Optimization of route choice, speeds and stops in time-varying networks for fuel-efficient truck journeys

David Watling[1],
Institute for Transport Studies, University of Leeds, Leeds, UK

Richard Connors,
Institute for Transport Studies, University of Leeds, Leeds, UK

David Milne,
Institute for Transport Studies, University of Leeds, Leeds, UK

Haibo Chen
Institute for Transport Studies, University of Leeds, Leeds, UK

A method is presented for the real-time optimal control of the journey of a truck, travelling between a pair of pick-up/drop-off locations in a time-varying traffic network, in order to reduce fuel consumption. The method, when applied during the journey, encapsulates the choice of route, choice of speeds on the links, and choice of stop locations/durations; when applied pre-trip, it additionally incorporates choice of departure time. The problem is formulated by using a modified form of space-time extended network, in such a way that a shortest path in this network corresponds to an optimal choice of not only route, stops and (when relevant) departure time, but also of speeds. A series of simple illustrative examples are presented to illustrate the formulation. Finally, the method is applied to a realistic-size case study.

Keywords: optimal routing, space-time network, fuel minimisation

1. Introduction

The impact of road traffic on energy efficiency is a major global policy concern. In 2016, carbon dioxide emissions from the transport sector accounted for 32% of all carbon dioxide emissions (BEIS, 2017b). In spite of major improvements in vehicle technology, between 2014 and 2015 transport sector emissions actually *increased* by 2%, primarily due to increases from passenger cars and Heavy Goods Vehicles (HGVs), due to increased vehicle kilometres travelled and higher use of fuel (BEIS, 2017a). Freight transport by all modes accounts for around a third of greenhouse gas emissions, with HGVs estimated to be responsible for up to a half of these (FTA LCRS, 2010)[2]. Therefore, the freight sector has a major role to play in alleviating such problems, particularly in the case of long trips by HGVs. The present paper is particularly concerned with the opportunities for improving fuel consumption (and thereby CO_2 emissions[3]) for large trucks

---

[1]Corresponding author: E: D.P.Watling@its.leeds.ac.uk T: +44 (0) 113 343 6612.

[2]We cite these figures as illustrative of the problem, but appreciate that there are varying views on the actual contribution of the road freight sector to the overall emissions (see, for example, CfIT, 2007).

[3]The strong link between fuel consumption and CO2 emissions has been noted by many authors, e.g. Mickūnaitis et al (2007).

on long journeys, which must operate under a variety of driving conditions that cannot be perfectly predicted in advance.

Fuel consumption is thus a social concern, but is it also something that is important for individual truck companies? Taveres et al (2009) cite evidence that fuel consumption plays a dominant role in the costs of municipal solid waste collection and transportation. A survey of forty shippers across Europe found that for long-distance haulage truck operations, there were no great differences between the factors affecting cost, with fuel cost the second largest at an average of 20.4% (Musso, 2001). While it thus plays a major role, these 'internal' fuel costs borne by the truck companies represent only part of the full societal cost of this consumption (the remainder referred to as external costs or externalities). Bektas et al (2019) suggest that the 'primary role of OR in green transportation as providing a set of decision-making tools, methods and approaches that ultimately yield win-win decisions' in which both the internal costs and the full costs are minimised. Alternatively, governments may choose to internalise the externalities through increased taxation on items such as fuel; in this vein, Austin (2018) estimated the impacts of internalising the external costs of US overland road freight transport, to be around 20 per cent of shipping costs. Thus even if truck companies do not currently consider reduced fuel use an imperative, they can be influenced to do so in the future by increased taxation, to address the policy concerns outlined in the opening paragraph.

The particular research described in the present paper considers how we might use real-time predictive control and information—derived from both the trucks themselves and the surrounding environment—in order to improve the fuel efficiency of the journeys they make. The motivation for such research is well explained by Eglese et al (2006), who report on a survey of over fifty users of routing/scheduling systems in the UK, in which an overwhelming majority reported problems with the credibility of the forecast travel times for the planned journeys, particularly in the light of peak traffic congestion or maintenance works. There is thus the potential to make improvements in planning journeys by time of day, by exploiting the increasing amount of real-time data available on time-dependent traffic conditions. A corollary is that this also applies when planning with an objective in mind other than travel time, if this objective is also sensitive to time-varying traffic conditions/speeds, as fuel consumption is. This potential for improvement is supported by the field experiments on light duty vehicles reported by Frey et al (2008), who found differences in fuel-use and emissions varying by in the order of 20% across different routes (for the same origin-destination movement), and when comparing congested with less congested conditions.

In using time-of-day varying, predictive information to reduce fuel consumption, it is possible to identify three broad, interacting scales of analysis at which we may make adaptations:

1. At the most detailed level there is the possibility to employ intelligent real-time control of an individual vehicle's powertrain system. At this level we are concerned with the impact of acceleration and deceleration rates, idling time, and stop-and-go frequencies, as well as the traffic patterns of the surrounding vehicles.

2. At the highest level, decisions are made on whether the effects of any unexpected circumstances (e.g. incident delays and any response to them such as re-routing) or control strategies (e.g. such as those that involve reducing speed) are sufficient to re-plan downstream deliveries/pickups on the truck's tour, such as re-scheduling delivery/pickup times or changing the order of deliveries/pickups on the tour.

3. At an intermediate level, we again consider the consequences of unexpected delays, which may be caused by incidents or unusual congestion levels, with the aim to potentially re-plan the way in which the current leg of the truck's tour will be made. By considering downstream impacts (up to the next delivery/pickup), decisions can be made as to whether to change the choice of route (i.e. the actual roads used) for the

current leg, to alter driver break-points on this leg, or to alter the 'target speed' on the current and subsequent links so as to avoid negative consequences downstream.

In the literature, there exist many previous analyses of type 1., concerned with methods for improving the fuel efficiency of powertrains (e.g. Damiani et al, 2014; Schepmann & Vahidi, 2016; Gao et al, 2015). Likewise, a growing body of literature exists on analyses of type 2., under the broad heading of "green vehicle routing", which aim to optimize the environmental efficiency of delivery tour sequences and schedules (e.g. Bektas & Laporte, 2011; Franceschetti et al, 2013; Demir et al, 2014; Molina et al, 2014; Dabia et al, 2016; Qian & Eglese, 2016; Herold, 2017; Eshtehadi et al, 2018) and the impacts of empty running of vehicles (Ozen & Tuydes-Yaman 2013). In the present paper we focus on an area that has relatively less attention, specifically level 3. problems as defined above. In this respect, we wish to build on studies exploring the impacts on individual trips of decisions such as route choice (Foytik & Robinson, 2015) and response to incidents (Ng et al, 2016).

The structure of the paper is as follows. We begin, in section 2, by identifying characteristic types of "downstream interaction" that motivate our study, and use this to set out our key assumptions. In section 3, we present the mathematical formulation of the problem, and specifically the equivalence to a form of space-time representation of the network. A series of simple examples are described in section 4 in order to illustrate the principles of the method. In section 5, the results of applying the method to a realistic-scale case study are presented. Finally, section 6 present conclusions and future research directions.

2. Motivation and assumptions

2.1 Motivation: Characterising downstream interactions

As explained in the introduction, the purpose of the optimization method we will develop is to consider the consequences of unexpected delays on a leg of delivery tour, in order to avoid negative downstream impacts on fuel consumption, with the possibility to change route[4], driver break-points, and/or the 'target speed' on current and subsequent links. A special case of this is where information is received prior to the start of the journey leg, providing an additional dimension of departure time adaptation. In order to motivate our subsequent work and to distinguish the types of intervention we have in mind, we define motivating examples of downstream interaction to which the truck may respond, and group the examples according to the type of response involved:

Type 1: Speed adaptation
1. Altering the current speed may affect whether, later in the journey, the truck passes a major urban area at a peak-time or not. If it passes during the peak it may be delayed by the regular peak-time congestion for that link, and so consume more fuel. In this case, a critical decision is whether slowing down earlier, which may reduce fuel consumption locally, may increase total fuel consumption for the whole journey by meaning that the truck will then experience peak-hour congestion downstream.

2. A major incident occurs on a downstream link of the currently-planned route of the truck, and there are no obvious diversion routes before the incident. However, given predictive information on how quickly the incident is likely to clear, a decision can be made as to whether to continue with the current planned speeds (which may be locally optimal), or

---

[4]One terminological clarification is that in the present paper, the word 'tour' is used to refer to the sequence of pickups/deliveries made by a truck (without any reference to which specific roads are used), and the word 'route' refers to a specific contiguous set of road links that allow one leg of a tour to be made. The focus of the present paper is on the latter, in spite of the term 'vehicle routing' being commonly used to refer to a choice of tour in the first sense.

whether by slowing down now the truck will arrive at the affected link after the incident has cleared, which in aggregate may reduce overall fuel consumption for the overall journey links.

Type 2: Re-planning of stops or departure time

3. A major incident occurs on a downstream link of the currently-planned route of the truck, there are no obvious diversion routes before the incident. Given predictive information on how quickly the incident is likely to clear, slowing down (as in 2)) is not expected to help. A decision then to be made is whether it would be better for overall fuel consumption for the truck to pull over at a service station to wait for some time before continuing the journey (perhaps bringing forward in time a planned regulatory break, if possible). Alternatively, if the information is received prior to the commencement of the journey leg, the same considerations may lead to the leg departure time being delayed.

4. The truck is currently in unexpected congestion and the congestion is predicted to continue ahead for some time. A decision to made then is whether it would be better for overall fuel consumption for the truck to pull over at a service station to wait for some time before continuing the journey.

Type 3: Re-routing

5. A major incident occurs on a downstream link of the currently-planned route of the truck, and there is a feasible diversion route ahead that avoids the incident location (but would normally be a sub-optimal choice); a decision to be made then is whether the truck should take the diversion in order to reduce fuel consumption.

In the cases described where it makes sense, for a fuel consumption objective, to make the adaptive responses proposed, there is a trade-off to consider with potential downstream impacts on other objectives or constraints. Thus there is a need to consider whether slowing down, stopping, deferring the departure time or diverting may have negative consequences such as:

- incurring additional costs (e.g. from hourly payments to drivers) or other consequences (e.g. on perishable goods being transported) due to extending the journey time;

- incurring a major annoyance or cost to a customer at the leg destination for the late arrival of some goods;

- arriving at a downstream location during a time-of-day-dependent or day-dependent truck restriction, which the truck would otherwise have avoided (as it would have passed before/after the restriction was operational, or would not have used the restricted route);

- extending the journey or the timing of stops such that constraints on drivers' working hours are violated.

The scenarios described in this section are characterised by the fact that decisions which may seem reasonable locally may have significant negative consequences later in the journey. Thus the problem needs to be considered on a whole journey (leg) level, forecasting the consequences for the whole leg, and cannot be solved by a purely local-level decision.

2.2 Assumptions
In our analysis it is assumed that several elements of input data are pre-determined:

- We are interested in the leg of the truck's journey from a given pickup/delivery location to the next pickup/delivery location on its tour (this latter being "the destination", as far as our analysis is concerned), this tour having been planned by some external process. If the optimization is applied prior to the commencement of the journey leg, then "the origin" is the starting pickup/delivery location; otherwise, if applied *en route*, then "the

origin" is the next downstream node on the current route, where a node may represent a stop location or a diversion opportunity.

- It is supposed that the vehicle type, weight, height, width is known, as well as its load for the particular leg of the tour under consideration. Also any other truck-specific characteristics required for the fuel consumption calculation are defined.

- A feasible road network is known, comprising links and nodes defining several feasible routes from the leg origin to leg destination, for the truck and load under consideration. The feasibility of the route takes account of any weight, height and width restrictions, and any permanent (i.e. not time-of-day or day-dependent) restrictions on truck usage.

- Any time-of-day or day-dependent restrictions that could potentially affect the truck/load under consideration are separately defined for each link of the road network.

- Fixed, physical aspects of the road network are defined, such as link lengths, gradients and any other characteristics relevant for the fuel consumption calculation.

- Feasible, potential *stop locations* are defined as nodes in the network, likely to be service stations. Feasibility needs to consider any restrictions on truck parking and size/weight.

- Any legislation or local agreements on driver restrictions are assumed to be known, specifically concerned with the duration between and length of any breaks required.

- A penalty function is defined for the leg destination, which gives an arrival time window (in which there is no penalty) and any penalties for increasingly early or late arrival outside that window. The penalty function represents the annoyance to the customer, but might have a monetary interpretation if any compensation is appropriate. Its functional form would be expected to follow the classic, piecewise linear form as used extensively in departure time choice modelling (extension to Small, 1982, to include an 'indifference band').

In addition to this fixed input data, real-time information is supposed to be available, providing a current prediction of the minimum travel time for the truck to traverse each link of the network, disaggregated by (future) entry time to the link. This may be thought of as providing a predicted 'speed timetable' for traversing each link (following the terminology of Eglese et al, 2006, though in our context they are treated as upper bounds rather than defined speeds). These predicted minimum travel times (maximum link traversal speeds) are exogenously determined, taking account of both historical information and any incidents through a journey time prediction algorithm. They take account of any fixed speed limits for a truck on the link and any constraints imposed by congestion; obviously they need to relate to predicted *truck* travel times/speeds, rather than the generally lower travel times (higher speeds) that would be predicted for general traffic.

Given this information, the objective of the optimization method is to choose a combination of the following *decision variables*:

- route;

- recommended 'target' speed for each link on the route;

- stop locations on the route;

- stop durations on the route;

- and, if applied pre-trip, the departure time;

- in order to trade-off *objectives* such as

- minimise fuel consumption for the leg;

- minimise the travel time on the leg;

- minimise the penalty for early/late arrival at the leg destination.

3. Formulation and Notation

3.1 Preliminary considerations

Section 2 concluded with a verbal definition of the form of optimization problem addressed by the present paper. In Section 3, a formal definition of this problem is given, cast in a particular (discretized) form strongly influenced by our choice of solution approach, and which in turn influences our choice of notation. Therefore it makes sense to explain the reasoning behind our choice of solution approach before introducing the notation and formal definition.

The optimization problem specified in section 2 clearly has elements of a time-dependent minimum cost path problem, assuming cost=fuel, for which a range of solution approaches have been proposed (e.g. Ziliaskopoulos & Mahmassani, 1993; Chabini, 1998; Wen et al, 2014; see also the recent review of point-to-point problems in Gendreau et al, 2015). This initially suggested to us the possibility to decompose the problem into two stages, of (i) a link-based problem of determining fuel-optimal link speeds and stop waiting times (as real decision-variables); and (ii) finding a time-dependent minimum fuel path (with a given link model relating speed to fuel). Congestion affects the fastest speed that the truck may traverse a link, and so these can be added as constraints in problem (i), and the constraints modified as real-time information emerges. There are, however, major difficulties with such an approach, since in fact problem (i) does not lend itself to a link-based formulation, due to the physical speed interdependencies between links that must not be violated; these constraints turn out to be the major difficulty. This issue arises fundamentally from a problem remarked on by Qian & Eglese (2016) in considering a node sequence $c_0 \rightarrow c_i \rightarrow c_j$ 'it may be better to travel faster/slower than the [locally] optimal speed with lower fuel efficiency from c_0 to c_i, so as to arrive at c_i earlier/later to avoid the congestion from c_i to c_j.' The analogous problem in our case is that for a vehicle travelling on link A and then link B, for a given entry time to link A, the chosen speed on link A will determine the link B entry time, and thus which time-dependent link B speed constraints are active. A corollary to this is that the optimal link speeds are *path-dependent*, even for a given destination; that is to say, in order to optimize speed on any link we need to know what sequence of links will be followed downstream from that link, as this will affect which downstream constraints are active to ensure a physically consistent route in space and time.

An alternative approach we considered was then one of explicit route enumeration; since large trucks were the focus of our research and these typically are restricted from using many roads due to height/weight/width constraints, then the number of possible routes available is typically relatively small compared with general traffic. Then for each route in turn we can solve the problem of optimal link speeds and stop durations. For this speed optimization sub-problem, we may apply the exact, quadratic time algorithm of Hvattum et al (2013), for each route in turn. While this approach provides an exact algorithm for the overall problem, it was considered unattractive since the requirement to enumerate routes makes it less scalable to problems with larger networks—which may arise when considering trans-national journeys or considering light duty vehicles with fewer height/weight/width constraints. This problem would be exacerbated further if expanded to include departure time choice prior to the leg commencement, for which extensive enumeration of (departure time and route) options would then be needed.

A third possible approach we considered was to adapt and build on several of the heuristic methods proposed for determining fuel-optimal *tours* with speed optimization in a time-dependent setting. These methods include the tabu search method proposed by Jabali et (2012) and the neighbourhood move heuristics proposed by Qian & Eglese (2016). We did not pursue this approach since we wished to identify an exact algorithm, which it seemed should be feasible

with the generally much greater computational simplicity of shortest path type problems as opposed to tour/routing problems.

Ultimately the approach adopted was based on a formulation that has had widespread application in the transportation domain, which is founded on the space-time framework proposed by Hagerstrand (1970), and is variously referred to a space-time prism, diachronic network, expanded space-time network or (the terminology we shall adopt) a Space-Time Extended Network (STEN). Such a framework has been adopted to solve diverse applications, such as trajectory tracking (Kuijpers & Othman, 2009), activity scheduling (Chen et al, 2013; Liao, 2016), choice of scheduled public transport services (Nuzzolo et al, 2001; Gentile & Noekel, 2016), routing of unmanned aerial vehicles (Zhang et al, 2015), and the modelling of multi-modal systems (Tong et al, 2019). There are several attractions of this framework for the present problem, as we shall show in the following sections, namely: (i) its ability to naturally encode the most complex constraints, namely those ensuring the physical feasibility of space-time routes for different potential choices of link speeds; (ii) its ability to provide an optimization framework to *simultaneously* optimize all decision dimensions (choice of speeds, route, stops, and where relevant departure time); and (iii) the fact that it leads to an exact algorithm. The major uncertainty we had when pursuing this approach was its computational tractability for the kinds of analysis we envisaged. In particular, we wished to be able to solve such problems to a high temporal resolution (in our later results we represent five days at a resolution of one minute), to avoid significant discretisation errors and to utilise real-time information at a high temporal resolution. However, this has serious implications for the size of our modified STEN, and so exploring its tractability became an important element of our computational tests, as we report later.

3.2 Creating the space-time network

As noted in section 2, we are considering a single truck making a single pick-up/delivery from a given origin to a given destination, for which the load, height and width are known and are fixed for the journey. We thus suppose that we can define in advance a network of permissible links for such a truck, i.e. links for which the load, height and width are permitted. This permissible network is defined as a graph $G=\{V,E\}$, where V is the set of vertices/nodes and E is the set of directed links joining the nodes. In the cases considered in the present paper the graph is a static entity, in that the permissible network neither varies by time of day nor is dependent on any past decisions[5]. It is, however, specific to the particular truck and load being routed on the single leg of the tour under consideration, such that the set E only contains permissible links for this truck, given any weight/height/width restrictions and any permanent bans on trucks.

An important element of our mathematical formulation will be the way in which 'time' is represented. We suppose that all times—whether we are referring to departure-time, clock-time, or travel-time—are measured in some discrete units that define the maximal resolution of the problem. These units may be minutes (and it may aid thinking to imagine that they are, initially at least), but from a formulaic point of view it is immaterial whether each one of these units is, say, 1 minute or 5 minutes. The choice of time unit is of course influential on the fidelity of the results, and a balance must be made between fidelity and computational efficiency (the latter favouring a low resolution). It is supposed that clock-time is measured in such discrete units past some reference time, denoted as time 0, and it is supposed that the set of such discrete clock-times covering the complete study time horizon is denoted by $T \subset \mathbb{N}$.

5In the more complex case, time-of-day bans on trucks may mean that the permissible link set E at any time may vary depending on the departure time of the truck and on any decisions on route/speed/stops made upstream from the current location. In cases, at any given current location along the route, reached at discrete clock time k, a forecast may be made of the permissible link set $E^{(k)}$ based on the latest forecasted downstream travel times, and then the described method applied to $E^{(k)}$ in place of E.

The routing problem that we consider applies to all points in the truck's journey, so includes both *en route* diversions and the pre-trip decisions at the origin of the trip leg. For the pre-trip decision, we consider two variants of our problem below: one in which the departure time from the trip origin is given, and one in which it is part of the optimization. Note that this is somewhat different from placing a potential stop location at the origin, since stops are always about delaying a journey, whereas departure time choice also allows the journey to be brought forward in time. The discussion below will first relate to the case where the departure time from the trip origin $d \in T$ is fixed, before later in the section describing the extension to departure time choice.

Suppose, then, we are at some given discrete clock-time $k \in T$, then there is various *internal* information we possess about the individual truck being routed, in particular: its current location; the route it is currently planning to follow; and historical information of its journey so far, e.g. travel time, fuel consumed, number and duration of stops. Let $v^{(k)} \in V$ denote the next downstream network node, given the route being followed by the truck at time k. This will be the next location at which any route/speed/stop decision is made, and will be the "pseudo-origin" for the optimization. In addition we assume *external* information is available, in the form of real-time predictive information on the shortest time needed to traverse each link of the network ("real-time predicted minimum travel times"), given predicted traffic congestion, predicted incident durations, and constrained by any relevant speed limits. These predicted minimum travel times will vary not only by clock-time k, but also by the (future) time at which a vehicle enters a downstream link. Thus, we use $\tau^{(k)}_{ijt,\min}$ to denote the minimum time to traverse the link from node $i \in V$ to $j \in V$ for a truck of this type entering the link at all future clock-times $t = k, k+1, ...$, given real-time predictive information available at time k. Note that importantly, these predicted minimum travel times are measured in the same discrete units as clock-time.

As noted in the discussion in section 3.1, a key issue is to ensure that the routes on which trucks are sent, and the speeds that they are set on the links of their journey, must together form a feasible space-time trajectory, whatever the objective of the optimization. As discussed there, our chosen approach is to achieve this by formulating the problem in the frame of a Space-Time Extended Network (STEN). A 'route' through a STEN moves in both space and time. In our problem, we have the flexibility to choose the route, speeds on links and waiting times at stop locations. However, to begin let us assume that choice of speed on the links is not a possible option, and then we could adopt a standard STEN representation, which can be created as follows from the original base network $G=\{V,E\}$, based on the predicted minimum travel times at any clock-time $k \in T$:

Step 1: Create a new vertex set $\hat{V}$ of by creating a space-time vertex (i, t) for each original node $i \in V$ and each clock-time $t \in T$.

Step 2: For each link $(i, j) \in E$, create a new link set $\hat{E}^{(k)}$ by adding a space-time travelling link from each pair of space-time vertices (i, t) to $\left(j, t + \tau^{(k)}_{ijt,\text{data}}\right)$, where $\tau^{(k)}_{ijt,\text{data}}$ is the travel time from i to j when leaving i at time t, using real-time predictive information at time k.

Step 3: Add to set $\hat{E}^{(k)}$ a space-time waiting link for each pair of space-time vertices (i, t) and $(i, t + 1)$.

We now adapt this basic STEN approach to fit with our problem specification:

- The speeds (and so travel times) are not fixed, but it is possible to travel at a lower speed than the maximum possible. We constrain choice of speed to a range ($[v_L, v_U]$ e.g. based on legal limits).

- For each link this defines a default minimum and maximum travel time: $\left[\tau^{(k)}_{ijt,v_U}, \tau^{(k)}_{ijt,v_L}\right]$.

- The available link travel times are constrained by real-time predictive information, whenever the maximum speed (or indeed minimum speed) are unattainable. The range

of   possible   travel   times   for   link $i-j$ at   time k is   then $\left[\max\left(\tau_{ijt,v_U}^{(k)}, \tau_{ijt,data}^{(k)}\right), \max\left(\tau_{ijt,v_L}^{(k)}, \tau_{ijt,data}^{(k)}\right)\right]$ which we denote $\left[\tau_{ijt,\min}^{(k)}, \tau_{ijt,\max}^{(k)}\right]$.

- It is assumed that we are interested in routing the truck from an origin/source node $i^* \in V$ to a destination/sink node $j^* \in V$, with the truck leaving at a given departure time $d \in T$, and in which we are indifferent to the arrival time at the destination (i.e. there is no penalty that describes the cost/inconvenience of arrival at the destination relative to the arrival time).

We refer to the graph produced by the procedure below as an ESTEN (for Extended STEN), created from base network $G=\{V,E\}$, the (origin, destination, departure-time) triple (i^*, j^*, d), and based on the predicted minimum and maximum travel times at any clock time $k \in T$ which are contained in the vectors $\boldsymbol{\tau}_{\min}^{(k)}$ and $\boldsymbol{\tau}_{\max}^{(k)}$, and will denote the resulting space-time network as the result of the mapping implied by the procedure:

$$\left(\tilde{V}, \tilde{E}^{(k)}\right) = \text{ESTEN}\left(V, E, i^*, j^*, d, \boldsymbol{\tau}_{\min}^{(k)}, \boldsymbol{\tau}_{\max}^{(k)}\right). \tag{1}$$

The procedure is as follows:

Step 1: Create a new vertex set $\tilde{V}$ by:

a. creating a space-time vertex (i,t) for each non-source node $i \in V \setminus \{i^*\}$ and each clock-time $t \in T$;

b. creating a single space-time vertex (i^*, d) for the given source and departure time;

c. creating a single, common, dummy ultimate sink vertex $j_{\text{SINK}} \notin V$ with a free (undefined) time, i.e. a space-time vertex $(j_{\text{SINK}}, \text{free})$.

Step 2: Create a new link set $\tilde{E}^{(k)}$ by:

a. creating a link for each non-source node $i \in V \setminus \{i^*\}$ and each link $(i,j) \in E$, by adding a space-time travelling link from each space-time vertex $(i,t) \in (V \setminus \{i^*\}, T)$ to $(j, t+\theta)$, for all integers θ such that

$$\tau_{ijt,\min}^k \leq \theta \leq \tau_{ijt,\max}^k$$

where $(\tau_{ijt,\min}^k, \tau_{ijt,\max}^k)$ are bounds on the travel time from i to j when leaving i at time t, using a combination of speed limits and real-time predictive information at time k.

b. creating a link from the space-time source node (i^*, d) and each link $(i^*, j) \in E$, by adding a space-time travelling link from (i^*, d) to $(j, d+\theta)$, for all integers θ such that

$$\tau_{i^*jd,\min}^k \leq \theta \leq \tau_{i^*jd,\max}^k \; ;$$

c. creating a link from each space-time sink node (j^*, t) $(t \in T)$ to the dummy ultimate sink node $(j_{\text{SINK}}, \text{free})$.

Step 3: Add to set $\tilde{E}^{(k)}$ a space-time waiting link for each non-sink space-time vertex, joining (i,t) and $(i, t+1)$ for $(i,t) \in (V \setminus \{i^*\}, T)$.

As noted earlier, the case we are considering above is for a given departure time from the origin. The method may also be extended so that it effectively embeds choice of departure time. This is achieved with two simple modifications:

- by a time-expansion of the source-node, which means we simply remove the special treatment of source nodes in steps 1(a)/(b) and 2(a)/(b);

- creating a single super-origin with links connected to the time-expanded source nodes.

- when associating costs with the space-time links (see section 3.3), associating an equivalent delay penalty cost with each of the space-time links from the time-expanded sink node (j^*, t) $(t \in T)$ to the dummy ultimate sink node $(j_{\text{SINK}}, \text{free})$, if the penalty is representing some kind of negative factor at the destination (lateness affecting customer satisfaction or associated with actual monetary costs of extended driver hours or opening a receiving location outside normal hours); and

- equivalently, for any penalties associated with the origin, associating equivalent penalties with the space-time links from the super-origin to the time-expanded source nodes.

3.3 Associating costs with the space-time network

The ESTEN created in section 3.2 provides a network representation of the feasible spatio-temporal trajectories, but as yet does not suggest any means of preferring one trajectory over another. The next step, then, is to specify how the 'costs' of the space-time links should be defined. If the objective is simply to arrive at the destination as quickly as possible, then it makes sense to define the cost of each space-time link equal to the (predicted) time taken to traverse the physical link, if entering at the time implied by the space-time link. Analysing such a case with the ESTEN proposed in section 3.2, while perfectly feasible, is somewhat unnecessarily complex, since it will always be optimal[6] to traverse the link at the minimum travel time for the given entry time, and so a regular STEN representation would suffice. (However, this provides a useful verification check of any ESTEN computer code, that it does indeed produce such an expected solution).

More generally, where the 'costs' may be specified to represent any objective, then we suppose that at the current clock-time $k \in T$, we can associate a cost c_{ijt}^k with each space-time link connecting space-time vertex $(i, t) \in \tilde{V}$ to $\left(j, t + \theta_{ijt}^k\right) \in \tilde{V}$ $\forall$ integer $\theta_{ijt}^k \in \left[\tau_{ijt,\text{min}}^k, \tau_{ijt,\text{max}}^k\right]$. For example, for an objective to minimise total fuel consumption along the route, then for such a link, since θ_{ijt}^k is the travel time, then $l_{ij} \big/ \theta_{ijt}^k$ and is the speed to move from the link entry to exit if l_{ij}

denotes the link length, and we may use a model such as the Comprehensive Modal Emission Model (CMEM) (Barth and Boriboonsomsin, 2008; Barth *et al* 2004; Scora and Barth 2006) to convert speed into fuel consumed, following a similar logic to that used by Franceschetti *et al* (2013).

In extended version of the ESTEN to represent departure time choice (as described at the end of section 3.2), then we may also consider an objective involving lateness at the destination, by:

- associating an equivalent delay penalty cost with each of the space-time links from the time-expanded sink node (j^*, t) $(t \in T)$ to the dummy ultimate sink node $(j_{\text{SINK}}, \text{free})$, to reflect some kind of negative factor at the destination (lateness affecting customer satisfaction or associated with actual monetary costs of extended driver hours or opening a receiving location outside normal hours); and

- equivalently, for any penalties associated with the origin, associating equivalent penalties with the space-time links from the super-origin to the time-expanded source nodes.

For all other space-time links, the costs are set to zero when representing a lateness objective.

6If the time-dependent link travel times are not consistent with FIFO at the link level, then it could be the case that slowing down on a link could reduce overall route travel time at the given departure time. However this could anyway be captured by waiting on a space-time waiting link, which (according to the description given in section 3.2) a standard STEN incorporates. In the cases we shall consider in the present paper, the travel times are FIFO-consistent and so this case does not arise.

3.4 Real-time, multi-objective optimization of truck journeys

Here we bring together the material in sections 3.1–3.3, in order to define the overall optimization problem to be solved. As noted earlier, we are routing/re-routing a truck dynamically as it traverses the network on a single leg: there is a single origin drop-off/pick-up and a single destination pick-up/drop-off, and a pre-fixed, planned departure time from the origin (though this can be deferred at a wait link to the network, in the light of real-time information). The decision variables are then (at any time/location in the journey):

- choice of route for the remainder of the journey from the current location to the destination;

- choice of speed along the links of this remaining route; and

- choice of wait time locations/durations for the remaining route.

At any given clock-time $k \in T$ the truck is either at the true origin of the trip leg, or it is following a route on its journey and there is a next downstream node $v(k)$ on that route. Without loss of generality, we assume that it is actually at node $v(k)$ at time k . We then apply the procedure for constructing the ESTEN (described in section 3.2) with the origin and departure time equal to the actual trip origin and actual departure time pair (i^*, d), or if the truck has started its journey then with the origin and departure time pair $(v(k), k)$ so as to route for the remainder of the trip. In practice, it may not be computationally efficient to *actually* reconstruct the ESTEN at each such intermediate point, but rather construct it once at the outset, and then have a variable indicator of the origin and departure time. However, *procedurally* it is easy to imagine that it is reconstructed, as then we can simply refer to 'the origin and departure time', and need not make any distinction below as to whether the truck is at the true origin of the trip leg or at an intermediate point. Therefore, with this explanation, we will simply describe the process as if the origin and departure time pair were (i^*, d), and leave it to the imagination that this can be replaced with $(v(k), k)$ at intermediate locations *en route*.

We thus are able, at any clock-time $k \in T$, to construct the ESTEN (see §3.2):

$$\left(\tilde{V}, \tilde{E}^{(k)}\right) = \text{ESTEN}\left(V, E, i^*, j^*, d, \boldsymbol{\tau}_{\min}^{(k)}, \boldsymbol{\tau}_{\max}^{(k)}\right). \tag{2}$$

from the base network $G=\{V,E\}$, the (origin, destination, departure-time) triple (i^*, j^*, d), and the predicted minimum and maximum travel times at clock-time k, namely $\boldsymbol{\tau}_{\min}^{(k)}$ and $\boldsymbol{\tau}_{\max}^{(k)}$. The resulting network $\left(\tilde{V}, \tilde{E}^{(k)}\right)$ is a static graph representation of our underlying time-dependent problem. The advantage of such a formulation is that any simple path through $\left(\tilde{V}, \tilde{E}^{(k)}\right)$ is a *feasible* triple of the decision variables: a feasible space-time route, feasible link travel times for links on this route, feasible link wait times for stop locations/durations on this route. In order to denote feasibility according to the time-extended network we shall write:

$$(\mathbf{i}, \mathbf{t}) \in \left(\tilde{V}, \tilde{E}^{(k)}\right) = \text{ESTEN}\left(V, E, i^*, j^*, d, \boldsymbol{\tau}_{\min}^{(k)}, \boldsymbol{\tau}_{\max}^{(k)}\right) \tag{3}$$

where $(\mathbf{i}, \mathbf{t})$ is a pair of vectors denoting a time-dependent route in $\tilde{E}^{(k)}$ utilising n links:

$$(\mathbf{i}, \mathbf{t}) \in \left\{\left((i_1, i_2, \dots, i_n), (t_1, t_2, \dots, t_n)\right) : \left((i_{r-1}, t_{r-1}), (i_r, t_r)\right) \in \tilde{E}^{(k)} \text{ for } r = 2, 3, \dots, n\right\} \tag{4}$$

and where additionally $i_n = j^*$ (the given destination), and where for routing from the trip origin i^* at departure time d then $i_1 = i^*$ and $t_1 = d$, or for routing from *en route* node $v(k)$ at time k then $i_1 = v(k)$ and $t_1 = k$.

The generic formulation of our time-dependent, multi-objective routing problem is then:

Minimise $(f_1(\mathbf{i}, \mathbf{t}), f_2(\mathbf{i}, \mathbf{t}), \dots, f_m(\mathbf{i}, \mathbf{t}))$ (5a)

    subject to

$$(\mathbf{i}, \mathbf{t}) \in \left(\tilde{V}, \tilde{E}^{(k)} \right) = \text{ESTEN} \left(V, E, i^*, j^*, d, \mathbf{\tau}_{\min}^{(k)}, \mathbf{\tau}_{\max}^{(k)} \right). \tag{5b}$$

In this formulation, we may naturally choose different definitions of the objective functions. For example, an objective to minimise travel time is written as:

$$f_1(\mathbf{i}, \mathbf{t}) = \sum_{r=2}^{n} (t_r - t_{r-1}) = t_n - t_1 \tag{6}$$

and one to minimise fuel consumption using a model relating speed to fuel consumed is:

$$f_2(\mathbf{i}, \mathbf{t}) = \sum_{r=2}^{n} g_{r-1,r} \left(\frac{l_{r-1,r}}{t_r - t_{r-1}} \right) \tag{7}$$

where $l_{r-1,r}$ is the length of the link joining node i_{r-1} to i_r , and $g_{r-1,r}(.)$ is a function that maps link speed (from node i_{r-1} to node i_r) onto the fuel consumed.

3.5 Algorithmic considerations

The space-time network created in section 3.2 provides a way of applying standard shortest path methods to a static graph representation of a time-dependent problem. Given such a representation, the Pareto optimal set may then be estimated using one of a number of existing special-purpose techniques, such as those based on solving a parametric family of single-objective, shortest-path problems (White, 1982; Henig, 1986) or special-purpose approximation schemes for multi-objective shortest path problems (Warburton, 1987). In many cases, a computationally simpler approach will likely be sufficient, such as the commonly-used weighting/scalarization method; in spite of its well-documented limitations, as Mavrotas (2009) notes, simpler methods such as this 'can provide a representative subset of the Pareto set which in most cases is adequate' for practical purposes. In this approach, the objectives are weighted into a single objective, thus requiring only a single objective shortest path problem to be solved, and the weights varied across a range to generate a subset of the full Pareto set. Indeed, since the intention that the problem we have defined will ultimately be implemented in a real-time system, the generation of a Pareto set is less useful (since there will not be the opportunity for some decision-makers to reflect on which they may prefer), and so in practice some kind of weighting method may indeed be needed.

In our later computational tests (section 5) we have indeed adopted the weighting/scalarisation approach, and solve the single objective shortest path problem by a (binary heap) implementation of the Dijkstra algorithm. For the ESTEN $\left(\tilde{V}, \tilde{E}^{(k)} \right)$, this algorithm has computational complexity $O\left(\left| \tilde{E}^{(k)} \right| \ln \left| \tilde{V} \right| \right)$, where $|A|$ denotes the cardinality of set A. We can in turn relate this complexity result to the input elements of the model. Let δ denote the number of minutes represented by one time unit in the notation of section 3.2, and let $T_{ijt,\min}^{k}$ and $T_{ijt,\max}^{k}$ denote the minimum and maximum journey time *in minutes* when leaving node $i \in V$ at time $t \in T$ to travel to node $j \in V$. The ESTEN procedure described, when these times in minutes are converted to the common time increment, will thus create $\left(T_{ijt,\max}^{k} - T_{ijt,\min}^{k} \right) \delta^{-1}$ space-time links in $\tilde{E}^{(k)}$ for each node j connected to space-time node $(i, t) \in \tilde{V}$. Letting $\alpha = \max \left(T_{ijt,\max}^{k} - T_{ijt,\min}^{k} : i, j \in V, t \in T \right)$, a worst case scenario is that for each link in E outgoing from node i, $\alpha \delta^{-1}$ space-time links will be created in $\tilde{E}^{(k)}$ for each space-time node $(i, t) \in \tilde{V}$. Let the full time horizon for the analysis be defined as the time-span needed for the earliest vehicle to depart and the latest vehicle to arrive, by whatever route and feasible travel time, and suppose this has duration β minutes. Thus $\tilde{V}$ contains $\beta \delta^{-1}$ copies of each node $i \in V$. Thus overall, $\left| \tilde{V} \right| = \delta^{-1} \beta |V|$ and $\left| \tilde{E}^{(k)} \right| \leq \delta^{-2} \alpha \beta |E|$, and so taking a worst case for $\left| \tilde{E}^{(k)} \right|$, the computational complexity of the Dijkstra algorithm for the ESTEN is $O(\delta^{-2} \alpha \beta |E| \ln(\delta^{-1} \beta |V|))$. This may be contrasted with a standard STEN, for which a *single* space-time link is created for each space-time node $(i, t) \in \tilde{V}$, and for which the Dijkstra algorithm would thus have computational complexity $O(\delta^{-1} \beta |E| \ln(\delta^{-1} \beta |V|))$. Thus, in contrast with a regular STEN, in the ESTEN both δ and α have a significant role in the computational effort. There then exists a trade-off between computational effort and both the temporal resolution (through δ) and avoiding sub-optimality (through

assuming high enough $T_{ijt,\max}^k$ values and hence high enough α, that the optimal link speed is feasible).

Since an ESTEN can thus be extremely large in comparison with the original network, especially at a high temporal resolution, it is then relevant to consider whether there are any sensible *network reduction strategies* that may be applied *a priori* to provide some computational advantage. Two strategies considered here (and reported on in section 5, with the unreduced network as a benchmark) are:

1. finding a bound on earliest time to reach each node from the origin if leaving at time d, by solving a shortest path problem with link times equal to lowest minimum travel time, and hence only creating outgoing space-time links that can feasibly be used;

2. restricting the range of link speeds permitted, which influences the travel-time range for each link, and hence the number of space-time links, for each space-time node.

4. Simple Examples

The simple examples presented below are intended to illustrate the concepts described earlier, and to highlight the contrast with conventional space-time networks. The examples are all based on the underlying road network (with associated node numbers) defined in Figure 1. The network represents the remaining permitted road segments downstream from the truck's current location, that lead to the destination of the pickup/delivery leg at node 3. Node 1 thus represents either the real origin of the pickup/delivery leg under consideration, or the next downstream node that the truck will reach on its current route. We refer to node 1 generically as "the origin" to cover both cases; in a situation in which it is relevant to distinguish the cases (e.g. in considering departure time choice from the actual origin of the trip leg) then we shall make this clear and explicit.

It is assumed that at the current clock-time, real-time predicted minimum and maximum link traversal times are available, varying by the clock-time at which the link is entered, and these are given in Table 1. The link traversal times are assumed to have been already converted to the discretized form described in section 3. Thus, for example, from Table 1 we are saying that the predicted traversal time for the link from node 1 to node 2, if departing node 1 at a clock-time of 5, is either 3, 4 or 5 time-units. Nodes 2 and 4 are also potential stop locations for breaking the (remainder of the) journey, considered in Example 3.
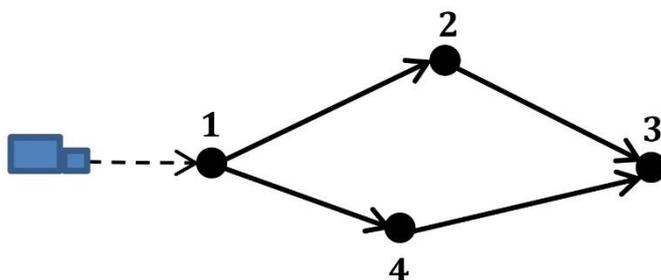

Figure 1.   Four-node, two-route downstream network

Table 1.   Min–max predicted link traversal time when entering link at given entry time

	Link entry time									
Link	1	2	3	4	5	6	7	8	9	10
1 → 2	1–2	1–2	2–3	2–3	3–5	3–5	2–4	2–3	2–2	1–2
2 → 3	1–2	1–2	1–2	2–3	2–3	4–5	3–4	2–3	1–2	1–1
1 → 4	1–2	1–2	1–3	1–3	2–3	2–3	2–4	2–3	1–3	1–2

| **4 → 3** | 1–2 | 1–2 | 2–4 | 3–5 | 2–3 | 2–3 | 2–3 | 2–3 | 1–2 | 1–2 |

Table 2.   Fuel consumed to traverse a link, as a function of link traversal time

		Link traversal time				
Link	1	2	3	4	5	6
1 → 2	4	3	2	4	5	8
2 → 3	7	6	5	8	9	11
1 → 4	5	4	5	5	6	8
4 → 3	8	7	9	11	12	14

Table 3.   Penalty for early/late arrival at the leg destination

Arrival clock-time at node 3	2	3	4	5	6	7	8	9	10	11
Schedule (Delay) Penalty	2	2	1	0	0	2	4	8	12	15

Table 2 defines the link-specific relationship between travel time and fuel consumed. Table 3 reflects any penalties for arriving outside a preferred arrival time window at the destination.

4.1 Example 1

In the first example, it is supposed there are no stop locations and the current clock-time is 2, at which time the vehicle is at node 1 (i.e. at space-time node (1,2)). The ESTEN representation of this problem is illustrated in Figure 2; note the distinctive feature relative to a conventional STEN, that there are multiple exit links from each space-time node, representing the different choices of link traversal speed (and hence travel time) that are permitted. Taking a pure objective of either travel time, fuel or schedule penalty minimisation, we obtain the weighted graphs in Figures 3–5, with the shortest weighted routes highlighted in each case.
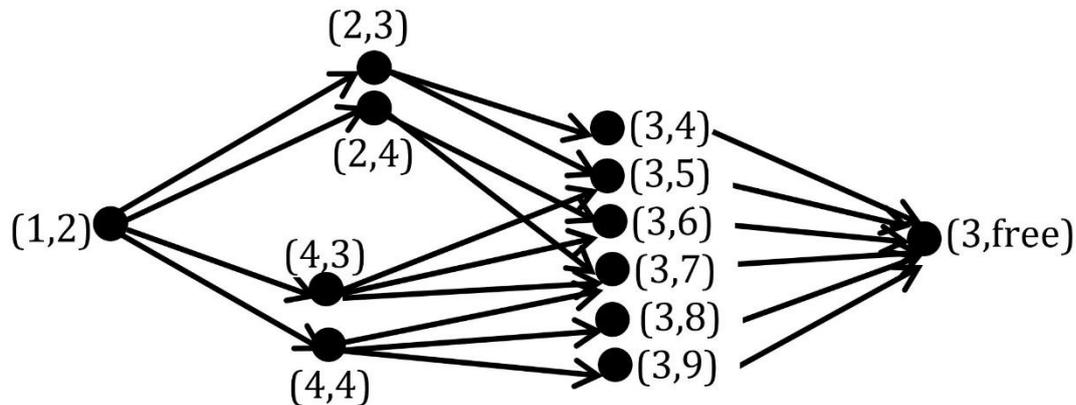

Figure 2.   ESTEN representation for four-node network (no stops)
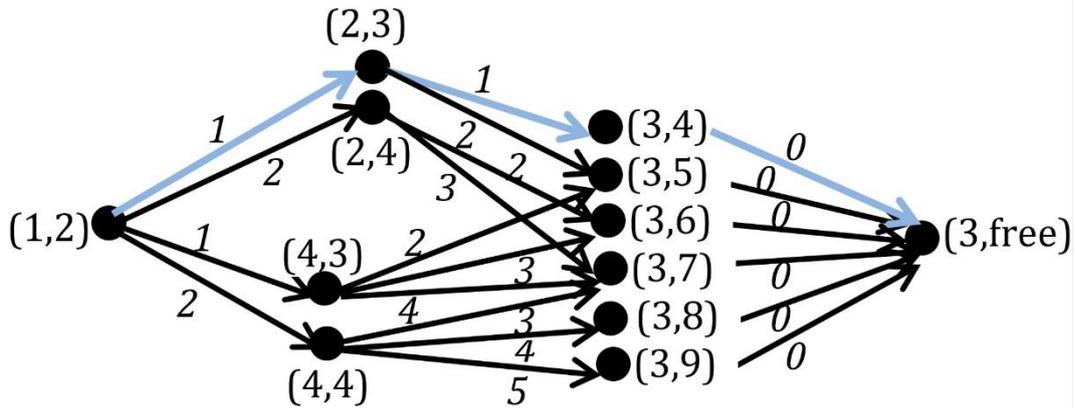
Figure 3.   Weighted ESTEN graph for travel time objective, with minimum travel time space-time route highlighted
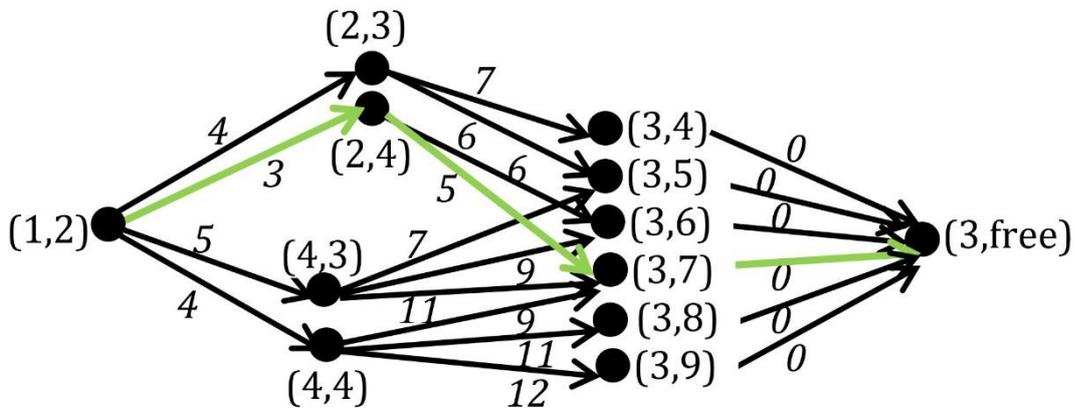

Figure 4.   Weighted ESTEN graph for fuel objective, with minimum fuel consumption space-time route highlighted
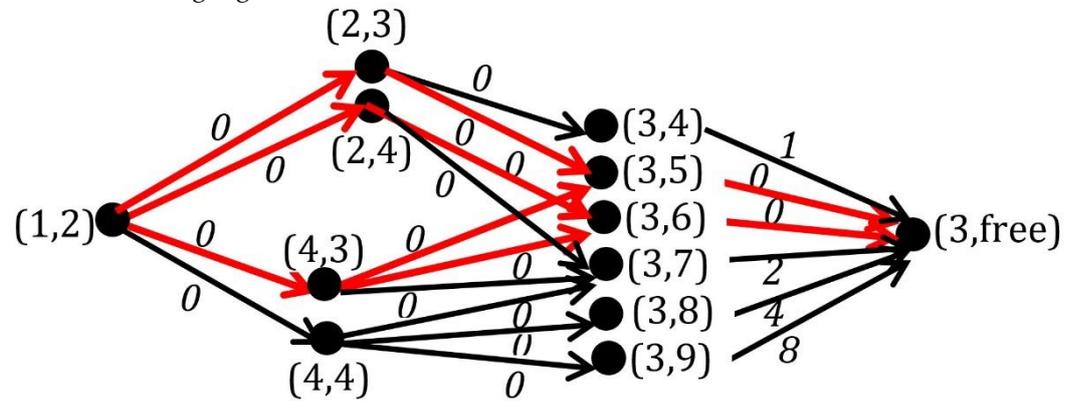

Figure 5.   Weighted ESTEN graph for schedule penalty objective, with minimum schedule penalty space-time routes highlighted

In this simple case it is possible to enumerate all (ten) feasible space-time trajectories, as given in Table 4, with the Pareto optimal solutions highlighted. In this case, all the Pareto solutions use the same physical route. The link traversal travel times (and hence speeds) can be inferred from the space-time trajectories, which clearly show the trade-off in accepting generally higher travel times for lower fuel consumption.

One other feature of this example is that the maximum travel time constraints are binding at some of the Pareto solutions. This has induced some "sub-optimality" (relative to a problem with no maximum travel time constraints), since it is simple to check that if we were to increase the maximum travel time to 3 for link 1 → 2 in time period 2, then the space-time trajectory (1,2) →

$(2,5) \rightarrow (3,8)$ has a travel time of 6, a fuel consumption of 7 and a schedule penalty of 4. This is a potentially desirable solution that was ruled out by the original, tighter maximum travel times, and illustrates the issue discussed in section 3.5 concerning the trade-off between constraining the space-time network size (for greater computation speed) and accepting the possibility of sub-optimality.

Table 4.   Enumerated space-time-feasible trajectories for four-node example (no stops), with Pareto optimal solutions highlighted

Space-time nodes	Travel time	Fuel consumed	Schedule Penalty
$(1,2) \rightarrow (2,3) \rightarrow (3,4)$	2	11	1
$(1,2) \rightarrow (2,3) \rightarrow (3,5)$	3	10	0
$(1,2) \rightarrow (2,4) \rightarrow (3,6)$	4	9	0
$(1,2) \rightarrow (2,4) \rightarrow (3,7)$	5	8	2
$(1,2) \rightarrow (4,3) \rightarrow (3,5)$	3	12	0
$(1,2) \rightarrow (4,3) \rightarrow (3,6)$	4	14	0
$(1,2) \rightarrow (4,3) \rightarrow (3,7)$	5	16	2
$(1,2) \rightarrow (4,4) \rightarrow (3,7)$	5	13	2
$(1,2) \rightarrow (4,4) \rightarrow (3,8)$	6	15	4
$(1,2) \rightarrow (4,4) \rightarrow (3,9)$	7	16	8

4.2 Example 2

In the case where node 1 represents the true origin of the trip leg, rather than an intermediate node *en route*, we may also have the additional option to optimize departure time. We assume that departure time is restricted to be between time 1 and 2, denoted by creating a space-time node $(1, 1 \leq \text{free} \leq 2)$, with the resulting ESTEN in Figure 6. By way of illustration the optimum strategy for a pure fuel consumption objective is highlighted in Figure 6.
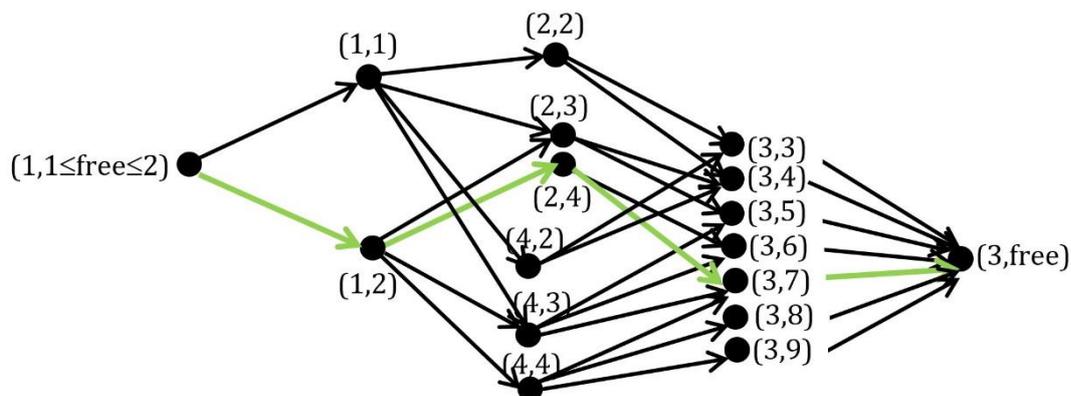

Figure 6.   ESTEN representation for four-node network with departure time choice (no stops), with optimum fuel consumption space-time route highlighted

4.3 Example 3

Returning to Example 1, we extend it to additionally consider stop locations, with maximum stop times of two units at any location. In addition, a trip feasibility constraint exists related to a driver's permitted continuous driving hours, meaning that at least one stop of at least one time unit must be taken during any trip where the time from beginning to end of the trip exceeds 4 time units. The ESTEN representation is shown in Figure 7, with the minimum fuel consumption path highlighted, where a stop is indeed included (as represented in the space-time link from $(2,4)$ to $(2,5)$).
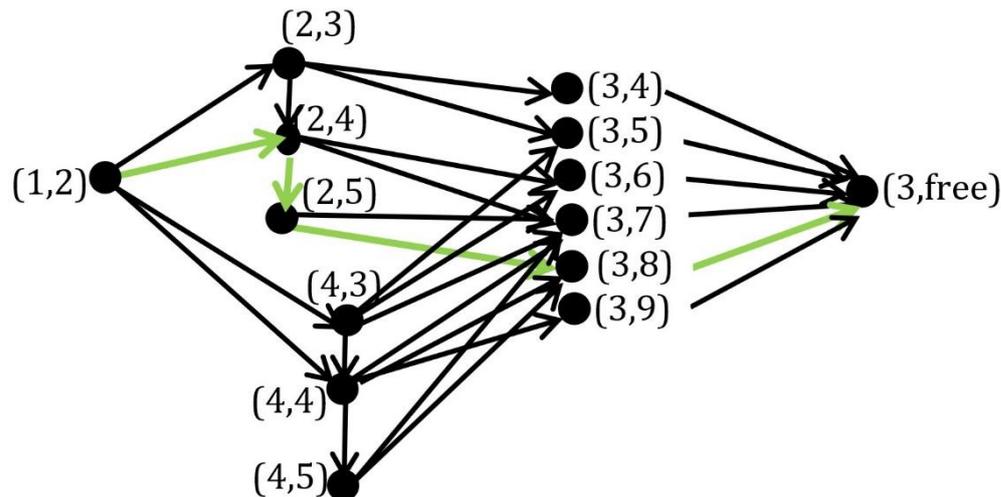
Figure 7.   ESTEN representation for four-node network (with stops), with minimum feasible fuel consumption space-time route highlighted

The Pareto optimal solutions are given in Table 5, with the final column added to highlight that the trip feasibility constraint (on required break times) has been additionally satisfied. Note that space-time trajectory $(1,2) \to (2,4) \to (3,7)$, which was Pareto optimal without the trip feasibility constraint (Table 4), now drops out. Note that the two solutions in Table 5 with stops (last two rows) are subtly different, with the first including a stop only to reduce fuel (relative to the no-stop $(1,2) \to (2,3) \to (3,4)$ solution), whereas the second requires a stop due to the trip duration, though this does in fact also reduce overall fuel consumed.

Table 5.   Pareto set for four-node example (with stops).

Space-time nodes	Travel time[7]	Fuel	Schedule Penalty	Trip Feasibility
$(1,2) \to (2,3) \to (3,4)$	2	11	1	✓
$(1,2) \to (2,3) \to (3,5)$	3	10	0	✓
$(1,2) \to (2,4) \to (3,6)$	4	9	0	✓
$(1,2) \to (2,3) \to (2,4) \to (3,6)$	3	10	0	✓
$(1,2) \to (2,4) \to (2,5) \to (3,8)$	5	8	4	✓

4.4 Example 4

In the fourth example, we consider a modification of Example 1 to represent an incident, to compare how strategies change. Firstly, we assume that a vehicle approaching node 1 and due to arrive there at time 2 receives information of an incident affecting link 1→2. An incident prediction model is used to forecast the impact and duration of the incident, with the new real-time predicted journey times given in Table 6. Secondly, we suppose that there is a potential stop location available at node 1, though (in contrast to Example 3) there is no *requirement* to make a stop.

Table 6.   Min–max predicted link traversal time when entering link at given link entry time for incident scenario

					Link entry time					
Link	1	2	3	4	5	6	7	8	9	10
1 → 2	1–2	6	6	4–5	3–5	3–5	2–4	2–3	2–2	1–2
2 → 3	1–2	1–2	1–2	2–3	2–3	4–5	3–4	2–3	1–2	1–1
1 → 4	1–2	1–2	1–3	1–3	2–3	2–3	2–4	2–3	1–3	1–2

---

7 Excluding time spent at stops, i.e. only time spent travelling.

| **4 → 3** | 1–2 | 1–2 | 2–4 | 3–5 | 2–3 | 2–3 | 2–3 | 2–3 | 1–2 | 1–2 |

The ESTEN representation is shown in Figure 8, with node 1 now expanded to include the possibility of stopping, and the network structure amended to account for the new information.
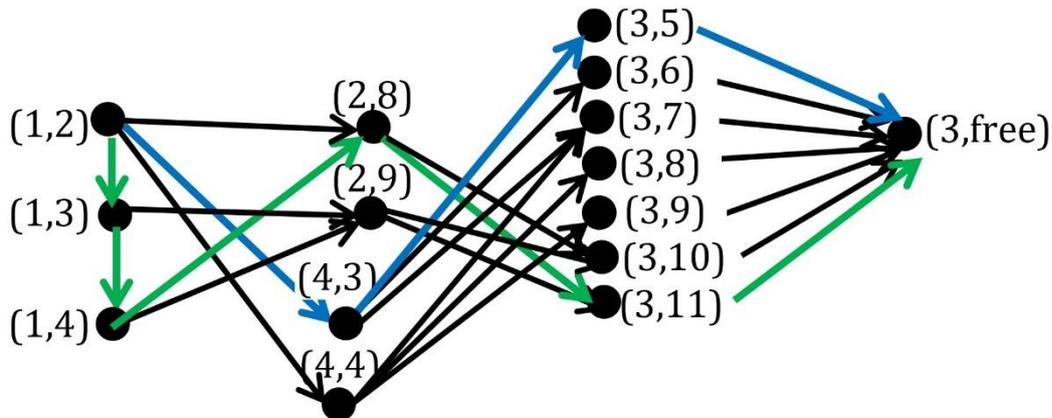

Figure 8.   Minimal fuel consumption (green) and minimal time/ schedule penalty (blue) space-time routes

Table 7.   Enumerated space-time-feasible trajectories for four-node example in incident scenario, with Pareto optimal solutions highlighted

Space-time nodes	Travel time[8]	Fuel	Schedule Penalty
(1,2) → (2,8) → (3,10)	8	14	12
(1,2) → (2,8) → (3,11)	9	13	15
(1,2) → (1,3) → (2,9) → (3,10)	7	15	12
(1,2) → (1,3) → (2,9) → (3,11)	8	14	15
(1,2) → (1,3) → (1,4) → (2,8) → (3,10)	6	10	12
(1,2) → (1,3) → (1,4) → (2,8) → (3,11)	7	9	15
(1,2) → (1,3) → (1,4) → (2,9) → (3,10)	6	12	12
(1,2) → (1,3) → (1,4) → (2,9) → (3,11)	7	11	15
(1,2) → (4,3) → (3,5)	3	12	0
(1,2) → (4,3) → (3,6)	4	14	0
(1,2) → (4,3) → (3,7)	5	16	2
(1,2) → (4,4) → (3,7)	5	13	2
(1,2) → (4,4) → (3,8)	6	15	4
(1,2) → (4,4) → (3,9)	7	16	8

The space-time feasible trajectories for this problem are enumerated in Table 7, with the pure fuel/time optimal solutions illustrated in Figure 8. Comparing with the non-incident scenario (Example 1), it is clear that following the physical route via node 2 (which was best for fuel and travel time without an incident) is not efficient for any of the three objectives, if no stop is made (first four rows in Table 7). The Pareto optimal solutions in the incident case provide an interesting trade-off to be made. From a fuel perspective it is most efficient for the truck to wait at node 1 until the worst of the incident impacts clear (fifth and sixth trajectories), and thence to continue via node 2. However, this incurs significant schedule penalty. Re-routing to take the route via node 4 (trajectory (1,2) → (4,3) → (3,5)) has a major advantage in terms of schedule penalty, yet consumes more fuel.

---

8 Excluding time spent at stops, i.e. only time spent travelling.

5. Case study and numerical results

We now apply the methods described above to a case-study involving a 40T truck travelling from the edge of the city of Glasgow to the Eurotunnel. The network identified is illustrated in Figure 9, comprising motorway-standard links suitable for a truck of such a size/weight; this physical network comprises 51 nodes and 66 directional links. The network includes stretches of road close to urban areas, so used by urban traffic and therefore potentially susceptible to time-of-day urban congestion problems.

In order to calibrate travel times for the 66 network links, the Google Maps live traffic API was polled every 15 minutes (for each link) for one week. Hence, we accrue *real-time predicted* travel time data for each link with entries such as:

link = 5, time = 07-Nov-2018 06:54:41, link travel time = 91.80 minutes.

While this is not actual travel time data, it is a useful proxy and will typically be close to the real data for the period in question. For each link, these "raw" travel time data are resampled on a regular discrete grid both for link entry times and link travel times. The discretisation timestep is chosen to be 1 minute. This allows even short links to be included in the network representation. Figure 10 illustrates this process: the 1min-discretised data (orange dots) interpolate the 15 minute google data (blue diamonds).

As an illustrative example of optimal routing, we consider a journey made by a truck travelling from Glasgow, with earliest start time of 05:00 on Wednesday 7th November 2018, and latest destination time (at the Eurotunnel) of midnight. These limits are encoded into the ESTEN by only creating space-time links from the origin and to the destination that are within these bounds. For each link entry time the *minimum* link travel time is the maximum of (i) the travel time interpolated from the Google data and (ii) travelling at the upper speed limit (default is 96km/h). The *maximum* link travel time is the maximum of (i) the travel time interpolated from the Google data and (ii) travelling at the lower speed limit (default is 40 km/h). Hence, in sufficiently congested conditions when the interpolated Google travel time data for a link corresponds to a link speed below 40km/h, the available link travel time is only the single value derived from the data. The CMEM (as described earlier) is used to relate travel times to fuel consumption, via the speed versus fuel consumption relationship, with parameter values adopted from Franceschetti et al (2013), but with vehicle mass set to be 40T.
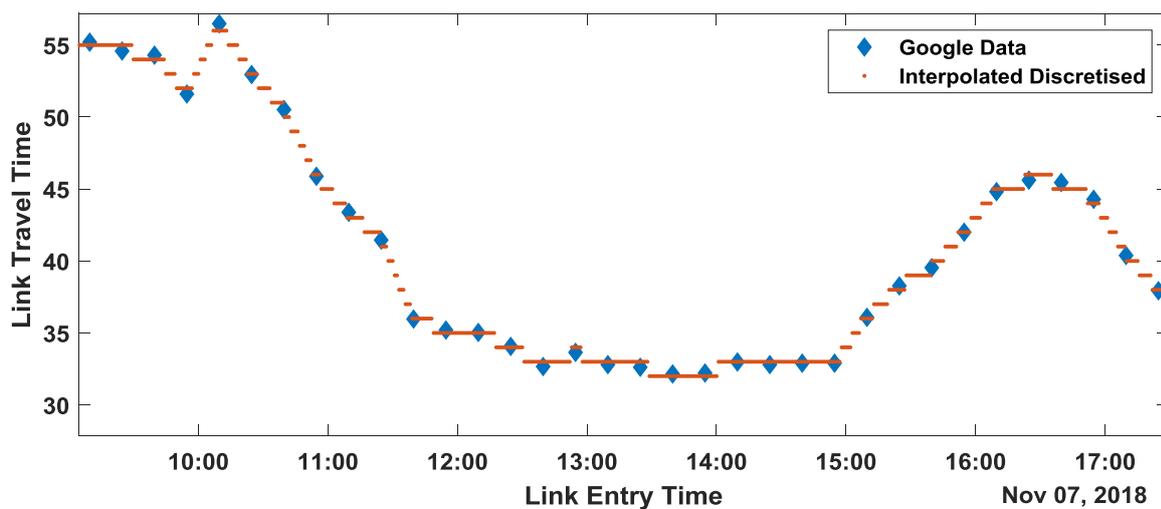
Figure 9.   Case study UK network

Figure 10.   Interpolation and discretisation of Google link travel time data for one example link.

The optimization (firstly, with respect to a single pure criterion) of departure time, link speeds, stops and route is then achieved by Dijkstra shortest path search in the ESTEN, as described in section 3.5. The performance statistics for the method are given in Table 8; the experiments were run on a standard desktop PC (Intel i7-4790 CPU 3.60GHz with 32GB RAM), written in Matlab

2019a and run under Windows 10. The main focus of our study was the space-time network created for the single day of 7.11.2018 (the top rows of the table), for which the network has over two million edges (with default speed range 40 - 96km/h). Despite its size, the Dijkstra algorithm ran in a fraction of a second, establishing the feasibility of the method for real-time applications. As an experiment, a space-time network was also created for a multi-day period (lower rows of Table 8), at the same one-minute temporal resolution. It is not unusual for international truck journeys to take several days, and certainly in such cases the question of where and when to take breaks is relevant (as is which days to depart and avoid), so this is not an artificial case. The space-time network in this case may have up to 45 million edges, yet again the optimal path can be computed in less than a second, providing evidence of the scalability of the method.

Table 8.   Computation times for different size ESTEN, all shortest paths using Dijkstra method. Shaded cells in italics show edge reduction by earliest node arrival time

Start Time	End Time	Speed Limits	Number Edges	Number Nodes	of ESTEN in RAM [MB]	CPU time [secs]		
						Build STEN	Shortest Path	Shortest Path Tree
7/11/18 05:00	7/11/18 23:59	20-110	7,164,558	80,323	110.0	6.28	0.148	0.16
			5,671,297	*80,323*	*87.2*	*4.91*	*0.109*	*0.14*
		40-96	2,368,044	69,256	36.7	2.63	0.055	0.11
			1,893,698	*69,256*	*29.4*	*1.67*	*0.039*	*0.047*
		50-60	703,711	67,012	11.2	1.59	0.023	0.031
			444,769	*67,012*	*7.3*	*0.80*	*0.008*	*0.047*
5/11/18 00:10	9/11/18 23:54	20-110	45,253,274	388,618	693.5	35.2	0.93	1.08
			43,738,328	*388,618*	*670.4*	*35.0*	*0.91*	*0.97*
		40-96	15,031,749	377,551	232.3	12.9	0.352	0.52
			14,517,356	*377,551*	*224.4*	*12.0*	*0.344*	*0.45*
		50-60	4,421,738	375,307	70.3	7.25	0.164	0.34
			4,202,921	*375,307*	*67.0*	*7.08*	*0.164*	*0.23*

Table 8 also illustrates the effect of the two network reduction strategies described in section 3.5. First, narrowing the range of link speeds to be considered (e.g. to 50-60km/h) reduces the available choice of link travel times and hence number of ESTEN links. Whatever range of speeds are considered, real-time data are still encoded into the ESTEN (by the definition of $[\tau^k_{ijt,\min}, \tau^k_{ijt,\max}]$ in section 3.2). To compare computational feasibility we include in Table 8 an increased range of allowed speeds (e.g. 20-110km/h for car or LGV), resulting in a larger ESTEN. Note that it is the lower speed limit that most strongly influences the number of ESTEN links. Consider a 10km long link with speed range $[v_L, v_U]$. With $v_U = 120$ km/h the minimum link travel time is 5 minutes, reducing to 4 minutes if $v_U = 150$ km/h. However, the slowest travel time (and consequently the number of ESTEN links) increases rapidly as v_L decreases as shown in Table 9.

Table 9.   Increase in ESTEN links with decreasing lower speed limit

v_L (km/h)	50	40	30	20	10
Max Link Travel Time (m)	12	15	20	30	60

Secondly, the number of ESTEN links can be reduced by only considering nodes from the earliest time they can be reached, using the earliest start time from the origin and the shortest travel time to each node (a single shortest path tree calculation). Shaded cells in Table 8 indicate the effect of this technique. This substantially reduced the number of ESTEN edges in the one-day example; the ESTEN timespan is 19 hours and, for example, 33 of the 51 nodes cannot be accessed in the first 4 hours, obviating the need for ESTEN links connecting them until after 09:00. This technique

is less effective in the multi-day example, since links can only be ignored during the first few hours from earliest possible departure after which time they could be accessed.

Figure 11 illustrates the results of applying this optimization (with stops not permitted in this illustration) for a criterion of minimizing travel time (left-hand pane) and a criterion of minimizing fuel (right-hand pane), based on the network constructed for 7.11.18. For a travel time minimisation objective, the optimal departure time is 14:16, with a travel time of 573 minutes (thereby arriving at 23:50), driving 822.4km and consuming 378.38 litres fuel. For a fuel consumption objective, on the other hand, the optimal departure time is much earlier at 08:02, but with a rather slower journey, taking 811 minutes and arriving at 21:34, but only using 325.60 litres of fuel on the shorter 747.3km journey.
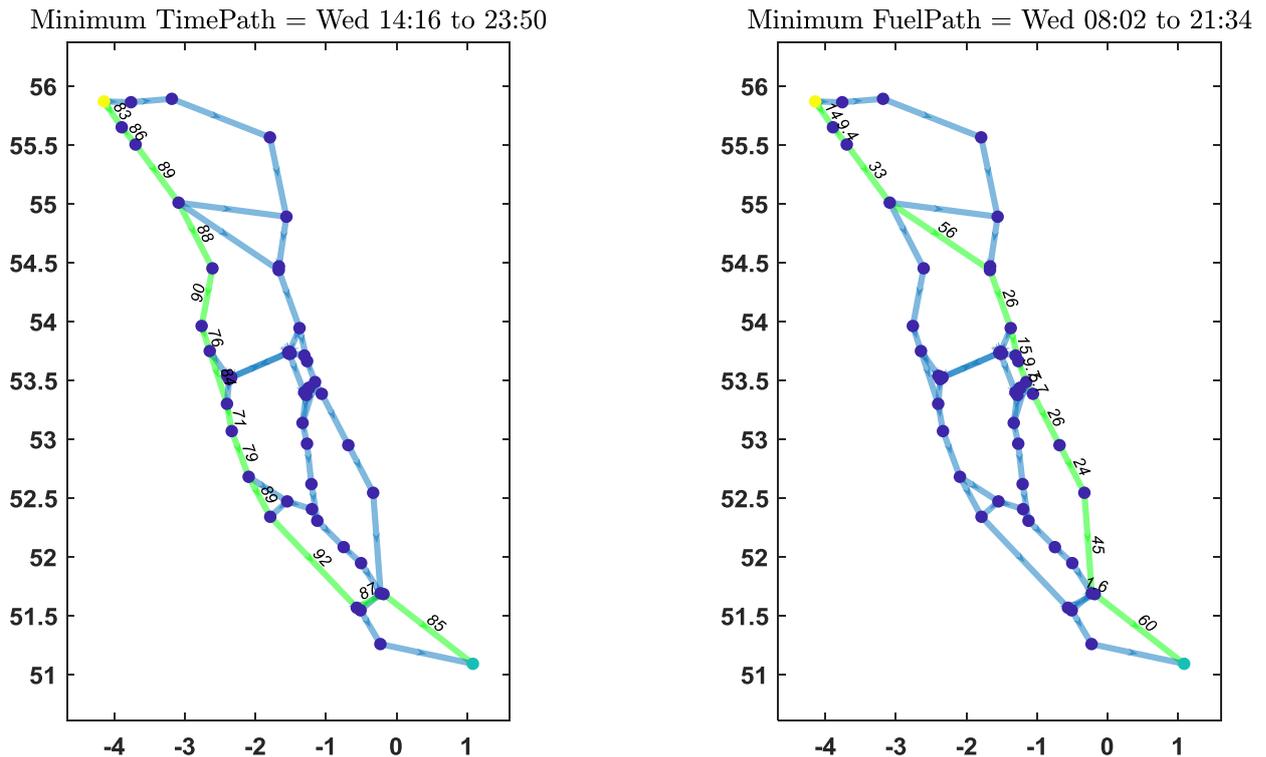

Figure 11.   Optimal routes from Glasgow to Eurotunnel, (left) minimum travel time [574 mins, 378.4 litres] and (right) minimum fuel-consumption [812 mins, 325.6 litres]].

Pursuing a single objective might hide attractive trade-offs between minimising fuel consumption (FC) and travel time (T). Here we follow the scalarization approach of considering weighted combinations of costs $\alpha T + (1 - \alpha)FC$, with $\alpha \in [0,1]$. The cases of $\alpha = 0$ and $\alpha = 1$ correspond to minimum travel time and minimum fuel consumption shown in Figure 11. We create STENs for each set of weighted link costs, and hence compute the trade-offs shown in Figure 12. Minimum fuel consumption is achieved by driving at the optimal speed (55,19 km/h) on route 1, with travel time of 812 minutes. Driving faster on this route decreases travel time with a concomitant increase in fuel consumption (blue stars). Real-time data limits the maximum attainable speeds on route 1, and hence the minimum route 1 travel time is 590mins. Travel time can be reduced further, to 574mins, by switching to route 2; not only is this route longer but needs to be traversed at higher speeds, hence consuming more fuel per km. The final incremental decrease in travel time therefore corresponds to a significant jump in fuel consumption, and this trade-off is only revealed by analysing the range of weighted costs.
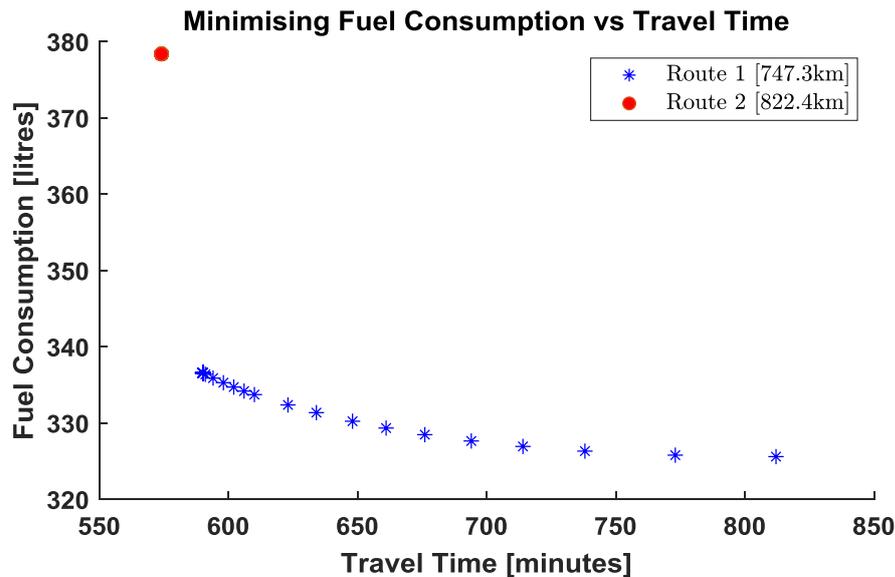
Figure 12.   Trade-off between minimising fuel consumption and travel time.

6.   Conclusions and Further Research

A method has been presented for optimizing fuel efficiency for truck journeys. The method takes into account the full downstream consequences of any decisions, and so differs from a method which would optimize only for the current local conditions. It is particularly applicable to time-varying networks, which can represent the spatio-temporal influences on long journeys, which potentially must pass busy urban areas during peak periods. The optimization is performed for the current estimates of predicted traffic conditions, and so is responsive to unexpected events and incidents. When run in pre-trip mode, it simultaneously optimizes the choice of departure time, route, link speeds, stop locations and stop durations, with respect to any of multiple criteria, e.g. travel time, fuel consumption. When run in on-line mode, the only difference is that departure time can no longer be selected, and the trip starting point is replaced by the next downstream node, otherwise it is identical to the pre-trip mode.

The problem is not readily written as a conventional optimization problem, due to the difficulty in specifying the time-varying linkages in the constraints; for example, the maximum speed that a link may be traversed will depend on the time at which it is traversed, which itself depends on the upstream decision regarding route/departure-time/speeds/stops, and so any attempt to write the constraints for a conventional solver will inevitably fail (aside from the complexity, solvers do not have a sense of 'ordering' of the decision variables, as implied by the constraint for a variable depending on the values of "upstream" variables). In order to resolve this problem, the constraints are written with respect to a novel form of space-time extended network, expanded so it not only includes choice of departure time, route, stops and stop durations, but also the possibility for the trucks to save fuel by travelling slower than the maximum speed possible on a link (the maximum being constrained by speed limits and time-dependent congestion). In this way, the optimization with respect to all factors is then equivalent to finding a shortest path in the space-time network.

Though in principle this solves the problem, the expanded network created is potentially enormous, much larger than (already large) conventional space-time networks due to edges added that represent choice of speed. Thus an important second stage of our work was to test the computational feasibility of the method in realistic-scale networks. Our experiments showed that, although the created networks had many millions of edges, the structure of these networks allowed shortest path determination in fractions of a second, thus demonstrating that the

methods were deployable in real-time applications. While our experiments were performed off-line, we replicated features of an on-line application by using the Google Maps API to interrogate the network for time-dependent, predictive travel times, and then mapping this information to the discretised form required for the space-time network approach.

There are several avenues for future research with this method. Firstly, the space-time network approach is highly amenable, we believe, to representing uncertainty in future traffic conditions through probabilistic modelling, which is especially important for longer-range downstream effects that might occur on very long truck journeys, or when difficult-to-forecast factors such as incident impacts come into play. Such functionality would naturally lead to the question of *how* to optimize under uncertainty, playing off the various risks involved. Secondly, it would be interesting to explore extensions of the method to accommodate additional sources of information that are influential on traffic, for example weather information. For this, it might be worthwhile to exploit cause-based stochastic /hierarchical models such as those in Sumalee and Watling (2008), which themselves may be amenable to a form of network representation. Thirdly, as we noted at the outset of our work, there are different levels at which decisions may be made, across the local, tactical and strategic levels. It would be interesting to explore the integration of these levels, both to understand when effects at a lower level might be fed up to a higher level of decision-making, and to derive methods to handle apparent conflicts between the levels.

Acknowledgements

References

Austin, D. (2018). Pricing Overland Freight Transport to Account for External Costs. *Journal of Transport Economics and Policy* **52**, 45-67.

Barth M., Boriboonsomsin K. (2008). Real-World Carbon Dioxide Impacts of Traffic Congestion. *Transportation Research Record* **2058**, 163–171.

Barth M., Scora G., Younglove T. (2004). Modal Emissions Model for Heavy-Duty Diesel Vehicles. *Transportation Research Record* **1880**, 10–20.

BEIS (2017a). *2015 UK Greenhouse Gas Emissions, Final Figures.* 2016 UK Greenhouse Gas Emissions, Provisional Figures. Statistical Release: National Statistics. February 7th 2017. Department for Business, Energy and Industrial Strategy.

BEIS (2017b). *2016 UK Greenhouse Gas Emissions, Provisional Figures*. Statistical Release: National Statistics. March 30th 2017. Department for Business, Energy and Industrial Strategy.

Bektas, T., Ehmke J.F., Psaraftis H.N., Puchinger J. (2019). The role of operational research in green freight transportation. *European Journal of Operational Research* **274**, 807–823.

Bektas, T., Laporte G. (2011). The Pollution-Routing Problem. *Transportation Research Part B* **45**, 1232–1250.

CfIT (2007). *Transport and Climate Change.* Advice to government from the Commission for Integrated Transport.

Chabini I. (1998). Discrete dynamic shortest path problems in transportation applications: Complexity and Algorithms with Optimal Run Time. *Transportation Research Record* **1645**, 170-175.

EJTIR **19**(4), 2019, pp.265-290

289

Watling, Connors, Milne, and Chen
Optimization of route choice, speeds and stops in time-varying networks for fuel-efficient truck journeys

Chen B.Y., Li Q., Wang D., Shaw S.-L., Lam W.H.K., Yuan H., Fang Z. (2013). Reliable Space–Time Prisms Under Travel Time Uncertainty. *Annals of the Association of American Geographers* **103**(6), 1502-1521.

Dabia S., Demir E., Van Woensel T. (2016). An exact approach for a variant of the pollution-routing problem. Transportation Science **51**(2), 607-628.

Damiani L., Repetto M., Pini Prato A. (2014). Improvement of powertrain efficiency through energy breakdown analysis. *Applied Energy* **121**, 252-263.

Demir E., Bektaş T., Laporte G. (2014). The bi-objective pollution-routing problem. *European Journal of Operational Research* **232**(3), 464-478.

Eglese, R., Maden W., Slater, A. (2006). A Road Timetable to aid vehicle routing and scheduling. *Computers & Operations Research* **33**, 3508–3519.

Eshtehadi R., Fathian M., Pishvaee M.S., Demir E. (2018). A hybrid metaheuristic algorithm for the robust pollution-routing problem. *Journal of Industrial and Systems Engineering*, in press.

Foytik P., Robinson R.M. (2015). Integrating Truck Emissions Cost in Traffic Assignment. *Transportation Research Record* **2503,** 119-127.

Franceschetti A., Honhon D., Van Woensel T., Bektaş T., Laporte G. (2013). The time-dependent pollution-routing problem. *Transportation Research Part B: Methodological* **56**, 265-293.

Frey H.C., Zhang K., Rouphail N. (2008). Fuel Use and Emissions Comparisons for Alternative Routes, Time of Day, Road Grade, and Vehicles Based on In-Use Measurements. *Environ. Sci. Technol.* **42**, 2483–2489.

FTA LCRS (2010). *Freight Transport Association, Logistics Carbon Reduction Scheme.* First Annual Report: Recording, reporting and reducing CO_2 Emissions from the logistics sector.

Gao Z., Smith D.E., Daw C.S., Edwards K.D., Kaul B.C., Domingo N., Parks II J.E., Jones P.T. (2015). The evaluation of developing vehicle technologies on the fuel economy of long-haul trucks. *Energy Conversion and Management* **106**, 766-781.

Gendreau M., Ghiani G., Guerriero, E. (2015). Time-dependent routing problems: A review. *Computers & Operations Research* **64**,189–197.

Gentile, G., Noekel K. (2016). *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems.* Springer.

Hagerstrand T. (1970). What about people in regional science? *Papers of the Regional Science Association*, **24**, 7-21.

Henig M.I. (1986). The shortest path problem with two objective functions. *European Journal of Operational Research* **25**(2), 281-291.

Herold D.M. & Lee K.H. (2017). Carbon management in the logistics and transportation sector: an overview and new research directions. *Carbon Management* **8**(1), 79-97.

Hvattum L.M., Norstad I., Fagerholt K., Laporte, G. (2013). Analysis of an exact algorithm for the vessel speed optimization problem. *Networks* **62**, 132–135.

Jabali O., van Woensel T., deKok A.G. (2012). Analysis of travel times and CO2 emissions in time-dependent vehicle routing. *Production and Operations Management* **21**, 1060–1074.

Kuijpers B., Othman W. (2009). Modeling uncertainty of moving objects on road networks via space–time prisms. *International Journal of Geographical Information Science* **23**(9), 1095–1117.

Liao, F. (2016). Modeling duration choice in space–time multi-state supernetworks for individual activity-travel scheduling. *Transportation Research Part C: Emerging Technologies* **69**, 16-35.

Mavrotas G. (2009). Effective implementation of the ε-constraint method in Multi-Objective Mathematical Programming problems. *Applied Mathematics and Computation* **213**, 455–465

Mickūnaitis V., Pikūnas A. & Mackoit I. (2007). Reducing fuel consumption and CO_2 emission in motor cars. *Transport* **22**(3), 160-163.

Molina J.C., Eguia I., Racero J., Guerrero F. (2014). Multi-objective vehicle routing problem with cost and emission functions. *Procedia Social and Behavioral Sciences* **160**, 254 – 263.

Musso, A. (2001). SOFTICE: Survey on Freight Transport Including a Cost Comparison for Europe. *Transportation Research Record* **1763**, 27-34.

Ng S.T., Cheu R.L. & Lee D.-H. (2006). Simulation Evaluation of the Benefits of Real-Time Traffic Information to Trucks during Incidents. *Journal of Intelligent Transportation Systems* **10**(2), 89-99.

Nuzzolo A., Russo F., Crisalli U. (2001). A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. *Transportation Science* **35**(3), 268-285.

optiTruck (2017). Optimal fuel consumption with Predictive Power Train Control and calibration for intelligent trucks. https://optitruck.eu .

Ozen M. & Tuydes-Yaman H. (2013). Evaluation of emission cost of inefficiency in road freight transportation in Turkey. *Energy Policy* **62C**, 625-636.

Qian J., Eglese R., 2016). Fuel emissions optimization in vehicle routing problems with time-varying speeds. *European Journal of Operational Research* **248**, 840–848.

Schepmann S. & Vahidi A. (2016). Ultracapacitor power assist with preview-based energy management for reducing fuel consumption of heavy vehicles. *International Journal of Powertrains* **5**(4), 375–394.

Small K.A. (1982). The scheduling of consumer activities: Work trips. *American Economic Review* **72**, 467-479.

Sumalee A. & Watling D. (2008). Partition-based Algorithm for Estimating Transportation Network Reliability with Dependent Link Failures. *Journal of Advanced Transportation* **42**(3), 213-238.

Tavares G., Zsigraiova Z., Semiao V., Carvalho M.G. (2009). Optimisation of MSW collection routes for minimum fuel consumption using 3D GIS modelling. *Waste Management* **29**, 1176–1185.

Tong L., Pan Y., Shang P., Guo J., Xian K., Zhou X. (2019). Open-Source Public Transportation Mobility Simulation Engine. DTALite-S: A Discretized Space–Time Network-Based Modeling Framework for Bridging Multi-agent Simulation and Optimization. *Urban Rail Transit* **5**(1), 1–16.

Warburton A. (1987). Approximation of Pareto Optima in Multi-Objective Shortest-Path Problems. *Operations Research* **35**(1), 70-79.

Wen L., Çatay B., Eglese R. (2014). Finding a minimum cost path between a pair of nodes in a time-varying road network with a congestion charge. *European Journal of Operational Research* **236**, 915–923.

White D. (1982). The Set of Efficient Solutions for Multiple Objective Shortest Path Problems. *Comput. Opns. Res.* **9**, 101-107.

Zhang J., Jia L., Niu S., Zhang F., Tong L., Zhou X. (2015). A Space-Time Network-Based Modeling Framework for Dynamic Unmanned Aerial Vehicle Routing in Traffic Incident Monitoring Applications. *Sensors* **15**, 13874-13898.

Ziliaskopoulos A. K., Mahmassani H. S. (1993). Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications. *Transportation Research Record* **1408**, 94–100.