IET Intelligent Transport Systems

Research Article

Improving the road network performance with dynamic route guidance by considering the indifference band of road users

ISSN 1751-956X
Received on 20th October 2014
Revised on 7th May 2015
Accepted on 8th May 2015
doi: 10.1049/iet-its.2014.0258
www.ietdl.org

Jaap D. Vreeswijk[1] ✉, Ramon L. Landman[2], Eric C. van Berkum[3], Andreas Hegyi[2], Serge P. Hoogendoorn[2], Bart van Arem[2]

[1]*Technology Centre, Imtech Traffic & Infra, Basicweg 16, 3821 BR Amersfoort, The Netherlands*
[2]*Transport and Planning, Delft University of Technology, Delft, The Netherlands*
[3]*Centre for Transport Studies, University of Twente, Enschede, The Netherlands*
✉ *E-mail: jaap.vreeswijk@imtech.com*

Abstract: When applying dynamic route guidance to improve the road network performance, the interests of the road authorities and the road users are likely to conflict. Road users are normally aiming at minimising their travel times by choosing the time-shortest route, whereas the authorities might want to send road users to a time-longer alternative to prevent phenomena like capacity drop and spill-back. In this study, the authors demonstrate how, in such a situation, bounded rationality and indifference bands (IB) of road users can be used to improve the network performance, such that the road users will not experience being disadvantaged by the advice. An interpretation and utilisation of the IBs in a route guidance control approach is proposed, which degrades the routes such that the network performance deterioration is prevented while the IBs of road users are respected. By means of a simulation test case, the potential gain in network performance of different IBs is evaluated. Results show a reduction in total travel time of 5% compared with user equilibrium, in case of an IB of 4 min for a trip of ~22 min, and up to 14% in case of an IB of 10 min.

1 Introduction

The increasing adverse effects of congestion indicate the need to apply dynamic traffic management (DTM) on a network level to improve the network performance. However, there exists a well-known conflict between realisation of system optimal conditions and user optimal conditions [e.g. 1–3]. Hence, to successfully implement DTM on a network level, a suitable trade-off between the interests of road authorities and those of drivers needs to be made. The main challenge addressed in this paper is to find and implement a control approach that implements road authorities' traffic management policies and steers the network towards the desired state without seriously violating drivers' interests.

Empirical research in the field of traffic psychology indicates that drivers have difficulty in assessing the quality of their chosen alternative [e.g. 4, 5] or are simply not willing to adapt their choice if the benefits of switching are below a certain threshold [e.g. 6, 7]. This perspective of individual decision making is known as bounded rationality and contributes to the indifference band (IB). On the basis of a simple example, using the green split of traffic lights as control mechanism, it was already demonstrated that the application of IBs can successfully steer a system towards its optimal state [8]. This paper therefore proposes the use of these psychological constructs to improve the network performance, in such a way that the user interests remain respected. To this aim, service level definitions are used to describe the quality of the network elements and the perceived quality from the perspective of the road user (in terms of average route speed or travel time). A (service level-oriented) route guidance approach will subsequently maintain a maximum travel time difference between a preferred route and an alternative that is in line with the size of the IB. In other words, road users are only redirected away from the bottleneck, if their perceived travel time difference between the preferred route and alternative lies within the users' IB.

Example: Blocking back within a route can be prevented (queue stabilised) by sending traffic to the alternative route. If no redundant capacity is available in that alternative, its quality will degrade (travel times increase). The IB is an indicator for the maximum acceptable travel time difference between both routes that is acceptable (i.e. non-observable and/or non-interested) from a user perspective. This in turn defines the achievable gain in network performance with respect to the user equilibrium situation and the situation in which no prescriptive route guidance is given.

The control approach should realise network states in line with the policy objectives, that is, the phenomena that decrease the network performance should be prevented in a systematic and comprehensible way without strongly violating the road users' interests (e.g. mode, route, departure time and arrival time). This is a challenge, because it is often acknowledged that the road users are generally most concerned with improving their own situation, disregarding the effect of their actions on the network performance and the intentions of traffic management policies. The inverse of this argument also holds; as the road authorities develop visions on how their network should function without explicit care for individual drivers.

By means of a simulation test case, we will demonstrate the use of IB to improve the network performance. The notion of IBs [6] is based on the observation that the drivers behave boundedly rational. For example, they make estimation errors that influence their perception of the situation. There is evidence that, such errors contribute to an IB that represent drivers' insensitivity to varying conditions. In this paper, we argue that the network performance can be improved by considering the IBs in traffic control while the expectation of individual road users remain protected. Note that although the network state moves from user equilibrium towards system optimal state, drivers are indifferent to this change because their perception of the old and new state is similar. Hence, they will not respond to the perceived change in traffic conditions because its change is not considered large enough.

The remainder of this paper will focus on two questions. First, what is known about the width of the IB and what values could be used for this study? This question involves the extent road users are insensitive to conditions that are suboptimal from their individual perspective. Second, what is the achievable network performance improvement when taking the IB into account when routing drivers away from a bottleneck in their preferred route.

The following section will give an extensive overview of the background of our work. Bounded rationality, IBs, perception error and service level-oriented route guidance will be discussed in detail. Next, we will formulate our approach from the theory and empirical evidence that is available. Through application of the approach in a test case, we will demonstrate the potential effect on the network performance. The final sections discuss the results and conclude.

2 Background

2.1 Bounded rationality

Many assumptions in conventional traffic modelling have been derived from standard economics. It is often assumed that the drivers are rational decision makers and, above all, are perfectly informed about the available choice of alternatives. Moreover, they can calculate the value of the different options available, that they are able to derive the optimal choice and that they are cognitively unhindered in weighting the implications of each potential choice [9, 10]. In other words, people make logical and sensible decisions and quickly adopt their choice to changing conditions. In reality, people have limited knowledge and constrained cognitive abilities leading to prejudiced reasoning and certain randomness in behaviour and choice outcomes [9, 11, 12]. Behavioural economics draw on the aspects of both (cognitive) psychology and economics, and study the motives and behaviours that explain deviations from rational behaviour [5, 13]. This perspective is known as bounded rationality or satisfying behaviour [4, 14], and also found its way into transport research, [e.g. 6, 15–18]. In summary, bounded rationality states that drivers do not necessarily make the most economical (or logical) choice.

2.2 Indifference bands

A well-known mechanism derived from the principles of bounded rationality, which has been used and validated in the field of transport, is the notion of IBs. According to the theory of IBs, drivers will only alter their choice when a change in the transport system or their trip, for example, the travel time, is larger than some individual-situation-specific threshold [10, 11, 19–21]. In the field of time psychology, this threshold is called the ‘comfort zone’ [22]. In addition, drivers supposedly do not update their choice (e.g. route, departure time, mode) when the difference in quality between two routes, for example, in travel time, is less than the same threshold.

There are many factors associated with IBs which explain why a change in network performance not necessarily leads to a behavioural response. Examples include limited awareness and disinterest [8]. Underlying reasons may be that a driver is not alert to changes because of the formation of habits, that a driver is not able to detect or ‘see’ the change because it is small or outside the driver’s periphery, that the driver is disinterested if the type of change is regarded insignificant, or simply due to the lack of (knowledge of) alternatives.

Multiple studies provide evidence that boundedly rational behaviours are neither random nor senseless; they are systematic, consistent, repetitive, and therefore predictable [5, 9, 23]. As a consequence, the IB can be estimated too and therefore used as an input variable for DTM. In several studies, an attempt was made to estimate the width of the IB. All studies acknowledged the existence of the phenomenon, but their estimations vary: 10 min [24], 18 min [25], 5–10 min or 17–22% [10]. From these figures, it is clear that no single, generic IB can be defined without the knowledge of the traffic conditions, trip lengths and so on. The IB is clearly situation specific.

2.3 Perception error

In literature, there is strong debate about discrepancies between drivers’ perception and the existing level of service standards [26]. The Highway Capacity Manual (HCM) proposed six levels of service ranging from ‘A’ very good service to ‘F’ very poor service which are separated by the threshold values of characteristic measures of traffic flow performance, such as traffic density, volume-to-capacity ratio and average speed. However, empirical evidence of below referenced studies shows that average drivers are unable to properly estimate the actual quality of the conditions they experience. Drivers’ perception of level of service appears widely variable, while usually only two or three levels of traffic conditions are distinguished.

In one study, drivers’ assessment of motorway traffic conditions, reported while waiting at traffic lights on a freeway exit were compared with the actual v/c ratio from the same time period [27]. This study showed that the drivers’ assessment of level of service is especially variable at moderate traffic conditions within the v/c interval of 0.55–0.70. Additionally, only the least tolerant drivers appear to distinguish levels of service A and B, whereas only the most tolerant drivers appear to distinguish levels of service D from E. The findings did not differ for driver and vehicle characteristics. On the basis of these results, three service levels were proposed: one level for the highest v/c values, one level for the medium-high v/c values and one level for all the other v/c ratios. Using the video clips taken from cameras on overpasses, another study with 195 individuals from five different occupational groups showed similar results [28]. Likewise, respondents of this study seemed to differentiate three levels of freeway traffic conditions. Additionally, compared with the HCM, they had lower tolerance for level of service (LOS) and a higher tolerance for worse LOS. For urban commuters, similar results were found as they appeared to be primarily concerned about the total trip time and its reliability to complete the journey in reasonable time [29]. As such, fine distinctions between LOS A through D did not seem applicable in the urban context.

Another stream of research investigated the perception of the level of service at signalised intersections. Study results (see [30] for a review) suggest that the drivers do not perceive level of service in a way consistent with the HCM criteria. Generally, two and perhaps three levels of service are perceived [31]. Lower levels of service were rated higher than expected, which suggests that the drivers may be more tolerant to longer delays (or used to them) than what is usually assumed. On the other hand, high levels of service, that is, A through C, are perceived as being very similar. Using a special and less rigid data clustering technique, it was concluded that drivers are able to differentiate six levels of service, but not in alignment with the existing HCM ones [32]. In this study, the service levels A and B were merged for a single level and level F was split into two.

A third stream of research looked at the accuracy of drivers’ perception of choice alternatives. Most studies observed that drivers’ perceptions become more accurate if the difference between alternative routes increases. It was found that drivers’ perceptions were on an average around 60% accurate [33]. Additionally, drivers were able to perceive travel speed better than travel time, although perception of the travel distance was least accurate. Several revealed preference studies showed that on a substantial percentage of trips, drivers do not choose the shortest route [34–37]. The number of trips varied from 25% to as much as 84%, depending on the route type (e.g. orbital or centre) and travel time difference between choice alternatives. Often the travel time difference is small (e.g. 30 s), but a substantial number of non-trivial travel time differences were found, ranging from 2 to 5 min or 8–25% of the average commute time [36, 37]. On the basis of the observation that drivers’ perceptions do not always correspond with their experiences, one could distinguish three types of choice behaviour [38]: (i) logical behaviour that reflects drivers choosing better perceived routes (perceive route A better and choose route A); (ii) cognitive behaviour reflecting drivers choosing a route in spite of not perceiving a difference between both routes, presumably to reduce mental working load (perceive no different, choose any route) and (iii) irrational behaviour that reflects drivers choosing worse perceived routes (perceive route A better and choose route B).

Finally, a recently adopted theory in transport research that is worth mentioning is prospect theory. The theory is derived from behaviour economics and is relevant in the context of this paper. It is based on the principle that the decisions are context dependent and alternatives are framed in terms of gains and losses relative to some common reference point, while losses weigh twice as much as gains of equivalent size [39, 40]. In line with this theory, it is arguable that the drivers are more likely to notice and respond to changes involving losses than changes involving gains, while the effect of any additional gains or losses decreases. The recently introduced theory of regret minimisation also builds upon these principles, that is, people anticipate and try to avoid the situation where a non-chosen alternative outperforms the chosen one [e.g. 41].

2.4 Usability of IBs and perception error

To define a service level structure for our test case, we base our design decision on the following conclusions. First of all, drivers may perceive the travel time of a route (PTT) differently than the actual travel time (ATT) as shown in Fig. 1. The dashed centre line represents the case of no perception error and equal PTT and ATT. In reality, drivers tend to overestimate (top-left) and underestimate (bottom-right) travel times depending on individual-situation-specific factors. There does not seem to be a general rule in the literature for drivers' overestimation and underestimation of travel times, probably because perception of travel time varies substantially between the routes depending on the route characteristics. To illustrate, a solid linear line is plotted for route x for which drivers systematically underestimate the travel time. From the viewpoint of the driver there is no difference between the travel times of both routes, whereas in reality there is. In practice, drivers' perception of travel time can be far more complex than a simple linear relation. An example is provided for route y by means of the dotted line for which low and high travel times are overestimated while moderate travel time is underestimated.

Perception errors based on the difference between the perception and the reality are a helpful indicator for the IB. This is shown in Fig. 2. For the purpose of illustration, we continue with the linear relation between PTT and ATT. In this example, travel time of route y is systematically underestimated, whereas travel time of route x is generally overestimated. These perception errors are indicated by PE[y] and PE[x], respectively. However, what matters the most to estimate the IB is the perception of route x relative to the perception of route y, indicated by PE[x–y]. In Fig. 2, the IB is the difference between the ATT of route x (ATT[x]) and the ATT of route y (ATT[y]), for which drivers perceive equal travel times (PTT[x, y]).

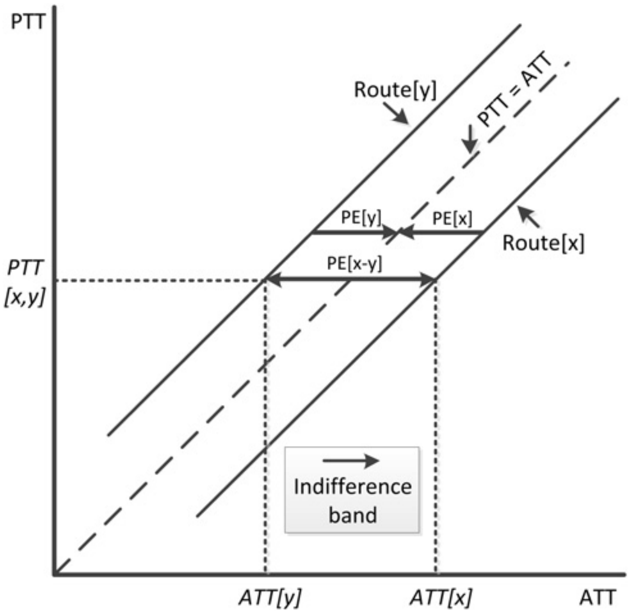

Fig. 2 *Perception errors (PE) and the IB*

Looking at service levels, the literature findings suggest that driver have more difficulty perceiving differences in high (i.e. A–B) and low levels of service (i.e. E–F) regimes than in moderate levels of service. Hence, it is reasonable to assume that the IB is wider for high and low levels of service than for moderate levels of service. Fig. 3 shows the level of service of route x against the level of service of route y. On the dashed centre line, the level of service of both routes is the same. Building upon Fig. 2, we assume again that due to perception errors, route x is generally perceived as

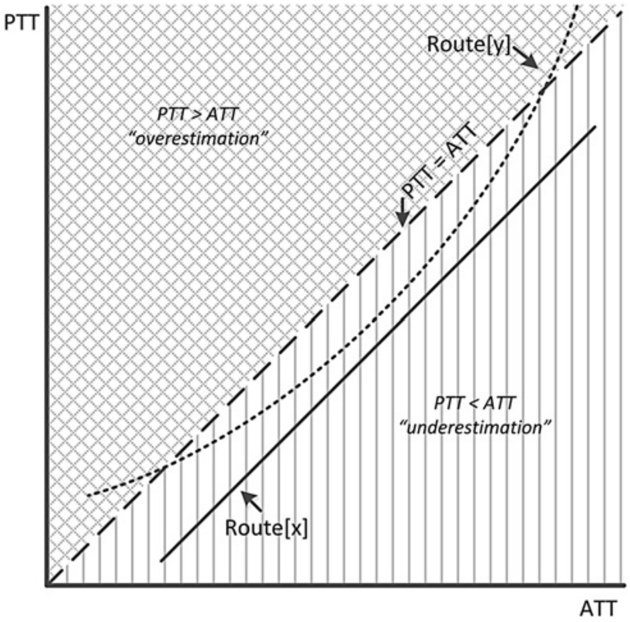

Fig. 1 *Perceived travel time (PTT) against ATT*

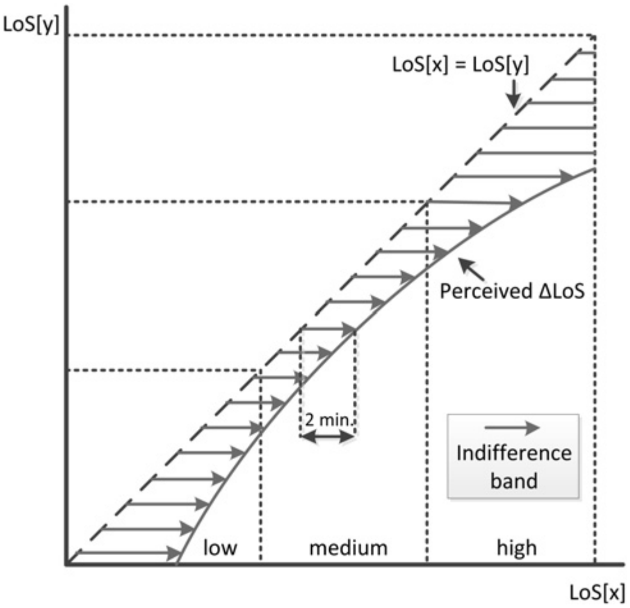

Fig. 3 *Level of service of route y (LoS[y]) against level of service of route x (LoS[x])*

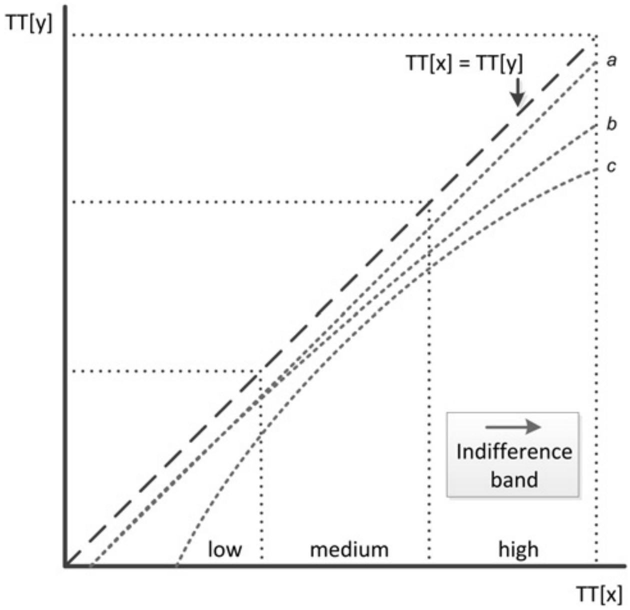

Fig. 4 *Width of the IB*

a Absolute value
b Percentage
c Continuous

being better than route *y* even though they are equal in reality. The perceived difference in level of service between both routes (ΔLoS) is plotted as the solid line with the suggested width for the three regimes low, medium and high. Consistent with Fig. 2, the horizontal lines represent the IB. On the basis of the literature, an appropriate width of the IB seems to be at least in the range of 2 min, whereas in certain circumstances, like in low and high levels of service, this width could increase to ∼10 min.

It was mentioned that the IB is situation specific, that is, route attributes are important in route choice and may influence drivers' perception. These attributes may vary over different routes and their exact influence on route choice may be hard to determine. Examples of route attributes are: directness, number of intersections, weather, information, congestion, presence of trees and so on.. Owing to the lack of situation-specific knowledge it might not be possible to estimate the width of the IB in the kind of details suggested in Fig. 4 or line *C* in the figure. One alternative is to assume that the IB can be represented by an absolute value which is equal for all regimes. Another alternative is to express the IB as a percentage of the ATT. Hence, in absolute sense, the IB increases with increasing travel times. Both cases are shown in Fig. 4 by the dotted lines *A* and *B*, respectively.

Finally, the notion of loss aversion implies that the service levels should have a different definition in case of degradation compared to recovery of service levels. This would require a specific service level structure for both the degradation and the recovery process. Roughly, the difference between upper and lower boundaries, choice alternatives and service levels for the degradation process would become half of these differences of the recovery process. However, to limit complexity, we would not consider asymmetry effects due to loss aversion in this paper. Instead, the reader may consult [42] for several numerical examples.

3 Service level-oriented route guidance

The approach that we adopt is a recently proposed service level-oriented route guidance approach that is able to systematically improve network outflow by preventing the negative effects of spill back and capacity drop. The control process will be presented, but details on the applied controller, a finite-state machine (FSM) in combination with feedback control laws, can be found elsewhere [43, 44].

The capacity of road infrastructure drops during the onset of congestion, because the flow out of the queue is smaller than the maximum achievable flow during free flow conditions. Blocking back of queues to upstream road infrastructure can cause hindrance to flows that do not need to pass the bottleneck. Both phenomena realise a decrease in the network outflow (or more total time spent by vehicles in the system) which can be prevented by guiding traffic away from the critical bottleneck to a suitable route alternative.

The dynamic route guidance approach controls the performance of two alternative routes by maintaining predefined target service levels. The critical performance conditions at which spill back occurs within a route are defined in terms of travel times. This can be done based on simulation or empirical data. The performance of the routes is then degraded stepwise towards this critical value by step sizes that remain well within the IB of road users (i.e. the performance difference between the routes is not noticeable by the road user). However, once a route reaches its critical travel time value, traffic is sent to the alternative, such that the travel time remains constant. To maximally postpone the occurrence of blocking back, the maximum performance difference is set equal to the maximum value of the IB of road users (for the specific situation).

Target service levels of a route are degraded and recovered during, respectively, over- and under saturated traffic conditions. Oversaturated means that the traffic demand for both routes is larger than their joint capacities, resulting in increasing congestion

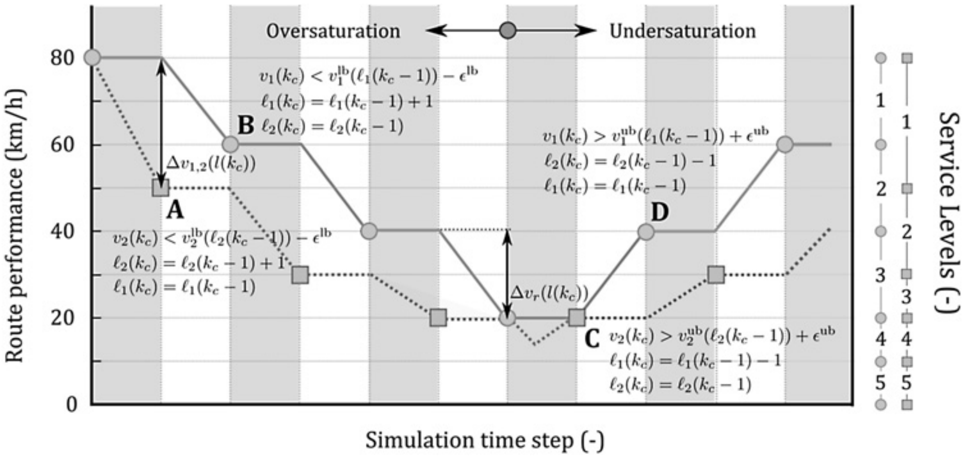

Fig. 5 *Process of degrading and recovering target service levels*

and decreasing service levels. If the demand for both routes is smaller than this joint capacity, routes are assumed to be undersaturated (even though congestion can still be present), resulting in performance recovery.

Service levels are defined as performance ranges, indicated by an upper boundary $v_r^{\mathrm{ub}}(l_r(k_\mathrm{c}))$ and a lower boundary $v_r^{\mathrm{lb}}(l_r(k_\mathrm{c}))$ of the travel time, with $r \in \{1, 2\}$ the route index, $l_r(k_\mathrm{c})$ the service level index at control interval k_c of route r. We assume that the preferred or main route between an origin is indicated with $r = 1$ and its alternative with $r = 2$. The service level upper boundaries are used as the target values to stabilise the performance of a route by sending traffic to the alternative (i.e. by adjusting the split fraction of the routable flow). Note from the table that, the boundaries of the same service level can be different for the different routes (i.e. any performance regime over the routes can be established), and that the level indices increase when the performance degrades.

3.1 *Degradation and recovery process*

The degradation and recovery process is briefly elaborated by means of Fig. 5. We assume all routes $r \in \{1, 2\}$ to initially perform within their first service level $l_r(0) = 1$ (i.e. both routes are in free flow conditions). During oversaturated conditions, the upper boundary of the main route's first service level $v_1^{\mathrm{ub}}(l_1(k_\mathrm{c}))$ is maintained and the performance of the alternative $v_2(k_\mathrm{c})$ is allowed to degrade until its first service level lower boundary $v_2^{\mathrm{lb}}(l_2(k_\mathrm{c}))$ in point A. Once this boundary is reached, the alternative's target service level at the current control interval k_c is increased to $l_2(k_\mathrm{c}) = l_2(k_\mathrm{c} - 1) + 1$ and the corresponding upper boundary value $v_2^{\mathrm{ub}}(l_2(k_\mathrm{c}))$ is maintained. The performance of the main route $v_1(k_\mathrm{c})$ is subsequently allowed to degrade until its first service level lower boundary $v_1^{\mathrm{lb}}(l_1(k_\mathrm{c}))$ in point B. Once reached, the service level of the main route is increased to $l_1(k_\mathrm{c}) = l_1(k_\mathrm{c} - 1) + 1$ and the corresponding value of the second service level is maintained $v_1^{\mathrm{lb}}(l_1(k_\mathrm{c}))$. As long as oversaturated conditions remain, this procedure will degrade the performance stepwise.

When the situation becomes under saturated, the route that is not kept at constant performance will recover until its active service level upper boundary $v_r^{\mathrm{ub}}(l_r(k_\mathrm{c}))$ is reached as can be seen in point C. Here, the performance of the alternative crosses its active performance upper boundary, hence the target service level of the main route is decreased to $l_1(k_\mathrm{c}) = l_1(k_\mathrm{c} - 1) - 1$ and the active upper boundary $v_2^{\mathrm{ub}}(l_2(k_\mathrm{c}))$ of the alternative maintained, so that the main route will further recover. If the main route crosses its performance upper boundary, the target service level of the alternative is decreased to $l_2(k_\mathrm{c}) = l_2(k_\mathrm{c} - 1) - 1$ and the upper boundary of the main route maintained, so that the alternative will recover.

The mechanism is designed such that the preferred route recovers before the alternative does, and that the target service levels of the routes never differ more than one service level index. With respect to the adoption of the psychological constructs, the following aspects of the service level definitions are important:

- The maximum performance difference between two routes per service level is determined by $\Delta v_{1,2}(l(k_\mathrm{c})) = v_2^{\mathrm{lb}}(l(k_\mathrm{c})) - v_1^{\mathrm{ub}}(l(k_\mathrm{c}))$
- The degradation step size within a service level of a route is determined by $\Delta v_r(l(k_\mathrm{c})) = v_r^{\mathrm{lb}}(l(k_\mathrm{c})) - v_r^{\mathrm{ub}}(l(k_\mathrm{c}))$

When maintaining route service levels, the boundary values are always translated into travel times. This prevents unrealistic and unfair travel time differences between choice alternatives, from as small variations in low speeds result in much larger travel time differences than small variations in high speeds.

The aim of service level-oriented route guidance approach is to guide rather than to inform drivers about delays in the network. Much research has been devoted to choice modelling, driver compliance and the influence of information, for reviews see [45–49]. We acknowledge that these are relevant aspects of route guidance. Clearly, the proposed control approach can only have an impact if the size of the controllable flow and the compliance rate are large enough. For this reason, we will leave this topic out of consideration and focus on the application of IBs in the service-level control approach. We believe that if a control approach is designed to respect the expectations of drivers it will be successful. In the remainder of this paper, when we refer to route guidance we refer to variable message signs (VMS) that inform drivers about the preferred route to a certain destination. No travel times or delays are shown, nor do the drivers receive any form of compensation or incentive to use the preferred route.

4 Test case

By means of a simulation test case, the potential to improve the network performance while respecting the threshold values of the IB is illustrated. To this aim, alternative indifferent bands are applied within the control approach to evaluate the corresponding network performance. It is also shown, how the IB is adopted into the service level-based control approach. Moreover, a comparison is made with system optimal route guidance that is realised by model predictive control (MPC) and user optimal route guidance realised by a predictive feedback control approach.

4.1 *Benchmark approaches*

4.1.1 MPC-based route guidance: MPC is used to solve the problem of realising system optimal route guidance. Optimal control signals are computed by numerical optimisation in each control interval. To this aim, a model is used to evaluate the system performance over a prediction horizon based on the current state of the system, the expected disturbances (e.g. demands) and some planned control trajectory. The corresponding performance of the system (e.g. the total time spent by vehicles in the system) is then evaluated by an objective function. The optimisation procedure minimises this objective function's value by means of a suitable optimisation algorithm. From the resulting optimal control trajectory, only the first element is applied to the process model. In the next control time step, a new optimisation is performed (with a prediction horizon that is shifted one control time step ahead) and of the resulting control trajectory the first element is again applied and so on. For more information on MPC see (Hegyi2004) and the references therein.

4.1.2 User equilibrium feedback: State feedback control is used to equalise the travel times over the route alternatives. By aiming at a zero-travel time difference, control signals are composed such that vehicles are routed towards the time-shortest route. By doing so, travel time in the shorter alternative will gradually increase again. Details on the control approach can be found in [50, 51].

4.2 *Arrival travel times*

To prevent the system from becoming unstable when using state feedback control, predicted (or arrival) travel times are used to determine the route guidance signals. Instability might result from the fact that the effect of a route guidance signal on the bottleneck location is delayed (i.e. the delay is determined by travel time between the VMS and bottleneck). Hence, for every control interval, a model is used to determine the predicted travel times of the routes.

4.3 *Applied traffic flow model*

The macroscopic first-order multi-class cell-based traffic flow model Fastlane [52] has been used for the process simulation, the state predictions of the service level-oriented and user equilibrium state feedback routing approach and for the optimisation procedure within the MPC route guidance approach. Fastlane propagates traffic flow destinations dependent through the network, enabling correct manipulation of flows by means of route guidance between an origin and destination pair. This also allows for proper simulation of the onset and dissolving of congestion, including the negative effects of the spill back phenomenon. The controllers manipulate the destination dependent split fractions within the node model. Moreover, no stochasticity is included, that is, scenarios are compared based on a deterministic single simulation run.

4.4 Performance indicators

The different control methodologies are evaluated based on the network performance indicator: the total time that vehicles have spent in the network (TTS). In the context of the above mentioned cell-transmission mode, the time spent by $N(k)$ vehicles in one time step k is $TN(k)$ and the total time that the vehicles spent in the network over a period $k = \{0, 1, \ldots, K\}$with K the total number of simulation time steps becomes

$$J_{\text{TTS}} = T \sum_{k=1}^{K} \sum_{m \in M} \sum_{c \in C_m} \rho_{m,\text{c}}(k) \lambda_{m,\text{c}} \tag{1}$$

with $\rho_{m,\text{c}}(k)$ the vehicle densities over the cells $c \in C_m$ of all links $m \in M$ in the network and $\lambda_{m,\text{c}}$ the corresponding cell lengths.

4.5 Test case layout

The applied traffic network and its characteristics are shown in Fig. 6. The VMS to distribute traffic is located in the north. Traffic moves from origin O_1 towards destinations D_1 in the east and D_2 in the south. Destination D_2 can be reached by a preferred route (main route) on the east side or the alternative on the west side. The main route is considered more important since a considerable part consists of a freeway section that is also used by other large traffic flows travelling towards destination D_1. Within each route, a bottleneck is located with a fixed capacity of 800 veh/h (e.g. representing an intersection) to realise congestion. Traffic is loaded into the network at origin O_1 over a 3 h simulation period.

The inflow at simulation time kT is interpolated from the pattern given in Table 1. From the total demand, 50% travels towards destination D_1 and 50% towards destination D_2. The compliance rate of traffic to a given advice is assumed to be 30% and the nominal split fraction at the node downstream the VMS towards destination D_2 over the main route is 50%.

4.6 Service level definition

The policy behind the test case is to increase the network production, with the restriction that the travel time difference over the routes

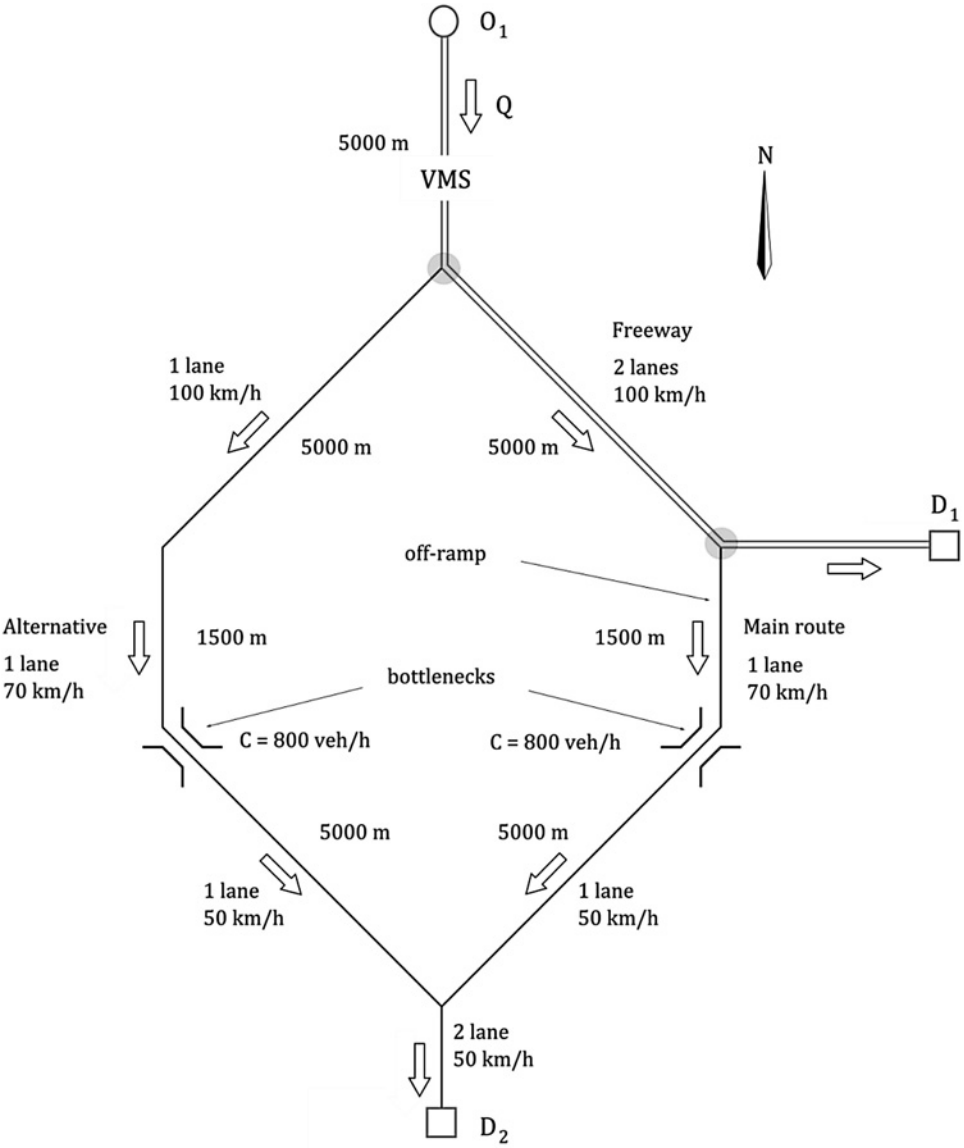

Fig. 6 *Test case network*

Table 1 Demand Q loaded at origin O_1

time	(hh:mm)	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00
demand	(veh/h)	2000	4000	4000	3500	2500	2500	0	0	0

should be less than the prevailing IB. The applied target service levels are given in Table 2. On the basis of a simulation experiment, the critical travel time at which the congestion in the main route spills back to the freeway is ~1100 s.

As long as the travel time within the main route is below this critical travel time, the routes can be degraded in an arbitrary number of steps that keep the travel time difference between the routes well within the IB. As can be seen in Table 2, the degradation step size of service levels 1–4 is chosen as 120 s, resulting in a maximum performance difference of 120 s over the routes (i.e. $\Delta\tau_{1,2}(l(k_c)) = \tau_2^{\text{lb}}(l(k_c)) - \tau_2^{\text{ub}}(l(k_c)) = 120$ for $l(k_c) = \{1, 2, 3, 4\}$.

If the travel time becomes longer than the critical value, the capacity at the freeway will drop because of the onset of congestions, resulting in decreased network performance. By sending vehicles to the route alternative, the travel time in the main route can be stabilised. This means that the critical travel time value is maintained once the main route degraded to service level 5. The IB that holds for the specific situation then determines the maximum acceptable travel time difference over the routes (i.e. the achievable gain in network performance without user interests being violated). In the test case, we study the potential gain in network performance by evaluating various absolute IBs (i.e. IB = {120, 240, 360, 480, 600, 720, 840} seconds). Table 2 illustrates a situation with a maximum value of the IB of 600 s, that is, the maximum difference between the lower bound of main route and the upper bound of the route alternative.

Table 2 Service level table for the test case with the first and second column of a route indicating the service level upper boundary (ub) and lower boundary (lb) in terms of travel time (s)

Levels	Main route		Alternative	
$l(k_c)$	$\tau_1^{\text{lb}}(l_1(k_c))$	$\tau_1^{\text{lb}}(l_1(k_c))$	$\tau_2^{\text{ub}}(l_2(k_c))$	$\tau_2^{\text{lb}}(l_2(k_c))$
1	630	750	630	750
2	750	870	750	870
3	870	990	870	990
4	990	1110	990	1110
5	1110	1230	1110	1710
6	1230	1350	1710	1830
7	1350	1470	1830	1950
…	…	…	…	…

Hence, once the main route is degraded to service level 5, the critical performance value of 1110 s (closest to the 1100 s value described earlier) is maintained and the alternative accepted to degrade until a travel time difference is established at 600 (i.e. $\Delta\tau_{1,2}(l(k_c)) = \tau_2^{\text{lb}}(l(k_c)) - \tau_1^{\text{ub}}(l(k_c)) = 600$ for $l(k_c) \geq 5$).

When the alternative has degraded to service level 6, blocking back is no longer prevented because of the IB constraint. To conclude, the specific tuning parameters (e.g. feedback gains) of the controller are chosen in line with the settings used in [44].

5 Results

5.1 Travel times and queue lengths

In Fig. 7, the travel times and corresponding queue lengths are given for the different control approaches per route. To realise system optimality in the test case, the MPC approach makes sure that the bottlenecks in the main route and alternative become active and released at the exact same time, and that the off-ramp queue in the main route does not spill back over upstream bifurcation point (line MPC main). As long as both bottlenecks are active and no other flows are hindered by spill back, it does not matter where the queues are located. In that respect, the MPC approach accepts a large travel time at the alternative (line MPC alt), resulting in a travel time difference over the main and alternative route that is larger than the IB.

For the user optimal solution, the travel times remain the same (line user equilibrium (UE) main/alt), however, the corresponding queue lengths indicate the disadvantage of this approach. As can be seen in Fig. 7*b* by the grey continuous line, the off-ramp queue within the main route spills back over the upstream bifurcation (off-ramp length is 1500 m) in an early stage, causing the capacity drop and hence decreased network performance.

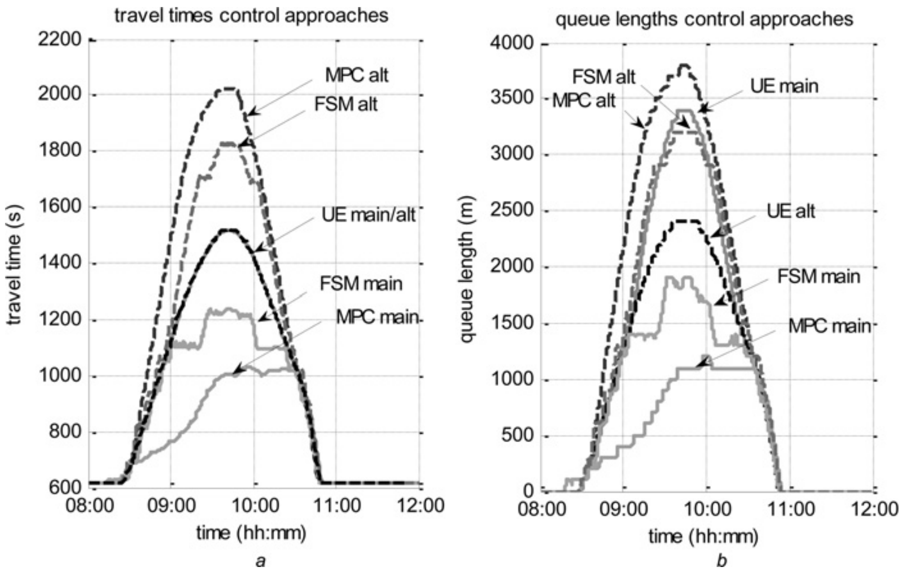

Fig. 7 *Different control approaches for the preferred route (main) and the alternative (alt) for*

a Travel times
b Queue lengths

Service level-oriented control that is realised by a FSM degrades the performance of the main route and alternative stepwise according the target service levels given in Table 2 (line FSM main and FSM alt). At 9:00 h the critical travel time is stabilised and the alternative is allowed to degrade until a travel time difference is realised of 600 s (i.e. the assumed IB). However, as can be seen by the orange continuous line in Fig.7*b*, spill back is not completely prevented within the main route after 9:20 h, since the queue length exceeds the off-ramp length of 1500 m. This result indicates that a travel time difference larger than 600 s is needed to completely prevent spill back from occurring. Shorter travel time differences will allow the main route queue to spill back over the bifurcation node in an earlier stage.

5.2 Travel times against IBs

In Fig. 8, the realised travel times over the main route and alternative are given for the FSM approach maintaining various predefined absolute IBs. The steps in the travel time data indicate the stepwise degradation and recovery of the route performance. The middle diagonal illustrates the user equilibrium situation and the other diagonals illustrate the acceptable relative deviation of the equilibrium situation (i.e. IBs in relive terms). Acceptable travel times over both routes therefore need to stay between the 0% diagonal and the formulated maximum IB definition (i.e. defined in either in relative or absolute terms).

The target service levels for degrading the main route to its critical performance value of 1110 s is the same for all IB settings. This can be seen by the strong overlap of data points until the travel time of 1110 s is realised within the main route. At this critical performance, the chosen absolute value of the IB (i.e. IB = {120, 240, 360, 480, 600, 720, 840}) is defined by the maximum deviation from the 0% diagonal. The applied absolute IBs directly determine the achievable network performance gain with respect to user equilibrium conditions. Note that relative IBs can be used as well to determine the maximum absolute acceptable travel time difference that can be maintained by the controller. Moreover, this type of plot can be used to assess if the resulting travel times from a route guidance approach satisfy the defined IBs.

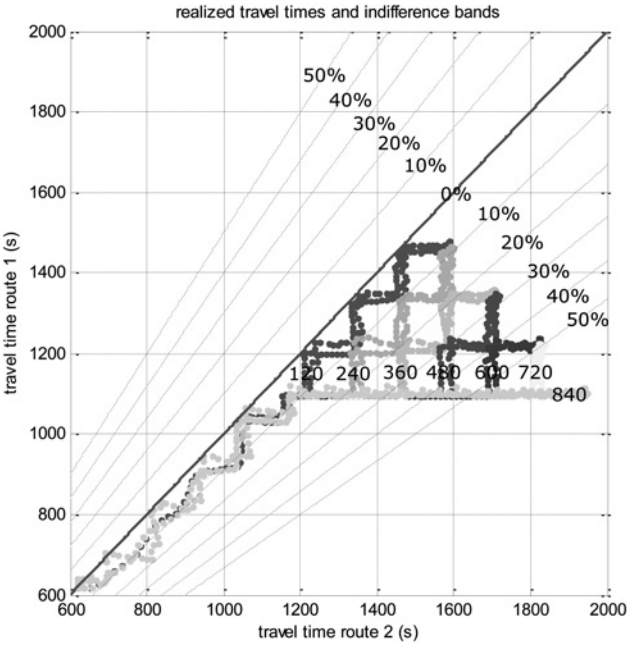

Fig. 8 *Realised travel times on main route and alternative route due to the service level-oriented control approach with adopted absolute IB values IB = {120, 240, 360, 480, 600, 720, 840}*

Diagonal lines additionally illustrate perceived IB boundaries in relative terms

Table 3 Network performance in TTS resulting from the UE, MPC and FSM-IB approach

	TSS_{D_1}, h	TSS_{D_2}, h	TSS_{TOT}, h	Decrease TSS_{D_1}
UE	787	1999	2784	—
MPC	661	1998	2660	16
FSM-IB-120	774	2011	2785	1.7
FSM-IB-240	748	2011	2759	5
FSM-IB-360	716	2011	2727	9
FSM-IB-480	699	2011	2710	11.2
FSM-IB-600	675	2011	2686	14.2
FSM-IB-720	668	2011	2679	15.1
FSM-IB-840	661	2011	2672	16

5.3 Network performance

In Table 3, the network performance of the user optimal approach, the system optimal approach and the service level-oriented approach that corresponds with the different IB settings are given. Both the user optimal and system optimal approaches realise the lowest TTS of traffic towards destination 2. The reason is that the optimal controller is able to determine the control signals that realise activation and release of the bottlenecks in the main route and alternative route at the same time. The user optimal solution in this specific case does the same by keeping travel times equal, since both routes have the same characteristics (length, speed). The service level-oriented approach realises little underutilisation (increase 0.6% TTS_{D_2}) in the under saturated phase when the bottleneck within the main route is released and the alternative route still has to recover from the lower bound performance of its first service level to free flow conditions. However, the total time spent of potentially hindered traffic to D_1 is of real interest, since hindrance to this flow strongly influences the network performance. The decrease of TTS to D_1 is therefore given in column 4.

The table shows for instance that an absolute IB of 4 min reduces the TTS of traffic that does not need to pass the bottleneck by 5%, whereas an IB of 10 min realises a 14% decrease of TTS to on-going traffic.

6 Conclusions

Road users have difficulty in assessing the quality of their chosen alternative. Building upon the notion of IBs, we have introduced a service level-oriented route guidance approach that utilises the inability to improve the network performance, without road user interests being violated.

Estimating the width of the IB is not trivial. It is situation specific and subject to drivers' perception of a route relative to drivers' perception of another route as well as reality. In case of insufficient knowledge to estimate the IB in great detail, we illustrated several other ways for interpretation and quantification of the IB. In this paper, the effect on the network performance and application of IBs in the route guidance approach was explored by means of a simulation test case. By applying absolute IBs ranging from 2 to 10 min, the test case showed network performance gains between 2 and 14%.

The IBs are easily adopted in the applied service level-oriented route guidance approach. The approach properly degrades and restores the performance of the controlled routes according to the defined target service levels (including the IBs). Hence, the behaviour of the control approach is comprehensible. As long as monetary incentives are not given to road users to make system optimal route decisions, the utilisation of IBs offers an acceptable trade-off between policy objectives of road authorities and the interests of individual road user. Although we focus on route guidance in this paper, we consider that this approach is suitable for any DTM system that influences the network performance in terms of travel times, delay times, traffic density, average speed and so on. This includes traffic lights, variable message signs, ramp metering, lane management and so on. As long as the IB is respected, driver response is assumed to be limited even if their situation declines. Conversely, if road authorities aim to change route choice, the effect of their measures should exceed the IB. Either way, the effectiveness of DTM is likely to increase when drivers' expectations are considered by means of the IB.

Finally, we would like to recommend several avenues for further research which, we were unable to capture within this paper. First of all, it would be interesting to assess the effects of day-to-day dynamics and driver learning on the performance of the control approach, in more complex and real-life situations. Secondly, the route guidance signal to drivers (e.g. travel time information, route advice) should be optimised to achieve high levels of compliance. In addition, in this study we assumed fixed driver compliance which in reality may vary and yield a different outcome in certain situations. It is recommendable to evaluate the effects of different user compliance rates. Finally, more empirical material is needed to estimate the width of the IB. At best, such estimate should provide a minimum width that is common for all cases and some direction for additional width in specific circumstances. It is particularly needed to understand what a realistic IB in any context is. For example, an indifference of 10 min for a trip of about 22 min (as used in this study) seems unrealistic. However, from a different viewpoint drivers in this network were used to 15 min of delay in comparison to free flow traffic in the case of user equilibrium. With delay in that order of magnitude, an IB of 10 min, or one as small as 4 min seems very reasonable.

7 Acknowledgments

The research presented in this article is on the intersection of two PhD researches. One of the two is part of the research program 'Traffic and Travel Behaviour in case of Exceptional Events', sponsored by the Dutch Foundation of Scientific Research MaGW-NWO. The other is funded by Imtech Traffic & Infra and University of Twente, both in The Netherlands.

8 References

1 Wardrop, J.G.: 'Some theoretical aspects of road traffic research', *Proc. Inst. Civ. Eng.*, 1952, **1**, pp. 325–378

2 Van Vugt, M.: 'Social dilemmas and transportation decisions' (Rijksuniversiteit Limburg, Maastricht, The Netherlands, 1996)

3 Koutsoupias, E., Papadimitriou, C.: 'Worst-case equilibria', *Comput. Sci. Rev.*, 2009, **3**, pp. 65–69

4 Simon, H.A.: 'A behavioural model of rational choice', *Q. J. Econ.*, 1955, **69**, pp. 99–118

5 Ariely, D.: 'Predictably irrational – revised and expanded edition' (HarperCollings Publishers, London, 2009)

6 Mahmassani, H.S., Chang, G.-L.: 'On boundedly rational user equilibrium in transportation networks', *Transp. Sci.*, 1987, **21**, (2), pp. 89–99

7 Mahmassani, H.S.: 'Dynamics of commuter behavior: recent research and continuing challanges', in Lee-Gosselin, M., Stopher, P. (Eds.): 'Understanding travel behavior in an era of change' (Pergamon Press, New York, 1996)

8 Vreeswijk, J., Bie, J., Van Berkum, E., Van Arem, B.: 'Effective traffic management based on bounded rationality and indifference bands', *IET Intell. Transp. Syst.*, 2013, **7**, (3), pp. 265–274

9 Avineri, E., Prashker, J.N.: 'Violations of expected utility theory in route-choice stated preferences', *Transp. Res. Rec.*, 2004, **1894**, pp. 222–229

10 Srinivasan, K.K., Mahmassani, H.S.: 'Role of congestion and information in trip-makers' dynamic decision processes experimental investigation', *Transp. Res. Rec.*, 1999, **1676**, pp. 44–52

11 Chorus, C.G., Timmermans, H.J.P.: 'Measuring user benefits of changes in the transport system when traveler awareness is limited', *Transp. Res. A: Policy Pract.*, 2009, **43**, (5), pp. 536–547

12 Guo, X., Liu, H.X.: 'Bounded rationality and irreversible network change', *Transp. Res. B: Methodol.*, 2011, **45**, (10), pp. 1606–1618

13 Avineri, E.: 'Choice architecture' and the design of advanced traveler information systems (ATIS)'. Proc. of the 13th IEEE – ITS Conf. – Workshop on Traffic Behavior, Modeling and Optimization, Madeira, 2010

14 Simon, H.A.: 'Models of bounded rationality: behavioral economics and business organization' (The MIT Press, Cambridge, 1982)

15 Chang, G.-L., Mahmassani, H.S.: 'The dynamics of commuting decision behavior in urban transportation networks', in Gerardin, B., (Ed.): 'Travel behaviour research' (Gower Publishing, Aldershot, England, 1989), pp. 15–26

16 Jayakrishnan, R., Mahmassani, H.S., Hu, T.Y.: 'An evaluation tool for advanced traffic information and management systems in urban networks', *Transp. Res. C*, 1994, **2**, (3), pp. 129–147

17 Manley, E.J., Orr, S.W., Cheng, T.: 'A heuristic model of bounded route choice in urban areas', *Transp. Res. C: Emerg. Technol.*, 2015, **56**, pp. 195–209

18 Khisty, C.J., Arslan, T.: 'Possibilities of steering the transportation planning process in the face of bounded rationality and unbounded uncertainty', *Transp. Res. C: Emerg. Technol.*, 2005, , **13**, (2), pp. 77–92

19 Chang, G.-L., Mahmassani, H.S.: 'Travel time prediction and departure time adjustment behaviour dynamics in a traffic system', *Transp. Res.*, 1988, **22B**, (3), pp. 217–232

20 Mahmassani, H.S., Liu, Y.H.: 'Dynamics of commuting decision behaviour under advanced traveller information systems', *Transp. Res. C: Emerg. Technol.*, 1999, **7**, (2–3), pp. 91–107

21 Jou, R.-C., Lam, S.-H., Liu, Y.-H., Chen, K.-H.: 'Route switching behavior on freeways with the provision of different types of real-time traffic information', *Transp. Res. A: Policy Pract.*, 2005, **39**, (5), pp. 445–461

22 Van Hagen, M.: 'Waiting experience at train stations' (University of Twente, Enschede, 2011)

23 Tversky, A., Kahnemann, D.: 'Advances in prospect theory: cumulative representation of uncertainty', *J. Risk Uncertain.*, 1992, **9**, pp. 195–230

24 Mahmassani, H.S., Chang, G.-L.: 'Dynamic aspects of departure-time choice behavior in a commuting system: theoretical framework and experimental analysis', *Transp. Res. Rec.*, 1985, **1037**, pp. 88–101

25 Van Knippenberg, C., Van Knippenberg, A.: 'Measurement of arrival times; a discussion of Mahmassani *et al.*, 1985', *Transp. Res. Rec.*, 1986, **1085**, pp. 47–49

26 Washburn, S.S., Ramlackhan, K., McLeod, D.S.: 'Quality-of-service perceptions by rural freeway travelers: exploratory analysis', *Transp. Res. Rec.*, 2004, **1998**, pp. 31–37

27 Papadimitriou, E., Mylona, V., Golias, J.: 'Perceived level of service, driver, and traffic characteristics: piecewise linear model', *J. Transp. Eng.*, 2010, **136**, (10), pp. 887–894

28 Choocharukul, K., Sinha, K.C., Mannering, F.L.: 'User perceptions and engineering definitions of highway level of service: an exploratory statistical comparison', *Transp. Res. A: Policy Pract.*, 2004, **38**, (9–10), pp. 677–689

29 Hostovsky, C., Wakefield, S., Hall, F.L.: 'Freeway users' perceptions of quality of service: comparison of three groups', *Transp. Res. Rec.*, 2004, **1883**, pp. 150–157

30 Zhang, L.: 'Signalized intersection level-of-service that accounts for user perceptions' (University of Hawaii, Manoa, 2004)

31 Pecheux, K.K., Pietrucha, M.T., Jovanis, P.P.: 'User perception of level of service at signalized intersections'. Proc. of the 79th Annual Meeting of the Transportation Research Board, Washington, DC, 2000a

32 Fang, F.C., Pecheux, K.K.: 'Fuzzy data mining approach for quantifying signalized intersection level of services based on user perceptions', *J. Transp. Eng.*, 2009, **135**, (6), pp. 349–358

33 Tawfik, A., Rakha, H.: 'Network route-choice evoluation in real-life experiment: necessary shift from network- to driver-oriented modeling', *Transp. Res. Rec.* 2012, **2322**, pp. 70–81

34 Jan, O., Horowitz, A.J., Peng, Z.R.: 'Using global positioning system data to understand variations in path choice', *Transp. Res. Rec.*, 2000, **1725**, pp. 37–44

35 Beckor, S., Ben-Akiva, M.E., Ramming, M.S.: 'Evaluation of choice set generation algorithms for route choice models', *Annals Oper. Res.*, 2006, **144**, (1), pp. 235–247

36 Zhu, S., Levinson, D.: 'Do people use the shortest path? An empirical test of Wardrop's first principle'. Proc. of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, 2012

37 Thomas, T., Tutert, S.I.A.: 'Route choice behavior based on license plate observations in the Dutch city of Enschede'. Proc. of the Seventh Triennial Symp. on Transportation Analysis: Tristan VII, 2010

38 Tawfik, A., Rakha, H., Miller, S.: 'Examining drivers experiences, perceptions, and choices in route choice behavior'. Proc. of the 13th IEEE – ITS Conf. – Workshop on Traffic Behavior, Modeling and Optimization, Madeira Island, 2010

39 Kahnemann, D., Tversky, A.: 'Prospect theory: an analysis of decision under risk', *Econometrica*, 1979, **47**, (2), pp. 263–291

40 Avineri, E., Bovy, P.H.L.: 'Identification of parameters for a prospect theory model for travel choice analysis', *Transp. Res. Rec.*, 2008, **2082**, pp. 141–147

41 Chorus, C.G.: 'Regret theory based on route choices and traffic equilibria', *Transportmetrica*, 2012, **8**, (4), pp. 291–305

42 Bie, J., Vreeswijk, J., Van Berkum, E.: 'Adopting prospect theory for modeling day-to-day traffic dynamics: risk aversion and equilibrium stability'. Fourth Int. Symp. on Dynamic Traffic Assignment, Martha's Vineyard, MA, 2012

43 Landman, R., Schreiter, T., Hegyi, A., Van Lint, H., Hoogendoorn, S.: 'Policy-based, service level-oriented, route guidance in road networks: a comparison with system and user optimal route guidance', *Transp. Res. Rec.*, 2012, **2278**, pp. 115–124

44 Landman, R.L., Hegyi, A., Hoogendoorn, S.P.: 'Service level-oriented route guidance in road traffic networks'. Proc. of the 14th Int. IEEE – ITS Conf., Washington, DC, 2011

45 Prato, C.G.: 'Route choice modeling: past, present and future research directions', *J. Choice Model.*, 2009, **2**, (1), pp. 65–100

46 Bonsall, P.: 'The influence of route guidance advice on route choice in urban networks', *Transportation*, 1992, **19**, (1), pp. 1–23

47 Chorus, C.G., Arentze, T.A., Timmermans, H.J.P.: 'Traveler compliance with advice: a Bayesian utilitarian perspective', *Transp. Res. E: Logist. Transp. Rev.*, 2009, **45**, (3), pp. 486–500

48 Chorus, C.G., Molin, E.J.E., Van Wee, B.: 'Use and effects of advanced Ttaveller information services (ATIS): a review of the literature', *Transp. Rev.*, 2006, **26**, (2), pp. 127–149

49 Han, Q., Dellaert, B., van Raaij, F., Timmermans, H.: 'Modelling strategic behaviour in anticipation of congestion', *Transportmetrica*, 2007, **3**, (2), pp. 119–138

50 Hegyi, A.: 'Model predictive control for integrating traffic control measures' (Delft University of Technology, Delft, the Netherlands, 2004)

51 Wang, Y., Papageorgiou, M., Messmer, A.: 'Predictive feedback routing control strategy for freeway network traffic', *Transp. Res. Rec.*, 2003, **1856**, pp. 62–73

52 Van Lint, H., Hoogendoorn, S., Schreuder, M.: 'Fastlane: new multiclass first-order traffic flow model', *Transp. Res. Rec.*, 2008, **2088**, pp. 177–187