Route information from a central route planner

G. Hek, G. Lunter, J.A.M. Schreuder, J. White

Abstract

We present a discussion of a problem posed by researchers of the company Ericsson, namely, to estimate the fraction of the road users in a road network that must participate in a central route planning scheme such that travel time predictions improve significantly. A road user who participates is expected to inform the central route planner of his intentions to travel from an origin to a destination and is expected to travel along the route advised by the planner.

The aim of this work is to derive a measure of travel time performance depending on the number of road users who are participating in the central route planner. The approach is mainly of a statistical nature.

Keywords

Central route planner, Traffic flow estimation, Traffic flow control, Nash equilibrium.

1 Introduction

Ericsson is interested in developing a "central route planner". The function of a central route planner (CRP) is to advise road drivers on journey routes. Specifically, before travelling from one location to another, a driver uses the telephone to query the central route planner, which tells the driver the fastest route to take, an estimated journey time and possibly other information such as reliability estimates or worst-case scenarios. Ericsson must decide how the central route planner will calculate the routes and times it distributes, and which of the various available sources of data providing information on traffic flow they should use in making these calculations.

In particular, Ericsson would use historical data on traffic densities (possibly correlated with variables such as the day of the week, season, and weather forecast). However, Ericsson has also considered using the number of user queries themselves, for a given particular route, in addition to this historical data. The traffic forecasts made by the central route planner may be improved by doing this. Ericsson would like a measure of the "improvement" and wants to know how this "improvement" would depend on the percentage of drivers using their service. Their main question to us was: "How many user queries does Ericsson need to significantly improve upon the historical data predictions?"

At the moment Ericsson knows very little about this type of problem, and wants some advice on various issues. Perhaps not surprisingly, given the economic importance of efficient road networks and the current traffic jam problems, behaviour of traffic on road networks has been studied greatly. In section **??** we give a sketchy overview.

Ericsson's problem has many different aspects, some of which are discussed in section **??**. Because of this scope, solving Ericsson's problem in full generality proved impossible. To get anywhere we had to make strong simplifications and restrictions, and ignore several important aspects of the problem. The simplifications we chose to make are discussed at the start of section **??**, and the rest of that section is devoted to a statistical model of the relation between CRP users and travel time.

One part of this model is a relation between number of drivers on the road, and average travel time. In section **??** we use a simplified version of a traffic model we found in the literature to obtain a reasonable-looking approximating formula.

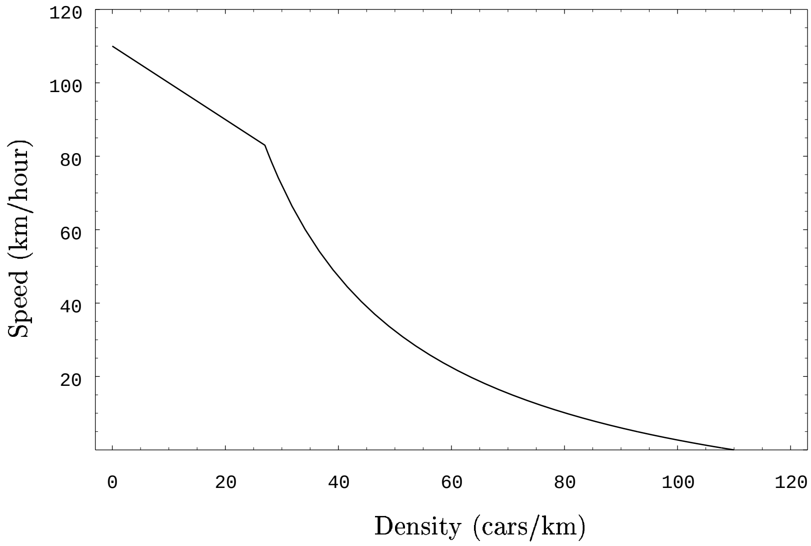


Figure 1: Smulders' function, relating density and equilibrium speed of traffic flow.

Section **??** gives concluding remarks. In our opinion, our analysis and the many different aspects of the problem that cannot be influenced indicate that a large number of users is necessary to significantly improve upon historical data predictions. We doubt whether such large numbers of drivers will cooperate in a central route planner system, and we think that Ericsson's idea will probably not be feasible.

2 Literature

The following is a selection of literature we used for this report. More references can be found in e.g. [**?**] and [**?**].

In [**?**] a traffic flow model is developed, inspired by fluid dynamics. The starting point is a set of partial differential equations. These equations are discretised, and terms are added to model the tendency of drivers to look ahead, and to adapt their speed to other traffic.

In [**?**] this model is further developed. The notion of "equilibrium speed" is introduced, and a relation between traffic intensity (cars/km) and equilibrium speed is proposed. This speed, as a function of intensity, is continuous and monotonically decreasing, with a sudden steep drop at a critical density, marking the onset of traffic jams; see figure **??**. This work also deals with dynamically influencing traffic flows. The analysis concerns a single highway segment. Influencing traffic flows is also studied in [**?**]. Methods for dynamic routing or dynamic traffic assignment, as opposed to static traffic assignment through signposts, are discussed.

A system theory approach to route planning, using both deterministic and stochastic models, can be found in [**?**]. The emphasis is more on the network, and less on modelling the traffic flow on single road segments, for which rather crude models are used.

Currently in the Netherlands, enough real-time information on traffic densities is available to make short-time density predictions feasible. A research group is using a model described in [**?**] of traffic on the Amsterdam ring, based on [**?**] and ideas from dynamic game theory, to predict densities a few hours in advance [**?**]. These predictions may be used to drive an electronic messaging system.

In [**?**] the idea of combining a toll-system (*rekeningrijden*) and advance booking was brought up. It contains elements of Ericsson's idea, and was investigated by the engineering consultants Niema, who concluded that the idea could be a "worthwhile contribution" to solving the traffic jam problem. Currently the idea is being discussed with several parties involved; see also [**?**].

3 Some definitions

We define here a few terms that we shall use throughout. A *user* is a road user who has telephoned Ericsson's central route planner, and is following the advice provided. A *network* is a directed graph of road segments (in practice, each edge usually has a complementing one going in the other direction). We use the word *road* to denote a single edge in the network, that is, a segment without crossings, on- or off-ramps. For each edge or abstract road we define the following quantities, some of which are time-dependent:

- Length, L; corresponds to length of road.
- Number, N; corresponds to number of cars on road.
- Density, $D = N/L$; corresponds to number of cars per kilometre on a given road. For each road, we assume homogeneity.
- Intensity, I, also called 'traffic flow'; corresponds to number of cars driving past a certain point per hour.
- (Average) speed, S; average number of kilometres cars travel per hour on a given road.
- Capacity; maximum intensity, reached for some optimal density and speed.
- Travel time, $T = L/S$.
- Equilibrium speed; speed of traffic in stationary state, at a certain density.

Detailed models differentiate between equilibrium speed and the actual current (average) speed. In the models we use we shall not make this distinction.

By *network flow parameters*, we mean capacity, density, intensity and average speed for each road in the network. Capacity clearly depends on the *type of road*, i.e. number of lanes, highway or not, speed limits, etc. It also depends on *road conditions*, by which we mean all variables that influence the capacity of a road such as weather conditions, daylight, construction work, accidents on the road (or on the road going in the other direction, which may create a *kijkfile* (spectator jam). Finally, *routing information* is the information a user obtains when he[1] calls the central route planner. This information includes the estimated time his planned journey will take, the time at which he has to leave, and the reliability of the estimate.

4 Aspects of the problem

In this section we identify and comment on several aspects of the traffic estimation and routing problem. Many of these aspects have not found a way into the proposed model; however, we believe they are all relevant, and important to keep in mind when a decision about follow-up research on the CRP is to be made.

4.1 Two main approaches to the traffic problem

Vaguely put, the goal of traffic routing is to make more efficient use of available network capacity. Two main approaches may be identified (see also [?]), which we dub the *top-down* and *bottom-up* approach.

- Top-down: Guides traffic so that the total capacity of the network is maximized at a global optimum;
- Bottom-up: Provides users with accurate information and predictions, enabling them to choose the most efficient (fastest) route, i.e. every user is in a local optimum.

[1] Whilst not making strained efforts for political correctness, we are aware of the existence of female drivers.

The first is the approach taken by the government, when they try to influence traffic, for instance by imposing speed limits. The second approach is taken by individual drivers, when they choose departure time and route in order to encounter fewer jams.

Note that the two approaches use different notions of optimality. Government is interested in average throughput, an individual user is interested primarily in his own travel time. And indeed, these differing notions give rise to different optimal configurations. For example, within some bandwidth, imposing a speed limit results in a higher road capacity (the number of cars per minute flowing through), although individual cars take longer to arrive. Other even more counterintuitive situations may occur; see section **??**.

4.2 Historical data and users information

We assume Ericsson has access to historical data on the network flow parameters. It is unclear to us whether this assumption is justified. We believe that a systematic, detailed and extensive database of past network usage is vital for predicting the traffic flow, and running a CRP service. Rijkswaterstaat routinely compares the traffic flow and speed of the flow with values of the recent past. Each traffic control center has a module for this.

The problem of predicting traffic flow using historical data alone is not trivial. A reasonable idea seems to identify independent, explaining variables, and use these to look up relevant past traffic situations. The independent variables would include at least the day of the week, the time of the day, the season and weather (forecasts). The database would provide an estimate of traffic intensity, as well as an estimate of road conditions, together with the resulting traffic densities.

Ericsson's extra source of information are the *user queries*. The information on users consists of the information about their current travel plans, and probably also of a database containing the travel histories of all users. This information may be used to improve the traffic intensity estimate. Using the database or a traffic model, this can subsequently be related to an improved traffic density estimate.

As the extra information contained in the user queries would almost certainly be related to network usage (intensity), rather than road conditions, it is important to compare how traffic flow varies with varying network usage, as compared to the variation due to varying road conditions. See section **??** for more remarks.

To determine the effect that using the extra information contained in user queries will have on the precision of estimated journey times, it is vital to look at the relationship between the number of drivers on a road, and the number of user queries (pertaining to this road). There may not in fact be a useful relationship between these quantities, for the following reasons.

Some proportion of the drivers may never call, for instance because they are bound to fixed departure times, or simply because of unfamiliarity with the system. People taking the same route regularly may not bother to call (often), especially if the proposed route and departure times do not vary much. An assumption on the relationship, for instance that a fixed percentage λ of the drivers call Ericsson, will therefore probably not hold in general, but might however hold for the group of occasional drivers.

The CRP can provide best routing information after all user queries are collected. This would mean that users must call twice to obtain the requested information, which is not practical. Moreover, they must plan their journey well in advance, which may not be possible or desirable for everyone. Alternatively, the routing information may be continuously updated, with the system always giving the latest predictions. This has the disadvantage of rewarding late callers, reducing the effectiveness of the system.

Users may also give unreliable information, particularly if this information must be given well in advance, or ignore the CRP's advice, causing further problems. User input will hardly influence the information they get from the CRP, so that there is little incentive for the user to be very precise.

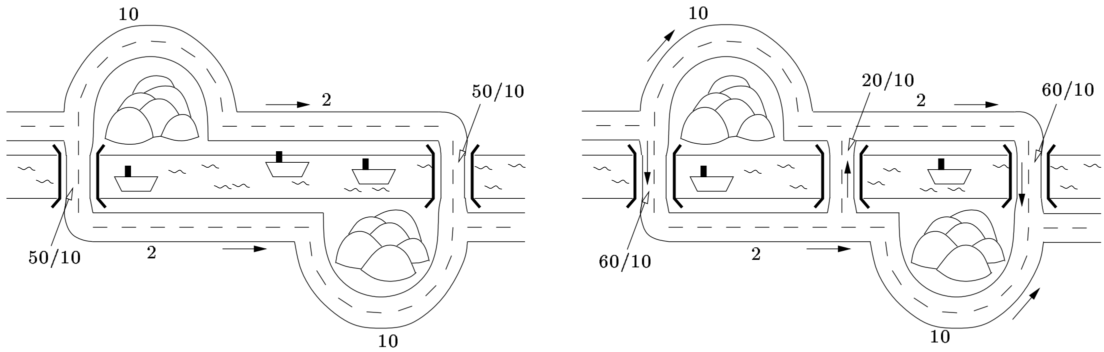


Figure 2: Braess' paradox. The numbers indicate the travel times for the corresponding edge; see text. Adding an edge, in this case a third bridge across the canal, leads to increased travel times for all drivers. Such paradoxical results have been observed in practice.

4.3 Nash equilibrium, Braess' paradox and the Prisoner's dilemma

One idea that came up during the discussions, is whether Ericsson, by advising their users properly, could "manage" traffic so that a more efficient use of the network would result, and hence better journey times for many of Ericsson's users. However, paradoxical situation may occur, as for instance noted by Frank Kelly [?]:

> [I]f drivers are provided with extra information about random delays ahead, the outcome may well be a new equilibrium in which delays are increased for everyone.

In this section we shall make some remarks about a similar paradox, which states that adding edges to a traffic network may similarly increase delays.

It is usually assumed that drivers choose routes so as to optimize their own situation. Assuming full information, this leads to a traffic situation where every driver uses a *locally optimal* route, meaning that choosing a different route will not result in a decreased journey time for this particular driver. This is called a *Nash equilibrium.* More than one Nash equilibrium may exist.

Given that Nash equilibria are locally optimal, it is perhaps surprising that the *globally optimal*, or most efficient, traffic situation may *not* be a Nash equilibrium. For this statement to be meaningful, we need a definition of "efficiency". Here we choose as efficiency measure, any (weighted) average of the journey times experienced by all drivers in a given traffic situation. More precisely, the statement is as follows. For certain networks, traffic situations exist where, compared to the most efficient Nash equilibrium, *every* driver has decreased journey time. These traffic situations are more efficient according to our efficiency measure.

The fact that a Nash equilibrium need not be optimally efficient, is closely related to *Braess' paradox* [?]. Braess found that adding an edge to a network may lead to a change in a Nash equilibrium, with *increased* delays for everyone, even though drivers have more routes to choose from. This phenomenon has actually been observed in practice [?].

An example network where Braess' paradox occurs is given in figure **??**; see also [?]. A network around two mountains and across a canal offers two alternative routes. Suppose that the delays on the various parts of the network are: 10 minutes to go around a mountain, 2 minutes to drive along the canal, and $n/10$ minutes to cross the canal, where n is the number of cars using the bridge. This term models the congestion, due for instance to the narrowness of the bridge. Suppose 100 cars want to cross the canal. In this case the unique Nash equilibrium is reached when cars distribute themselves equally among the two available routes. The associated delay is $10 + 2 + \frac{50}{10} = 17$ minutes for each route.

Suppose now that a third bridge is constructed, as indicated in the right panel in figure **??**. Two new routes are available, one going around both mountains, the other avoiding going around either. The first will not be used, but the second route is fast, with initially $\frac{50}{10}+2+0+2+\frac{50}{10} = 14$ minutes delay, so that the previous traffic situation is no longer a Nash equilibrium. The new unique Nash equilibrium is reached when 20 cars take the new route, with the remaining 80 divided equally among the two old routes. The delay in this situation is 18 minutes for all routes, longer than before.

The previous traffic situation, with 17 minutes delay, is still possible in the new network configuration. However, it is not a Nash equilibrium and will not spontaneously occur in practice: since the shorter route is faster, with only 14 minutes delay, taking this route is beneficial for the individual driver, even though it is detrimental to the "community".[2] This situation is analogous to the classical Prisoner's dilemma.

One way to induce drivers to move towards the non-Nash optimally efficient situation is to artificially increase the "cost" of the short route, for instance by imposing a monetary fine (essentially the content of the NIEMA proposal, see [?, ?]). Drivers will weigh the benefit of a decreased delay against the monetary cost associated to the quicker route. A new equilibrium will set in, which is a Nash equilibrium associated to the weighted graph, with weights that include both the delay along the edge, and the monetary cost involved. With fines chosen appropriately, such a Nash equilibrium can correspond to an optimally efficient traffic situation.

We conclude that the two approaches, top-down or global, and bottom-up or local optimizing (see also section ??), result in different optimal solutions. Efforts for globally optimizing network usage are best done at the government level. Ericsson, on the other hand, is interested in the local problem of predicting travel time and the best possible route, in order to provide a service to users. For this reason, we abandoned the idea of global traffic management, and instead focused on predicting traffic densities, in order to find the fastest route for individual users.

4.4 Finding the shortest route

Once road speeds are estimated, finding the fastest route is relatively easy. It can be basically done with Dijkstra's shortest path algorithm (mind the datahandling!) for directed weighted graphs, adapted to take into account that the weights of the edges (the travel time along this edge) vary with time. Provided that the weights do not vary too quickly (as otherwise waiting before taking an edge may get you to your destination quicker than leaving immediately – Dijkstra's algorithm does not handle this correctly), this can be done easily. Such an algoritm has polynomial time complexity. For an application of Dijkstra's algorithm in a real-life situation see [?].

4.5 The explaining variables of traffic jams

To provide reliable information, we want to predict how quickly a user can travel on the road network. This depends on the user (whether he is driving a truck or a car, whether he will drive fast or slow if he has an option), and on traffic conditions (the maximum attainable speed on each leg of his journey). Traffic conditions depend on several things, such as the type of road, the road conditions, the number of drivers on the road, the type of vehicles (cars or trucks), and previous traffic conditions (a jam takes time to dissolve; above criticality, a jam reinforces itself).

Let us focus on two main variables, demand and road conditions. Since the central route planner bases its estimate on an improved estimate of traffic density, and since Ericsson's additional source of information only carries further information on the first variable, demand, it is important to know something about the relative importance of these two variables for traffic density estimates.

A very simple first approach to answer this question empirically could be as follows. Instead of looking at the average speed, which varies over the roads of the network and may be difficult to measure, we look at the length of a traffic jam. This may be regarded as a stochastic variable. The lengths of traffic jams are already being measured and broadcast on the radio. Correlating these with variables related to demand and road conditions then gives information on how these variables explain the lengths of traffic jams. Real-time measurements of intensity are already done at some points in the Dutch road network (see [?]).

[2] Choosing routes according to a globally efficient traffic situation is, in Hofstadter's language, following a "super-rational" strategy (see [?, Part VII]). His experiments indicated that rational people do not follow such strategies.

A potential problem in this approach is that the relationship between traffic jam lengths and demand is probably highly nonlinear: below a certain threshold, jams are very unlikely to occur. Because of this nonlinearity, ordinary linear correlation might not be the best way to measure the dependence between these variables.

5 A statistical approach

Here we present a model to answer Ericsson's question quantitatively, in a simplified setting.

5.1 Simplifying assumptions

All network flow parameters are important for e.g. simulations. ¿From the users' perspective, the average speed (or travel time) are mainly of interest. In order to predict the total travel time for a single user, Ericsson needs to be able to predict the travel time on each leg of the network, at every instant of time.

Once these estimates are known, finding the fastest route is relatively trivial. Therefore we shall focus on the precision of the journey time estimate, which depends on the "reliability" of network flow parameter estimation. Many of the variables that influence this are inherently stochastic in nature (such as the occurrence of accidents), so a statistical approach seems to be natural.

In order to simplify further, we will not consider the whole network or discuss the correlations in time mentioned, but focus on the problem of predicting the travel time on a *single* given road.

As final major simplification, we assume that the network flow parameters (such as intensity) are independent of time. In practice, this means that we shall consider a short time interval, where conditions can be regarded as being constant (but see section **??**).

5.2 The model

In this model, we consider a single road. Let N denote the number of cars on this road and T the time taken to travel along the road. As mentioned previously, we assume there is a deterministic relation between the density and the speed, so that T is some function of N:

$$T = g(N) \tag{1}$$

say. It is reasonable to assume g is monotonically increasing and hence injective (see also section **??**). Some of these N drivers call Ericsson; say U users. The problem is now to estimate N given U.

First we need to model the distribution of N itself. All we know of N is that it is a discrete variable. We assume N has a Poisson distribution with parameter μ, say.

Furthermore, we assume that the probability of a driver being a user is λ. In other words, the conditional distribution of U given $N = n$ is a binomial distribution with n trials and parameter λ. The actual value of λ may be deduced from historical (users) data. Then U is Poisson distributed with parameter $\mu\lambda$.

We have historical data for a particular road, i.e. with frequencies $f_1, \ldots, f_k$ there are $n_1, \ldots, n_k$ cars on the road. The travel times are then $t_1, ..., t_k$ where $t_i = g(n_i)$.

In practice many n_i-values can occur, while the historical data may be limited. It is therefore probably a good idea to use *intervals* $[n_i, n_{i+1})$ of some appropriate length, instead of points. See also section **??**.

Suppose we observe $U = u$ say, and we know

$$P(N = n_i | U = u) = \frac{P(U = u | N = n_i) P(N = n_i)}{P(U = u)}$$

by Bayes Theorem. Now $P(U = u | N = n_i)$ is given by the above assumption, $P(N = n_i) \simeq$

$f_i / \sum f_i$ by our assumption on the historical data, and

$$P(U = u) = \sum_{n_j \geq u} P(U = u | N = n_j) P(N = n_j).$$

As n_i varies, this gives us the posterior distribution of N, given $U = u$. Hence, using that g is injective (see also figure **??**),

$$P(T = t_i | U = u) = P(g(N) = g(n_i) | U = u) = P(N = n_i | U = u) \tag{2}$$

gives the posterior distribution of T given $U = u$. We finally estimate T from a given u by choosing the value t_i which maximizes (**??**).

Let us now use our knowledge of the distributions of N and U. We have

$$P(N = n) = \frac{e^{-\mu} \mu^n}{n!}, \quad P(U = u | N = n) = \binom{n}{u} \lambda^u (1 - \lambda)^{n-u} \quad \text{and} \quad P(U = u) = \frac{e^{-\mu\lambda} (\mu\lambda)^u}{u!},$$

and a little algebra yields the posterior distribution of N:

$$P(N = n_i | U = u) = \frac{e^{-\mu(1-\lambda)} (\mu(1 - \lambda))^{n_i - u}}{(n_i - u)!},$$

a Poisson distribution with parameter $\mu(1 - \lambda)$, translated by u. It reaches its maximum at $N = u + \mu(1 - \lambda)$, which is also its mean. Its variance is $\mu(1 - \lambda)$. Now, $T = g(N)$, so that

$$\text{var}(T | U = u) \approx \text{var}(N | U = u) \left(g'(E(N | U = u))\right)^2 = \mu(1 - \lambda) \left(g'(u + \mu(1 - \lambda))\right)^2 \tag{3}$$

(Note that the quality of this approximation depends on the smoothness of g. In our case g is nondifferentiable, see section **??**, and (**??**) will be an underestimate just below critical densities.) An estimate of the variance of T without user information is $\mu g'(\mu)^2$. With user information, this changes to (**??**). The difference can serve as a measure of the improvement of our estimate.

5.3 An alternative approach: Continuous distributions

A major drawback of the discrete approach is that we need to choose appropriate lengths for the intervals $[n_i, n_{i+1})$. A convenient way to avoid this is to model N by a continuous distribution, say $\mathcal{N}(\mu, \sigma)$. The facts that μ is large and that many different values N can occur validate this choice. Since negative values of N make no sense, we assume that $\sigma \ll \mu$, so that the probability of such values occurring is negligible. We assume again that the probability of a driver being a user is λ, in other words, that U is binomially distributed with parameter λ and N, given $N \in \mathbf{N}$. By the law of large numbers

$$B(N, \lambda) \sim \mathcal{N}(N\lambda, \sqrt{N\lambda(1 - \lambda)}).$$

The rest of the analysis follows the previous section, and we shall not give the details.

6 The dependence of travel time on road usage

In the previous section, we used an unspecified function g to describe the dependence of the travel time T on the number of users N of a road segment; see (**??**). In this section we use a simple model of traffic dynamics to arrive at a candidate for g.

Many different models of traffic flow can be found in the literature. They may be characterized, crudely, as *microscopic* (individual cars, see references in [**?**, **?**]), *mesoscopic* (densities and average speeds over segments a few hundred meters in length, again see [**?**, **?**]) and *macroscopic* (on the level of networks, see [**?**]).

We focus on the mesoscopic level. The existing models are too detailed for us, and we make a few extra simplifications.

6.1 Traffic model

The model we describe here is a simplified version of the models used in [?, ?, ?]. The main simplification is that we consider a single road segment, on which we assume that homogeneous conditions prevail.

In section ?? we mentioned the relation between density and equilibrium speed proposed in [?]. We shall assume that traffic flow is always in equilibrium, so that the relation between densities (supplemented by variables describing the road condition) and speed (hence intensity) is deterministic.

Input of the model is a function $A(t)$ describing the influx of cars on the road segment per time unit, in cars per hour. The output is a function $D(t)$ describing the instantaneous density, in cars per kilometre. The density increases due to the influx of cars, and decreases due to the outflux, which is equal to the intensity (in cars/hour). The intensity is a function of the density, namely Smulders' function (see figure ??) multiplied by the density. Denoting the length of the road segment by L, this leads to the following model:

$$L\frac{\mathrm{d}D}{\mathrm{d}t} = A(t) - D(t)\cdot S(D(t)) \tag{4}$$

Here S is Smulders' function. In principle, this function depends on various parameters, like road conditions, type of traffic etcetera. For simplicity we shall ignore this and use the following formula:

$$S(D) = \begin{cases} v_{\text{free}}\left(1 - \frac{D}{D_{\text{jam}}}\right) & \text{if } D \leq D_{\text{crit}} \\ v_{\text{free}} D_{\text{crit}}\left(\frac{1}{D} - \frac{1}{D_{\text{jam}}}\right) & \text{if } D > D_{\text{crit}} \end{cases}$$

(See figure ??, and [?, p. 30] for a motivation.) The various parameters are

$$v_{\text{free}} = 110 \text{ km/h}, \quad D_{\text{jam}} = 110 \text{ cars/km}, \quad D_{\text{crit}} = 27 \text{ cars/km}$$

Because we assume that conditions on the road segment are homogeneous, the length of the segment is an important parameter of the model (??). We used the value $L = 30$ km, which led to reasonable results.

6.2 Road usage

The influx of cars at a certain instant, per unit of time, is given by $A(t)$. As a model for road usage we take

$$A(t) := N\sqrt{\frac{\alpha}{\pi}}e^{-\alpha t^2}, \tag{5}$$

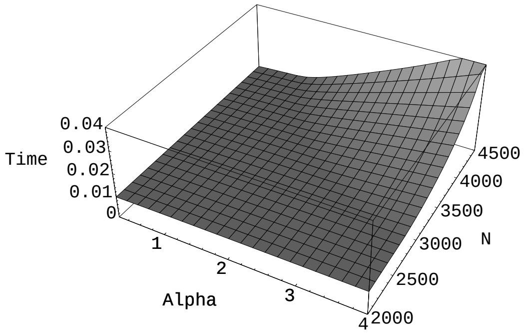
a bell curve, where N is the total number of cars passing the road, and α is related to the width of the bell curve. Both parameters influence the development of a jam.

6.3 Analysis

We are interested in the throughput of the road segment. One statistic related to this is the *average time* it takes to travel through the segment. The time spent is the time of exit minus time of entry; the average time spent on the segment for all cars is therefore

$$\begin{aligned} T &= \frac{\int_{-\infty}^{\infty} t\,(\text{outflux}(t) - \text{influx}(t))\,\mathrm{d}t}{\int_{-\infty}^{\infty} \text{influx}(t)\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} t\,(D(t)\cdot S(D(t)) - A(t))\,\mathrm{d}t}{\int_{-\infty}^{\infty} A(t)\mathrm{d}t} \\ &= \frac{1}{N}\int_{-\infty}^{\infty} -Lt\frac{\mathrm{d}D(t)}{\mathrm{d}t}\mathrm{d}t = \frac{L}{N}\int_{-\infty}^{\infty} D(t)\mathrm{d}t, \end{aligned}$$

where we integrated by parts, with vanishing boundary terms. This is a first candidate for the function $g(N)$.



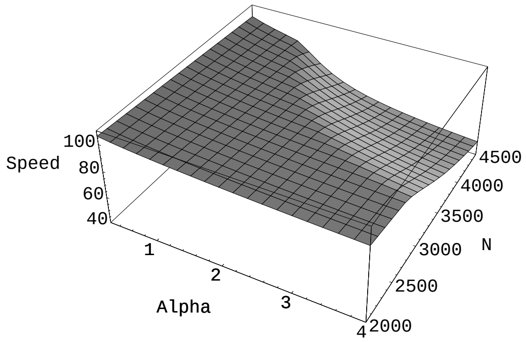


Figure 3: The average travel time, and average speed, on a road described by traffic model (**??**) under influx (**??**), for varying parameter values N and α.

A different but related statistic is the *average speed.* It may be calculated by noting that an intensity of cars $D(t) \cdot S(D(t))$ experience a speed $S(D(t))$; the average speed therefore is

$$V = \frac{1}{N} \int_{-\infty}^{\infty} D(t) \left(S(D(t))\right)^2 \mathrm{d}t$$

For parameter values $N \in [2000, 4500]$ and $\alpha \in [\frac{1}{4}, 4]$, both statistics have been plotted in figure **??**. In both plots a 'ridge' can be seen, along which T and V seem to have discontinuous derivatives. The corresponding curve in the $\alpha - N$ plane is related to the onset of jams.

6.3.1 Critical curve

We now try to obtain an estimate of this curve of critical parameter values. The pair (α, N) is critical when the density reaches, but not increases beyond, the critical density D_{crit}, at some time t_{crit}. At this moment $\frac{\mathrm{d}D}{\mathrm{d}t} = 0$. Plugging this into (**??**) we get that t_{crit} satisfies

$$A(t_{\text{crit}}) = D(t_{\text{crit}})S(D(t_{\text{crit}})) = D_{\text{crit}}S(D_{\text{crit}}) = I_{\text{crit}}, \tag{6}$$

where the critical intensity I_{crit} is defined by $I_{\text{crit}} := D_{\text{crit}}S(D_{\text{crit}}) = 27S(27) = 2241$ cars/h. Equation (**??**) has two solutions, one negative and one positive. Since D lags A and reaches its maximum after A does, only the positive solution is relevant.

The remaining condition is that $D(t_{\text{crit}}) = D_{\text{crit}}$. To solve this equation we need to solve the differential equation (**??**). To simplify the latter, first note that $0 \leq D \leq D_{\text{crit}}$ globally. In this range $S(D)$ depends linearly on D, and varies by approximately 25%. We approximate $S(D)$ by a constant S_{avg}. This constant is chosen somewhere between 110 and 83 km/h, but with a bias towards the lower value since $S(D)$ affects the differential equation more when D is larger. Then (**??**) becomes linear,

$$L\frac{\mathrm{d}D(t)}{\mathrm{d}t} + S_{\text{avg}}D(t) = A(t),$$

and can be solved by variation of constants, $D(t) = (1/L)\int_{-\infty}^{t} e^{S_{\text{avg}}(u-t)/L} A(u)\mathrm{d}u$. The important parameter here is S_{avg}/L, which in our case is approximately 3. This justifies the use of the following estimate which is more useful for our purpose, and which is valid for large parameter values S_{avg}/L:

$$D(t) = \frac{1}{S_{\text{avg}}} A\left(t - \frac{L}{S_{\text{avg}}}\right) + \frac{L^2}{2S_{\text{avg}}^3} A''\left(t - \frac{5L}{3S_{\text{avg}}}\right) + \cdots \tag{7}$$

Using only the first term, we get the following condition for t_{crit}:

$$A\left(t_{\text{crit}} - \frac{L}{S_{\text{avg}}}\right) = S_{\text{avg}}D_{\text{crit}}$$

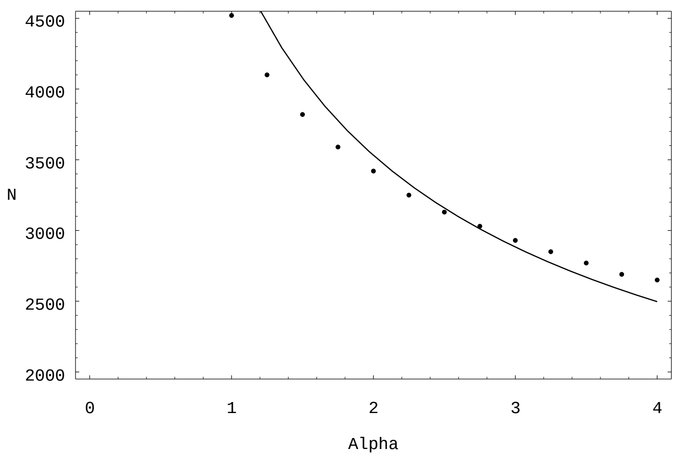


Figure 4: Curve of critical values in the $N-\alpha$ plane: jams start to occur upon crossing this curve from the left. Dots are obtained by numerically solving (**??**); the curve is the approximation (**??**),(**??**).

Expanding to first order, and using (**??**), we can rewrite this as $A'(t_{\text{crit}}) = (1/L)S_{\text{avg}}(I_{\text{crit}} - D_{\text{crit}}S_{\text{avg}})$. ¿From (**??**) we find that $A'(t)/A(t) = -2\alpha t$, and using this the condition becomes

$$2\alpha t_{\text{crit}} = \frac{1}{L}S_{\text{avg}}\left(\frac{D_{\text{crit}}S_{\text{avg}}}{I_{\text{crit}}} - 1\right) \tag{8}$$

Given α, we can solve (**??**) for t_{crit}, and then

$$N_{\text{crit}} = I_{\text{crit}}\sqrt{\frac{\alpha}{\pi}}e^{\alpha t_{\text{crit}}^2} \tag{9}$$

Fitting the resulting curve to the curve obtained by numerical integration, we find good agreement at $S_{\text{avg}} = 95$ km/h. The critical curve is plotted in figure **??**.

6.3.2 Average time with jams

We now analyze what happens when we cross the critical curve. We assume that the time during which the density exceeds the critical value is small, compared to the total time interval considered. Let N and α be on the critical curve, and let t_{crit} denote the instant at which critical density is reached, and $I_{\text{crit}} := D_{\text{crit}}S(D_{\text{crit}}) = D(t_{\text{crit}})S(D(t_{\text{crit}}))$ the critical intensity. We increase $A(t)$ by a factor ϵ, that is, we set

$$A(t) := (1+\epsilon)N\sqrt{\frac{\alpha}{\pi}}e^{-\alpha t^2}$$

For small ϵ, the time spent in the jam-regime $D \geq D_{\text{crit}}$ will be small. This justifies approximating I and A by linear models,

$$\begin{aligned} I(D(t)) &= I_{\text{crit}} - \gamma(D(t) - D_{\text{crit}}) \\ A(t) &= (1+\epsilon)I_{\text{crit}}(1 - \beta(t - t_{\text{crit}})), \end{aligned}$$

where we used that $A(t_{\text{crit}}) = I_{\text{crit}}$ when $\epsilon = 0$. Here $\gamma = \frac{\mathrm{d}I}{\mathrm{d}D}$ as $D \searrow D_{\text{crit}}$, and $\beta = -\frac{\mathrm{d}A}{\mathrm{d}t}/I_{\text{crit}}$ at $t = t_{\text{crit}}$. For $\epsilon > 0$ the critical density will be reached for $t_1 < t_{\text{crit}}$. Using $D(t) \approx (D_{\text{crit}}/A(0))A(t - t_{\text{crit}})$ (equation (**??**)), valid when $D \leq D_{\text{crit}}$, we find the approximation

$$D(t) \approx (1+\epsilon)D_{\text{crit}}\left(1 + \frac{A''(0)}{2A(0)}(t - t_{\text{crit}})^2\right) = (1+\epsilon)D_{\text{crit}}(1 - \alpha(t - t_{\text{crit}})^2)$$

for the case $\epsilon > 0$ and $D \leq D_{\text{crit}}$. Solving $D(t) = D_{\text{crit}}$ we find

$$t_1 = t_{\text{crit}} - \sqrt{\frac{\epsilon}{\alpha(1+\epsilon)}}$$

For convenience we now choose translated variables t' and D', so that $t' = 0$ corresponds to t_1, and $D' = 0$ to D_{crit}. Setting up the differential equation for the jammed regime in these variables, we get

$$L\frac{\mathrm{d}D'}{\mathrm{d}t'} = (1+\epsilon)I_{\text{crit}}\left(1 - \beta\left(t' - \sqrt{\frac{\epsilon}{\alpha(1+\epsilon)}}\right)\right) - I_{\text{crit}} + \gamma D' = a - bt' + \gamma D'$$

where $a = I_{\text{crit}}(\epsilon + (1+\epsilon)\beta\sqrt{\epsilon/\alpha(1+\epsilon)})$ and $b = I_{\text{crit}}(1+\epsilon)\beta$. This equation can be solved by variation of constants again, yielding

$$D'(t') = \left(\frac{a}{L}t' - \frac{b}{2L}t'^2\right)e^{t'\gamma/L} \tag{10}$$

The solution is valid for $D' \geq 0$, that is, between $t_1' = 0$ and

$$t_2' = 2\sqrt{\frac{\epsilon}{\alpha(1+\epsilon)}} + \frac{2\epsilon}{(1+\epsilon)\beta} \tag{11}$$

We are interested in the value $\int D(t)\mathrm{d}t$. Without jams this would be N/S_{avg}, see (**??**). With jams the value becomes larger, due to two effects. First of all, from the second term in (**??**) it is seen that the time interval where $D \geq D_{\text{crit}}$ is longer than it would be without jams. Secondly, the density in this interval is larger than it would be without jams. Integrating (**??**) over the appropriate interval, and truncating at degree ϵ^2, we the following formula:

$$T = \frac{L}{N}\int_{-\infty}^{\infty} D(t)\mathrm{d}t = \frac{L}{S_{\text{avg}}} + \begin{cases} 0 & \text{if } N < N_{\text{crit}} \\ \frac{LD_{\text{crit}}}{N}\frac{2\epsilon}{(1+\epsilon)\beta} + \frac{16\beta I_{\text{crit}}}{3N}\sqrt{\frac{\epsilon^3}{\alpha^3(1+\epsilon)}} & \text{if } N \geq N_{\text{crit}} \end{cases} \tag{12}$$

where $N = (1+\epsilon)N_{\text{crit}}$. It turns out that for reasonable parameter values, the last term, measuring the large-density effect, is the least important one. As a final improvement, we replace the constant S_{avg} by a number that is 110 for $N = 0$, and linearly decreases to S_{avg} when N reaches its critical value. The resulting curve, and the numerically obtained statistic T, are plotted in figure **??**.

We conclude that the expression (**??**) may serve as a good approximation of $g(N)$ below and around critical densities.

7 Conclusions

Ericsson's problem has many different aspects, which makes it impossible to give a precise answer to their question. Instead we have tried to provide an overview of these aspects, which helped us to subsequently formulate and analyze a model problem.

We selected relevant literature and described some of the recent research in the area. Traffic routing problems and traffic density predictions, as well as traffic flow models, have been studied greatly. It turned out that even advance booking had been investigated, an idea which is closely related to Ericsson's ideas. We indicated a few problems that may arise when a CRP is implemented. Our main concerns are, that too few drivers may cooperate for traffic density estimates to improve significantly, that users may give unreliable information or ignore the CRP's advice, and that road conditions (which cannot be predicted well in advance) may influence the traffic densities more than the demand.

To get a sense for the relation between user cooperation and reliability of travel time estimates we modelled a single road segment, under strong assumptions. Here we used the variance of travel

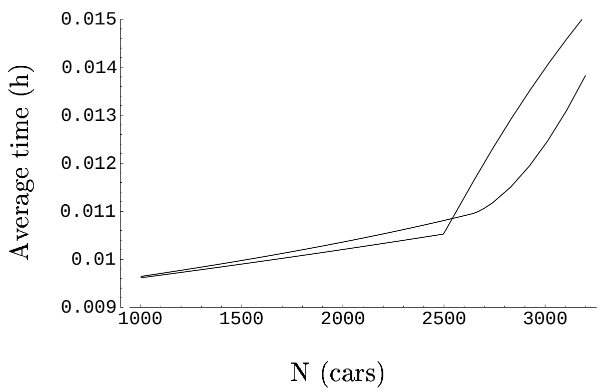


Figure 5: Average time to travel through the segment, for fixed $\alpha = 4$, as a function of N. The smooth curve has been determined numerically, the other curve corresponds to (??).

time estimates as a measure of reliability. We also analyzed a simplified traffic model to find a relation $T = g(N)$ between road usage N and average travel time T. The fraction λ of the drivers that use the CRP appears to influence the travel time predictions in two ways. If the number of cars on a road is not near a certain critical number, the variance of the travel time depends more or less linearly on the fraction λ (in this case the derivative $g'(N)$ is roughly constant when N changes due to CRP advices). Far from critical situations, the effect of users information will therefore only be noticeable when many drivers become users. If the number of cars on a road is near the critical number, $g'(N)$ changes drastically with small variations of N. This may cause a higher order dependence of the variance on λ, and means that user information becomes more useful. It is however hard to predict in advance whether the situation on a road will be near criticality.

Moreover, a change in N, caused by the CRP's advice, may also let $g'(N)$ increase. This can result in a less reliable estimate than the estimate without users information. Partly this is due to our notion of reliability; however, paradoxical situations may occur, related to Braess' observation. It would be interesting to study this, and identify when reduced reliability, or increased travel time, can occur as a result of providing users with better information. Another topic that seems interesting and relevant to study is the dependence of jams on road conditions versus road usage.

References

[Bra68] D. Braess. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmenforschung*, 12:258–268, 1968.

[Cre79] M. Cremer. *Der Verkehrsfluß auf Schnellstraßen.* Springer Verlag, 1979.

[dR94] E. de Romph. *Dynamic traffic assignment models.* PhD thesis, Technical university Delft, 1994.

[Hof85] Douglas R. Hofstadter. *Metamagical themas.* Basic Books, 1985.

[Kel91] Frank Kelly. Network routing. *Philosophical Transactions of the Royal Society A*, 337:343–367, 1991. Also available at `http://www.statslab.cam.ac.uk/~frank/CP/-CPREAD/nr.html`.

[Mie00] Mark Mieras. Alstublieft reserveren voor een autoritje. Intermediair 3, January 20, 2000.

[Nie00] Niema. `http://www.autoritadvies.nl`, 2000.

[Pay71] H. J. Payne. Models of freeway traffic and control. In *Mathematical models of public systems. Simulation Council Proceedings*, pages 51–61, 1971.

[RB94] B. Ran and D. E. Boyce. *Dynamic Urban Transportation Network Models*, volume 417 of *Lecture Notes in Economics and Mathematical Systems*. Springer-verlag, 1994.

[Sch98] Jan Schreuder. A strategic approach for the ambulance covering of the province of friesland. Memorandum No. 1462, University of Twente, Enschede, 1998.

[Smu96] S. A. Smulders. *Control of freeway traffic flow*, volume 80 of *CWI Tract*. CWI, 1996.

[Tim90] New York Times. What if they closed 42nd street and nobody noticed?, December 1990.

[vS99a] J. H. van Schuppen. Dynamisch routeren. Technical report, CWI, 1999.

[vS99b] J. H. van Schuppen. Implementation of traffic model. Private communication, 1999.

[Wai] Mark Wainwright. A small road network. `http://www.expert.demon.co.uk/mark/-essays/roads.html`.