, the additional time is expressing the, sometimes considerable, amount of time a traveller is willing to assign to a journey in order to avoid being late. And even then, all travellers will on some occasions be late, because situations with extraordinaire long travel times occur.

In order to cope with the travel time variability and, by time, be able to include effects of variability in project appraisal, travel time variability is subject of a number of research projects. Two different approaches characterise many of these projects: The scheduling approach, which originates in the travellers' planning of a journey, and the mean-variance approach, which originates in a system approach with observations of travel time on road segments. Even though, the origins are very different, recent research indicates that the two approaches need not contradict.

This paper starts out by reviewing some of the main results of both the scheduling and the mean-variance approach. Afterwards a methodology, which originates in the scheduling approach, is presented. This methodology is focusing on the traveller minimising the cost related to travel time by choosing an optimal travel time, which depends on the aversion of being late. Also, the methodology differs from former work by making no restricting assumptions on the distribution of travel time and by relating these distributions to the amount of traffic on the road sections rather than time of day. Finally, the methodology is applied to three different road sections in the Copenhagen area in order to illustrate the potentials of the methodology. The analysis shows interesting relationships for the travel time distributions, and it indicates potentials of including effects of travel time variability in project appraisal.

Theory of travel time and variability

It is well known that travel time varies by time of day, week and year. The sources for this variability are many, so it is important to have clear definitions when measuring travel time and travel time variability. The most predominant definition is stated in the work by Department for Transport (DfT) in the UK, see e.g. DfT (2003). According to DfT

- **Travel time** is the average time used for completing a specific journey or passing a given road section. The travel time consists of a free flow travel time and a time element of congestion, which is the expected delay. In this way the travel time includes e.g. effects of rush hours and seasonal effects.
- **Travel time variability** is the unpredicted variation in the time used to complete a specific journey or to pass a given road section. The travel time variability results from incidents and day-to-day variation, which is due to short term changes in either supply or demand for transport.

By this definition the primary effects of congestion are focused on travel time. However, empirical evidence indicates that congestion affects travel time variability as well, i.e. in the form of increased effects of both incidents and short term changes in supply and demand. Finally, congestion may affect the frequency of incidents.

In project appraisal, the cost associated to travel time and travel time variability can be measured in a number of ways. Although, there are two dominating approaches, the scheduling approach presented among others by Noland & Small (2002) and the mean-variance approach presented in e.g. DfT (2003).

Scheduling

The scheduling approach is summarised in the literature review of Noland and Small (2002), which refers a number of sources with the eldest dating back to Gaver (1968). The approach originates in the travellers' situation when dealing with travel time variability without reflection on the sources of the variability. Instead, the focus is on the fact that travel time variability makes the time necessary to complete a journey uncertain, so at some times the traveller will be early and at other times late compared to a preferred arrival time. Based on values associated with being either early or late the time element of travel cost can be measured by

$$C = \alpha\, T + \beta\, SDE + \gamma\, SDL + \theta\, D_L \qquad (1)$$

where T is the average travel time, SDE the expected scheduled delay early, SDL the expected scheduled delay late and D_L is a dummy for late arrival. Furthermore, α is the value of travel time, β the value of being early, γ the value of being late, and θ is a fixed penalty for late arrival. The relationship among the cost parameters are $\beta < \alpha < \gamma$.

Underlying this formulation is a distribution of travel times. The formulation includes both average travel time and travel time variability. However, it is possible to split the elements in the formula to obtain a formulation, which is consistent with the abovementioned split in travel time and travel time variability, see formula (2).

Based on the distribution of travel times, the actual valuation of α, β, γ and θ, and the preferred arrival time it is possible to determine the optimal travel time and by that the optimal departure time, which minimises the cost associated with the journey. Assuming an exponential distribution of travel times, Noland & Small (1995) derives the optimal expected travel cost, $C*$, corresponding to the optimal departure time as

$$C* = \alpha(T_f + T_x + b) + \theta P_L^* + b\left\{\beta \ln\left[\frac{\theta + b(\beta+\gamma)}{b(\beta - \alpha\Delta)}\right] - \frac{\theta(\beta - \alpha\Delta)}{\theta + b(\beta+\gamma)} - \alpha\Delta\right\} \quad (2)$$

where T_f is free flow travel time and T_x is expected delay (due to recurrent congestion) so that T_f and T_x are reflecting the travel time in the definition above. Furthermore, b is the unexpected delay (due to non-recurrent congestion) as well as the variance of the exponential distribution, and Δ is a measure for the change in profile of the recurrent congestion.

This expression gives a formula for calculating the optimal cost related to travel time and variability. It does not seem very operable and consequently, by assuming $\theta = 0$ and $\Delta = 0$, Polak (1996) has derived the following more simple expression

$$C* = \alpha T + \beta \ln\left(1 + \frac{\gamma}{\beta}\right) b \quad (3)$$

Where T and b are mean and variance respectively in the distribution of travel time. This expression might be interpreted as a parallel to the mean-variance approach presented below.

Mean-variance approach

The second approach is the mean-variance approach used in a number of studies by the UK DfT and most recently applied to data from the Stockholm road pricing experiment by Eliasson (2006). The mean-variance approach is based on measuring travel time and variability on road sections, i.e. a system approach, consequently having more focus on operability and the origins of the variability. As a result, the mean-variance approach does not focus on how travellers act in a system with travel time variability.

The mean-variance approach focuses on the causes of the travel time variability. In this way the approach distinguishes between variation due to incidents and day-to-day variation. Expected effects of congestion are, as stated in the first definition of this section, included in the specification of the travel time.

The different reports from DfT finds that cost of travel time and variability can be approximated by a linear expression of the travel time and its standard deviation.

$$C = \alpha t + \beta \sigma \quad (4)$$

where t is the travel time, σ is the standard deviation, and α and β are parameters.

The approach was applied to data from the Stockholm road pricing experiment. Here, Eliasson (2006) shows that the cost of travel time variability counts for an additional 15% compared to expected travel time. Also, the cost of travel time variability counts for 15% of the total benefits of the road pricing scheme.

Methodology

The methodology applied here for the determination of cost of travel time and its variability is based on the scheduling approach originating in the travellers' choice of departure time for a given journey. For each journey a traveller is considering when to leave home (or the origin) in order to get to the destination on the preferred arrival time. In some cases the preferred arrival time is fixed, i.e. it is very important to be there on time, and in other cases the preferred arrival time is 'preferred' but not critical. Only in rare occasions is a traveller truly indifferent with respect to arrival time.

In order to determine the 'best' departure time the traveller is now estimating the travel time for a given journey based on experience and on the present time of day and year, i.e. the traveller is making (maybe unconsciously) assumptions about the distribution of travel times. The difference between the 'best' departure time and the preferred arrival time is the optimal travel time, which minimises the costs of travel as expressed in formula (1).

Below this process of choice is put into mathematical expressions, and the preferred travel time is determined based on distributions of travel time and averseness of being late.

Time related cost

The cost of travel time and variation with respect to a preferred arrival time is according to Noland & Small (2002)

$$C = \alpha\, T + \beta\, SDE + \gamma\, SDL + \theta\, D_L$$

Empirical evidence shows that travel time as well as possible early or late arrival depends on the distribution of travel time. However, it is not the clock-time, which determines the effects of travel time but rather the amount of traffic on that specific time. Thereby, the methodology is related to flow as can be seen in the examples in the application part of this paper. Here, it becomes evident that distributions of travel time and variability are very much related to the actual flow level, as is seen for the traditional speed-flow relationship.

So far the referenced literature has dealt with special cases of these distributions e.g. the exponential distribution. However, it is possible to obtain the results with no other assumption than the distribution being a probability distribution. With $f_\nu(t)$ denoting the distribution of travel time for flow level ν the size of T , SDE , SDL and D_L can be written as four integrals expressing the expected values depending on the distribution. As both SDL and D_L refers to late arrival this is expressed in one integral in formula (5), the cost associated with a travel time (t) at flow level ν .

$$C_\nu(t) = \int_{\underline{t}}^{\infty} \alpha(\tau) f_\nu(\tau) d\tau + \int_{\underline{t}}^{t} \beta(t-\tau) f_\nu(\tau) d\tau + \int_{t}^{\infty} (\gamma(\tau - t) + \theta) f_\nu(\tau) d\tau \qquad (5)$$

The first integral is expressing the cost of travel ($\alpha\, T$) with $\underline{t}$ denoting the free flow travel time, and the second integral the cost of early arrival ($\beta\, SDE$) with the share of the distribution sooner than t . Finally the last integral is expressing the cost of late arrival ($\gamma\, SDL$) including the penalty ($\theta\, D_L$) for the share of the distribution later than t .

As before, α is the value of travel time, β the value of being early, γ the value of being late, and θ is a fixed penalty for late arrival, with $\beta < \alpha < \gamma$ as the relationship among the cost parameters.

Optimal travel time

Determining the preferred departure time means minimising this cost function. The associated travel time is denoted t^* and can be derived by differentiating $C_v(t)$ with respect to t. Assuming $\theta = 0$, the following rather simple expression can be obtained

$$\frac{dC_v(t)}{dt} = \int_{\underline{t}}^{t} \beta\, f_v(\tau) d\tau - \int_{t}^{\infty} \gamma\, f_v(\tau) d\tau$$

$$= \beta\, F_v(t) - \gamma\,(1 - F_v(t)) = 0 \qquad (6)$$

where $F_v(t)$ is the cumulative distribution corresponding to $f_v(t)$. Manipulating (6) shows that t^* can be determined by solving

$$F_v(t^*) = \frac{\gamma}{\beta + \gamma} \qquad (7)$$

The interpretation of this is that the optimal travel time, t^*, can be determined as a quartile of the distribution of travel times. By that, it is sufficient to know the travel times associated with specific quartiles of the distribution. Which quartile to use depends on the relative values of β and γ, i.e. on the travellers' averseness of being early relative to being late. If for instance γ is twice the value of β, the 67%-quartile is used and if γ is four times the value of β the 80%-quartile is used.

Even though this methodology incorporates both distribution of travel time and travellers' averseness of being late, the traveller will still be late from time to time. If the traveller repeats the same journey a number of times, using the 50%-quartile (the median) means that the traveller will be late half the times, whereas using the 80%-quartile the traveller will 'only' be late 20% of the times.

With the optimal travel time, t^*, the expected cost of the journey is

$$C_v(t^*) = \alpha\, T^* + \beta\, DSE_{t*} + \gamma\, DSL_{t*} \qquad (8)$$

The practical effects of these results are illustrated for a pilot study in the next section.

Empirical results

Even though several of the references for travel time variability are dated back 10 years or more, travel time variability is rarely included in appraisal of infrastructure projects. The empirical results to be presented here show that inclusion of travel time variability may improve the potential benefits of project appraisal.

Data and basic relations

The methodology described above is applied to three different types of roads based on data from a pilot study in and around Copenhagen. The three road sections include

- a **ring motorway** around Copenhagen from north to south-west with a speed limit of 110 km/h. This is the M3, which has marked peaks morning and afternoon with slightly more northbound traffic in the morning and southbound in the afternoon. It is a four lane motorway, which is currently being extended to 6 lanes. The entire section is 18 km.

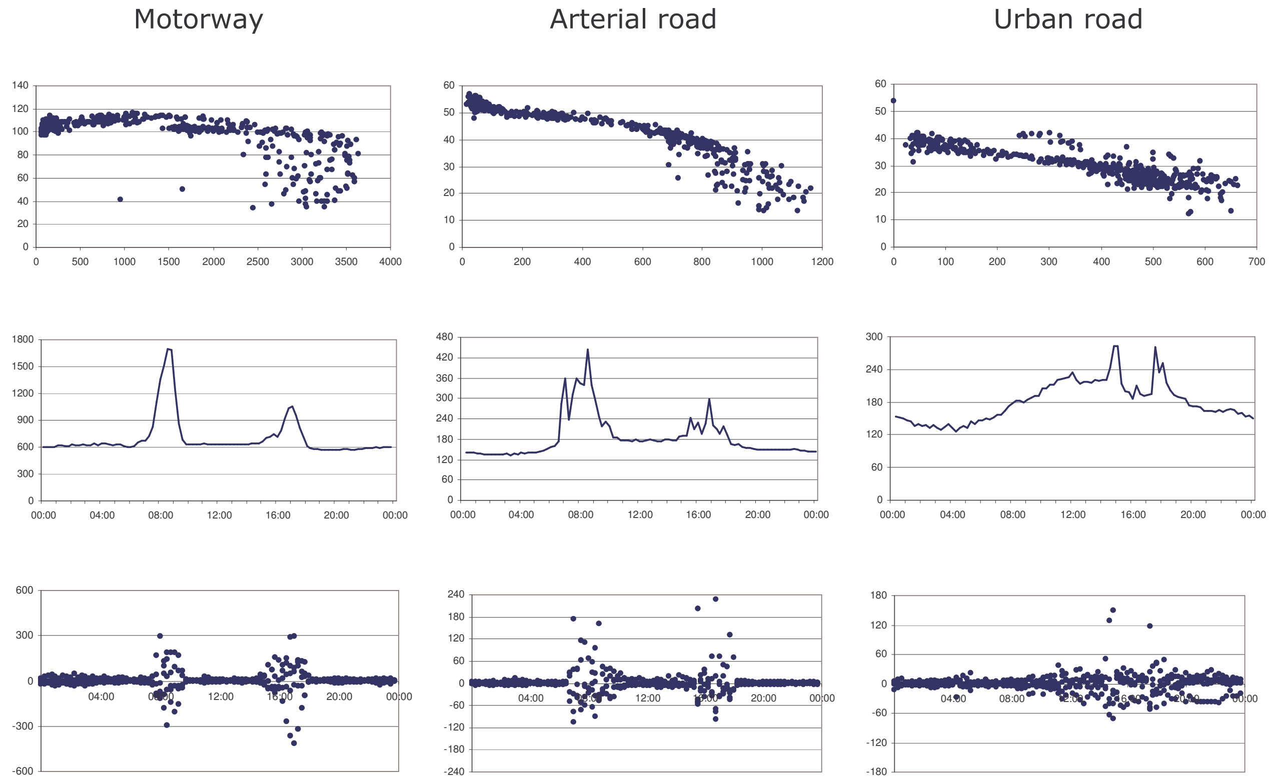
- an **arterial road**, which is the extension of a motorway towards the centre of Copenhagen. The road section has clear directions in the peaks, inbound in the morning and outbound in the afternoon. It is a four lane arterial road with a speed limit of 60 km/h. In the inbound direction it is reduced from two lanes to one due to a right turn lane causing congestion each morning. The length of the examined road section is 2 km and includes 7 signalled intersections.

- an **urban road** near the centre of Copenhagen. This is characterised by no marked peak but shopping traffic throughout the day with lots of goods delivery. It has two wide lanes, which are often reduced in width due to stopping lorries. The length of the road section is 1.4 km and includes 6 signalled intersections and a number of un-signalled intersections.

The data collection took place for one week in May 2001. For each type of road there were simultaneous registrations of flow and travel time for each direction and lane separately. For the arterial and urban road sections flow was registered by electronic traffic counts (including speed) with less than 500 m in between. Flows were recorded on 15 minute intervals for the entire week. Travel times were recoded for selected intervals of time on different weekdays by registration of license plates when vehicles entered and left the road section. These travel times ware used to calibrate formulas for calculating travel time based on speed and flow as well as green times at intersections. For the ring motorway electronic traffic counts with speed registration is placed before and after every ramp and in between, is the distance exceeds 1 km. Registrations were aggregated for 15 minute intervals for the entire week. Travel times were calculated based on these registrations. Formulas for calculation have been calibrated against observed travel time prior to the data collection. The analysis here is based on the five weekdays only.

Figure 1 below illustrates some of the main characteristics of these three road sections in the form of speed-flow relations, average travel time and variability over the day. Please note, that the speed in the speed-flow relationships is average travel speed including stops at intersections. Each of the relationships shows the expected form with respect to bending and spreading of the observations. As an example, the well-known backwards bending is clear for the motorway, whereas it is more blurred for the arterial and urban roads due to the effect of intersections, which are restricting the capacity for these roads.

Figure 1 **Illustration of travel time and variability for one direction of each type of road. The units for the first group of figures are flow and travel speed, for the second group time of day and average travel time in seconds, and for the third group time of day and travel time variability in seconds.**



Note: The data represents the northbound direction for the motorway, the inbound direction for the arterial road and outbound direction for the urban road.

Regarding travel time and variability figure 1 shows a clear relationship between higher average travel time and higher travel time variability in the peaks. This is a well known relationship for travellers using the transport system. The relationship for average travel time (or speed) is usually expressed by the speed-flow relationship whereas the corresponding relationship for travel time variability is not widely included in analyses.

Figure 2 **Travel time variability for the arterial road as a function of flow. Units are flow and travel time variability in seconds.**

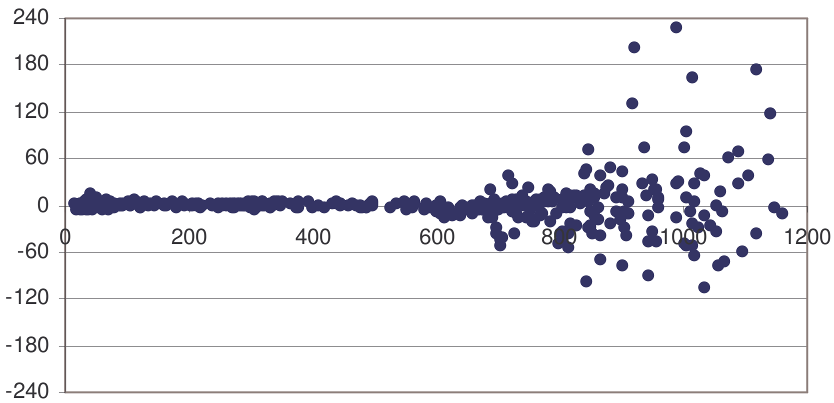


Figure 2 shows that there is a clear relation between increasing flow and increasing travel time variability even when the average travel time is removed, including the part which is due to congestion. The data in figure 2 represents the same observations as for the arterial road above, but now the travel time variability is plotted against the observed amount of traffic. Furthermore, the figure shows very little travel time variability for flows below 650 vehicles per hour. The variability increases significantly when the flow exceeds 8-900 vehicles per hour, corresponding to the capacity of the restricting intersections.

Travel time distributions

In order to analyse the travel time variability in relation to flow the distributions of travel time for given levels of flow have been drawn. These distributions correspond to $f_v(t)$ in the methodology above, and they include the entire travel time including both average travel time and variability.

Figure 3 Distributions of travel time for the arterial road. Units are travel time in seconds and probability.

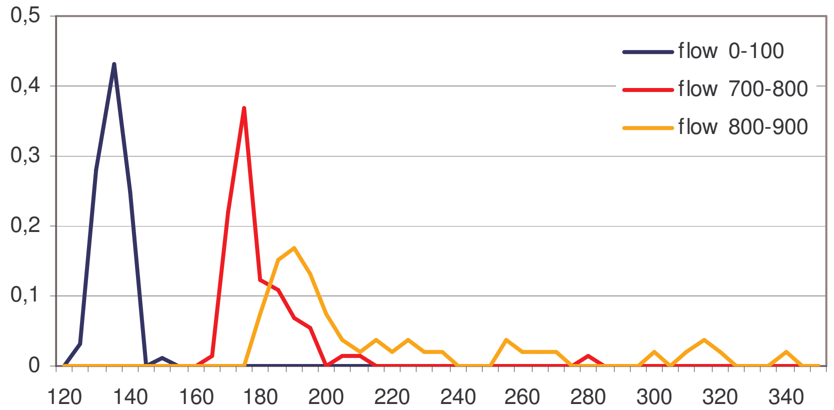


Figure 3 is worth paying some attention as it illustrates the dynamics of congestion and thereby the main problem for travellers.

1. When flow levels are low relative to the capacity of a road or road section, the experienced traveller will have a good sense of the necessary travel time. The distribution is narrow and close to the free flow travel time of the road. Some travellers may even observe travel times below free flow as they disregard the speed limits. In this case the average travel time is 135 sec (54 km/h) with maximum observed time going up to 150 sec (an additional 11%).

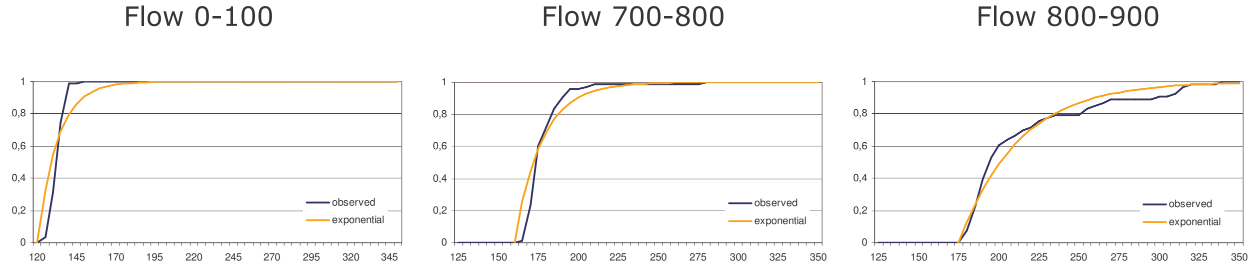
2. When traffic is building up travel times increase, as it is no longer possible to travel at speed limit. In the distribution the minimum observed travel time increases. Furthermore, there are occasions of longer travel times, which show in the distribution as a longer right tail. Now the average travel time is 180 sec (40 km/h) with maximum observed time going up to 280 sec (an additional 56%).

3. When the traffic reaches the level of the capacity of the restricting intersections the shortest travel time becomes even higher. Furthermore, the distribution of travel times becomes rather flat with a long thick right tail. This makes it very difficult to predict the actual travel time, and for travellers with a preferred arrival time it is necessary to assign considerable time in order not to be late. In this case the average travel time is 216 sec (34 km/h) with maximum observed time going up to 340 sec (an additional 57%).

Similar relationships of flow and travel time distributions can be seen for the motorway and urban road.

In order to be able to use these observations in practice it is important to be able to generalise some of these relationships like it is done for the speed-flow relationship. As a step in that direction the distributions of figure 3 has been fitted to exponential distributions in figure 4.

Figure 4 **Fitting observed distributions to exponential distributions for the arterial road. Units are travel time and probabilities.**



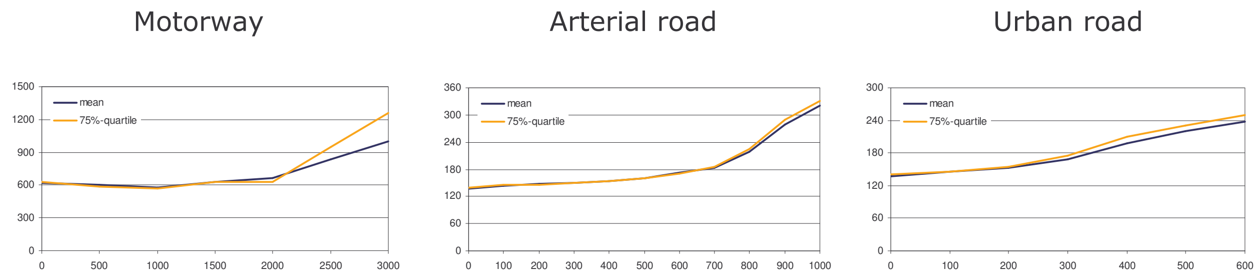
The figures show that the distributions of travel time fit well with an exponential distribution. Especially, for an arterial road just below capacity (the figure in the middle). For very low levels of flows the observed distribution is narrower than the exponential distribution, whereas it is wider and has a longer right tail for higher flow levels. Similar patterns of fit have been found for the other road types.

Cost of travel time variability

Calculating the cost of travel time variability requires that the optimal travel time is determined first. According to formula (7) the optimal travel time can be found based on valuation of being early and late. Noland & Small (2002) tests a number of model formulations regarding valuation of earliness and lateness, and when θ is not included the results indicate that γ is around 3 times the value of β. In practice this means that the travellers prefer to be up to 15 min early rather than 5 minutes late, and that the 75%-quartile of the travel time distribution should be used to determine the optimal travel time.

Figure 5 shows optimal travel times for given flow levels for the 75%-quartile for each of the three road types. These are compared to the average travel time, which is used traditionally.

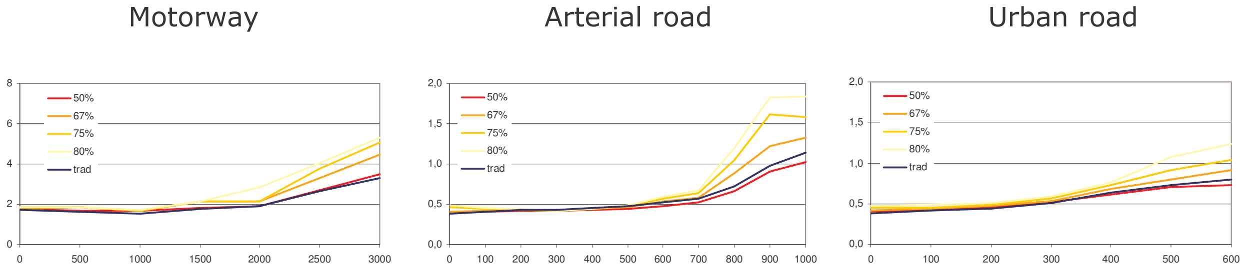
Figure 5 **Mean and optimal travel times for different levels of flow. Units are flow and travel time in seconds.**



For the arterial and urban roads there is very little difference in the travel time for the mean and the 75%-quartile. The main reason being, that the distributions have very long thin tails which tend to increase the mean. When compared to the median, the 75%-quartile shows travel times that are approximately 15% longer. For the motorway the figure shows a more significant difference with optimal travel times being up to 25% higher than mean travel times and above 50% higher than the median. Common for all road types are that the travel time is almost identical when the level of flow is well within the capacity of the road.

The cost associated with travel time and variability can now be determined based on the optimal travel time. The relations between the different time elements are given by β and γ, which are varying for different quartiles, whereas the value of travel time is set to 10 $ per hour. With these assumptions figure 6 shows how cost of travel time and variability in increasing with higher quartiles. This is not surprising, since both the longer optimal travel times and the higher values of late arrival contributes to the cost.

Figure 6 Cost of travel time and variability compared to traditional travel cost. Units are flow and travel time costs in $ for passing the road section.



As for the former figures, figure 6 shows that there is little difference between the different quartiles for lower levels of flow, whereas the differences increases as the flow level tends to reach the capacity of the road segments. For the arterial road the cost seems to decrease for the highest level of flow, but this can be related to relative to the form of the speed-flow curve. Better estimates may be obtained by measuring density rather than flow.

For all three road types the cost is compared to the traditional travel cost, which is valued by 10 $ per hour for free flow travel time and 15 $ per hour for delay, which by definition is all time exceeding free flow travel time. Figure 6 shows that the traditional approach has the same pattern as the suggested methodology with the traditional cost most similar to the 50%-quartile (the median).

Finally, the average time related cost of a traveller is generated based on the current distribution of traffic across flow groups. These costs reflect the effect of travel time variability compared to the traditional approach and the effects of different levels of risk averseness.

Table 1 Average time related cost with the current distribution of traffic. Unit is $ per km.

	Traditional	50%-quartile	67%-quartile	75%-quartile	80%-quartile
Urban road	0.40	0.40	0.44	0.48	0.52
Arterial road	0.28	0.27	0.31	0.35	0.38
Motorway	0.11	0.12	0.13	0.14	0.15

Table 1 shows that by including the effects of travel time variability, the average cost increases considerably. When using the 75%-quartile as indicated by the valuation of β and γ in Noland & Small (2002) the costs increases by 18% for the urban road and 25% for the arterial road and the motorway compared to the present methodology for evaluating the cost of travel time.

With such increases in time related cost by the inclusion of travel time variability and with the rather steep cost increases for higher levels of flow, there is potential for obtaining higher benefits of projects, which in one way or the other reduces the congestion and improves the use of the infrastructure.

Conclusions

The paper has presented the two dominating approaches for measuring the cost related to travel time and variability. Based on the scheduling approach the paper has presented a methodology, which uses the valuation of travel time, early arrival and late arrival to determine the optimal travel time for a traveller given a preferred arrival time and distributions of travel time. As a general result the optimal travel time can be determined from the valuations of early and late arrival and the cumulative distribution of travel times. In this methodology the distributions of travel time are related to flows, but otherwise the methodology does not make any further assumptions as to the form of the distributions. However, the methodology can still be improved in a number of ways. E.g. θ should be included in formulas, and the general concept might improve by more investigation into the effects of building up and dismantle of queues and the effects this has on the distributions. Furthermore, the empirical findings may improve by measuring the level of traffic in density rather than flow.

The methodology is applied to data from a Danish pilot study, which includes corresponding observations of travel time and flow. Analysing these data illustrates some important relations between flow and travel time. One important relation is that travel time variability increases with increasing levels of flow. In an effort to describe the patterns of the travel time distributions, the observed distributions are fitted with exponential distributions. In this field there is still some work before it might be possible to express the parameters of the exponential distribution by general variables of the road sections. These generalisations are necessary, if the methodology should be used in broader perspectives and for e.g. forecasting, as the present form is rather data requiring.

Applying the methodology on the data shows how the optimal travel time and the cost increases with risk averseness. Furthermore, it makes it possible to compare the suggested approach with the traditional approach. Using the valuation of earliness and lateness suggested by Noland & Small (2002) on the road sections in the pilot study indicates that time related cost may increase by 25% when taking travel time variability into account. Since many infrastructure projects have little effect on the actual travel time and much more effect on the travel time variability, using methodologies as the suggested may increase the benefits of such projects considerably.

Acknowledgements

The data used for this work was made available by a Danish research project on congestion financed by the Danish Ministry of Transport and Energy. The first steps for the work were taken in that project by the author in collaboration with Professor Birgitte Sloth, University of Southern Denmark, and Director Niels Buus Kristensen, Danish Transport Research Institute.

References

Brems, Buus Kristensen & Sloth (2002) Congestion Costs, proceedings to European Transport Conference, Oct 2002, Cambridge, UK

DfT (2002) Journey time variability – Modelling and appraisal of journey time variability, report prepared by Arup, UK

DfT (2003) Frameworks for Modelling the Variability of Journey Time on the Highway Network, report prepared by Arup, UK

DfT (2004) Advancing methods for evaluating network reliability, report prepared by ITS, Leeds University, UK

Eliasson (2006) Forecasting travel time variability, proceedings to European Transport Conference, Sept 2006, Strasbourg, France

Gaver (1968) Headstart strategies for combating congestion, *Transportation Science*, Vol. 2, pp. 172-181

Noland & Small (1995) Travel-time uncertainty, departure time choice, and the cost of morning commutes, *Transportation Research Record*, Vol. 1493, pp. 150-158

Noland & Small (2002) Travel time variability: a review of theoretical and empirical issues, *Transport Reviews*, Vol. 22 (1), pp. 39-54.

Polak (1996) An overview of the recent literature on modelling the effects of travel time variability. Working paper, Centre for Transport Studies, Imperial College, London, UK