

Enriching the tactical network design of express service carriers with fleet scheduling characteristics

W. J. M. Meuffels · H. A. Fleuren ·
F. C. A. M. Cruijssen · E. R. van Dam

Published online: 16 December 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Express service carriers provide time-guaranteed deliveries of parcels via a network consisting of nodes and hubs. In this, nodes take care of the collection and delivery of parcels, and hubs have the function to consolidate parcels in between the nodes. The tactical network design problem assigns nodes to hubs, determines arcs between hubs, and routes parcels through the network. Afterwards, fleet scheduling creates a schedule for vehicles operated in the network. The strong relation between flow routing and fleet scheduling makes it difficult to optimise the

Statement of Contribution: This paper concerns the tactical network design in road transportation of express carriers. The research is inspired by practical considerations not yet dealt with in literature. In particular, this is related to the cost function used in traditional network design models, which does not correctly reflect operational constraints of an express carrier. Our research enriches the cost function of the tactical network design model by the inclusion of fleet scheduling characteristics (such as vehicle capacities, vehicle balancing, and drivers' legislations) that fleet planners in practice encounter in their job every day. Significant cost reductions in express networks can be accomplished by the use of this new and richer tactical network design model.

W. J. M. Meuffels (✉)
ORTEC, Groningenweg 6k, P.O. Box 490, 2800 AL Gouda, The Netherlands
e-mail: ineke.meuffels@ortec.com

W. J. M. Meuffels · H. A. Fleuren · E. R. van Dam
Department of Econometrics and Operations Research, Tilburg University, Warandelaan 2,
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: fleuren@uvt.nl

E. R. van Dam
e-mail: edwin.vandam@uvt.nl

F. C. A. M. Cruijssen
Department of Logistics, Decision and Informations Sciences, Wageningen University,
Hollandseweg 1, 6706 KN Wageningen, The Netherlands
e-mail: frans.crujssen@wur.nl

network cost. Due to this complexity, fleet scheduling and network design are usually decoupled. We propose a new tactical network design model that is able to include fleet scheduling characteristics (like vehicle capacities, vehicle balancing, and drivers' legislations) in the network design. The model is tested on benchmark data based on instances from an express provider, resulting in significant cost reductions.

Keywords Express service carriers · Freight transportation · Tactical hub network design · Integer programming · Fleet scheduling · Heuristics

1 Introduction

Express service carriers provide time-guaranteed deliveries of parcels. Direct transport from sender to receiver is the fastest way of transport but this is in general not cost efficient. Therefore, express carriers operate a network in which parcels of many customers are consolidated. Parcels of several senders are consolidated at nodes (in practice called depots, terminals, etc.), transported to other nodes via the line-haul network and finally delivered to the consignees. We will now briefly describe how the express supply chain is organised. Then a description of network design is given followed by a discussion on fleet scheduling. At the end of this introduction, our research goals are stated.

1.1 Express supply chain

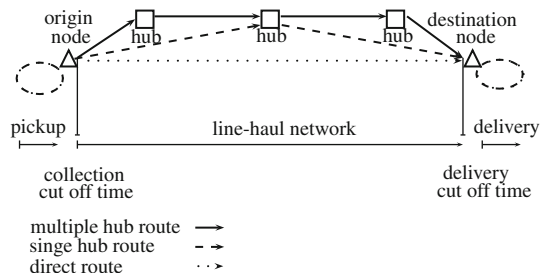
The first node at which a parcel arrives after pickup is called the *origin node* (or *origin*) of the parcel; the node from where the parcel is delivered to the consignee is called the *destination node* (or *destination*) of the parcel. The transport of parcels between origin node and destination node is called *line-haul*. Origin and destination node form an *od-pair*. For these od-pairs, several *services* are offered, defined in terms of promised delivery dates and times of the parcels. Parcels of an od-pair with the same service can always be transshipped together during line-haul transport. The number of parcels of one service to be transshipped between two nodes is called the *flow* of the *origin-destination service pair* (*od-service pair*); the total flow of parcels to be transported between two nodes is called the flow of the od-pair.

Cut off times form the connection between the pickup and delivery process and the line-haul process and guarantee the on-time delivery of parcels. That is, all parcels of one service collected in the pickup process have to be processed and loaded into line-haul vehicles before the *collection cut off time* of the corresponding service; the line-haul transport starts afterwards. The line-haul vehicles have to arrive at the destination nodes before the *delivery cut off time* of the corresponding service. The line-haul vehicles are unloaded after arrival at the destination node and parcels are processed such that the final delivery to consignees can start afterwards. Flows in the line-haul network are either directly transported between nodes or consolidated at *hub* locations. A hub is a sorting

centre serving nodes and other hubs. Hubs in the express network are crucial in making fast and reliable connections. A *direct route* between nodes can be established if there is enough flow to create a (nearly) full vehicle load between two nodes. A direct route can also be used when none of the hub routes is able to meet the service requirements of the corresponding od-pair. A *hub route* is a route from node to node visiting hubs in between; note that hub routes result in detours of flow routing.

Carriers can use ground or air modes in their line-haul transport. Generally, road transport is preferred because of the lower cost involved. Air transport is used to establish services that cannot be offered by ground transport. Considering cost, it is clear that fleet cost dominate in the design of air networks. In road networks, fleet cost is an important cost component though other cost components (like handling cost) are important as well, and the trade off between these cost components determines the final network. Besides, the cost of a single aircraft is in general much higher than the cost of a vehicle, so that it is more costly to use an additional aircraft compared to the use of an additional vehicle. This difference is also illustrated in the design of both networks: while models on air network design focus on fleet routing, road network design focuses on flow routing. The resulting flight schemes are in general the same each day independently of flow size while road schemes slightly differ per day, i.e. the main schedule in a road network is fixed, though in case of large flows it is possible to gain some additional vehicle capacity (but against a higher, but still profitable, price). A second difference in the design of both networks is the capacity of the fleet: aircraft capacities are mostly higher than vehicle capacities, so that in air networks each route is performed by one aircraft, while in road networks multiple vehicles may be scheduled along the same route. Single aircraft routing in combination with a constant aircraft fleet often leads to the introduction of sort windows at hubs in the air network in order to guarantee services: that is, all aircrafts arrive before the start of the sorting process and leave after sorting has taken place. On the contrary, vehicles in road networks arrive and leave hubs at several moments in time. Finally, air networks often face additional restrictions on the possibilities of transport (like time slots at airports, runway constraints, aircraft landing constraints, ...) that are not found in road networks. In this paper, we focus on road transport and our modelling approach is sophisticated to such a network. An overview of a typical express supply chain for road transport is given in Fig. 1.

Fig. 1 Express supply chain



1.2 Hub network design

Consolidation at hub locations was introduced in the literature by O’Kelly (1986). The construction of a line-haul network is better known as the *hub network design problem*. Generally, there are two decision levels in hub network design problems. The *strategic* hub network design problem of express carriers decides on the number and location of hubs in the line-haul network. The *tactical* hub network design problem concerns the assignment of nodes to hubs, determines arcs (i.e. line-hauls) between hubs, and routes flows through the network.

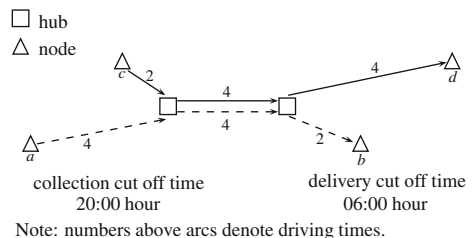
In general, strategic and tactical network design discussed in the literature focus on minimisation of the sum of unit transport cost. It is generally assumed that consolidated transport between hub locations benefits from economies of scale such that unit transport cost of inter-hub flows can be discounted. The main restrictions in both strategic and tactical network design are *flow conservation* and *service commitment*. Flow conservation requires that all flow has to be transported between nodes; service commitment requires that flows are transported within predefined time limits. It is often assumed that the hub network is complete when a link between every hub pair is established, and that no direct routes are allowed (Alumur and Kara 2008). Besides, some literature assumes *capacitated* hub locations that can only deal with a limited amount of flow (e.g. Aykin 1994; Melkote and Daskin 2001).

1.3 Fleet scheduling

After tactical network design, vehicle schedules need to be created such that the flow can be transported. An important aspect of fleet scheduling is the inclusion of *waiting times* (Kara and Tansel 2001): a vehicle can only depart once the flow scheduled on that vehicle has arrived and been processed. In particular, waiting times are important in case of the last vehicle moving via a certain arc. Flows can only be consolidated when there is enough time available for consolidation. This is illustrated in Fig. 2: the cut off times imply that there are only 10 h available to transport flows (a, b) and (c, d); as a result, consolidation of inter-hub flows is not possible. Note that cut off times not only define the available time of transport, but also define the moment of transport.

Another important aspect of fleet scheduling is *vehicle balancing*: since express carriers operate on a daily basis, the number of incoming and outgoing vehicles should be balanced for every node. A third aspect in fleet scheduling that needs

Fig. 2 Consolidation not possible



attention concerns *drivers' legislations*. Maximum driving times and prescribed breaks may not be violated. If the driving time between two locations exceeds the maximum driving time of one driver, a second driver is required resulting in additional cost. In the network design literature, the problem of fleet scheduling and balancing is referred to as the *fleet scheduling problem*.

1.4 Research goals

This paper concerns the tactical network design in road transport of express carriers. The research is inspired by practical considerations not yet dealt with in the literature.

1.4.1 Cost function: plainly linear

The first extension on the existing literature concerns the cost function, which in practice turns out to be more complex than generally seen in the literature. In the latter, the cost function results from unit transport cost and inter-hub transport is discounted. However, O'Kelly and Bryan (1998) claim that the inclusion of an exogenously determined discount applied to all inter-hub arcs regardless of the differences in the flows travelling across them, oversimplifies the problem. The authors claim that the cost has to be presented by a non-linear function such that marginal travel cost decreases as flows increase. The non-linear cost function is afterwards approximated by a piece-wise linear cost function.

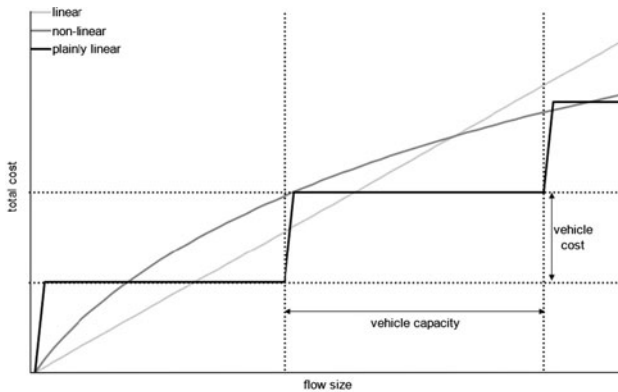
We agree with O'Kelly (1986) that hub consolidation results in economies of scale compared to direct driving, but like O'Kelly and Bryan (1998) we disagree with the traditional discounting of inter-hub transport only. As in O'Kelly and Bryan (1998), we observe that unit transport cost decreases as flows increase, however, we will not apply a discounting of flows but determine vehicle movements explicitly both towards hubs and in-between hubs. In this approach, we assume that only one vehicle type is available, which hardly limits practical applications.¹ Besides, note

¹ There are several reasons that support this assumption. First, because unit cost of medium size vehicles are about 85% of large size vehicles, while the capacity is halved. This means that medium size vehicles will only be used in case of very small flows. Besides, if medium size vehicles are scheduled at (approximately) full vehicle loads, daily fluctuations may require a second medium size vehicle, resulting in higher transport cost in the end. A second reason to apply only one vehicle type comes from subcontracting of transport, which occurs regularly at the express carrier in scope. For subcontracting, vehicle movements are combined in so-called *tours* that can be driven by the same vehicle. There are several classifications of these tours, and the better the tours, the lower the cost of subcontracting. We will not elaborate on the details of tour generation since it is behind the scope of this research. However, an example of a low-cost tour is a 1-day tour (total driving time about 9 h) that starts and ends at the same location and can be driven by one driver satisfying drivers breaks prescribed by regulation. It is easy to understand that if we would use medium trucks towards hubs and large trucks between hubs, tour generation becomes more difficult since separate tours need to be created towards hubs or in-between hubs. The cost of subcontracting these tours is in general much higher than the cost savings achieved by using trucks of medium size and hence we assume that only one vehicle type is available.

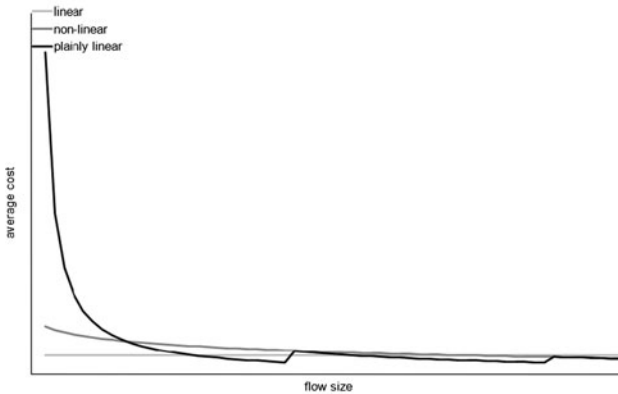
that the model that is proposed can easily be extended to relax this assumption when needed (more routes need to be generated in that case).

The cost in our network design incorporates the plainly linear cost function, since we explicitly determine vehicle movements. This approach is applied to all arcs in the network, so it is not limited to inter-hub arcs only. The discounting of only inter-hub arcs was also questioned by Podnar et al. (2002), who proposed a discounting applied to all arcs traversing flows larger than a certain threshold. Since the dispatching of vehicles is subject to the size of the flow in our network design, we do not need to make any further assumptions on this.

An illustration of the two functions found in literature and the plainly linear cost function proposed in this paper, can be found in Fig. 3, where Fig. 3a shows the total cost and Fig. 3b shows average cost as a function of the total flow. In this figure, ‘linear’ refers to the cost function found in the traditional literature, ‘non-linear’ refers to the cost function proposed by O’Kelly and Bryan (1998), and ‘plainly linear’ reflects the vehicle dependent cost function.



(a) Total cost.



(b) Average cost.

Fig. 3 Comparison of cost functions

1.4.2 Cost function: additional cost components

Besides, we improve the reflection of real-world cost made by express carriers by the inclusion of some other cost components. One of these is vehicle balancing cost. Crainic (2002) describes the need to move empty vehicles because of the imbalances that exist in trade flows that result in discrepancies between vehicle supply and demand in various zones or nodes in the network. Since balancing cost forms a substantial part of the total costs in an express network, we include this cost in our network design. A second cost component that we add to our design concerns the cost of a second driver. Drivers legislations may not be violated, so additional cost is made when a second driver is required. The last cost component that we take into account concerns variable handling cost at hub locations. Note that we do not need to include fixed hub cost, since hub locations are given in the tactical network design.

1.4.3 Note: road versus air networks

It should be noted that (some of) the cost aspects discussed above are captured in the literature on design of air networks. However, to the best of our knowledge, no literature on the design of road networks has included these cost components in their modelling.

Express carriers offering next day services face tight time constraints. The literature discusses the usage of a cover radius (Kara and Tansel 2003), which is a bound on transport time. However, the available time to transport flows depends on the service definition. The tactical network design model presented in this paper uses cut off times to derive the available time to transport flows. In this way multiple services can be included. However, during network design it is not checked whether flows can be combined in a truck. This is done in the heuristic that is run afterwards.

1.4.4 Assumptions on routing

Routes that are allowed in our model can be varied, so long as service requirements can be satisfied with respect to the cut off times. We therefore do not have to assume a complete hub network, nor exclude direct routing. Besides, nodes are not restricted to be connected to a *single* hub node, so we allow *multiple assignments* of nodes to hubs. However, each node is directly assigned to a hub node, that is, no *stopovers* at other nodes are made.

1.4.5 Assumptions on hub nodes

Finally, we assume that hub locations can handle a limited amount of flow because hub locations are fixed and given in the tactical hub network design.

The remainder of the paper is organised as follows. Section 2 gives an overview of the literature on hub network design. The modelling approach is presented in Sect. 3 Two network design models are presented, a traditional model and a new model. A fleet scheduling heuristic is used to derive the final network cost so that

the two models can be compared. The models are tested on data instances of an express provider. The results are presented in Sect. 4 Finally our conclusions and directions for further research are given in Sect. 5.

2 Related literature

This section briefly discusses the literature on the hub network design problem. Recent overviews on hub network design in express networks are given by Alumur and Kara (2009). Overviews on hub network design in general are given by ReVelle et al. (2008) and Melo et al. (2009).

Hub consolidation was introduced in the literature by O’Kelly (1986). In this work, O’Kelly introduced the concept of economies of scale on inter-hub flows: the idea is that flows between hubs might enjoy a discounted transport rate arising from the greater volume on these arcs. This is modelled by discounting unit transport cost for inter-hub flows. The first strategic hub network design model is a quadratic model presented by O’Kelly (1987). Afterwards, several researchers studied strategic and tactical hub network design and several variants of the problem are proposed. The strategic hub network design selects the locations of hubs in the network such that the sum of unit transport cost is minimised (O’Kelly 1992; Aykin 1994, 1995; O’Kelly et al. 1996), the largest transport time is minimised (Kara and Tansel 2001), the number of hubs is minimised (Kara and Tansel 2003; Tan and Kara 2007; Yaman et al. 2007; Alumur and Kara 2008), or the total freight to be delivered to customers within a certain time bound is maximised (Yaman et al. 2008). This paper focuses on the tactical hub network design. The remainder of this section concerns the literature on the tactical hub network design problem.

Kuby and Gray (1993) consider the tactical network design in air transport examining tradeoffs and savings involved with stopovers and feeders towards a single air hub location. The authors observed that in real-world practices direct flights towards an air hub occur only occasionally: most flights stop over at several cities along their routes, and often feeder routes with smaller planes transfer loads to larger planes at intermediate cities. Therefore, a mixed-integer program is developed to design the least-cost single-hub air network including stopovers and feeders. In this, it is assumed that the hub location is already determined. The authors conclude that substantial improvements in cost, miles flown, load factor and number of aircraft can be achieved by using stopovers and feeders in the hub network, and that it is unrealistic to assume a network with only direct flights.

The tactical hub network design in air transport is further examined by Barnhart and Schneur (1996). Pick up and delivery aircraft routes and schedules are derived towards a single hub node. Each aircraft route begins at the hub, visits a set of destination nodes followed by an idle period, then visits a set of origin nodes before returning to the hub. The idle time in between can be used for ferrying (i.e. repositioning of aircrafts). Earliest pick up and latest delivery times are used at the nodes. Associated with the hub is a cut off time, which is the latest time an aircraft

may arrive at the hub. Three service levels are defined in these models: next-day service (24 h), second day service (48 h) and deferred service (3–5 days).

A system that determines aircraft routes, fleet assignments and package routings simultaneously has been described by Armacost et al. (2004). Like Barnhart and Schneur (1996), pick up and delivery routes towards a single air hub are derived including time windows for pick up and delivery. Armacost et al. (2002, 2004) use a composite variable formulation to solve a comparable model.

Multiple hub road networks are considered by Lin (2001). The author observes that vehicle balancing and drivers official work rules are important operating constraints in a cost-effective line-haul operating plan. The work afterwards considers the flow routing problem only, assuming that hub locations and fleet schedules satisfying the operational constraints are given. This problem can then be compared to a capacitated multi-commodity flow problem. Cost taken into consideration is unit transportation cost at the arcs and unit handling cost at hubs. To satisfy the service commitment constraint and capture for connectivity issues (see Fig. 2) three hub windows are defined at which sorts can occur. Two algorithms, a Lagrangian Relaxation and implicit enumeration algorithm with ε -inequality are used to solve the flow routing problem.

Lin and Chen (2004) considers the integrated flow routing and fleet scheduling problem of an air-ground express carrier. Clusters of nodes are created, in which each such cluster contains a hub location. Afterwards, secondary fleet routes are derived to transport parcels between nodes and hubs, and primary fleet routes are derived for transportation between hubs. Given primary and secondary fleet routes, flow routes are assigned to these fleet routes such that the service commitment constraint is satisfied. Connectivity issues are solved by assuming hub sorts to take place at given moments in time. Cost taken into consideration is fixed fleet cost, fleet transportation cost and location handling cost. Furthermore, it is assumed that there is insufficient demand under tight time restrictions to fill up vehicles or aircrafts, so that only one vehicle or aircraft can be dispatched on each fleet route.

Lin and Chen (2008) considers the integration of flow routing and fleet scheduling in a network which may contain stopovers and directs. Cost components taken into account are fixed fleet cost, fleet transportation cost, balancing cost, and location handling cost. Again, fleet routes are derived and each such route can be performed by one vehicle or aircraft. Compared to their work in Lin and Chen (2004), no clustering operation is performed although the model will still assign a node location to one hub. However, the hub used for inbound operations can differ from the hub used for outbound operations, though all inbound (outbound) flow will use the same route to (from) the hub location. Hub sorts are presented to deal with connectivity issues in order to satisfy service commitment. A feasible fleet route plan is determined and afterwards flow routes are derived.

A comparison of the discussed literature on the important design aspects can be found in Table 1. In this table, *NDtrad* and *NDnew* refers to the models that are discussed in the next section.

Table 1 Comparison of the literature

	Design configuration									
	General	Hub network	Direct routes	Node-hub assignment	Stoppers (at non-hubs)	Max.nr. hub touches	Hub capacity	Hub sort window	Consolidation effect	Service commitment
<i>Ndtrad</i>	Tactical	Road	No	Multiple	No	2	Uncapacitated	No	Discount inter-hub flows	Yes
<i>Ndnew</i>	Tactical	Road	Yes	Multiple	No	Unlimited	Capacitated	No	Based on fleet size	Yes
Alumur and Kara (2008)	Strategic	Road	No	Single	No	3	Uncapacitated	No	Discount inter-hub flows	Yes, by a time bound
Armacost et al. (2004) [(2002)]	Tactical	Air	No	Single	Yes	1	Capacitated	Yes	Based on fleet size	Yes
Aykin (1995) [(1994)]	Strategic	Road	Yes	Multiple	No	2	Uncapacitated [capacitated]	No	Discount from/to and inter-hub flows	No
Barnhart and Schneur (1996)	Tactical	Air	No	Single	Yes	1	Uncapacitated	Yes	Based on fleet size	Yes
Kara and Tansel (2003) [(2001)]	Strategic	Road [air]	No	Single	No	2	Uncapacitated	No	Discount inter-hub time	Yes, by a time bound
Kuby and Gray (1993)	Tactical	Air	No	Single	Yes	1	Uncapacitated	No	Based on fleet size	Yes
Lin (2001)	Tactical	Road	No	Multiple	No	2	Capacitated	Yes, multiple	Based on fleet size	Yes
Lin and Chen (2008) [(2004)]	Tactical	Road/air	Yes [no]	Single	Yes	Unlimited	Uncapacitated	Yes, multiple	Based on fleet size	Yes
Melkote and Daskin (2001)	Strategic	Road	No	Single	No	2	Capacitated	No	None	No
O'Kelly (1986, 1987)	Strategic	Road	No	Single	No	2	Uncapacitated	No	Discount inter-hub flows	No
O'Kelly (1992)	Strategic	Road	No	Single	No	2	Uncapacitated	No	Discount inter-hub flows	No
O'Kelly and Bryan (1998)	Strategic	Road	No	Multiple	No	2	Uncapacitated	No	(dependent) Discount inter-hub flows	No
O'Kelly et al. (1996)	Strategic	Road	No	Multiple	No	2	Uncapacitated	No	Discount inter-hub flows	No
Podnar et al. (2002)	Strategic	Road	Yes	Multiple	Yes	Unlimited	Uncapacitated	No	Discount flows above threshold	No
Tan and Kara (2007)	Strategic	Road	No	Single	No	2	Uncapacitated	No	Discount inter-hub time	Yes, by a time bound
Yaman et al. (2008) [(2007)]	Strategic	Road	No	Single	No [yes]	2	Uncapacitated	No	Discount inter-hub time	Yes, by a time bound

Table 1 continued

Cost components							
	Fixed hub cost	Variable hub cost	Labour cost	Fixed transport cost	Unit transport cost	Balancing cost	
<i>Ndirad</i>	No	No	Yes	Yes	Yes	No	
<i>Ndnew</i>	No	Yes	Yes	Yes	Yes	Yes	
Alumur and Kara (2008)	Yes	Yes	No	Yes	Yes	No	
Armacost et al. (2004) [(2002)]	No	Yes [no]	Yes	Yes	Yes	Yes	
Aykin (1995) [(1994)]	No [yes]	No	No	No	Yes	No	
Barnhart and Schneur (1996)	No	No	Yes	Yes	Yes	Yes	
Kara and Tansel (2003) [(2001)]	No*	No*	No*	No*	No*	No*	
Kulby and Gray (1993)	No	No	Yes	Yes	Yes	No	
Lin (2001)	No	Yes	No	Yes	Yes	No	
Lin and Chen (2008) [(2004)]	No	Yes	No	Yes	Yes	Yes [no]	
Melkote and Daskin (2001)	Yes	No	No	Yes	Yes	No	
O’Kelly (1986, 1987)	No	No	No	No	Yes	No	
O’Kelly (1992)	Yes	No	No	No	Yes	No	
O’Kelly and Bryan (1998)	No	No	No	No	Yes	No	
O’Kelly et al. (1996)	No	No	No	Yes	Yes	No	
Podnar et al. (2002)	No	No	No	No	Yes	No	
Tan and Kara (2007)	No*	No*	No	No	Yes	No	
Yaman et al. (2008) [(2007)]	No*	No*	No*	No*	No*	No*	

* Model does not minimise on cost

3 Modelling

The modelling presented in this section solves the network design and fleet scheduling problem in two steps. First, a tactical network design model is run to derive flow routes. The tactical network design models that are used are discussed in Sect. 3.1. Two models are proposed, the first model is a traditional model that discounts economies of scale on inter-hub flow routing, and will be used for benchmarking. The second model is new and includes fleet scheduling characteristics in network design. In order to compare the results, a fleet scheduling heuristic is solved to determine the network cost (Sect. 3.2). The fleet scheduling heuristic uses the flow routes found by one of the network design models. However, the heuristic can also be applied on existing routes of an express provider. Afterwards, a balancing model is run to derive the repositioning of vehicles. Final output of the model is fleet schedules and network cost. Figure 4 gives an overview of the modelling approach.

3.1 Network design model

The network design starts with a set of locations L containing hub locations $H \subset L$ and nodes $N \subset L$. Without loss of generality it is assumed that each location is either a hub location or a node, i.e. $H \cap N = \emptyset$. Node i offers services s to customers guaranteeing a delivery time of parcels received at the node before the collection cut off time c_{is}^o ; in order to satisfy the service, the parcel has to be delivered at the destination node j before the delivery cut off time c_{js}^d . It is assumed that services between nodes are only offered to the customers if their service requirements can be met. That is, the available time between collection cut off time and delivery cut off time of the od-service has to be larger than the driving time between these locations. The total flow of parcels of service s from node i to node j is denoted by f_{ijs} .

Recall that we assumed that there is only one vehicle type available to transport flows. The capacity of this vehicle is equal to v units of flow. Note that vehicle capacity and flow need to be expressed in the same unit (e.g. weight, volume,

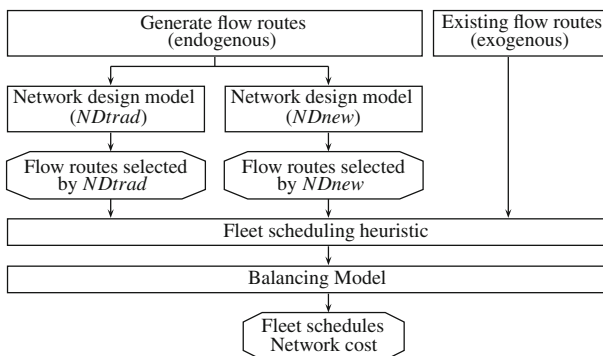


Fig. 4 Overview of the modelling approach

parcels, etc.). Vehicles move via the arcs A of the network; the start location of an arc is denoted by $s_a \in L$ and the end location is denoted by $e_a \in L$. The distance of arc $a \in A$ is given by d_a .

Drivers' legislations should be taken into account when determining the drivers' cost, since maximum driving times and prescribed breaks may not be violated. When the driving time between two locations exceeds the maximum driving time of one driver, a second driver is required and this cost has to be incorporated. We follow the European Regulations (EUR-lex 2006) that prescribe a maximum driving time of 9 h and an uninterrupted break of no less than 45 min, after a driving period of 4.5 h. The total costs of a vehicle moving via arc a is denoted by C_a and includes vehicle transport cost and (second) drivers' cost. All required cost information is available.

For each pair of nodes i, j with a positive flow $\sum_s f_{ijs} > 0$, routes are generated. A route $r \in R$ is created via the arcs a of the network; the parameter u_{ra} equals 1 if route r uses arc a and 0 otherwise. Since route r starts at a node location i and ends at a node location j it can only be used to satisfy services of the corresponding pair of nodes. Besides, the route can only be used for service s of the node pair if it can leave node i after the collection cut off time (c_{is}^o) and arrives at node j before the delivery cut off time (c_{js}^d) taking transport time and hub sorting time into account. This results in a parameter p_{ijsr} that equals 1 if a route can be used to serve service s of od-pair i, j and 0 if it cannot. See Fig. 5 for an illustration of the route generation.

A *direct route* $i \rightarrow j$ is a route that only uses a node-node arc (i.e. $s_a, e_a \in N$). It is not allowed to pass nodes other than the origin and destination node of the route, i.e. a route $i \rightarrow j \rightarrow k$ with $i, j, k \in N$ is not allowed. A *single hub route* $i \rightarrow h_1 \rightarrow j$ is a route that uses a single node-hub arc (i.e. $s_a \in N$ and $e_a \in H$) and a single hub-node arc (i.e. $s_a \in H$ and $e_a \in N$). A *multiple hub route* $i \rightarrow h_1 \rightarrow \dots \rightarrow h_2 \rightarrow j$ can pass more than one hub and uses one node-hub arc, one or more hub-hub arcs (i.e. $s_a, e_a \in H$) and one hub-node arc. Routes are categorised by their number of hub touches n ($n = 0, 1, 2, 3, \dots$); a route with n hub touches is referred to as type Hn route. Note that $H0$ refers to a direct route, $H1$ refers to a single hub route and $Hn, n > 1$ denotes multiple hub routes. The modeller can

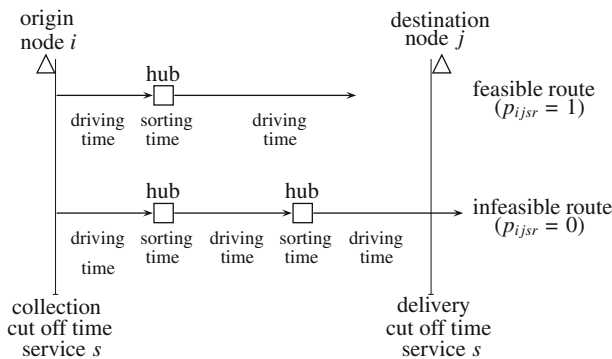


Fig. 5 Route generation, route feasible?

indicate which routes should be taken into account in the model. In general, when we say that we include Hn -routes, all possible routes with n hub touches are generated for each service of an od-pair. However, note that routes that cannot meet the requirements of an od-service are not included in the set of routes that are fed into one of the models.

It is assumed that each service of an od-pair can be satisfied by at least one of the routes generated. If none of the hub routes is able to meet the service requirements, the flow has to be routed directly from origin node to destination node. This flow that has to be routed directly because of tight time constraints is denoted by f_{ijs}^D and can be determined in a preprocessing phase of one of the network design models presented in the next sections. If some services of an od-pair i, j have to be routed directly while others can be routed via a hub route, it is possible to allow these services to use this direct route as well. In this case, either all flow of these services can be routed via this direct route or only part of the flow can use this route. This is discussed in more detail in the sections below. The remaining flow for which a route has to be determined by one of the network design models is denoted by f_{ijs}^R and equals $f_{ijs} - f_{ijs}^D$. An overview of the parameters is given below.

L	Set of locations, index l
$N \subset L$	Set of nodes, index i, j
$H \subset L$	Set of hub locations, index h
S	Set of services, index s
c_{is}^o	Origin cut off of node i service s
c_{js}^d	Destination cut off of node j service s
f_{ijs}	Total flow from node i to node j of service s
f_{ijs}^D	Flow from node i to node j of service s which has to be routed via a direct route
f_{ijs}^R	Flow from node i to node j of service s for which a route needs to be determined
A	Set of arcs, index a
s_a	Start location of arc a
e_a	End location of arc a
d_a	Distance of arc a
v	Capacity of a vehicle
C_a	Cost of one vehicle v moving via arc a
R	Set of routes, index r
u_{ra}	1 if route r uses arc a and 0 otherwise
p_{ijsr}	1 if od-pair i, j can use route r for service s 0 otherwise

Sections 3.1.1 and 3.1.2 present the network design models that are used to determine the flow routes.

3.1.1 Network design model: traditional model

This section discusses a traditional model of the tactical network design problem of an express provider. We call this model *traditional* since it reflects aspects that generally are incorporated in network designs seen in the literature (see Table 1). However, two aspects differ from traditional models, namely service commitment and the node-hub assignment. We chose to satisfy the service commitment constraint here, since service commitment has highest priority in the express business, and network design that does not satisfy this restriction is of no value to the express company. Moreover, the new network design also has to satisfy this requirement, so that a comparison can only be made if this requirement is also incorporated in the traditional model. Due to this requirement and the short time to satisfy services, it is not always possible to connect a node to only one hub, and hence we allow multiple assignments from nodes to hubs. Unit transport cost of a vehicle moving via arc a follow from dividing the cost of one vehicle moving via that arc by the capacity of the vehicle, i.e. $\frac{1}{v}C_a$. To incorporate economies of scales on inter-hub flow routing, a factor α_a is included such that $\alpha_a \leq 1$ for inter-hub arcs and $\alpha_a = 1$ for non-hub arcs.

As described above, it is possible that some flow has to be routed directly because of tight time constraints. It is possible that some flow of an od-pair has to be routed directly while other services of the od-pair can be satisfied by a hub route. In that case, we assume that all flow of the od-pair is routed directly. The parameters f_{ijs}^D and f_{ijs}^R are updated accordingly. Note that either f_{ijs}^D or f_{ijs}^R is equal to 0.

The network design model chooses one route for each service s of od-pair i, j with a positive flow f_{ijs}^R . The variable x_{ijsr} equals 1 if od-service i, j, s uses route r . The flow conservation constraint can now be modelled as

$$\sum_r p_{ijsr} x_{ijsr} = 1, \quad \forall i, j \in N, s \in S, f_{ijs}^R > 0.$$

The total costs of the network design can be formulated as

$$\sum_{ijsr} \sum_a \left(\frac{1}{v} \alpha_a u_{ra} C_a \right) f_{ijs}^R x_{ijsr}.$$

An overview of the model and additional parameters and variables is given below. This network design model is referred to as *NDtrad*.

Parameters

α_a discount factor on arc a for economies of scale

Variables

x_{ijsr} 1 if od-pair i, j uses route r for service s ,
0 otherwise

NDtrad-model

$$\min \sum_{ijsr} \sum_a \left(\frac{1}{v} \alpha_a u_{ra} C_a \right) f_{ijs}^R x_{ijsr} \tag{1}$$

$$\sum_r p_{ijsr} x_{ijsr} = 1, \quad \forall i, j \in N, s \in S, f_{ijs}^R > 0 \tag{2}$$

$$x_{ijsr} \in \{0, 1\}, \quad \forall i, j \in N, s \in S, r \in R. \tag{3}$$

Note that the model enumerates on all feasible routes, selecting the least cost route for each od-pair and service type combination. The traditional model is formulated in this way so that it can be compared to the model formulation of the new network design model presented in the next section.

3.1.2 Network design model: new model

Instead of incorporating a scaling factor for economies of scales, an upper bound on economies of scales can be obtained by determining the minimum number of vehicles required to transport the flows. This network design model selects a route for each service of an od-pair; the routes that can be selected need to satisfy the service requirements of the corresponding service of the od-pair. Since each chosen route is feasible, the model results in a minimum number of vehicles to transport the flows. If time constraints are tight, more vehicles are needed to transport the flows. In case of loose time constraints, the number of vehicles determined by the network design model is sufficient to transport the flows. The model therefore results in an upper bound on achievable economies of scale.

If some flow of an od-pair must be routed directly because of tight time constraints the remaining capacity on the used vehicles are available for transporting flow of the od-pair that could be routed via a hub route (i.e. f_{ijs}^R). However, this flow only uses this direct route if there is enough time for consolidation. The parameters f_{ijs}^R and f_{ijs}^D are updated accordingly. Note that f_{ijs}^R and f_{ijs}^D can be larger than 0 at the same time.

The network design model again chooses one route for each service s of od-pair i, j with $f_{ijs}^R > 0$. As in Sect. 3.1.1, the variable x_{ijsr} is used to denote that od-pair i, j , service s uses route r . The flow conservation constraint is again modelled as

$$\sum_r p_{ijsr} x_{ijsr} = 1, \quad \forall i, j \in N, s \in S, f_{ijs}^R > 0.$$

The total vehicle capacity on each arc has to be sufficient to transport the flow using that arc. By y_a^R we denote the number of vehicles needed to transport the flows f_{ijs}^R via arc a . This results in the constraint

$$\sum_{ijsr} u_{ra} f_{ijs}^R x_{ijsr} \leq v y_a^R, \quad \forall a \in A.$$

The parameter \bar{y}_a^D denotes the number of vehicles required to transport direct flows, and equals $\lceil f_{ijs}^D / v \rceil$. The required number of repositioning vehicles moving via arc a is denoted by y_a^B . The vehicle balancing constraint now becomes

$$\sum_{a|s_a=l} (\bar{y}_a^D + y_a^R + y_a^B) = \sum_{a|e_a=l} (\bar{y}_a^D + y_a^R + y_a^B), \quad \forall l \in L.$$

In practice, the amount of flow that can pass through a hub is limited to the capacity of the hub. We assume that hub h can handle at most Q_h units of flow. Note that it is never optimal to handle flows more than once in a hub, so that the restriction of capacitated hub locations is non-restrictive if $Q_h \geq \sum_{ijs} f_{ijs}^R$. Since routes are generated, it is known if hub h is passed by a route r ; this is denoted by the parameter q_{rh} that equals 1 if route r uses hub h and is equal to 0 otherwise. The hub capacity constraint is modelled as

$$\sum_{ijsr} f_{ijs}^R q_{rh} x_{ijsr} \leq Q_h, \quad \forall h \in H.$$

The total costs are the sum of the variable hub cost and the cost of vehicles moving via the arcs of the network. The hubs that are passed by a route are known so that the variable cost of one unit of flow using route r can be derived. This cost is denoted by C_r^H . Some express providers subcontract vehicle movements (a discussion of subcontracting can be found in Krajewska and Kopfer 2009). As a result, repositioning vehicle movements are sometimes bought at a lower rate when subcontractors can use the movement for other purposes. Now, the total costs of the network follow as (with repositioning vehicles discounted by a factor $\gamma \leq 1$)

$$\sum_r C_r^H f_{ijs}^R x_{ijsr} + \sum_a C_a (\bar{y}_a^D + y_a^R + \gamma y_a^B).$$

An overview of the model and additional parameters and variables is given below. This network design model is referred to as *NDnew*.

Parameters

- γ discount factor of repositioning vehicles moving via arc a
- Q_h maximum amount of flow which can pass through hub h
- q_{rh} 1 if route r uses hub h and 0 otherwise
- C_r^H variable hub cost of using route r
- \bar{y}_a^D number of vehicles moving via arc a to transport flows f_{ijs}^D

Variables

- x_{ijsr} 1 if od-pair i,j uses route r for service s , 0 otherwise
- y_a^R number of vehicles moving via arc a to transport flows f_{ijs}^R
- y_a^B number of repositioning vehicles moving via arc a

NDnew-model

$$\min \sum_{ijsr} C_r^H f_{ijs}^R x_{ijsr} + \sum_a C_a (\bar{y}_a^D + y_a^R + \gamma y_a^B) \tag{4}$$

$$\sum_r p_{ijsr} x_{ijsr} = 1, \quad \forall i, j \in N, s \in S, f_{ijs}^R > 0 \tag{5}$$

$$\sum_{ijsr} u_{ra} f_{ijs}^R x_{ijsr} \leq v y_a^R, \quad \forall a \in A \tag{6}$$

$$\sum_{a|s_a=l} (\bar{y}_a^D + y_a^R + y_a^B) = \sum_{a|e_a=l} (\bar{y}_a^D + y_a^R + y_a^B), \quad \forall l \in L \tag{7}$$

$$\sum_{ijsr} f_{ijs}^R q_{rh} x_{ijsr} \leq Q_h, \quad \forall h \in H \tag{8}$$

$$x_{ijsr} \in \{0, 1\}, \quad \forall i, j \in N, s \in S, r \in R \tag{9}$$

$$y_a^R, y_a^B \in N_0, \quad \forall a \in A. \tag{10}$$

3.2 Fleet scheduling heuristic

The network design models of Sect. 3.1 determine a route for each od-service i, j, s . The fleet scheduling heuristic presented in this section determines the real number of vehicles required to transport the flow and derives fleet schedules. Post-processing determines repositioning cost once fleet schedules are created. This heuristic is used to test relative performance of *NDtrad* and *NDnew*.

The fleet scheduling heuristic uses the following rules for vehicle departures via arc a :

- a vehicle can depart if its departure is critical for the service requirements of one of the od-services for which flow is loaded on the vehicle;
- a vehicle can depart if all flow to be transported via arc a is available;
- a vehicle can depart if it has a full vehicle load.

The heuristic uses an *event list* E of possible departures. All flow is assumed to be available at the origin node at the collection cut off time. These collection cut off times c_{is}^o are the first possible departure times that are added to the event list. The second group of possible departure times that are added are the so-called *critical departure times*. All flow has to be available at the destination node before the delivery cut off time of its corresponding service. Since the flow route x_{ijsr} of od-service i, j, s is known, the latest departure time at each arc in the route can be determined via backwards computing, by starting at the delivery cut off time taking into account transport time and sorting time. The latest departure time of od-service i, j, s at location s_a of arc a is called the critical departure time, denoted by t_{ijsa}^{crit} . The last group of events, the availability time of flow at hub locations, results from the arrival of a vehicle: flow that arrives at a hub location needs further transport and this transport is possible after sorting. The time at which arrived flow can leave the hub location is called the *availability time*; the availability time of od-service i, j, s to be further transported via arc a is denoted by t_{ijsa}^{avail} . Every time a vehicle is scheduled to depart, flow can be transported. If there is more flow available than the

capacity of the vehicle, flow with the earliest critical departure time at the corresponding departure location has highest priority to use this vehicle and is transported to the next location.

The heuristic starts with the first event time e in the event list. Then it checks: (1) do there exist arcs with flow having reached a critical departure time? If there is some flow, vehicles are scheduled to depart and the flow is transported. If there is none, the next question is: (2) do there exist arcs for which all flow has arrived? If there exists such an arc, vehicles are scheduled and flows are transported to the next location. Finally, it is checked: (3) do there exist arcs at which a full vehicle can be loaded? If this is the case, a vehicle departs and the flow arrives at the next location. Afterwards, e is removed from the event list and the next event in the event list is considered. The heuristic terminates when all flow has arrived at its destination node.

Note that vehicle departures are caused because of flow arrivals in step (2) and (3): in step (2), the last flow to be transported via an arc has arrived, and in step (3), flow arrives resulting in a full vehicle load. However, in step (1), a departure does not need to be instigated by the arrival of flow. It might be that some flow is waiting for other flows to arrive, but at some moment (the critical time) it can no longer wait. Then, a vehicle is scheduled to transport this flow. However, this vehicle could already leave at the moment the last flow, which is transported by this vehicle, arrived. This time is referred to as time e^* . Note that e^* can be the availability time of the flow that causes the critical departure, or the availability time of other flow that arrived at this arc (after the arrival of the flow causing the critical departure). The vehicle is scheduled at time e^* , which can be earlier than the critical event time (and therefore also earlier than the current event time, i.e. $e^* < e$). If this vehicle arrives at a hub location, the flow needs further transport. Recall that this flow becomes available for further transport at time $t_{ij\text{sa}}^{\text{avail}}$. Now notice that it is possible that the flow becomes available before the current event time e , because the vehicle might have been scheduled before this time. Since this could impact vehicles already scheduled between $t_{ij\text{sa}}^{\text{avail}}$ and e , these vehicle departures need to be reconsidered. Therefore, the heuristic turns back in time so that the event list restarts at $e = e^*$. All vehicle departures scheduled after e^* are cancelled and the flow is

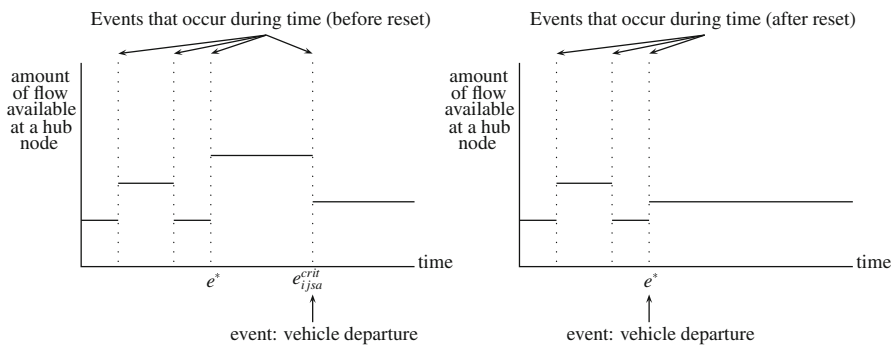
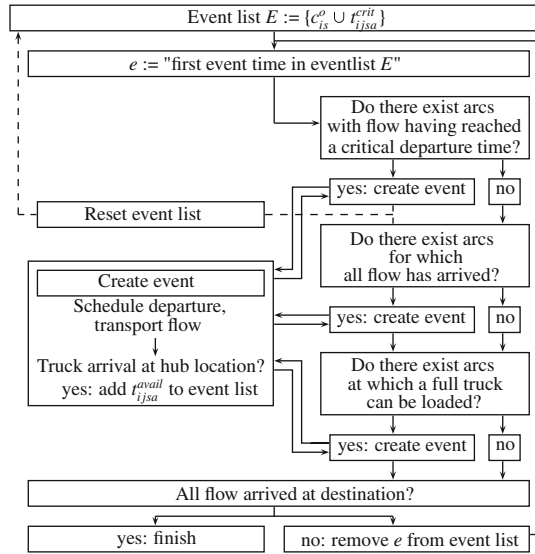


Fig. 6 Example: reset event list at critical time

Fig. 7 Fleet scheduling heuristic



pushed backwards accordingly. This step is referred to as a *reset* of the event list. An example of a reset is given in Fig. 6.

An overview of the heuristic is given in Fig. 7.

3.2.1 Postprocessing: vehicle balancing

Vehicle balancing cost is not determined in the network design model *NDtrad*; the network design model *NDnew* determines vehicle balancing cost but due to tight time constraints, the real number of vehicles required to transport the flows can be higher. Therefore, vehicle balancing cost needs to be determined.

The fleet scheduling heuristic results in a number of vehicles moving via each arc; this is denoted by \bar{y}_a . The required number of balancing vehicles moving via each arc (i.e. y_a^B) needs to be derived. The balancing constraint becomes

$$\sum_{a|s_a=l} (\bar{y}_a + y_a^B) = \sum_{a|e_a=l} (\bar{y}_a + y_a^B), \quad \forall l \in L.$$

The repositioning cost needs to be minimised so that the resulting model becomes

Parameters

- γ discount factor of repositioning vehicles moving via arc a
- \bar{y}_a number of transportation vehicles moving via arc a

Variables

- y_a^B number of repositioning vehicles moving via arc a

Balancing model

$$\min \sum_a \gamma C_a y_a^B \tag{11}$$

$$\sum_{a|s_a=l} (\bar{y}_a + y_a^B) = \sum_{a|e_a=l} (\bar{y}_a + y_a^B), \quad \forall l \in L \tag{12}$$

$$y_a^B \in N_0, \quad \forall a \in A. \tag{13}$$

3.3 Remark

Cut off times are used to determine the available time to transport flows in the network design models. However, the moment of transport is not taken into account (i.e. the possibility to combine flows in time is not checked during network design). Note that including time moments in modelling flow routes would dramatically increase the number of routes possible, since no assumptions are made on departure moments at hub locations. Therefore, the fleet scheduling heuristic is run to estimate the possibility to combine flows in time; if flows cannot be combined, additional vehicles are required. The resulting cost after fleet scheduling are therefore in general higher than the cost found after network design, but the difference in cost depends on the routes given to the heuristic. As a result, a suboptimal solution of the network design model could give lower cost after fleet scheduling than the optimal solution of the network design model.

4 Computational study

The research was inspired by practical considerations of an express carrier. This section presents the results of the models applied to modified instance data of the express service carrier.

Data instances were created for two geographies (Geography *A* and *B*) which are based on actual countries. Data instances define the number and location of nodes and hubs, the cut off times and the services offered between nodes. An overview of the characteristics of the geographies can be found in Table 2. Geography *A* has 31 nodes and Geography *B* has 37 nodes; in both geographies, four hubs are available. Note that the largest distance and the average distance between od-pairs are larger in Geography *B* than in Geography *A*. There is a positive flow between each pair of nodes in Geography *B* while there are only 750 od-pairs with a positive flow in Geography *A*. In the latter, there is no flow between 180 od-pairs. However, the total flow in Geography *A* is larger than in Geography *B*.

In both geographies, two services were defined: services s_1 and s_2 . In both geographies, 80% of the total flow is of service s_1 and the remaining 20% has service s_2 . Parcels with an s_1 -service are available at the origin node before 20:00 h and have to arrive at the destination node before 07:00 h in both geographies. In Geography *A*, s_2 -parcels are available at the origin node before 21:00 h and have to arrive at the destination node before 06:00 h. In Geography *B*, s_2 -parcels are available at the origin node before 20:00 h and have to arrive at the destination node

Table 2 Overview characteristics geographies

	Geography A	Geography B
nr. of nodes	31	37
nr. of hubs	4	4
nr. of od-pairs	750	1,332
Largest distance	909	1,428
Average distance	372	626
Total flow	7,50,000	3,50,000
% s_1 flow	80%	80%
% s_2 flow	20%	20%
$c_{is_1}^o$	day1-20:00	day1-20:00
$c_{is_1}^d$	day2-07:00	day2-07:00
$c_{is_2}^o$	day1-21:00	day1-20:00
$c_{is_2}^d$	day2-06:00	day3-07:00

before 07:00 h two days later. Note that s_2 is a faster service in Geography A, but a slower service in Geography B.

For both geographies, three cases are constructed varying in the demand for each service. For every geography, the total demand is the same in each case, however the geographical spread differs. The first case, *Case 1*, describes the situation in which there is equal demand for each service (i.e. f_{ijs_1} is the same for each i, j and f_{ijs_2} is the same for each i, j). *Case 2* considers the situation with moderate fluctuations in demand for each service. Finally, *Case 3* describes the situation with strong differences in demand for each service. The latter can be interpreted as a situation where a group of nodes represents net senders generating large flows to be transported to net receivers, while there is only small demand vice versa. In both geographies, 15 nodes are indicated as net senders and the remaining nodes are net receivers. Note that the equality in demand as assumed in *Case 1* does not often occur in reality. Fluctuations as created in the second and third set of cases, occurs regularly, where the intensity of these fluctuations is subject to the country in scope. For example, fluctuations as in *Case 2* are found in countries as Germany or France. On the other hand, intensive fluctuations as in *Case 3* occur for example in Turkey, where the economic activity in the West is much stronger than the economic activity in the East.

Balancing movements are discounted by 10% of a transport movement cost (i.e. $\gamma = 0.90$) and variable hub cost is € 0.05 per kg (i.e. $C_r^H = 0.05$). Hub capacities are assumed to be non-restrictive and 60 min of sorting time is needed at each hub location.

Section 4.1 compares the results found by the traditional and the new network design model. Sensitivities of the new network design model are discussed in Sect. 4.2. Section 4.3 concludes with some remarks on computation time, routes generated, and vehicle types used.

4.1 Comparison of the results

This section compares the results found by using *NDtrad*-routes or *NDnew*-routes. The *NDtrad* model is used for benchmarking; we therefore chose α in each case such that the *NDtrad* model gave lowest cost (considered values of α are $\alpha = 0.0, 0.2, 0.4, 0.6, 0.8$ or 1.0) to ensure a fair comparison. Of course, in practice α is input. The resulting values of α are 0.8 in all cases of Geography A, $\alpha = 0.00$ in Cases B1 and B2; Case B3 shows lowest cost when $\alpha = 0.20$. The results of *NDnew* are found by including direct routes, 1-hub routes, 2-hub routes, and 3-hub routes (i.e. we generate all feasible routes visiting respectively at most zero, one, two, or three hub locations). 4-hub routes are excluded since none of these routes could meet the service requirements. Note that only 1-hub and 2-hub routes are included in the traditional model, since we assumed that only two hub touches are allowed. Since we do not exclude any hub-hub arc, this assumption is non-restrictive, because the triangle inequality indicates that a 2-hub route is always cheaper than a 3-hub route.

Table 3 shows for each instance the results found by using *NDtrad*- or *NDnew*-routings and the percentage difference between them. The first column in the table displays the total cost; afterwards the total number of vehicle movements and the

Table 3 Results comparison: cost overview

	Tot. cost (euro)	Tot. nr. mov.	Tot. distance (kms)	Avg. nr. hub touches per KG	Avg. hub throughput
<i>CaseA1—NDtrad</i>	2,24,636	359	1,55,511	1.03	1,93,406
<i>CaseA1—NDnew</i>	2,19,307	350	1,55,425	0.89	1,66,632
<i>CaseA1—difference</i>	-2.4%	-2.5%	-0.1%	-13.8%	-13.8%
<i>CaseA2—NDtrad</i>	2,28,519	368	1,57,661	1.07	2,00,490
<i>CaseA2—NDnew</i>	2,16,042	353	1,53,637	0.86	1,60,534
<i>CaseA2—difference</i>	-5.5%	-4.1%	-2.6%	-19.9%	-19.9%
<i>CaseA3—NDtrad</i>	2,36,317	392	1,65,476	1.04	1,94,508
<i>CaseA3—NDnew</i>	2,19,351	363	1,56,540	0.86	1,61,651
<i>CaseA3—difference</i>	-7.2%	-7.4%	-5.4%	-16.9%	-16.9%
<i>CaseB1—NDtrad</i>	3,07,510	582	2,92,951	1.11	96,772
<i>CaseB1—NDnew</i>	3,06,751	596	2,93,830	0.99	86,367
<i>CaseB1—difference</i>	-0.2%	2.4%	0.3%	-10.8%	-10.8%
<i>CaseB2—NDtrad</i>	3,14,175	606	2,96,546	1.28	1,12,322
<i>CaseB2—NDnew</i>	3,08,065	599	2,98,396	0.82	71,541
<i>CaseB2—difference</i>	-1.9%	-1.2%	0.6%	-36.3%	-36.3%
<i>CaseB3—NDtrad</i>	3,10,906	597	2,99,770	0.93	81,586
<i>CaseB3—NDnew</i>	3,07,796	596	2,97,650	0.87	75,872
<i>CaseB3—difference</i>	-1.0%	-0.2%	-0.7%	-7.0%	-7.0%
<i>CaseGeography A—difference</i>	-5.0%	-4.7%	-2.7%	-16.9%	-16.9%
<i>CaseGeography B—difference</i>	-1.1%	0.4%	0.1%	-18.0%	-18.0%
<i>CaseOverall—difference</i>	-3.0%	-2.1%	-1.3%	-17.5%	-17.5%

corresponding distance driven are presented. The last two columns of the table show the average number of hub touches per kg of flow, and the average hub throughput respectively. A cost breakdown in the three cost components (i.e. balancing cost, variable hub handling cost, and transport cost) is displayed in Fig. 8a. The division of flow over the kind of routes can be found in Table 4 and Fig. 8b.

Comparison of the cost shows that in all cases cost can be reduced by using the routes proposed by the new network design. On average, total cost can be reduced with 5.0% in Geography A and with 1.1% in Geography B. Recall that our new cost function includes three cost components: transport cost, variable hub cost, and balancing cost. The main cost savings are achieved by reducing variable hub cost and balancing cost: on average, 17.5% of the variable hub handling cost can be saved and 18.1% of the balancing cost can be saved. transport cost decreases in half of the cases, while it increases in cases A1, B1, and B2.

We see a decrease in number of vehicle movements in all cases except for Case B1. On average, we see that the total number of vehicle movements is reduced by 2.1%. The resulting total distance driven is on average reduced by 2.1%.

The changes in cost and vehicle movements are caused by changed flow routings. In general, the new routings show less hub touches. On average, the traditional model results in 18.4% of direct routes, 55.5% of 1-hub routes, and 26.1% 2-hub routes; the new model results in 25.1% direct routes, 63.0% 1-hub routes, 10.8% 2-hub routes, and 1.2% 3-hub routes (see Table 4). The resulting average number of hub touches and the average hub throughput are reduced with 17.5% when using the

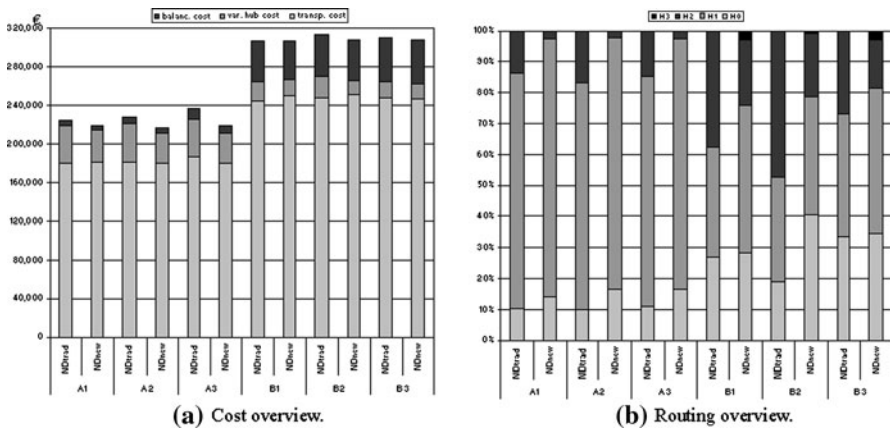


Fig. 8 Results comparison *NDtrad*-routes and *NDnew*-routes

Table 4 Results comparison: routing overview

	H0 (%)	H1 (%)	H2 (%)	H3 (%)
<i>NDtrad</i> —average	18.4	55.5	26.1	0.0
<i>NDnew</i> —average	25.1	63.0	10.8	1.2

new network design model. Less hub routing immediately implies less hub handling cost. Balancing cost are reduced due to the inclusion of direct routes.

The results of this section are referred to as the ‘base’ results of the *NDnew* model.

4.2 Sensitivities *NDnew*-routings

This section shows the sensitivities of the results of *NDnew*-routings to kind of routes, hub capacities, variable hub cost, transport cost, and balancing cost. The results are compared based on cost and route usage. The “Cost overview”-figures that are shown, present cost divided in transport cost, variable hub cost, and balancing cost. The “Routing overview”-figures show the percentage of flow that is routed per kind of route (i.e. direct route, 1-hub route, 2-hub routes, or 3-hub routes).

4.2.1 *NDnew* routings: sensitivity to kind of routes

Here we show the results found by *NDnew*-routing, when varying the kind of routes. The kind of routes that can be included are direct routes, 1-hub routes, 2-hub routes, and 3-hub routes. Note that the ‘base’ results assume that all kind of routes may be used. The results are averages over the cases and are displayed in Fig. 9.

When we only allow direct routes, the cost are €712,653 which is about 2.6 times² as high as the cases in which hub routes are allowed. However, note that including hub routes reduces transport cost but on the other hand leads to increasing handling cost and balancing cost. Including 2-hub routes reduces the cost with 4.7% when direct routes are allowed and with 6.2% when direct routes are not allowed. Apparently, inter-hub flow routing is profitable. However, the inclusion of 3-hub routes has only small impact: on average, 0.4% of the cost can be saved. When we compare the results of allowing direct routes to the results in which direct routes are not allowed, it can be seen that the cost on average are 1.1% higher if direct routes are forbidden. Besides, transport cost remain almost at the same level but both variable hub handling cost and vehicle balancing cost are higher.

Concluding, hub routing leads to a large cost saving. However, in the used geographies, the cost effect of including 3-hub routes is only small when 1-hub and 2-hub routes are included. When allowing direct routings together with hub routings, cost can be further reduced due to lower handling cost and balancing cost. This implies that it is favorable to use more direct routes than in the traditional model.

4.2.2 *NDnew* routings: sensitivity to hub capacities

This section shows the results of limiting hub capacities. Recall that the ‘base’ results assume non-restrictive hub capacities. The results are averages over the cases and are displayed in Fig. 10.

² If we compare to a situation in which only medium size vehicles are used in a network with only directs, this factor is slightly lower, namely 2.2.

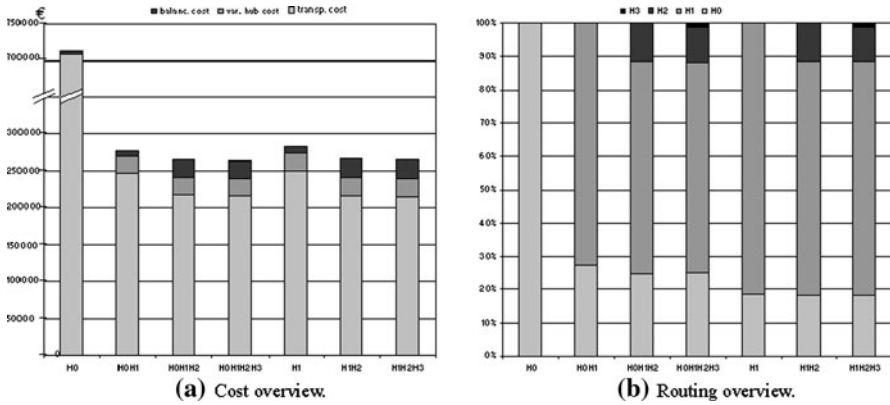


Fig. 9 Results *NDnew*—Sensitivity kind of routes

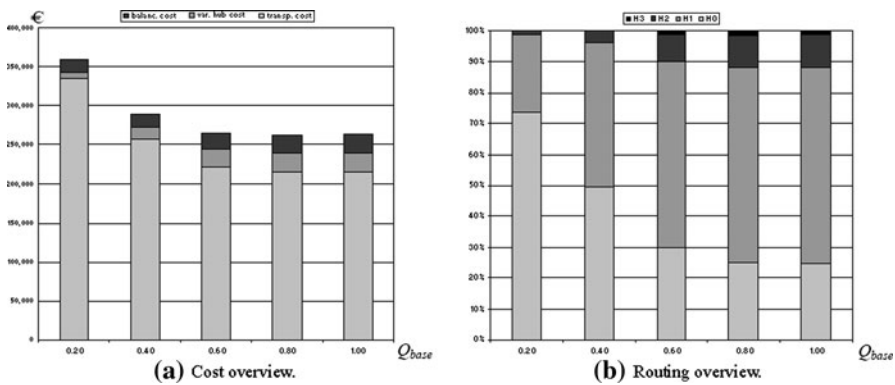


Fig. 10 Results *NDnew*—Sensitivity hub capacities

Since maximum hub capacities were unknown, these capacities were first derived as follows. In each case, the maximum hub capacity was derived as the maximum hub throughput found for one of the hubs in the ‘base’ results. We refer to this as Q_{base} . Afterwards, fractions θ of these hub capacities were taken as bound on the maximum hub capacity, i.e. $Q_h = \theta Q_{base}$ for all hubs h . This was done for fractions $\theta = 0.20, 0.40, 0.60, 0.80$ and 1.00 of the maximum hub capacity.

It is easily understandable that cost decreases when θ increases. This is caused by decreasing transport cost. However, balancing cost and variable hub cost increase. In case only 20% of the hub capacity is available, cost are 37% higher than in the ‘base’ results. However, the cost level stabilizes as soon as 60% of the hub capacity is available. Apparently, the flow can be spread more evenly over the hub locations so that cost savings from consolidation can still be achieved. This can also be seen from the routings: only an additional 5.1% of the flow is routed directly in case $\theta = 0.60$ when comparing to the ‘base’ results.

Concluding, it can be said that there is a strong relation between hub routing and hub capacities. In particular, when hub capacities are restrictive, more direct routing is used, resulting in higher cost. However, cost stabilizes as soon as 60% of the hub capacity is available. Note that fixed hub cost is left out of consideration, although this is likely to depend on capacity.

4.2.3 *NDnew routings: sensitivity to variable hub cost*

The sensitivity of the results was tested against varying variable hub cost C_v^H . The results are averages over the cases and are displayed in Fig. 10.

From the results of varying hub cost it can be concluded that there is a strong relation between variable hub cost and flow routing: when variable hub cost increases, hub routing is less preferred. As a result, required hub capacities are strongly impacted by variable hub cost: there is an overcapacity of 29.8% when variable hub cost increases to €0.25 (compared to €0.05).

4.2.4 *NDnew routings: sensitivity to transport cost*

Finally, the influence of the varying transport cost C_a was investigated. Again, the results are averages over the cases and are displayed in Fig. 11.

The model was run for increasing transport cost C_a . These cost were increased to $1.5C_a$, $2.0C_a$, $2.5C_a$, $3.0C_a$, and $3.5C_a$. Compared to the ‘base’ results (i.e. $1.0C_a$), we see that all cost components increased. It was expected that increasing transport cost would result in more hub routing, since consolidation of flows reduces the number of vehicle movements. But the results show that only a small percentage of the flow uses more hub routing: when transport cost are 3.5 times as high, direct routing decreases only with 2.8% and 1-hub routing decreases with 1.7%. That means that direct routing and 1-hub routing is still attractive even when transport

Fig. 11 Results *NDnew*—Sensitivity variable hub cost

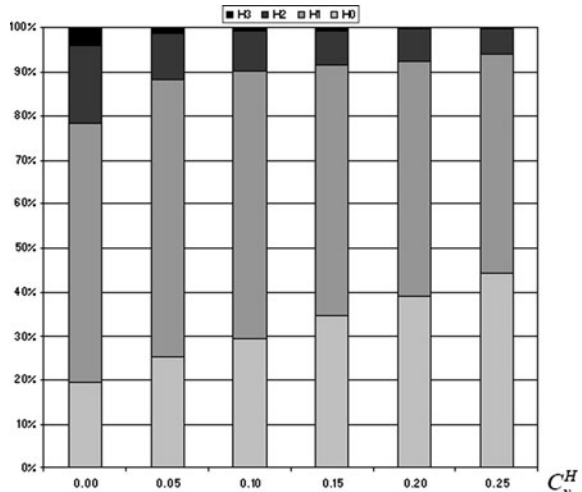
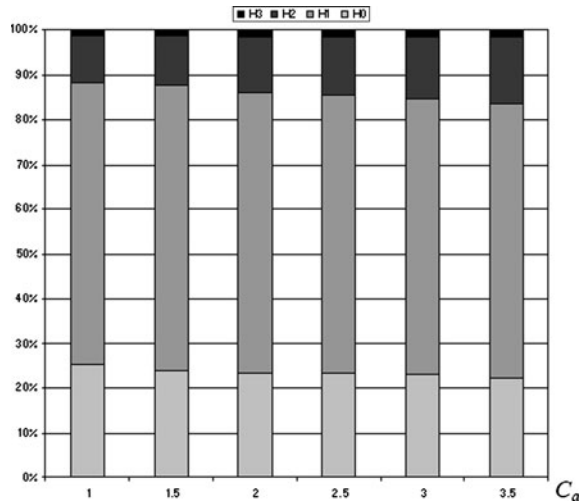


Fig. 12 Results ND_{new} —
Sensitivity transport cost



cost is high. Finally note that hub capacities need to increase: when we multiply C_a with 3.5, we see an increased hub capacity of 8.3% compared to Q_{base} .

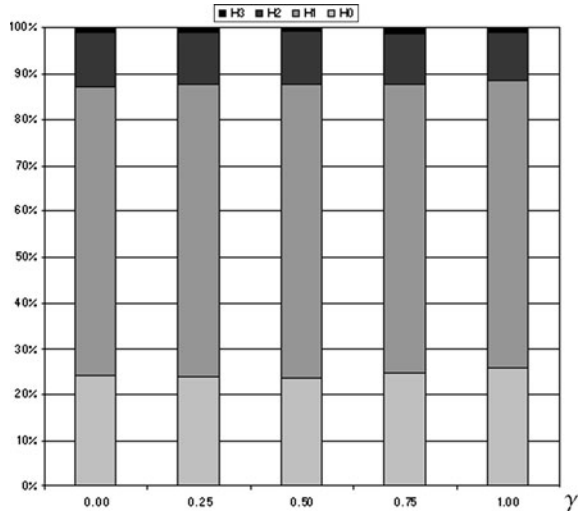
Two effects need to be taken into consideration to explain the impact of transport cost on hub routing: first, more hub routing results in higher total transport cost due to increasing variable hub cost and the detour of flow; second, more hub routing results in lower total transport cost as a result of more consolidation. Note that varying transport cost has both a positive and negative effect on the total transport cost. The results show that there is only small impact of increasing transport cost on hub routing. That indicates that the cost savings of more consolidation are only small compared to the increasing cost due to the detour of flow routing and the increasing variable hub cost. As a result, direct and 1-hub routing is still profitable even when transportation cost strongly increases (Fig. 12).

4.2.5 ND_{new} routings: sensitivity to discounting of balancing cost

Finally, the sensitivity of the results was tested against varying discounting of balancing cost γ . In the 'base' results, it is assumed that balancing cost is discounted with 10% ($\gamma = 0.90$). The results are again averages over the cases and are displayed in Fig. 13.

The results are shown for decreasing discounting of balancing cost. Considered values of γ are 0.00, 0.25, 0.50, 0.75, and 1.00. From the cost it can be seen that cost increases when γ increases, but the largest differences are caused by increasing balancing cost. However, there is a small increase in transport cost when balancing cost increases. Besides, observe that more flow is routed directly. This can be explained as follows. Suppose some flow can be routed via a hub route, consolidated with other flows so that it does not generate additional transport cost (it only generates variable hub cost). However, due to imbalances of flow, a repositioning vehicle is required to drive between origin node and destination node. On the other hand, the flow can be routed directly, but in that case a vehicle has to be scheduled

Fig. 13 Results *NDnew*—Sensitivity discounting of balancing cost



resulting in additional transport cost. No repositioning vehicle is required in that case. It is obvious that the first option, routing via hub locations, is preferred when repositioning is (strongly) discounted; the last option is preferred otherwise.

Concluding, it can be said that there is only a small dependency between direct routing and balancing cost.

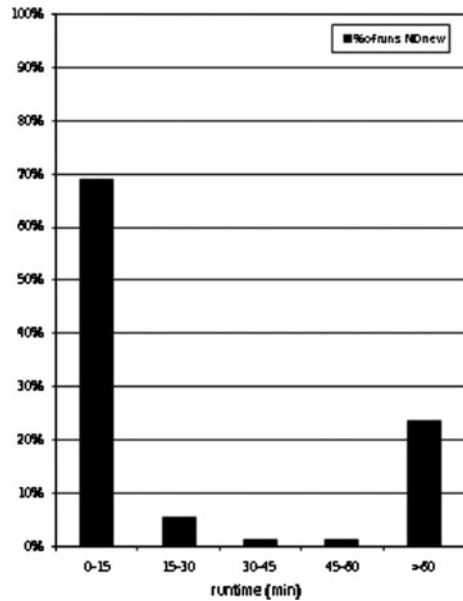
4.3 Remarks

The cases in this section are run on an Intel(R) Core 2 CPU, 2.00 GHZ, and 2.00 GB of RAM. The traditional network design model is run in a few seconds, an overview of run times of the new network design is given in Fig. 14 and shows that 70% of the cases were run within 15 minutes of computation time. 25% of the cases take longer than one hour of computation time, but this is mainly the result of restricting the capacity of flows that can pass through a hub. For all these cases, the best solution found within an hour is within one percent of the optimal solution. The fleet scheduling heuristic is on average run in 11 min.

The number of routes that are generated a priori, depends on the type of routes that are included. The maximum number of Hn -routes that are included is equal to $S * N * (N - 1) * \prod_{h=1}^n (H - n + 1)$. However, no routes are generated if there is no flow between an od-pair, and also routes that do not meet the service restrictions are excluded. The maximum number of routes to be generated in the new network design of Geography A equals 76,260 but the actual number of routes fed into the model equals 6,615. In Geography B, the maximum number of routes equals 1,09,924 and the actual number of routes equals 50,141. For the traditional network design, these numbers (maximum versus actual) are 29,760 and 4,772 in Geography A and 42,624 and 22,030 in Geography B.

In the introduction we discussed the use of medium size vehicles towards hubs and large size vehicles between hubs. Some tests are done based on this assumption,

Fig. 14 Results *NDnew*—Run times



to validate the comparison between the traditional and new network design model. All tests that are done, show that the cost savings achieved by using medium trucks in the new network design are at least as high as the cost savings achieved by the traditional network design. This could also be expected, since the new network design uses less consolidation compared to the traditional network design.

This section showed the results of applying the network design models on modified data of an express service carrier. The next section states our conclusions and recommendations for further research.

5 Conclusions and directions for further research

This paper proposed a new tactical network design model for express carriers.

The model was tested on modified instance data of an express carrier. Test cases were created for two geographies, and for each such geography three test cases were generated varying in the geographical spread of demand. In each test case, cost savings could be achieved if routes proposed by the new network design were used instead of the traditional routes. The first geography showed an average cost saving of 5.0% and the second geography showed an average cost saving of 1.1%. The main cost savings were achieved by reduced variable hub cost and reduced balancing cost.

Furthermore, the sensitivity analyses showed that the cost was 2.6 times as low when consolidation is used to transport flows compared to only direct driving. These savings can still be achieved even when only 60% of the hub capacity is available. Of all cost components, variable hub cost influences hub routing the most:

increasing variable hub cost leads to a strong decreasing hub routing. Higher balancing cost leads only to a small increase in direct routing; higher transport cost results in a small increase in (multiple) hub routing.

This article showed cost reductions by including fleet scheduling characteristics in the tactical network design of express service carriers. The models were tested on modified instance data of two geographies. The results of the geographies differ; it should be further investigated how characteristics of a geography affect the routings. Furthermore, it would be interesting to consider the network design of multiple countries at once (e.g. the network design of Europe). We suggest a column generation approach to reduce computation time when scaling up to data instances of this size.

Final fleet schedules were derived after flow routing. Further cost reductions are expected if fleet schedules and flow routings are determined simultaneously. This article focused on the tactical network design of express carriers. More research needs to be done to show the impact of fleet scheduling on the strategic network design of express carriers.

Acknowledgments We would like to thank handling editor, Professor David Pisinger, and two anonymous referees for their helpful comments and suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alumur S, Kara B (2008) Network hub location problem: the state of the art. *Eur J Oper Res* 190(1):986–1003
- Alumur S, Kara B (2009) A hub covering network design problem for cargo applications in Turkey. *J Oper Res Soc* 60(10):21
- Armacost A, Barnhart C, Ware K (2002) Composite variable formulations for express shipment service network design. *Transp Sci* 36(1):1–20
- Armacost A, Barnhart C, Ware K, Wilson A (2004) UPS optimizes its air network. *Interfaces* 34(1):15–25
- Aykin T (1994) Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *Eur J Oper Res* 79:501–523
- Aykin T (1995) The hub location and routing problem. *Eur J Oper Res* 83:852–863
- Barnhart C, Schneur R (1996) Air network design for express shipment service. *Oper Res* 44(6):852–863
- Crainic T (2002) A survey of optimization models for long-haul freight transportation. *Handbook of transportation science*, 2nd edn. Kluwer, p 11
- EUR-lex (2006) Regulations (EC) No 561/2006 of the European Parliament on the Council of 15 March 2006 on the harmonisation of certain social legislation to road transport
- Kara B, Tansel B (2001) The latest arrival hub location problem. *Manage Sci* 47(10):1408–1420
- Kara B, Tansel B (2003) The single-assignment hub covering problem: models and linearizations. *J Oper Res Soc* 54:59–64
- Krajewska M, Kopfer H (2009) Transportation planning in freight forwarding companies—tabu search algorithm for the integrated operational transportation planning problem. *Eur J Oper Res* 197:741–751
- Kuby M, Gray R (1993) The hub network design problem with stopovers and feeders: the case of Federal Express. *Transp. Res. -A* 27A((1):1012

- Lin C (2001) The freight routing problem of time-definite freight delivery common carriers. *Transp Res B* 35:525–547
- Lin C, Chen S (2004) The hierarchical network design problem for time-definite express common carriers. *Transp Res B* 38:271–283
- Lin C-C, Chen S-H (2008) An integral constrained generalized hub-and-spoke network design problem. *Transp Res E* 44:817–848
- Melkote S, Daskin M (2001) Capacitated facility location/network design problem. *Transp Res A* 35:515–538
- Melo MT, Nickel S, Saldanha da Gama F (2009) Facility location and supply chain management—a review. *Eur J Oper Res* 196:401–412
- O’Kelly M (1986) The location of interacting hub facilities. *Transp Sci* 20(2):92–106
- O’Kelly M (1987) A quadratic integer program for the location of interacting hub facilities. *Eur J Oper Res* 32:393–404
- O’Kelly M (1992) Hub facility location with fixed costs. *Pap Reg Sci J RSAI* 71(3):293–306
- O’Kelly M, Bryan D (1998) Hub location with flow economies of scale. *Transp Res B* 32(8):605–616
- O’Kelly M, Bryan D, Skorin-Kapov D, Skorin-Kapov J (1996) Hub network design with single and multiple allocation: a computational study. *Location Sci* 4(3):125–138
- Podnar H, Skorin-Kapov J, Skorin-Kapov D (2002) Network cost minimization using threshold-based discounting. *Eur J Oper Res* 137:371–386
- ReVelle C, Eiselt H, Daskin M (2008) A bibliography for some fundamental problem categories in discrete location science. *Eur J Oper Res* 184:1264–1277
- Tan P, Kara B (2007) A hub covering model for cargo delivery systems. *Networks Special Issue on Multicommodity Flows and Network Design* 4(1):28–39
- Yaman H, Kara B, Tansel B (2007) The latest arrival hub location problem for cargo delivery systems with stopovers. *Transp Res B* 41:906–919
- Yaman H, Karasan O, Kara B (2008) Release time scheduling in cargo delivery. Available online: www.bilkent.edu.tr/~hyaman/release.htm, pp 80–89

Author Biographies

W. J. M. Meuffels studied Econometrics and Operations Research at Tilburg University, the Netherlands. In 2007, she graduated and started as an external PhD student working at ORTEC. ORTEC is one of the largest providers of advanced planning and optimisation software solutions and consulting services. As consultant, she advises a large express carrier, which also inspires the work she is doing during her research project on the topic of time definite networks.

H. A. Fleuren is professor in the application of Operations Research (OR) in practice at Tilburg University and he has his own consultancy OR Coach. The application of OR can be very rewarding in many organizations but it is much, much more than just the introduction of some algorithm(s). His main conviction is that the practical application of quantitative techniques, for operational as well as strategic problems, can save a lot of resources (time, money, material, etc). During his career he has been working at the edge of science and practice. Until 2006 he was director of the Center for Applied Research at Tilburg University. From 1995 until 1999 he was manager of the Operations Research group at the consultancy bureau Centre for Quantitative Methods (CQM BV) in Eindhoven, The Netherlands. From 1988 until 1995 he was senior consultant at CQM with specialization in vehicle routing, supply chain planning and discrete event simulation of production systems. On the subject of vehicle routing he has written a dissertation in 1988 at the Twente University of Technology. Further he holds a degree in both Information Science and applied Operations Research.

F. C. A. M. Cruijssen studied Economics and Operations Research at Tilburg University. In 2006, he completed his PhD study (cum laude) on the topic of horizontal cooperation in transport and logistics. He has published papers in international academic journal on the topics of, amongst others, horizontal cooperation, network design, new logistics concepts, and benchmarking. He holds academic positions at Wageningen University (Operations Research and Logistics) and Antwerp University (Industrial Economics). Furthermore, Frans is partner at the logistics consultancy firm ArgusI. At ArgusI he specializes in projects on collaboration, network design, simulation and management games.

E. R. van Dam graduated in Technical Mathematics in Eindhoven, the Netherlands in 1992. He obtained his PhD degree from Tilburg University, the Netherlands in 1996, and afterwards held a postdoctoral position at Queen's University in Kingston, Ontario. Since 1998 he has been affiliated with the department of Econometrics and Operations Research at Tilburg University. He was a research fellow of the Royal Netherlands Academy of Arts and Sciences (2001-2006), and is a full professor in Mathematics and Operations Research since 2008. His research interests are in combinatorics, algebra, and their applications in optimization.