Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 186 (2008) 990–1007
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Discrete Optimization

Vehicle routing with dynamic travel times: A queueing approach

T. Van Woensel [a,*], L. Kerbache [b], H. Peremans [c], N. Vandaele [d]

[a] TU/e, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] HEC School of Management, Paris, France
[c] UA, University of Antwerp, Antwerp, Belgium
[d] KULeuven, Catholic University of Leuven, Leuven, Belgium

Received 23 November 2005; accepted 26 March 2007
Available online 7 April 2007

Abstract

Transportation is an important component of supply chain competitiveness since it plays a major role in the inbound, inter-facility, and outbound logistics. In this context, assigning and scheduling vehicle routes is a crucial management problem. In this paper, a vehicle routing problem with dynamic travel times due to potential traffic congestion is considered. The approach developed introduces mainly the traffic congestion component based on queueing theory. This is an innovative modeling scheme to capture travel times. The queueing approach is compared with other approaches and its potential benefits are described and quantified. Moreover, the optimization of the starting times of a route at the distribution center is evaluated. Finally, the trade-off between solution quality and calculation time is discussed. Numerous test instances are used, both to illustrate the appropriateness of the approach as well as to show that time-independent solutions are often unrealistic within a congested traffic environment, which is usually the case on European road networks.

Keywords: Vehicle scheduling; Dynamic travel times; Queueing

1. Introduction

Routing problems have been largely studied due to its interest in different applications in logistics and supply chain management. Not surprisingly, transportation is an important component of supply chain competitiveness since it plays a major role in the inbound, inter-facility, and outbound logistics. Transportation costs represent approximately 40–50% of total logistics costs and 4–10% of the product selling price for many companies [11]. As such, transportation decisions directly affect the total logistic costs. The passage of the transportation deregulation acts in the 1980s in the USA and in the 1990s in the EU drastically changed the business climate, within which the transportation managers operate. Within the EU, the competition is becoming intense between transporters since they often operate at transnational levels and must provide higher levels of service with lower costs to meet the various needs of customers. In this context, assigning, scheduling and routing the fleet of a transportation company is a crucial management

* Corresponding author. Tel.: +31 40 247 5017.
E-mail addresses: t.v.woensel@tm.tue.nl (T. Van Woensel), Kerbache@hec.fr (L. Kerbache), herbert.peremans@ua.ac.be (H. Peremans), Nico.Vandaele@kuleuven-kortrijk.be (N. Vandaele).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2007.03.012

issue. From an operations research point of view, these operational decisions can be formulated as a vehicle routing problem. In general, the vehicle routing problem (VRP) aims to construct a set of shortest routes for a fleet of vehicles of fixed capacity. Each customer is visited exactly once by one vehicle which delivers the demanded amount of goods to the customer. Each route has to start and end at a depot, and the sum of the demands of the visited customers on a route must not exceed the capacity of the vehicle. In this paper, it is assumed that there is only one depot from where the routes start and end for each vehicle, a homogeneous fleet consisting of several vehicles with fixed capacity, while each customer's demand is predetermined i.e. static and deterministic demand.

Formally, the vehicle routing problem can be represented by a complete weighted graph $G = (V, A, c)$ where $V = \{0, 1 \ldots n\}$ is a set of vertices and $A = \{(i,j) : i <> j\}$ is a set of arcs. The vertex 0 denotes the depot; the other vertices of V represent cities or customers. The non-negative weights c_{ij}, which are associated with each arc (i,j), represent the cost (distance, travel time or travel cost) between i and j. For each customer, a non-negative demand qd_i and a non-negative service time d_i is given ($d_0 = 0$ and $qd_0 = 0$). The objective is then to find the minimum cost vehicle routes where the following conditions hold: every customer is visited exactly once by exactly one vehicle; all vehicle routes start and end at the single depot; every vehicle route has a total demand not exceeding the vehicle capacity Q; every vehicle route has a total route length not exceeding the maximum length L [28]. If it seems reasonable to assume that the service time at each vertex (customer) is known in advance, it is definitely not the case for the travel time between two vertices. In fact, the travel times are the result of a stochastic process related to traffic congestion. Clearly, travel times depend greatly on the different number of vehicles occupying the road and on their speeds. In this paper, the VRP problem considered deals with dynamic travel times. In this case, the non-negative weights c_{ij}^p associated with each arc (i,j), represent the travel time between i and j starting in time zone p.

The main contributions are the following:

(i) Time-dependent travel times (modeled using a queueing approach) are introduced in the routing problem. More specifically, the stochastic and dynamic nature of travel times is captured using queueing theory applied to traffic flows. A major advantage of using these queueing models is that the real-life physical characteristics of the road network can be mapped immediately onto the parameters of the queueing model. Moreover, the inherent stochasticity of travel times can explicitly be taken into account via the analytical queueing models. We show that our approach is superior compared to the time-independent situation and has significant gains compared to other time-dependent modeling approaches. On top of this, in most real-life cases, speed are not available or are only available for a small part of the road network. Flows, which are much easier to measure and estimate, are however more available (e.g. in The Netherlands the expected loads of the road network are reported [40]). The queueing approach offers in this case a valuable model for translating flows into speeds.

(ii) The effect of the time discretization on both the solution quality and the calculation times is evaluated in great detail. Our approach allows for any number of time zones in which the time is discretized. We compare our approach with 144 time zones of each 10 minutes with other approaches found in the literature (i.e. Ichoua [23] with 3 zones; no zones or time-independent). In combination with this time granularity, we also analyze the effect of different road types on the solution quality obtained. We found that a higher number of time zones, improves the solution quality. Moreover, this effect is even greater for different road types.

(iii) In a dynamic environment, a truck can decide to leave earlier or later to avoid periods of (anticipated) high congestion. The value of such a decision is quantified in this paper. We found that optimizing the starting times for a solution has a significant effect on the obtained solution quality.

(iv) Finally, the basic trade-off between solution quality and calculation time is evaluated in detail. Based on the results, we found that the extra calculation time for large datasets is significant but is certainly worthwhile if one takes into account the improvement in solution quality. The analysis done gives important insights to a decision maker to decide whether he wants to invest the extra needed calculation time.

This paper is organized as follows: in Section 2, the literature background on our variant of the vehicle routing problems is briefly presented. Section 3 is devoted to the modeling and determination of the travel time distribution. Section 4 addresses the experimental design, the input data and the solution approach used to cope with the dynamic vehicle routing problem. In Section 5, computational results are presented based on some standard datasets. The managerial implications are discussed in more detail in Section 6. Finally, conclusions are presented in Section 7.

2. Literature review

Despite numerous publications dealing with efficient scheduling methods for vehicle routing, very few addressed the inherent stochastic nature of this problem. Most research in this area has focused on dynamic routing and scheduling that considers the variation in customer demands. However, there has been limited research on routing and scheduling with congestion dependent travel times. The motivation for using time-dependent models is that the vehicles in the routing problem operate in a traffic network, which will be congested depending upon the time of the day. If it seems reasonable to assume that the service time at each vertex (customer) is known in advance, it is definitely not the case for the travel time between two vertices. In fact, the travel times are the result of a stochastic process related to traffic congestion. Clearly, travel times depend greatly on the different number of vehicles occupying the road and on their speeds.

The literature related to vehicle routing with dynamic travel times is rather scarce [14,23,30,22,31,5–7]. The reason is that the dynamic VRP is much harder to model and to solve. The travel time function is modeled through discretizing the day into a very limited number of time intervals (e.g., morning, midday and afternoon) with a distinct associated fixed mean speed, which is set in an arbitrary way. For instance, Brown et al. [10] and Shen and Potvin [33] used a rough approximation of travel time by manually resequencing the route taking into account congestion. Ichoua et al. [23] used a model based on discrete travel time by adding correction factors to model the congestion. To reduce run times and complexity, the authors do their simulations and experiments with three time slices. For each link, they add a weighting factor representing the expected congestion on that link for that time slice. The speed differences due to congestion are then modeled using correction factors on the weights of the links. Any information on how to link the weights of the arcs in the model with the crucial real-life physical characteristics is missing.

Bertsimas and Simchi-Levi [8] provide an interesting survey of static vs. dynamic routing problems (e.g. the travel time being a function of the time of the day due to traffic congestion or that the time in which demand occurs is a renewal process) as well as deterministic vs. stochastic routing problems (i.e. depending whether some of the characteristics follow a certain probability distribution or not; typical examples are probabilistic demand quantities or probabilistic travel times). Moreover they examine the robustness and the asymptotic behavior of the known algorithms. Bertsimas and Simchi-Levi [8] argue that analytical analysis of the vehicle routing problem offers new insights into the algorithmic structure and it makes performance analysis of classical algorithms possible. Moreover, it leads to a better understanding of models when integrating vehicle routing with other issues like inventory control. Furthermore, they point out that dealing with stochasticity in the VRP provides insights that can be useful when constructing practical algorithms for the VRP within a dynamic and stochastic environment.

3. Modeling travel times

One of the key issue in this paper, is computing the travel times on the arcs depending on the time period. In general, the travel time from city i to city j during time period p, T_{ij}^p can be determined as follows:

$$T_{ij}^p = \frac{d_{ij}}{v_{ij}^p}.$$

Hence, to determine the travel time on arc (i,j), one needs information on the length of (i,j) and on the travel speed on that arc at time bucket $p : v_{ij}^p$. The distance is already available in the classical VRP models, but the speed still needs to be obtained.

Different alternatives are available in the literature to model speeds. Usually these models start from the observation that the speed for a certain time period tends to be reproduced whenever the same flow is observed. Based on this observation, it seems reasonable to postulate that, if traffic conditions on a given road are stationary, there should be

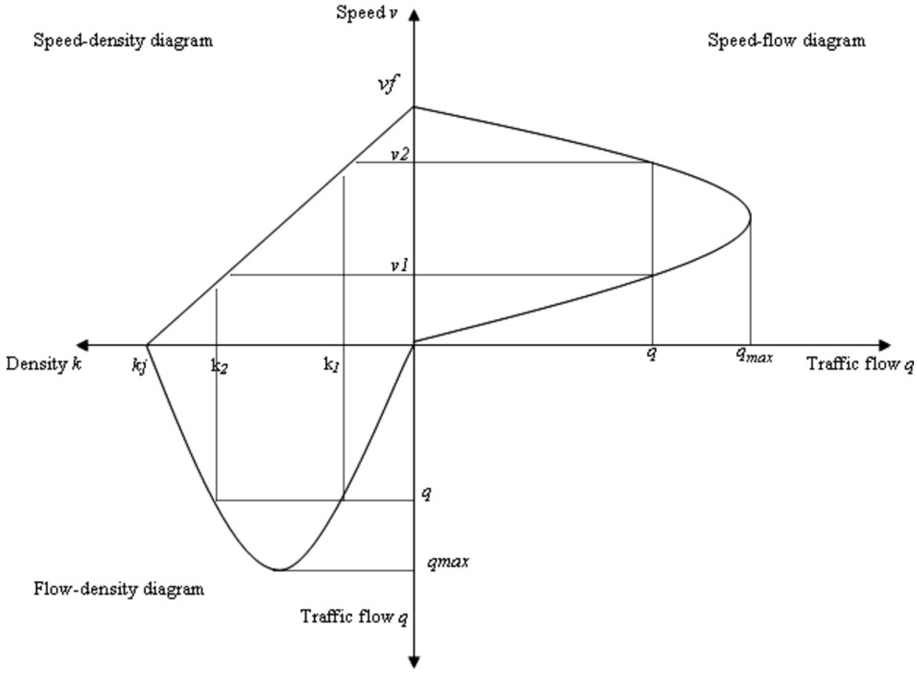

Fig. 1. The relations between the speed-flow, the speed-density, and the flow-density diagrams.

a relationship between flow, speed, and density. This relationship results in the concept of speed-flow-density diagrams. These diagrams describe the interdependence of traffic flow (q), density (k) and speed (v). The seminal work on speed-flow diagrams was the paper by Greenshields in 1935 [19]. Using well-known formulas of queueing models, these speed-flow-density diagrams can be constructed (Fig. 1).

Fig. 1 illustrates that, although every speed v corresponds with one traffic flow q, the reverse is not true. There are two speeds for every traffic flow: an upper branch (v_2) where speed decreases as flow increases and a lower branch (v_1) where speed increases. Intuitively it is clear that, as the flow moves from 0 (at free flow speed v_f) to $q_{\max}$, congestion increases but the flow rises because the decline in speed is over-compensated by the higher traffic density. If traffic tends to grow past $q_{\max}$, flow falls again because the decline in speed more than offsets the additional vehicle numbers, further increasing congestion [12]. The flow-density diagram and the speed-density diagrams are an equivalent representation and can be interpreted in the same way. Traditionally, these speed-flow-density diagrams are modeled empirically: speed and flow data are collected for a specific road and econometrically fitted into curves [12]. In this paper, these specific relationships are obtained using queueing models for traffic flows [35,20]. The basic difference is that the queueing approach structures and models the underlying processes that result in the speed-flow-density diagrams, while the other approaches observe the outcome of the underlying processes and fit a model on this outcome. As such, the queueing approach is much more flexible to react on changes (e.g. by changing the input parameters of the queueing model).

3.1. Queueing approach

Vandaele et al. [35] and Heidemann [20], showed that queueing models[1] can also be used to model traffic flows and thus offering a more analytical approach, useful for sensitivity analysis, forecasts, etc. Jain and Smith [24] describe in their paper a state-dependent $M/G/C/C$ queueing model for traffic flows. Also a lot of research is done on a travel

[1] In this paper, queueing models are referred to using the Kendall notation, consisting of several symbols – e.g. $M/G/1$. The first symbol is shorthand for the distribution of inter-arrival times, the second for the distribution of service times and the last one indicates the number of servers in the system.

time-flow model originating from Davidson [13]. The model is based on some concepts of queueing theory but a direct derivation has not been clearly demonstrated [2,3].

Vandaele et al. [35] developed different queueing models that can be used to model traffic flows. The $M/M/1$ queueing model (exponential arrival and service rates) is considered as a base case, but due to its specific assumptions regarding the arrival and service processes, it is not useful to describe real-life situations. Relaxing the specifications for the service process of the $M/M/1$ queueing model, leads to the $M/G/1$ queueing model (generally distributed service rates). Relaxing both assumptions for the arrival and service processes results in the $GI/G/m$ queueing model. Moreover, following Jain and Smith [24], a special case of the $GI/G/m$ queueing model is derived: a state-dependent $GI/G/m$ queueing model. This model assumes that the service rate is a (linear, exponential, etc.) function of the traffic flow. In this case vehicles are served at a certain rate, which depends upon the number of vehicles already on the road.

In our queueing approach to traffic flow analysis, roads are subdivided into segments, with length equal to the minimal space needed by one vehicle on that road (Fig. 2). Define k_j as the maximum traffic density (i.e. maximum number of cars on a road segment). This segment length is then equal to $1/k_j$ and matches the minimal space needed by one vehicle on that road. Each road segment is then considered as a service station, in which vehicles arrive at a certain rate λ and get served at another rate μ [35,39,20].

Following Heidemann [20], the arrival rate λ is defined as the product of the traffic density k and the free flow speed v_f, or $\lambda = k \times v_f$. Similarly, the service rate μ is defined as the product of free flow speed v_f with the maximum traffic density k_j, or $\mu = k_j \times v_f$. Vandaele [35] and Heidemann [20] showed that the speed v can be calculated by dividing the length of the road segment $\left(\frac{1}{k_j}\right)$ by the total time in the system (W)

$$v = \frac{1/k_j}{W}. \tag{1}$$

Table 1
The specific form of W_q for each queueing model

$GI/G/m$	W_q
KLB	$\frac{r}{(1-r)}\frac{(c_a^2+c_s^2)}{2}\frac{1}{k_j v_f}\exp\left[\frac{-2(1-r)(1-c_a^2)^2}{3r(c_a^2+c_s^2)}\right]$
K	$\left(\frac{c_a^2+c_s^2}{2}\right)\left(\frac{r^{(\sqrt{2(m+1)}-1)}}{m(1-r)}\right)\left(\frac{1}{k_j v_f}\right)$
W	$\phi\left(\frac{c_a^2+c_s^2}{2}\right)\left(\frac{1}{k_j v_f}\right)W_{qM/M/k}$
	With ϕ a correction factor defined in Whitt [37] and $W_{qM/M/m}$ the formula for the waiting time in an $M/M/m$ queue

The total time in the system W in formula 1 is different depending upon the specific queueing model used. The total time spent in the system W equals the sum of the waiting time W_q and the service time W_p. Table 1 shows the specific form of W_q for the general queueing models. For the $GI/G/m$ queueing models, no exact solutions are available and one must rely on approximations. Here, three approximations are considered: the Kramer–Lagenbach–Belz (KLB) approximation [26] is widely used but is limited to single servers only. To cope with multiple lanes, the heavy traffic or Kingman approximation (K) [25] and the Whitt (W) approximations [37] with multiple servers are used. In this paper, the $GI/G/m$ queueing models with the Whitt approximations are used.

Results show that the developed queueing models can be adequately used to model traffic flows [40]. Moreover due to the analytical character of these models, they are very suitable to be incorporated in other models, e.g., the VRP. In general, formula 1 can be rewritten in the following basic form (see [38] for the details):

$$v = \frac{v_f}{1+\Omega}. \tag{2}$$

Formula 2 shows that the speed is only equal to the free flow speed v_f if the factor Ω is zero. For positive values of Ω, v_f is divided by a number strictly larger than 1 and speed is reduced. The factor Ω is thus the influence of congestion on speed. High congestion (reflected in a high Ω) leads to lower speeds than the maximum. The factor Ω is a function of a number

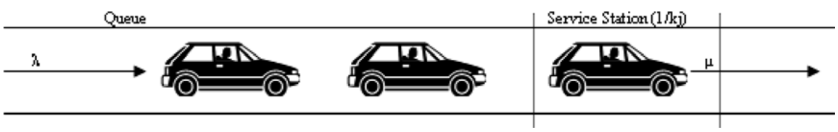

Fig. 2. Queueing representation of traffic flows.

of parameters depending upon the queueing model chosen: the traffic intensity r, the coefficient of variation of service times c_s and coefficient of variation of inter-arrival times c_a, the jam density k_j and the free flow speed v_f. High coefficients of variation or a high traffic intensity will lead to a value of Ω strictly larger than zero. Actions to increase speed (or decrease travel time) should then be focussed on decreasing the variability or on influencing the traffic intensity, for example by manipulating the arrivals (arrival management and ramp metering).

The major strength of using the queueing models is that, given the physical characteristics of the road network, it can immediately be linked with the parameters of the queueing model. In practice, the jam density and the free flow speed is fixed for a given arc (i,j), leaving only the coefficients of variation to represent the traffic conditions (e.g., bad weather, etc.). The flow q is a parameter that is determined empirically over time, allowing the determination of realistic velocity profiles as a function of time. Analytical queueing models based on traffic counts model the behavior of traffic flow as a function of the most relevant determinants (e.g. free flow speeds, jam density, variability due to weather, etc.). An empirical validation of the queueing approach is provided in Van Woensel and Vandaele [40]; validation based on simulation results is provided in Van Woensel et al. [41]. Consequently, the travel times can be modeled much more realistically using these speeds (i.e. expressed in kilometer per hour) and are directly related to the physical characteristics and the geographical location on the arc.

For a more detailed discussion of the queueing models and their results, the interested reader is referred to Vandaele et al. [35] and Van Woensel et al. [39].

3.2. Towards travel times

To compute the travel times, one should note that in the time-dependent case, the travel speeds are no longer constant over the entire length of the arc. More specifically, one has to take into account the change of the travel speed when the vehicle crosses the boundary between two consecutive time periods. For example, the speed changes when going from time period p to time period $(p+1)$ from v_{ij}^{p} to $v_{ij}^{(p+1)}$. The time horizon is discretized into P time periods of equal length Δ_p with a different travel speed associated to each time period p $(1 \leqslant p \leqslant P)$. The travel speeds are obtained using the above discussed queueing models for traffic flows. Formally, the travel time T_{ij}^{p0} going from customer i to customer j, starting at some time p_0, must satisfy the following condition:

$$\int_{p_0}^{p_0+T_{ij}^{p_0}} [v^p]dp = d_{ij}$$

with v^p denoting the speed in time period p and d_{ij} the distance traveled. Solving this integral for T_{ij}^{p0} and making use of the discrete time horizon, results in:

$$d_{ij} = \Delta_p(\varphi v^{p_0} + v^{p_0+1} + \cdots v^{p_0+(k-2)} + \phi v^{\text{last}}).$$

Rewriting as a function of the time slices, gives:

$$T_{ij}^{p_0} = \varphi\Delta_p + (k-2)\Delta_p + \phi\Delta_p.$$

The travel time is thus the sum of the following components:

(i) The fraction of travel time still available in the first time zone, given by $(\varphi\Delta_p)$ with φ the fraction parameter $(0 \leqslant \varphi \leqslant 1)$.
(ii) The duration of the $(k-2)$ intermediate time zones passed: $(k-2)\Delta_p$.
(iii) The fraction of the travel time in the last time zone, given by $(\phi\Delta_p)$, with ϕ the fraction parameter $(0 \leqslant \phi \leqslant 1)$.

Concluding, in total k time buckets are crossed. The number k is totally defined by the equations in the paper and is a function of the distance i to j and the speeds in the different time buckets. In practice, the other values are computed incrementally: Starting at time p_0 (part of the first time bucket), one knows the fraction of time left in the first time bucket and consequently if $\varphi = 1 - \frac{p_0}{\Delta_p}$. Then a number of time buckets Δ_p is added as necessary to reach the destination city j. Of course, one will spend in the last time bucket only part of the time or the fraction ϕ. This fraction is totally depending upon the residual distance that needs to be traveled in the last time bucket. Using this incremental procedure, the travel time $T_{ij}^{p_0}$ from customer i to customer j, starting at time p_0 can be determined easily based on the distance d_{ij} and the speed v^p for the different time periods p.

4. Experimental design, input data and solution approach

This section describes in detail the experimental design followed, the input data used (with regards to speeds, model parameters, etc.) and gives details on the solution approach based on the tabu search heuristic.

We compare three different speed approaches: first, no speed effects are taken into account. This is actually a variant of the time-independent model in which every vehicle travels with the same speed. Secondly, we split up the day in three big parts (morning, middle and evening) and determined the speed that minimizes the total sum of squared errors for this part of the day. This resembles almost exactly the situation described by Ichoua et al. [23], with the difference that the arc weights are now speeds rather than an arbitrary set of numbers that try to approximate the congestion behavior. Finally, we use the queueing model to set speeds for every 10 minutes interval. We optimized the queueing model to obtain the best parameter setting in a similar way as presented by van Woensel et al. [40]. Basically, a Theil coefficient is minimized such that the queueing model speeds are as close as possible to the observed real speeds.

4.1. Experimental design

The experimental design involves three steps: first, we calibrate the input used in the optimization step based on real-life data. We use a dataset with real-life observations (see next input data section in the paper) as input. Based on this data, we selected the best parameter settings for each of the three speed approaches that fitted the data best (*calibration phase*). Secondly, we generate solutions (i.e. routes) using the tabu search heuristic and for each of three speed approaches. Thirdly, these solutions are re-evaluated using a different validation dataset for a different comparable day (one week later) at the same counting point (i.e. *validation phase*). This resembles the situation where we plan a route using some average traffic information, then look at the realization of this route, and finally, evaluate and compare these routes on the realizations. This last then gives the value of the solution obtained for the three different speed approaches but evaluated using the same measure. It should be clear that all solutions are obtained using the same heuristics and its implementation but with different speed patterns as input data. Fig. 3 gives a structured overview of the experimental design used in this paper.

Experimental design
For three travel time approaches (time-independent; three time zones and queueing model), the following steps are performed:
- *Step 1: Calibration using observed data day 1:*
  - Time-independent: determine mean speed over full day
  - Three time zones: determine mean speed over each time zone
  - Queueing model: determine the optimal parameter settings
- *Step 2: Generate solutions with Tabu Search heuristic*
- *Step 3: Recalculate solution using observed speed-flow data day 2*
- *Step 4: Compare the different recalculated solutions:*
  - Based on solution quality
  - Based on starting times
  - Based on calculation times

Fig. 3. Structured overview of the experimental design.

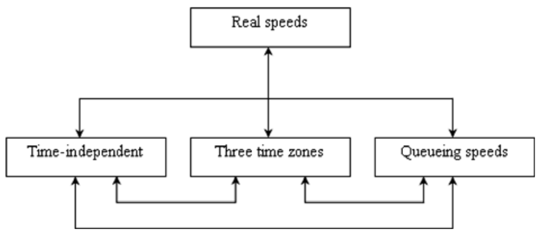

Fig. 4. The different comparisons between the speed patterns.

Fig. 4 shows the different comparisons made in this paper: we look at time-independent versus three time zones; time-independent versus the queueing approach; three time zones versus the queueing approach. This latter comparison basically tries to assess the contribution of using a more fine-grained speed function, which is based on the result of the different queueing models. We do this setup for a number of problem instances from the benchmark set described in Augerat [4].

4.2. Input data

A limited dataset collected by the ministry of transportation of the Flemish Government (Belgium) is used to calibrate and validate the model. The original dataset contains minute-per-minute observations the number of trucks, the number of passenger cars and their speeds for four counting points in Belgium for two weeks (which is the longest period for which this type of date is stored on this level of detail). One counting point of this set (Ternat) is selected. All trucks and passenger cars are converted to a vehicle equivalent, i.e. cars are 1 vehicle equivalent and trucks are 2 vehicle equivalents (see e.g. [12]). The concept of the vehicle equivalents is based on observations of freeway conditions in which the presence of heavy vehicles (e.g. trucks, buses, etc.) create less-than-ideal conditions, including longer and more frequent gaps of excessive lengths both in front of and behind these heavy vehicles. Physical space taken up by a heavy vehicle is typically two to three times greater in length than that of a passenger car (Transportation Research Board [9]). All minute-per-minute observed flows and speeds are aggregated to a 10-minute time interval and aggregated over the different lanes. More specifically, the minute-by-minute speeds are averaged over 10 minutes and the minute-by-minute flows are summed over 10 minutes (see for similar procedures e.g. the Transportation Research Board ([9])). These preparation steps provide us with the real-life speed-flow data to be used in the calibration step. Fig. 5 shows the real-life speed data, the time-independent speed used, the three time zone speeds used and the queueing speeds used. In all three approaches, the data shown is the best fit for the observed real-life speed data. We tried to optimize the time zones after careful observation of the data and found the following slots for the three time zone approach: 6.30AM to 9.00AM; 9.00AM to 4.30PM; 4.30PM to 7PM; 7PM to 6.30AM. Obviously, the identification of the relevant and best time zone boundaries is one of the difficulties of applying this approach, which is circumvented with the 10-minutes approach.

Similar to Ichoua et al. [23], we take into account multiple road types e.g. to represent highways versus rural roads. Due to the lack of real-life data on both speeds and flows for different road types, we used the same observed speeds and flows as in Fig. 5 by applying a re-scaling to represent different road types. If there is only one road type, the original series is used; if there are two road types, both the original as 60% of the original series is used; if there are three road types, the original, 60% of the original and 30% of the original series was used. In the two road types setting, all even to even node arcs are of the faster type. In the two road type setting, all even to even node arcs are of the fast type, all odd to odd node arcs are of the middle type and everything else is slow. Table 2 shows the different input parameters used considering both the number of road types as the specific speed approach. The values for the time-independent case and the three time zone case are then the values that minimized the total sum of errors (which results in the mean for the respective time zone). The parameter settings for the queueing approach are obtained via the same procedures as described in van Woensel et al. [40].

The obtained solutions using the above input data are all re-evaluated using the speed-flow pattern observed one week later for the same weekday and the same counting point. The detailed data comprising all the speed-flow data are available from the authors upon request.

4.3. Solution approach

A solution strategy based on local search is proposed. According to Aarts and Lenstra [1], local

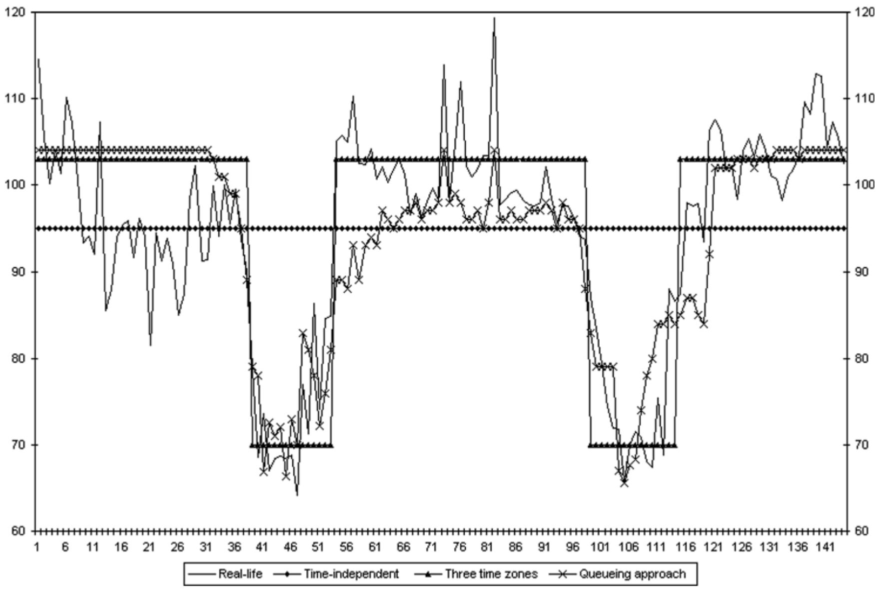

Fig. 5. The different speed patterns.

Table 2
The input parameters for the different road types and the different speed approaches

		One road type	Two road types	Three road types
Time-independent	Road 1	95	95	95
	Road 2	–	70	70
	Road 3	–	–	50
Three time zones	Road 1	103/70	103/70	103/70
	Road 2	–	62/42	62/42
	Road 3	–	–	30/21
Queueing approach	Road 1	$c_a = 0.75$; $c_a = 0.75$; $k_j = 40$; $v_f = 105$		
	Road 2	–	$c_a = 0.75$; $c_a = 0.75$; $k_j = 35$; $v_f = 70$	
	Road 3	–	–	$c_a = 0.75$; $c_a = 0.75$ $k_j = 25$; $v_f = 50$

search is a solution process that tries to improve a given initial solution by making relatively small changes in several steps in the solution space. The quality of the solutions is determined with the cost function of the problem. Local search techniques will result in a good but not necessarily optimal solution within reasonable computing time. In this paper, the tabu search heuristic is used for obtaining solutions for the vehicle routing setting presented. The success of this methods is due to several factors: general applicability of the approach, flexibility for taking into account specific constraints in real cases and ease of implementation [32].

Tabu Search can be described as a local search technique guided by the use of adaptive or flexible memory structures. Tabu search was first proposed by Glover [17,18] and involves the examination of all the neighbors of a solution of which the best is selected. To prevent cycling, solutions that were recently examined are forbidden and inserted in a constantly updated tabu list. For our tabu search implementation the following references where used as a basis: Gendreau et al. [15,16] and Hertz et al. [21]. The only change made to this basic algorithm is to replace distance by dynamic travel time. We programmed our implementation in JAVA.

Each solution is checked for 2-optimality and is improved if possible. A route is 2-optimal if it is not possible anymore to improve the route by exchanging two arcs. Unlike in the static VRP where the gain is calculated based on distances, in the dynamic VRP, the gain is calculated in terms of travel time. As the evaluation is done in terms of travel times, the arrival times at a certain node will be affected by an exchange of two arcs. Therefore, the potential gains of the solution has to be re-evaluated. Note that in principle, one does not need to re-evaluate the complete solution as the travel times of the arcs traversed before the exchanged ones, do not change. Neither are the subtours not involved in the arc swap, affected by this operation.

In addition to the 2-opt improvement, extra improvement heuristics are performed taking into account explicitly the time-dependent nature of the problem. First, all the different tours that make up a complete VRP solution, are checked whether it would be advantageous in terms of travel time to break up the tour in two parts by adding the depot. This procedure is repeated until no more improvements can be realized or if the maximum number of trucks is exceeded. Secondly, all starting times of the different tours that make up a complete VRP solution, are shifted in time to evaluate the effect of the start time on the total travel time. In case of improvement, the starting time of the associated tour is updated. The rationale behind this optimization is that in a dynamic reality, a truck can decide to leave earlier or later to avoid periods of (anticipated) high congestion.

5. Results

The sets from Augerat [4] are used in this paper. These problem sets contain between 32 and 80 customers in addition to the depot. The duration of a time zone for this experiment is set equal to 10 minutes. Note that the choice of the time settings is purely arbitrarily, i.e. in the extreme case, time zones of 1 minute can be considered. All capacities of the trucks are set to 100 (following Augerat [4]). All coordinates are multiplied with a constant factor of 10 to ensure that multiple time zones are covered over longer distances. Of course, this operation does not effect in any way the comparison presented here between the static and the dynamic VRP. Again, the starting time is allowed to be different (comparable to the real-life decisions of leaving earlier or later due to congestion) in the dynamic case: trucks are allowed to start their routes anywhere between 6AM and 11AM.

In the tabu search, the length of the tabu list is set equal to 30 recent moves (which is defined as the 2 cities being swapped in the 2-opt and the shift in the starting zones). After each accepted move, the oldest move in the list is replaced by the newly accepted move. Intensification is achieved by successively restarting the *TS* heuristic with the obtained starting times from the previous run. This is continued until no improvement is achieved for 2 successive runs. Diversification is realized by letting the neighborhood space increase depending upon the objective value obtained in the previous iteration (i.e. the time zones can be shifted by either 1, 3 or 6 zones depending upon the solution quality from the previous iterations). Each iteration involves a large number of swap operations and almost exhaustive neighborhood scans. An iteration ends when no better solution through the multiple neighborhood scans is found. As such, a significantly large amount of solutions is evaluated within each iteration. Depending upon the specific area of the solution space and the problem set used, on average between 8000 and 50,000 solutions are evaluated per iteration. As such, we are capable of escaping from local optima.

We verified our implementation by comparing the results obtained from the tabu search implementation with complete enumeration for some very small problems (10 customers). The difference in the heuristic solutions and the exact solution found by enumeration was observed to be acceptable for all cases (i.e. less than 5%).

The results for different set of experiments are presented here. We describe the results as described in the experimental design. First the time-independent approach, the three time zone approach and the queueing approach are compared with each other. Note that all solutions obtained are recalculated using the speed data of the second day, which allows for these comparisons. Secondly, we look at the starting time and its influence on the solution quality. Finally, we analyze and discuss the calculation times issue for the different cases. The detailed results data are available from the authors upon request. All results are obtained on a Dell Latitude D810 with an Intel Pentium M with 2 Ghz 797 MGhz and 1 GB of RAM.

5.1. Different road types

Referring to Table 2, we generated solutions for the different compositions of the roads: one road, two road types and three road types. A high level overview of the results is reported in Table 3. The detailed results are presented in Tables 6–8. In these tables, TID refers to time-independent, TD_3 refers to time-dependent with three time zones and TD_{QM} refers to time-dependent with travel times resulting from the queueing model and with 144 time zones. Table 2 is read as follows: if we compare the time-independent case with the time dependent with three time zones (TID $\leftrightarrow$ TD_3), the gains in travel time by a more advanced modeling of the travel times equals 2.87% in the one road case, 8.35% in the two roads case and 27.42% in the three road case. These latter results are similar to the ones reported in Ichoua [23], confirming the validity of the results reported (see Table 9).

From the last row in Table 2, it is clear that in general the time-dependent cases TD_3 and TD_{QM} perform much better than the time-independent one. Moreover, TD_3 seems to be less effective than TD_{QM}: in the three road case for TD_3 27.42% of the travel time can be reduced compared to TID;

Table 3
High level overview of the results

	TID $\leftrightarrow$ TD_3			TID $\leftrightarrow$ TD_{QM}			TD_3 $\leftrightarrow$ TD_{QM}		
	Average (%)	Minimum (%)	Maximum (%)	Average (%)	Minimum (%)	Maximum (%)	Average (%)	Minimum (%)	Maximum (%)
One road	2.87	0.008	4.31	16.12	12.65	21.03	13.64	10.60	18.32
Two roads	8.35	0.92	17.39	20.62	16.01	27.23	13.33	7.51	18.48
Three roads	27.42	6.67	66.26	40.12	25.95	72.00	17.17	8.67	30.86
Overall average	12.88	–	–	25.62	–	–	14.72	–	–

Table 4
High level overview of TD_3^* and TD_3^{**} versus TD_{QM} and TID

	$TID \leftrightarrow TD_3^*$			$TD_3^* \leftrightarrow TD_{QM}$			$TD_3^{**} \leftrightarrow TD_{QM}$		
	Average (%)	Minimum (%)	Maximum (%)	Average (%)	Minimum (%)	Maximum (%)	Average (%)	Minimum (%)	Maximum (%)
One road	14.41	9.90	18.98	1.99	0.13	6.09	0.19	−1.62	4.82
Two roads	13.83	6.73	22.86	7.79	1.95	14.47	6.20	−0.84	13.65
Three roads	30.94	12.13	68.07	12.89	4.35	29.09	11.95	1.65	25.57
Overall average	19.73	–	–	7.56	–	–	6.11	–	–

Table 5
Calculation times versus solution quality

	32k5			46k7			80k10		
	Solution (minutes)	Time (seconds)	ΔTD_{QM} (minutes)	Solution (minutes)	Time (seconds)	ΔTD_{QM} (minutes)	Solution (minutes)	Time (seconds)	ΔTD_{QM} (minutes)
TID	2282.97	12	440.40	2699.80	30	554.83	4819.54	250	1110.35
TD_3	2060.85	20	218.28	2562.55	102	417.57	4218.83	686	509.65
TD_3^*	1967.55	20	124.98	2418.62	102	273.65	3984.93	686	275.74
TD_3^{**}	1910.54	20	67.97	2369.52	102	224.54	3932.68	686	223.49
TD_{QM}	1842.57	43	–	2144.97	107	–	3709.19	1818	–

Table 6
Detailed results for one road types

Nr.	Set	TID	TD_3	TD_{QM}	$TID \leftrightarrow TD_3$ (%)	$TID \leftrightarrow TD_{QM}$ (%)	$TD_3 \leftrightarrow TD_{QM}$ (%)
1	32k5	1411.18	1410.09	1193.97	0.08	15.39	15.33
2	33k5	1189.08	1167.82	989.43	1.79	16.79	15.28
3	33k6	1373.38	1322.37	1111.06	3.71	19.10	15.98
4	34k5	1389.43	1352.71	1181.26	2.64	14.98	12.67
5	36k5	1433.14	1392.45	1211.51	2.84	15.46	12.99
6	37k5	1238.82	1215.32	1025.76	1.90	17.20	15.60
7	37k6	1735.43	1692.34	1490.41	2.48	14.12	11.93
8	38k5	1312.77	1274.31	1100.35	2.93	16.18	13.65
9	39k5	1464.20	1426.26	1250.83	2.59	14.57	12.30
10	39k6	1482.88	1444.04	1261.61	2.62	14.92	12.63
11	44k6	1697.99	1643.40	1422.55	3.22	16.22	13.44
12	45k6	1720.12	1651.90	1436.12	3.97	16.51	13.06
13	45k7	2052.24	1984.26	1773.84	3.31	13.57	10.60
14	46k7	1669.38	1631.63	1374.04	2.26	17.69	15.79
15	48k7	1890.73	1855.84	1617.37	1.85	14.46	12.85
16	53k7	1840.64	1761.21	1531.53	4.31	16.79	13.04
17	54k7	2076.19	2009.33	1780.88	3.22	14.22	11.37
18	55k9	1994.91	1922.18	1599.08	3.65	19.84	16.81
19	60k9	2449.03	2366.50	2055.38	3.37	16.07	13.15
20	61k9	1960.07	1895.02	1547.93	3.32	21.03	18.32
21	62k8	2325.20	2266.38	1994.06	2.53	14.24	12.02
22	63k9	2848.09	2786.09	2487.79	2.18	12.65	10.71
23	63k10	2434.04	2355.90	1997.00	3.21	17.96	15.23
24	64k9	2528.85	2466.40	2164.59	2.47	14.40	12.24
25	65k9	2192.42	2105.65	1767.25	3.96	19.39	16.07
26	69k9	2139.91	2068.54	1768.19	3.34	17.37	14.52
27	80k10	3205.16	3081.42	2749.67	3.86	14.21	10.77

Table 7
Detailed results for two road types

Nr.	Set	TID	TD_3	TD_{QM}	TID ↔ TD_3 (%)	TID ↔ TD_{QM} (%)	TD_3 ↔ TD_{QM} (%)
1	32k5	2282.97	2060.85	1842.57	9.73	19.29	10.59
2	33k5	1848.42	1831.32	1540.06	0.92	16.68	15.90
3	33k6	2100.63	2058.22	1726.65	2.02	17.80	16.11
4	34k5	2177.39	2059.32	1776.59	5.42	18.41	13.73
5	36k5	2345.93	2206.44	1960.01	5.95	16.45	11.17
6	37k5	2030.56	1803.34	1667.93	11.19	17.86	7.51
7	37k6	2757.13	2512.84	2256.48	8.86	18.16	10.20
8	38k5	2174.01	2103.11	1714.43	3.26	21.14	18.48
9	39k5	2314.12	2232.88	1939.42	3.51	16.19	13.14
10	39k6	2327.23	2188.20	1900.04	5.97	18.36	13.17
11	44k6	2736.38	2603.97	2217.35	4.84	18.97	14.85
12	45k6	2705.91	2454.86	2100.97	9.28	22.36	14.42
13	45k7	3114.25	2770.37	2488.81	11.04	20.08	10.16
14	46k7	2699.80	2562.55	2144.97	5.08	20.55	16.30
15	48k7	2930.75	2777.68	2352.18	5.22	19.74	15.32
16	53k7	3010.86	2691.99	2323.08	10.59	22.84	13.70
17	54k7	3360.44	2776.01	2541.90	17.39	24.36	8.43
18	55k9	2992.32	2804.27	2458.29	6.28	17.85	12.34
19	60k9	3593.76	3243.38	2775.26	9.75	22.78	14.43
20	61k9	3098.28	2797.49	2280.79	9.71	26.39	18.47
21	62k8	3748.52	3269.32	2727.63	12.78	27.23	16.57
22	63k9	4423.91	3684.72	3336.60	16.71	24.58	9.45
23	63k10	3726.63	3303.96	2859.98	11.34	23.26	13.44
24	64k9	4038.39	3566.54	3009.72	11.68	25.47	15.61
25	65k9	3435.84	3146.13	2715.89	8.43	20.95	13.68
26	69k9	3220.36	3028.65	2704.79	5.95	16.01	10.69
27	80k10	4819.54	4218.83	3709.19	12.46	23.04	12.08

for the TD_{QM} even 40.12% can be reduced. These numbers are relatively large as in the TID case for 3 road types, no information with regards to fast or slow roads is available. The two other time-dependent approaches that take more of this information into account, perform significantly better. The results show that the routes improve significantly when one explicitly takes into account the dynamic character of the problem. Using the dynamic congestion information results in routes that are (considerably) shorter in terms of travel time. Moreover, significant extra gains in travel time can be obtained when utilizing the queueing approach.

5.2. Starting time effect

In the TD_{QM} case, starting times are optimized (comparable to the real-life decisions of leaving earlier or later due to congestion). Trucks are in this case allowed to start in any time bucket of 10 minutes between 6AM and 11AM. In the TD_3 case and the results reported so far, we did not allow for a starting time optimization. The rationale was that we used Ichoua [23] as a basic reference for this case, and in that paper no starting times were optimized. In order to get a more fair comparison, we also optimized the starting times in the TD_3 case. As such, the results presented here are an extension of the type of approach as introduced by Ichoua [23].

In the three time zones, the only possible starting times available are the beginning times of each of these three zones (this is the lowest detail in this approach). In our case, this means that the set of possible starting times for TD_3 is $\mathbb{T} = \{6{:}30\text{AM}, 9{:}00\text{AM}, 4{:}30\text{PM}\}$. We refer to TD_3^* when describing the results based on the starting times of the above set. Alternatively, one could apply the same 10 minute granularity as in TD_{QM} to the TD_3 case. This would then mean that the starting times can now be any 10 minute time interval between GAM and 11AM. This last case is referred to as TD_3^{**}. A high level overview of the results is reported in Table 4. The detailed results are presented in Table 7.

Table 8
Detailed results for three road types

Nr.	Set	TID	TD_3	TD_{QM}	TID ↔ TD_3 (%)	TID ↔ TD_{QM} (%)	TD_3 ↔ TD_{QM} (%)
1	32k5	7363.05	2484.38	2061.59	66.26	72.00	17.02
2	33k5	2401.21	2241.04	1656.68	6.67	31.01	26.08
3	33k6	2762.73	2177.86	1891.77	21.17	31.53	13.14
4	34k5	3101.71	2380.36	2009.03	23.26	35.23	15.60
5	36k5	3549.95	3121.44	2158.02	12.07	39.21	30.86
6	37k5	2939.74	2062.19	1666.83	29.85	43.30	19.17
7	37k6	4796.64	3125.47	2852.03	34.84	40.54	8.75
8	38k5	3154.90	2045.50	1868.14	35.16	40.79	8.67
9	39k5	3204.12	2645.69	2314.85	17.43	27.75	12.50
10	39k6	3245.37	2445.50	2158.71	24.65	33.48	11.73
11	44k6	3857.20	3321.77	2474.52	13.88	35.85	25.51
12	45k6	3863.20	2771.21	2225.58	28.27	42.39	19.69
13	45k7	4353.77	3389.57	2871.22	22.15	34.05	15.29
14	46k7	3927.58	2976.49	2359.30	24.22	39.93	20.74
15	48k7	4148.00	3161.48	2609.32	23.78	37.09	17.47
16	53k7	4234.11	3370.64	2515.82	20.39	40.58	25.36
17	54k7	5032.51	2926.40	2531.97	41.85	49.69	13.48
18	55k9	3996.90	3451.74	2959.89	13.64	25.95	14.25
19	60k9	4773.19	3474.47	3067.95	27.21	35.73	11.70
20	61k9	4425.99	3276.39	2503.44	25.97	43.44	23.59
21	62k8	6246.74	3303.62	2809.53	47.11	55.02	14.96
22	63k9	6091.65	3935.50	3460.89	35.40	43.19	12.06
23	63k10	5264.98	3805.39	3097.85	27.72	41.16	18.59
24	64k9	6694.29	4013.52	3195.65	40.05	52.26	20.38
25	65k9	4895.06	3685.20	2988.91	24.72	38.94	18.89
26	69k9	4217.87	3292.95	2898.20	21.93	31.29	11.99
27	80k10	6659.51	4611.97	3864.31	30.75	41.97	16.21

5.3. Calculation times

Calculation times are important as in a real-life planning environment, one does not have the luxury of spending too much time for obtaining a solution. On the other hand, waiting longer and as such, significantly improve the solution might be beneficial too. Therefore, a trade-off between calculation time and solution quality is necessary.

In this section, we analyze this trade-off in detail for a small dataset(32k5), a medium-sized dataset (46k7) and a large dataset (80k10) for the two roads case. The basic insights are illustrated for these few datasets, but the results shown hold in general for all datasets and all road cases. Table 5 shows the basic results with regards to the calculation times for each of the different settings. The column ΔTD_{QM} gives the difference between the TD_{QM} solution and the respective other solutions (TID, TD_3, TD_3^* and TD_3^{**}). For example, focusing on the 32k5 set, we see that the difference between the TID solution and TD_{QM} solution is 440.40 minutes. Additionally, the required computation time is given for each set and each setting. Again for the 32k5 set, the TID solution ran 12 seconds, while the TD_{QM} solution took 43 seconds.

In general, the solution quality improved significantly when using the TD_{QM} approach compared to either TID,TD_3, TD_3^* and TD_3^{**} approaches. Comparing the TD_{QM} setting with the best alternative TD_3^{**}, we still see significant gains of 67.97 minutes of travel times for the 32k5 set, gains of 224.54 minutes of travel times for the 46k7 set and gains of 223.49 minutes of travel times for the 80k10 set. The gains comparing TD_{QM} with the other settings are even greater. On the other hand, the improved solution quality comes at the price of longer calculation times. The TD_{QM} takes two to three times more calculation times than the TD_3 cases and three to seven times more compared to the TID case. The increases in calculation times are however still very reasonable compared to the solution quality improvements.

Fig. 6 shows for the 32k5 set the evolution of the solution quality compared to the calculation time needed. The TID solution is obtained after 12 seconds; the solutions for the TD_3, TD_3^* and TD_3^{**} are obtained after 20 seconds. The TD_{QM}

Table 9
Detailed results for TD_3^* and TD_3^{**}

Nr.	Set	One road			Two roads			Three roads		
		$TID \leftrightarrow TD_3^*$ (%)	$TD_3^* \leftrightarrow TD_{QM}$ (%)	$TD_3^{**} \leftrightarrow TD_{QM}$ (%)	$TID_3^* \leftrightarrow TD_3^*$ (%)	$TD_3^* \leftrightarrow TD_{QM}$ (%)	$TD_3^{**} \leftrightarrow TD_{QM}$ (%)	$TID \leftrightarrow TD_3$ (%)	$TD_3^* \leftrightarrow TD_{QM}$ (%)	$TD_3^{**} \leftrightarrow TD_{QM}$ (%)
1	32k5	9.90	6.09	4.82	13.82	6.35	3.69	68.07%	12.30	12.71
2	33k5	15.10	1.98	0.00	6.73	10.67	10.26	12.13	21.49	25.57
3	33k6	17.77	1.62	−0.36	9.73	8.95	7.82	27.09	6.08	4.58
4	34k5	13.66	1.53	−0.16	11.26	8.05	6.80	25.00	13.64	12.50
5	36k5	12.96	2.87	1.11	7.44	9.74	6.54	14.28	29.09	19.12
6	37k5	13.78	3.96	2.22	15.84	2.40	−0.84	32.51	15.98	15.08
7	37k6	12.81	1.50	−0.27	13.10	5.82	4.47	37.09	5.49	3.50
8	38k5	14.64	1.81	0.00	7.80	14.47	13.65	38.09	4.35	1.65
9	39k5	12.81	2.02	0.25	7.36	9.53	8.80	21.36	8.12	7.70
10	39k6	13.73	1.38	−0.54	11.69	7.54	5.33	28.41	7.09	4.63
11	44k6	13.94	2.65	1.05	8.11	11.82	11.94	15.61	23.98	23.49
12	45k6	15.13	1.62	−0.21	16.19	7.36	6.04	32.22	15.00	15.80
13	45k7	13.45	0.13	−1.62	16.16	4.68	2.49	25.17	11.87	11.79
14	46k7	15.52	2.57	0.57	10.41	11.31	10.47	27.01	17.70	18.51
15	48k7	12.76	1.95	0.00	10.27	10.56	8.62	25.74	15.29	15.25
16	53k7	16.13	0.79	−0.87	15.66	8.52	6.87	23.20	22.64	22.05
17	54k7	13.20	1.18	−0.58	22.86	1.95	0.42	44.78	8.89	7.00
18	55k9	18.27	1.93	0.00	13.86	4.63	2.14	19.97	7.47	6.55
19	60k9	14.57	1.76	−0.11	16.64	7.37	6.19	31.36	6.36	4.04
20	61k9	18.98	2.52	0.49	19.46	8.60	7.32	32.44	16.28	17.50
21	62k8	12.80	1.65	−0.08	18.03	11.23	10.61	49.77	10.46	9.05
22	63k9	11.36	1.45	−0.30	21.32	4.14	1.71	37.97	8.42	5.69
23	63k10	16.19	2.10	0.17	17.46	7.02	4.78	31.56	14.03	13.51
24	64k9	13.26	1.32	−0.32	16.58	10.66	9.28	43.28	15.83	16.97
25	65k9	16.88	3.02	1.17	15.56	6.39	5.04	29.22	13.74	13.74
26	69k9	16.63	0.89	−0.97	12.76	3.73	0.97	27.65	5.02	3.52
27	80k10	12.91	1.49	−0.24	17.32	6.92	6.03	34.51	11.39	11.15

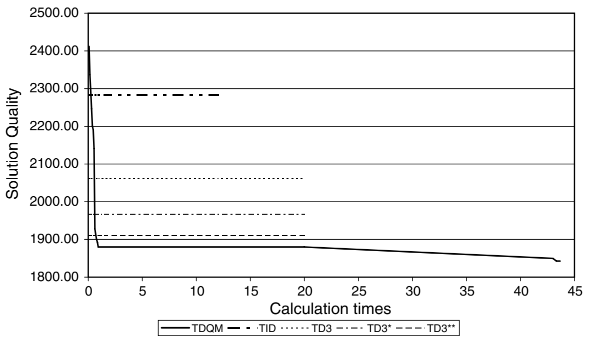

Fig. 6. The evolution of the solution quality for TD_{QM}.

approach stops after 43 seconds. Clearly, the TD_{QM} approach generates better solutions than the other approaches. Comparing solutions versus calculation time, we observe that with a calculation time of 12 seconds (similar to TID), TD_{QM} gives a better solution value (i.e. TID = 2282.97 versus $TD_{QM}^{12\text{ seconds}} = 1879.95$). A similar conclusion holds for the TD_3, TD_3^* and TD_3^{**}: after 20 seconds, the TD_{QM} results in a better solution. Concluding, this means that for the same calculation times as the TID, TD_3, TD_3^* and TD_3^{**} approaches, the TD_{QM} gives superior solutions.

Concluding, investing more time in the calculation of the routes pays off. The TD_3, TD_3^* and TD_3^{**} approaches are close to each other in terms of performance; the TD_{QM} approach reduces travel times even further at the expense of an acceptable increase in calculation times. Finally, one should be aware that our software is an academic implementation and not necessarily programmed as efficiently as it could be to be used in practice.

6. Managerial insights

The routing problem discussed in this paper is motivated by the fact that in many circumstances traffic conditions can not be ignored in order to carry out a realistic optimization. Results show that the total travel times can be improved significantly when explicitly taking into account congestion during the optimization. Explicitly using the information of congestion results in routes that are (considerably) shorter in terms of travel time. The assumption that everything in transportation goes according to a schedule is unrealistic, which will result in a planning gap (i.e. the difference between the planned route and the actual route). The main consequence of taking into account a congestion function in the route planning is that the planning gap will decrease as the planned route is then much more realistic compared to the classic VRP route optimized in terms of distances. Computational results suggest that in systems where the speeds differences are considerable, it is crucially important to explicitly consider the variability due to traffic congestion in the model. The impact of dynamic components will be even more important when relating the approach to urban contexts (see also Taniguchi et al. [34] for more insights on the impact of city logistics on routing decisions).

6.1. Taking into account time-dependency

The capability of taking into account dynamic travel times is extremely valuable, not only because speeds profiles do affect the objective function of the optimization, but also, as demonstrated above, the best solutions for the static problem applied in a dynamic context, are in general suboptimal. Compared to the actual realization in real-life, the planning gap (difference between plan and actual) can be reduced substantially when taking into account a manifestation of the traffic on the roads. Consequently, due to the resulting reduced planning gap, less time and effort needs to be invested in replanning (in real-time) during the day. This paper aims at obtaining a routing solution that performs well in the face of the extra complications due to congestion, which eventually leads to a better solution in practice. These more realistic solutions have the potential to reduce real operating costs for a broad range of industries daily facing routing problems.

In the current paper the fleet size is assumed to be unlimited. However, the benefits of a more realistic estimate of the true travel times will become even more important when the fleet size is limited (i.e. only a number of trucks are available). Indeed, many more opportunities and improvement possibilities exist for the limited fleet size problem when taking into account time. For example, a manager might take advantage of a travel time reduction of 25% by letting the same truck do two tours on a day (e.g. morning and afternoon).

6.2. Advantages of the queueing approach

Traditionally, traffic flows are modeled empirically, using e.g. origin-destination matrices [12]. The objective of this approach is mainly explorative and explanatory and gives an empirical justification of the well-known speed-flow and speed-density diagrams. An alternative approach is to collect data on traffic flows are and fit into curves [12]. The power of both two approaches lies in the description and explanation of traffic flows. Compared to these descriptive models, this paper presents a more operational approach using queueing theory. The speed-flow-density diagrams are constructed analytically. The advantage is that the underlying processes themselves are modeled rather than the output of these processes. This of course then easily allows for any kind of optimization, sensitivity analysis, what-if analysis as is shown in this paper.

Analytical queueing models based on traffic counts also allow to model the variance of the travel times. For a given time period p, the variance of the travel time can be determined as follows (using the formulae from Section 3):

$$\mathrm{Var}(T_{ij}^{p}) = \mathrm{Var}\left(\frac{d_{ij}}{v_{ij}^{p}}\right) = d_{ij}^{2} * \mathrm{Var}(k_j * W^{p})$$
$$= d_{ij}^{2} * k_j^{2} * \mathrm{Var}(W^{p}).$$

The distance from i to j, d_{ij} and the jam density of the road k_j are assumed to be known. The variance of the total time in the system W, can be obtained using the two moment approximations from Whitt. As there is no exact form for the variance of the waiting time, one needs to rely on approximations to obtain the variance of the waiting time [36]. These approximations have already proven their value and usability in production management (see e.g. Lambrecht et al. [27]; Whitt [36]; and others). When including the variance of the travel time, the potential applications are vast: it gives a manager a powerful tool to incorporate and take into account congestion uncertainty in his optimization. The heuristics proposed here are currently being extended to include this variance (see [29] for more information). As such, this is an interesting and unique perspective left for future research.

7. Conclusions and future research

In this paper, a dynamic vehicle routing problem with time-dependent travel times due to traffic congestion was presented. Recently, the problem considered has received increasing attention due to its relevance to real-life problems. The approach developed here introduces the traffic congestion component in the standard VRP models. The traffic congestion component was modeled using a queueing approach to traffic flows. By making use of this analytical approach to traffic flows, the necessary data to model congestion is easily obtained and very limited, which opens the door for real-life applications.

Results showed that the total travel times can be improved significantly when explicitly taking into account congestion during the optimization. Explicitly making use of the time-dependent congestion results in routes that are (considerably) shorter in terms of travel time. Additionally, we found that a higher number of time zones, improves the solution quality. Moreover, this effect is even magnified if one considers different road types. We found that adapting the starting times for a solution has a significant effect on the obtained solution quality. Finally, we found that the extra calculation time needed is for large datasets significant but is certainly worthwhile as the solution quality also greatly improved.

Another future research path involves applying the developed approach to the dynamic VRP with the addition of time windows. Time windows are convenient to customers due to the relative certainty of delivery and the ease to schedule the deliveries to suit the work pattern. On the other hand, time windows are inconvenient to delivery companies as they limit flexibility. Time windows can then be used as an indicator of customer service level, which can be violated if this violation allows a sufficiently large reduction in transportation costs. Combined with the queueing approach developed in this paper, the objective function as a whole can be evaluated based not only on the mean travel time but also on the variability in the travel time on the route. Hence, given a certain time window at the customer, one can also obtain approximate probabilities of meeting the time window, which can then be used to set the customer rates accordingly.

Finally, although the current paper is situated more in the part of the deterministic class, it is a first incursion into the almost unexplored stochastic domain as the analytical queueing framework described here is expected to be an interesting approach towards stochastic VRPs.

Acknowledgements

The authors would like to thank the anonymous referee and the editor for their valuable input during the process of (re-)writing the paper. Also many thanks to Christophe Lecluyse (University of Antwerp) for helping with implementing the heuristics.

References

[1] E.H.L. Aarts, J.K. Lenstra, Local Search in Combinatorial Optimization, Wiley, 1997.
[2] R. Akçelik, Travel time functions for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function, Australian Road Research 21 (3) (1991) 49–59.
[3] R. Akçelik, Relating flow, density, speed and travel time models for uninterrupted and interrupted traffic, Traffic Eengineering and Control 37 (9) (1996) 511–516.
[4] P. Augerat, J.M. Belenguer, E. Benavent, A. Corber, D. Naddef, Separating capacity constraints in the CVRP using tabu search, European Journal of Operations Research 106 (1998) 546–557.
[5] D. Bertsimas, G. Van Ryzin, A stochastic and dynamic vehicle routing problem in the euclidian plane, Operations Research 39 (1991) 601–615.
[6] D. Bertsimas, G. Van Ryzin, Stochastic and dynamic vehicle routing problems in the euclidean plane with multiple capacitated vehicles, Operations Research 41 (1993) 60–76.
[7] D. Bertsimas, G. Van Ryzin, Stochastic and dynamic vehicle routing with general demand and interarrival time distributions, Applied Probability 25 (1993) 947–978.
[8] D.J. Bertsimas, D. Simchi-Levi, A new generation of vehicle routing research: Robust algorithms, addressing uncertainty, Operations Research 44 (2) (1996) 286–304, March–April.
[9] Transportation Research Board, Traffic flow theory: A state-of-the-art report. Technical report, Transporation Research Board, 1996.
[10] G.G. Brown, C.J. Ellis, G. Lenn, W. Graves, D. Ronen, Real time, wide area dispatch of mobile tank trucks, Interfaces 17 (1987) 107–120.
[11] J.J. Coyle, E.J. Bardi, J.J. Langley Jr, The Management of Business Logistics, West Publishing, 1996.
[12] C.F. Daganzo, Fundamentals of Transportation and Traffic Operations, Elsevier Science, 1997.
[13] K.B. Davidson, The theoretical basis of a flow-travel time relationship for use in transportation planning, Australian road research 8 (1) (1978) 32–35.
[14] B. Fleischmann, M. Gietz, S. Gnutzmann, Time-varying travel times in vehicle routing, Transportation Science 38 (2) (2004) 160–173.
[15] M. Gendreau, A. Hertz, G. Laporte, A tabu search heuristic for the vehicle routing problem, Management Science 40 (10) (1994) 1276–1290.
[16] M. Gendreau, G. Laporte, R. Séguin, A tabu search heuristic for the vehicle routing problem with stochastic demands and customers, Operations Research 44 (3) (1996).
[17] F. Glover, Tabu search, part, ORSA Journal of Computing (1989) 190–206.
[18] F. Glover, Tabu search, part II, ORSA Journal of Computing (1990) 4–32.
[19] B.D. Greenshields, A study of traffic capacity, in: Highway Research Board Proceedings 14 (1935) 448–477.
[20] D. Heidemann, A queueing theory approach to speed-flow-density relationships, in: Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France, Transporation and traffic theory, 1996.
[21] A. Hertz, G. Laporte, M. Mittaz, A tabu search heuristic for the capacitated arc routing problem, Operations Research 48 (1) (2000) 129–135.
[22] A.V. Hill, W.C. Benton, Modeling intra-city time-dependent travel speeds for vehicle scheduling problems, European Journal of Operational Research 43 (4) (1992) 343–351.
[23] S. Ichoua, M. Gendreau, J.-Y. Potvin, Vehicle dispatching with time-dependent travel times, European Journal of Operational Research 144 (2003) 379–396.
[24] R. Jain, J. MacGregor Smith, Modeling vehicular traffic flow using M/G/C/C state dependent queueing models, Transportation Science 31 (1997) 324–336.
[25] J.F.C. Kingman, The single server queue in heavy traffic, in: Proceedings of the Cambridge Philosophical Society, vol. 57, 1964, p. 902–904.
[26] W. Kraemer and M. Lagenbach-Belz. Approximate formulae for the delay in the queueing system GI/GI/1, in: Congressbook of the Eight International Teletraffic Congress, pages 235–1/8, Melbourne, 1976.
[27] M.R. Lambrecht, P.L. Ivens, N.J. Vandaele, ACLIPS: A capacity and lead time integrated procedure for scheduling, Management Science 44 (11) (1998) 1548–1561, Part 1 of 2, November.
[28] G. Laporte, The vehicle routing problem: An overview of exact and approximate algorithms, European Journal of Operational Research 59 (3) (1992) 345–358.
[29] C. Lecluyse, T. van Woensel, H. Peremans, Stochastic vehicle routing with random time dependent travel times, submitted for publication.
[30] C. Malandraki, M.S. Daskin, Time dependent vehicle routing problems: Formulations, properties and heuristic algorithms, Transportation Science 26 (3) (1992) 185–200.
[31] C. Malandraki, R.B. Dial, A restricted dynamic programming heuristic algorithm for the time dependent traveling salesman problem, European Journal of Operational Research 90 (1996) 45–55.
[32] M. Pirlot, General local search methods, European Journal of Operational Research 92 (1996) 493–511.
[33] Y. Shen, J.-Y. Potvin, A computer assistant for vehicle dispatching with learning capabilities, Annals of Operations Research 61 (1995) 189–211.
[34] E. Taniguchi, R.G. Thompson, T. Yamada, R. Van Duin, City Logistics: Network Modelling and Intelligent Transport Systems, Pergamon, 2001.
[35] N. Vandaele, T. Van Woensel, A. Verbruggen, A queueing based traffic flow model, Transportation Research D 5 (2) (2000) 121–135, February.
[36] W. Whitt, The queueing network analyzer, The Bell System Technical Journal 62 (9) (1983) 2779–2815.
[37] W. Whitt, Approximations for the GI/G/m queue, Production and Operations Management 2 (2) (1993) 114–161.
[38] T. Van Woensel, Modelling Uninterrupted Traffic Flows, a Queueing Approach. Ph.D. Dissertation, University of Antwerp, Belgium, 2003.
[39] T. Van Woensel, R. Creten, N. Vandaele, Managing the environmental externalities of traffic logistics: The issue of emissions, POMS journal, Special issue on Environmental Management and Operations 10 (2) (2001).

[40] T. Van Woensel, N. Vandaele, Empirical validation of a queueing approach to uninterrupted traffic flows, 4OR, A Quarterly Journal of Operations Research 4 (1) (2006) 59–72.

[41] T. Van Woensel, B. Wuyts, N. Vandaele, Validating state-dependent queueing models for uninterrupted traffic flows using simulation, 4OR, A Quarterly Journal of Operations Research 4 (2) (2006) 159–174.