Multi-objective optimization of a two-echelon vehicle routing problem with vehicle synchronization and 'grey zone' customers arising in urban logistics

Anderluh, Alexandra; Nolz, Pamela; Hemmelmayr, Vera; Crainic, Teodor Gabriel

Published in:
European Journal of Operational Research (EJOR)

DOI:
10.1016/j.ejor.2019.07.049

Published: 01/01/2021

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Multi-objective optimization of a two-echelon vehicle routing problem with vehicle synchronization and ‘grey zone’ customers arising in urban logistics

Alexandra Anderluh, Pamela C. Nolz, Vera C. Hemmelmayr, Teodor Gabriel Crainic

1. Introduction

Increasing urbanization combined with growing transport volumes currently constitute two major challenges in the field of city logistics. In the European Union (EU), already 75% (2017) of all inhabitants live in urban areas and this number is going to increase by 0.5% annually (Bank, 2018). Road freight transport is still increasing in the EU, with an average growth rate of 0.9% since the year 2000, leading to a total of 1722 billion ton kilometers in 2015, which is of importance because total road transport (passenger and freight) causes 72.9% of all transport-related greenhouse gas (GHG) emissions (Commission, 2017). In addition, road freight transport contributes to noise and congestion, which can be summarized as disturbance, that negatively affects people living and working near heavily used streets.

The volume of European road freight transport is further boosted by the increasing number of e-commerce users, which rose from 28% in 2009 up to 45% in 2016 and causes smaller and smaller order volumes as well as an increasing number of home deliveries with short delivery times (Statista, 2017). These developments make it difficult to supply citizens with all required goods without, at the same time, deteriorating the quality of life due to increasing traffic and growing amounts of GHG emissions, noise and congestion.

One way of dealing with these negative effects of urban freight transport is the use of small emission-free vehicles for the required goods’ deliveries. Especially vehicles like cargo bikes, cargo tricycles or small electric vehicles can be used to deliver goods in densely populated city areas without contributing much to GHG emissions and noise. Even congestion can be reduced, because for...
example, cargo bikes, can use different parts of the road infrastructure (e.g., bike lanes or one-way streets in both directions). They have the additional benefit of more easily supplying city zones where traffic is severely limited, like for example centers of historical cities. Nevertheless, two drawbacks of these vehicles have to be considered. First, they have a limited load capacity. Second, the operational distance is restricted (FGM-AMOR, Outspoke, ECF, & CTC, 2014; Gruber, Ehler, & Lenz, 2013). Therefore, vehicles like cargo bikes cannot efficiently be used for the total amount of urban freight but must be combined with conventional types of vehicles, like vans or city freighters, which can transport a higher amount of goods over a longer distance.

A two-echelon vehicle routing problem (2eVRP) addresses the planning of a system with two vehicle fleets on two echeleons. When vehicles of both echelons are allowed to deliver goods to customers, it has to be decided which customers are visited by vehicles of which echelon. One way of dealing with this issue is pre-assigning all customers to either the first echelon, which consists of the area outside of the inner-city center to the outskirts of the city, or the second echelon, which consists of the inner-city center. This can lead to solutions in which customers located near the borderline, which separates second-echelon from first-echelon customers, are serviced by a second-echelon vehicle (SEV) although a first-echelon vehicle (FEV) passes already close by or vice versa. In contrast to this idea dealt with by Anderluh, Hemmelmayr, and Nolz (2017), we address a more general problem by deciding about the assignment of these customers near the borderline (we call them ‘grey zone’ customers) in the solution procedure. Modeling this problem, which allows us to explore the impact of establishing such a ‘grey zone’ on the solution, is the first contribution of this paper.

We focus on optimizing the two-echelon system, that is, building routes, assigning customers to echeleons and inserting required synchronized meetings between vehicles of different echeleons, to achieve the goals of economic efficiency and social and environmental benefits. The economic objective expressed in costs is taken into account as a first objective in the model. To consider also objectives of citizens and municipal authorities, external effects (GHG emissions and disturbance of citizens caused by noise and congestion) of freight transport are included. Hence, factors of all three aspects of sustainability are included in the respective multi-objective optimization model. Because of the problematic process of assigning appropriate cost factors to GHG emissions and disturbance (Musso & Rothengatter, 2013), we include up to three objective functions in our optimization problem. We express GHG emissions in kg CO₂e (carbon dioxide equivalent) emitted and disturbance as a factor based on the number of citizens affected by road traffic. The metaheuristic solution procedure for the multi-objective problem is based on a large neighborhood search (LNS) procedure in combination with an ε-constraint method to approximately set the efficient (Pareto-optimal) solutions for up to three objectives. Combining LNS with the heuristic rectangle splitting method introduced by Matl, Hartl, and Vidal (2019), which we extended to three objectives, is the second contribution of the paper.

The impact of the city layout (different location of the city center with respect to the total city area and the location of the first-echelon depot) on the two-echelon city distribution scheme is tested by evaluating three different city structures reflected by 18 newly generated artificial test instances. To our knowledge such an investigation has not been done before and is the third contribution of this paper.

The remainder of the paper gives an overview of related literature in Section 2. Section 3 describes the problem in detail, while in Section 4 the problem is formulated as a mixed integer linear program. Section 5 explains the heuristic solution method and Section 6 deals with the computational results obtained by testing artificial test instances and a realistic test instance based on Vienna. Finally, Section 7 concludes the paper.

2. Literature review

Considering multiple objectives in routing problems is not a new field of research. Already in the 1980s, a standard vehicle routing problem (VRP) with the overall aim to minimize total distance, maximize fulfillment and minimize the deterioration of goods was solved by a goal programming approach (Park & Koelling, 1986). In the following 20 years, numerous papers tackled related problems, which are summarized by Jozefowiez, Semet, and Talbi (2008) in an extensive literature review on multi-objective VRPs.

Since then, the number of papers dealing with multiple objectives has increased continuously. Braekers, Hartl, Parragh, and Tricore (2016) give an extensive review on different kinds of VRP by building on the review by Eksioglu, Vural, and Reisman (2009). Vehicle routing literature from 2009 to mid 2015 is classified based on numerous criteria including the considered objectives. 161 out of 327 surveyed articles tackle more than one objective; nevertheless, no detailed information on how the problem is dealt with is included.

In recent years numerous papers have focused on environmental criteria as an additional objective in routing problems. These papers consider either emissions or fuel consumption (Abad, Vahdani, Sharifi, & Etebary, 2018; Alexiou & Katsavounis, 2015; An-droustopoulos & Zografos, 2017; Demir, Bektaş, & Laporte, 2014; Eskandarpour, Ouelladj, Hatami, Juan, & Khosravi, 2019; Ghannadpour & Zarrabi, 2019; Gupta, Heng, Ong, Tan, & Zhang, 2017; Kumar et al., 2016; Poonthbal & Nadarajan, 2018; Ramos, Gomes, & Barbosa-Póvoa, 2014; Sawik, Faulin, & Pérez-Bernabeu, 2017; Soleimani, Chaharlang, & Ghaderi, 2018; Toro, Franco, Echeverri, & Guimarães, 2017; Tricore & Parragh, 2017; Wang et al., 2018).

Besides, external social criteria (noise, congestion, disturbance) are added as a further objective by Nolz, Absi, and Feillet (2014), Sawik et al. (2017) and Govindan, Jafarian, and Nourbakhsh (2018). The fact that there may exist a trade-off between emissions as an environmental objective and disturbance as a social objective is especially investigated by Grabenschweiger, Tricore, and Doerner (2018).

Despite this vast number of papers considering multiple objectives in routing problems, the multi-objective 2eVRP has gained little attention yet. The two-echelon vehicle routing problem deals in general with the distribution of goods in two steps: from a city distribution center (depot) to intermediate facilities (satellites), which represents the first echelon of the problem, and from the satellites to the customers, which represents the second echelon. A comprehensive survey on 2eVRP-literature is conducted by Cuda, Guastaroba, and Speranza (2015), and Cattaruzza, Absi, Feillet, and González-Feliú (2017) provide an overview of multi-level distribution systems in city logistics, but neither review focuses on multi-objective problems.

Although some recently published papers focus on electric vehicles in 2eVRP (Breunig, Baldacci, Hartl, & Vidal, 2019; Jie, Yang, Zhang, & Huang, 2019), multi-objective 2eVRPs have not been considered much. Soysal, Bloehof-Ruwaard, and Bektaş (2015) consider a time-dependent 2eVRP and combine different objectives like distance traveled, travel time, vehicles and emissions in one weighted objective function. Li, Zhang, Lv, and Chang (2016) deal with a time-constrained 2eVRP occurring in linehaul-delivery systems and consider an objective function consisting of different parts, but no Pareto front is provided. Wang, Shao, and Zhou (2017) focus on a 2eVRP with environmental aspects by formulating the objective function as the sum of drivers wage, fuel
cost, and handling cost. A matheuristic based on variable neighborhood search is used to solve this problem. Marinelli, Colovic, and Dell’Orco (2018) take into account environmental effects besides economic cost in a 2eVRP. They solve the problem by means of a dynamic programming approach considering either the one or the other objective.

In contrast to these papers, in which all objectives considered are aggregated in one objective function or are dealt with separately, Esmaili and Sahraeian (2017) minimize total customer waiting time and total travel cost in a 2eVRP. GHG emissions are included as a constraint in the problem, which is solved by an additive weighing method. This is one method to find efficient – Pareto-optimal – solutions to a problem with multiple objectives. For a detailed description of multi-objective optimization we refer to Ehrhoff (2005). Wang et al. (2018) deal with a bi-objective two-echelon waste collection problem in which total cost and the number of vehicles are minimized, and a Non-dominated Sorted Genetic Algorithm-II (NSGA-II) is combined with a Clarke and Wright savings algorithm to find efficient solutions. A collaborative multiple centers 2eVRP is investigated by Wang et al. (2018), in which aggregated operating cost and carbon dioxide emissions are simultaneously minimized by using a NSGA-II. In addition to that, they deal with the distribution of cost savings between cooperating companies.

Another method to find efficient solutions to a multi-objective optimization problem is the $\epsilon$-constraint method already developed by Haines (1971). This method is based on the idea that only one of the objectives of the multi-objective optimization problem is minimized (or maximized) and all other objectives are used as additional constraints with decreasing (or increasing) $\epsilon$-values as right hand sides of the constraints. The resulting single-objective problem can rather easily be addressed by applying exact or heuristic solution procedures, and the set of efficient solutions to the underlying multi-objective problem can be obtained. Because of the parametrization of the $\epsilon$-values, this method belongs to the pool of scalarization techniques (we refer to Ehrhoff (2005) and Antunes, Alves, and Climaco (2016) for further details).

Besides the difficulty of the $\epsilon$-constraint method to determine appropriate $\epsilon$-values, other challenges when used with a heuristic solution method have to be considered: (i) a potentially Pareto-optimal solution can become dominated by a solution later found in the procedure, (ii) when terminated prematurely, entire regions can be left without testing for Pareto-optimal solutions and (iii) the appropriate setting of $\epsilon$-values is challenging and instance-dependent.

Matl et al. (2019) try to master these challenges by introducing a different method to split the solution space of a bi-objective optimization problem. The heuristic rectangle splitting (HRS) starts with determining minimum and maximum values for both objectives, which forms a rectangle. This rectangle can then be split temporarily into a feasible and an infeasible half. The solution to the corresponding sub-problem allows to discard areas which cannot contain Pareto-optimal solutions. In each iteration the largest remaining rectangle is chosen to be split and the feasible half is tested for a Pareto-optimal solution.

To our knowledge no paper currently deals with a multi-objective 2eVRP with vehicle synchronization and customer deliveries on both echelons considering a ‘grey zone’. In addition, the $\epsilon$-constraint method has not been used for solving such a problem yet and to our knowledge HRS has been tested only for bi-objective problems. Therefore, in this paper, we focus on a multi-objective 2eVRP with vehicle synchronization between echelons and customer deliveries on both echelons including a ‘grey zone’. We consider the economic objective together with the environmental objective in this problem and address it by means of a metaheuristic, which embeds a large neighborhood search into a heuristic cubic splitting. The latter is our extension of the heuristic rectangle splitting suggested by Matl et al. (2019) to handle three objectives. Hence, we are able to consider the social objective beside the formerly mentioned two objectives.

3. Problem description

The basic problem considered deals with the combined and synchronized usage of two types of vehicles on two echelons in a city distribution scheme (2eVRPSyn). In our 2eVRPSyn it is assumed that all vehicles start and end their routes at their respective depots. FEVs (e.g., vans or light duty vehicles) deliver goods from the depot of the first echelon (the place where all goods are stored) to customers outside of the city center and/or to transshipment points without storage facilities (called satellites). SEVs (e.g., cargo bikes or small electric vehicles) start at the depot of the second echelon without any goods loaded and must meet immediately after leaving the second-echelon depot with FEVs at satellites to get goods, before they can start serving customers.

As only a limited amount of waiting time is allowed at a satellite for both types of vehicles due to economical considerations, a FEV and a SEV must meet at a certain satellite at approximately the same point in time to perform the loading. So, for any synchronized meeting at a satellite either a FEV or a SEV may wait for a limited amount of time, if necessary. This waiting time is minimized in the objective function as part of the time-related cost (see objective (1) in the mathematical model in Section 4).

From a satellite, SEVs deliver goods to customers in the city center. Whenever a SEV has delivered all goods loaded, it can meet again with a FEV at a satellite for loading purposes and continue its multi-echelon route as long as the maximum route duration, reflecting a working day, is not exceeded.

Summarizing the synchronization aspect, it can be stated that whenever a SEV has no (more) goods loaded to serve customers, the maximum route duration is not exceeded and there are still unserved customers, it has to meet with a FEV at any appropriate satellite to reload goods (see Fig. 1). A satellite can be used more than once for such a synchronized meeting and in such a meeting it is also possible that one FEV meets with more than one SEV at the same time. Each second-echelon route includes at least one synchronized meeting with a FEV right after leaving the second-echelon depot. For such a meeting the starting time of the SEV is chosen in such a way that the waiting time for both vehicles is zero. Whenever an additional synchronized meeting is required for a SEV, then waiting times may occur. In Anderlüh et al. (2017), the 2eVRPSyn is solved with all customers preassigned to echelons. Customers located inside the city center (see green dashed circle in Fig. 2) are preassigned to the second echelon, whereas all other customers are preassigned to the first echelon. This fixed pre-assignment of customers may result in solutions where some customers (see customer A on the left of Fig. 2) are part of a first-echelon route although they may better fit into a second-echelon route and vice versa (see customer B on the left of Fig. 2).

To overcome these drawbacks, we assume a so-called ‘grey zone’ (see grey-shaded zone on the right of Fig. 2) – an area between the inner-city center and the area outside of it – where customer deliveries can be made by vehicles of both echelons. So, the assignment of these ‘grey zone’ customers is part of the solution procedure and should contribute to improve the solution quality compared to the one obtained by a complete pre-assignment of customers.

This assumption stems from discussions with experts as well as considerations about the usability of the assumed SEVs. Especially in European cities with a historic city center, we face numerous narrow one-way streets, pedestrian zones and the absence of parking space for conventional delivery vehicles in the inner-
city center. Therefore, this area seems appropriate for deliveries by SEVs only. In contrast, the area near the outskirts of a city is too far away to be supplied by SEVs assuming an action radius of 3-5 km around the second-echelon depot for SEVs like cargo bikes (FGM-AMOR et al., 2014; Gruber et al., 2013).

Hence, all customers located outside this radius are preassigned to FEVs. All remaining customers located inside the ‘grey zone’ can be served either by FEVs or SEVs (see customers A and B on the right of Fig. 2, which are now served in a better way than on the left of Fig. 2).

In this paper, we focus not only on costs as the objective of the optimization problem at hand. We also take external effects into account as additional objectives. Such modeling provides the means to analyze the various cost elements and decide on the strategy to adopt for both individual companies and the system regulator (e.g., municipal authorities).

Costs are assumed as the sum of time-related and distance-related variable cost for each route plus fixed cost for each vehicle used. The time-related cost include travel time, service time and waiting time occurring at satellites.

In addition, we consider GHG emissions as a climate-relevant externality and disturbance (noise and congestion as negative impacts on inhabitants) as health-relevant and therefore a social externality based on ideas described in Musso and Rothengatter (2013).

The environmental effect is assumed to be distance-based. More detailed emission models can be found in the literature we refer the interested reader to Demir, Bektas, & Laporte, (2011) but as the focus of this paper is not on the detailed calculation of GHG emissions, this simplification seems appropriate.

Yeh (2013) evaluates disturbance (or externalities, as it is called in this paper) for specific territories and modes of transport in Greater Paris. This is done by a gravity indicator related to traffic intensity and population density. Our method how disturbance along traversed paths can be calculated for a city like Vienna is described in Section 5.5.

Thus, the ‘grey zone’ makes sense also with respect to the multiple objective functions, especially because we do not only consider economic cost, but also environmental and social objectives. Specifically, when schools and hospitals are located in the ‘grey

---

**Fig. 1.** Synchronization along the routes of two first-echelon vehicles (FEV_1 and FEV_2) and a second-echelon vehicle (SEV_1).

**Fig. 2.** Two-echelon city distribution scheme; on the left all customers preassigned to echelons; on the right customers within ‘grey zone’ without pre-assignment to echelons. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)
zone’, the use of eco-friendly vehicles seems more appropriate. The resulting problem is a multi-objective two-echelon vehicle routing problem with vehicle synchronization and ‘grey zone’ customers (MO2eVRPSynCZ).

4. Mathematical model

The problem is defined on a graph $G = (V, A)$. The vertex set $V$ consists of the set of depot nodes $V_d$, the set of customer nodes $V_c$ and the set of satellite nodes $V_s$ (the notation used is summarized in Table 1).

The set $V_d$ contains one depot node $V^m_d$ for each echelon $m \in \{1, 2\}$. For the first echelon a depot is assumed as storage facility and starting point of all first-echelon vehicles, and for the second echelon a depot is assumed as starting point for second-echelon vehicles. As this depot is located in the inner-city center and prices per m² are high, only one location is assumed here too. For each depot a start and end node is created ($v^m_0 \cup v^m_1$). The extended set of depot nodes is denoted by $V^m_d$.

The set $V_c$ consists of three subsets of customer nodes $V^m_c$ with $m \in \{0, 1, 2\}$, with $m = 1$, meaning that only first-echelon vehicles can go there. The same holds for $m = 2$ and the second echelon. If $m = 0$, vehicles of both echelons may go to these nodes, which we denote as ‘grey zone’ customers.

The set $V_s$ is extended to $\hat{V}_s$ by duplicating every physical satellite node $v_s$ as many times as there are customer nodes in $V^m_c \cup V^2_s$ which can be served by second-echelon vehicles, yielding $|V^m_s| + |V^2_s|$. This is due to the fact that every physical satellite node may be used more than once during the execution time of the routes and also by more than one vehicle of the same echelon at the same time. So, the extended set of vertices can be denoted by $V_s = V^m_s \cup V^2_s \cup \hat{V}_s$, with all nodes in the set $\hat{V}_s$ reachable by vehicles of both echelons. On the other hand we can define the sets $V^m = V^m_c \cup V^m_s \cup V^2_c \cup V^2_s$, with $m \in \{1, 2\}$, that is, the set of all vertices reachable by a first-echelon ($m = 1$) or a second-echelon ($m = 2$) vehicle. The arc set $A$ consists of all feasible arcs in the problem. So, following the above vertex set definition, we define two mode-specific sets of arcs $A^m = \{ (i, j) \mid i, j \in V^m \}$ with $m \in \{1, 2\}$.

Each node $i \in V^m_e$ has its specific loading or service time $\lambda_i \geq 0$ and each node $i \in V^m_s$ has its specific demand $d_i > 0$.

The set of vehicles $F$ consists of a set of first-echelon vehicles $F^1$ and a set of second-echelon vehicles $F^2$. Customers in the ‘grey zone’ can be visited by any vehicle, i.e., $F^0 = F^1 \cup F^2 = F$. Each vehicle $k$ has a certain capacity $Q^k$ as well as a certain distance $\delta^k_{ij}$ and travel time $r^k_{ij}$ for each feasible arc $(i, j)$. For the calculation of costs each vehicle $k$ has a specific vehicle cost $c^k_\ell$ per unit of distance traveled, driver cost $c^k_d$ per unit of time used and fixed cost $c^k_f$ per vehicle used. The total cost of vehicle $k \in F$ for traversing a feasible arc $(i, j)$ is represented by $c^k_{\ell ij}$. Eventually, for each vehicle $k$, distance-based GHG emissions $e^k_{\ell ij}$ and disturbance $\rho^k_{\ell ij}$ for traversing a feasible arc $(i, j)$ are known. As we assume homogeneous sets of vehicles, all these parameters are the same for vehicles of the same echelon. The maximum route duration is given by $t_{\text{max}}$ and the maximum permitted waiting time at any satellite $s \in \hat{V}_s$ for each vehicle $k$ is defined by $w_{\text{max}}$.

Furthermore, we define as decision variables the binary variable $x^1_{ij} = 1$ if vehicle $k$ travels from node $i$ to node $j$, and continuous variables $w^1_{kij} \geq 0$, specifying the waiting time occurring for vehicle $k$ at satellite $s$, $t^1_{kij} \geq 0$, specifying the arrival time of vehicle $k$ at node $i$, and $u^1_{kij} \geq 0$, specifying the load of vehicle $k$ after serving node $i$. In addition, we use the constant $M_k$, which represents the total demand of all customers. The optimization problem at hand can be formulated as a mixed integer linear problem with the objectives:

$$
\text{min} \sum_{k \in F} \left( \sum_{ij \in V^m} \left( (r^k_{ij} + \lambda_j) c^k_\ell + \delta^k_{ij} c^k_d + \sum_{s \in \hat{V}_s} w^k_{sij} c^k_f \right) \right) + \sum_{ij \in V^m} c^k_{\ell ij} x^1_{ij}
$$

$$
\text{min} \sum_{k \in F} \sum_{ij \in V^m} c^k_{dij} x^1_{ij}
$$

$$
\sum_{k \in F} \sum_{ij \in V^m} \sum_{s \in \hat{V}_s} \theta^k_{sij} x^1_{ij}
$$

Subject to:

$$
\sum_{j \in V_c} x^k_{ij} = \sum_{j \in V^m_c} x^k_{ij}, \forall k \in F, \forall m \in \{1, 2\}
$$

$$
\sum_{j \in V_c} x^k_{ij} \leq 1, \forall k \in F, \forall m \in \{1, 2\}
$$

$$
\sum_{h \in \hat{V}_s} x^k_{ij} = \sum_{h \in \hat{V}_s} x^k_{ij}, \forall i \in V_c, \forall k \in F
$$

$$
\sum_{k \in F} \sum_{h \in \hat{V}_s} x^k_{ih} = 1, \forall m \in \{0, 1, 2\}, \forall i \in V^m_c
$$

$$
\sum_{k \in F} \sum_{h \in \hat{V}_s} x^k_{si} = \sum_{j \in V^m_c} x^k_{ij}, \forall s \in \hat{V}_s, \forall k \in F
$$

\[
\sum_{k \in F^0} \sum_{i \in V^m} x_{ik}^k \leq 1, \forall s \in \bar{V}_s, \forall m \in \{1, 2\} \tag{9}
\]

\[
\sum_{k \in F^1} \sum_{i \in V^0} x_{ik}^k = \sum_{k \in F} \sum_{i \in V} x_{ik}^k, \forall s \in \bar{V}_s \tag{10}
\]

\[
u_i^k + d_j \leq u_i^k + M_d(1 - x_{ij}^k), \forall k \in F, \forall i \in V_k, \forall j \in V_c \tag{11}
\]

\[
0 \leq u_i^k \leq Q_k, \forall k \in F, \forall i \in V_k \tag{12}
\]

\[
u_i^k \leq M_d(1 - x_{i1}^k), \forall k \in F^2, \forall i \in V_k, \forall s \in \bar{V}_s \tag{13}
\]

\[
u_i^k \leq M_d(1 - x_{i1}^k), \forall k \in F^3, \forall i \in V_k, \forall s \in \bar{V}_s \tag{14}
\]

\[
u_i^k + u_i^k \leq u_i^k + M_d(1 - x_{i1}^k), \forall k \in F, \forall i \in F^1, \forall l \in F^2, \forall i \in V_l, \forall s \in \bar{V}_s \tag{15}
\]

\[
t_k^i \leq \sum_{j \in V_k} x_{ij}^k t_{ij}^{max}, \forall i \in V_k, \forall k \in F \tag{16}
\]

\[
t_k^i = 0, \forall k \in F^1 \tag{17}
\]

\[
t_k^i + \tau_{ij}^k + \lambda_i - t_k^j \leq (1 - x_{ij}^k) t_{ij}^{max}, \forall i \in V_k \setminus \bar{V}_s, \forall j \in V_k, \forall k \in F \tag{18}
\]

\[
t_k^i + \tau_{ij}^k + \lambda_s + w_s^k - t_k^j \leq (1 - x_{ij}^k) t_{ij}^{max}, \forall s \in \bar{V}_s, \forall j \in V_k, \forall k \in F \tag{19}
\]

\[
w_k^j \geq \sum_{i \in V_k} t_k^i - \sum_{i \in F^1} t_s^i, \forall s \in \bar{V}_s \tag{20}
\]

\[
w_k^j \geq \sum_{i \in F^1} t_k^i - \sum_{i \in \bar{V}_s} t_s^i, \forall s \in \bar{V}_s \tag{21}
\]

\[
w_k^j \leq \sum_{j \in V_k} x_{ij}^k t_{ij}^{max}, \forall s \in \bar{V}_s, \forall k \in F \tag{22}
\]

\[
x_{ij}^k \in \{0, 1\}, \forall i, j \in V_k, \forall k \in F \tag{23}
\]

\[
t_k^i, u_i^k, w_s^k \geq 0, \forall i \in V_k, \forall s \in \bar{V}_s, \forall k \in F \tag{24}
\]

Objective function (1) minimizes total transportation cost consisting of time-related and distance-related variable cost and vehicle-related fixed cost. Total GHG emissions caused by all arcs used are minimized in objective function (2), while objective function (3) minimizes the total disturbance caused by all arcs traversed.

Constraints (4) and (5) ensure that each vehicle leaving its depot also returns to that depot and that each vehicle is used at most once. Constraints (6) and (7) guarantee that each customer is visited exactly once by a vehicle. Constraints (8) and (9) ensure that each cloned satellite is used at most once by vehicles of the same type and constraint (10) guarantees that, if it is used as supply point by a second-echelon vehicle, it has to be visited by a first-echelon vehicle too. Constraint (11) ensures for each vehicle that the load after visiting a node is equal to the load before minus the demand of the respective node. Constraint (12) ensures that the load of a vehicle at each visited node does not exceed the vehicle’s capacity. Constraint (13) guarantees that each vehicle is empty when reaching the depot at the end of its route, whereas (14) ensures that every second-echelon vehicle is empty when visiting a satellite. Constraint (15) ensures that the load of the second-echelon vehicle at a satellite is added as demand for the first-echelon vehicle which visits this satellite. Constraint (16) restricts the maximum route duration of a vehicle, while constraint (17) enforces the start time of each first-echelon vehicle at the depot to be zero. Constraint (18) ensures the scheduling of nodes for all vehicles except satellites. The scheduling for satellites is enforced by constraint (19). Constraints (20) and (21) determine the waiting time of a first-echelon/second-echelon vehicle at a node as the difference between the arrival time of the respective vehicle at this node minus the arrival time of the corresponding vehicle of the other echelon visiting this node. Constraint (22) limits the synchronization of first-echelon and second-echelon vehicles at cloned satellites within a maximum permitted time span. Eventually, constraint (23) imposes binary values on the main decision variable, while constraint (24) ensures non-negative values for all other decision variables.

5. Solution method

The solution procedure is a metaheuristic to find good solutions to the problem with respect to one objective in an affordable amount of time. This is done by using a LNS concept, which has already proven to be very effective for addressing routing and scheduling problems in general (Pisinger & Ropke, 2010). Our metaheuristic integrates a LNS into a multi-objective method to find solutions along the Pareto front. The multi-objective method used in our metaheuristic is based on the heuristic rectangle splitting suggested by Matl et al. (2019), which is applicable for bi-objective problems and which we extend in this paper to a heuristic cuboid splitting applicable to problems with three objectives.

An overview of all parts of our metaheuristic, which are described in the following subsections, is provided in Fig. 3.

5.1. Construction of initial solutions

The metaheuristic solution procedure starts with finding a feasible initial solution s for the economic objective of the problem defined in objective function (1). The main idea of building an initial solution to the problem is based on Anderlhub et al. (2017), where a sequential construction procedure is used. We start with the construction of second-echelon routes for all preassigned second-echelon customers vehicle by vehicle. This is done by a GRASP. The next customer for a route is chosen randomly out of a restricted candidate list, in which promising second-echelon customers, which are not yet assigned to a route, are included.

Whenever a meeting of a SEV with a FEV is required, i.e., just after leaving the second-echelon depot or when a SEV runs out of load and has time and remaining customers to continue its route, the cheapest available satellite is inserted into the route as a dummy satellite. This is done to have an approximation of the detours required for the synchronized meetings with FEVs. As soon as all second-echelon customers are routed, we can calculate the demand required as well as the arrival time of the respective SEVs at these satellites.

This information is then used for the construction of the first-echelon routes in an analogous way as for the second echelon. After assigning all preassigned first-echelon customers to FEVs, the remaining ‘grey zone’ customers have to be dealt with. ‘Grey zone’ customers are inserted one by one randomly in the already built first-echelon or second-echelon routes at the best insertion position. If no feasible insertion position can be found in existing routes, a new first-echelon or second-echelon route is built depending on the insertion cost.

As soon as all ‘grey zone’ customers are routed, all dummy satellites are removed and the routes are improved by a local
search. Within each route a 2-opt operator is applied. The swap operator exchanges two nodes within the same route (intra-route) as well as between different routes of the same echelon (inter-route). As additional inter-route operator, a relocate operator is used, which moves one node to another route of the same echelon.

Then satellites are inserted into the second-echelon routes by a dynamic programming approach. After the satellite insertion route segments between satellites are improved by a 2optseg-operator applicable segment-wise. Based on the demand and arrival time at each satellite inserted in a second-echelon route, the respective satellites are then inserted into first-echelon routes by a simple
best-fit approach starting with the satellite with the earliest arrival time of a SEV at its cheapest position.

Whenever a FEV and a SEV, which are required to meet at a satellite, do not arrive there at exactly the same point in time, the vehicle which arrives first has to wait until the other arrives. If this waiting time exceeds a given limit (maximum waiting time) the respective insertion position for the satellite is infeasible and the next insertion position is checked. In case, no feasible insertion position can be found, an additional first-echelon route is created, which includes only the satellite. After each insertion of a satellite in a first-echelon route, the 2optseg-operator is applied to this route. In addition, all routes have to be updated with respect to demand, arrival time at satellites and total duration of the route, before the next satellite can be inserted. If the solution is feasible, it is used as initial solution for the metaheuristic; otherwise it is discarded and a new solution is built.

For further details on the local search operators as well as on the satellite insertion procedures we refer the interested reader to Anderluh et al. (2017).

5.2. Large neighborhood search

The initial solution $s$ is improved by a LNS heuristic based on the work by Pisinger and Ropke (2010) and Hemmelmayr, Cordeau, and Crainic (2012). The main principle of LNS is to remove part of the solution by different destroy-operators and put the respective nodes in a pool of nodes which needs to be reinserted into the remaining part of the solution. For this reinsertion, different repair-operators are used. We use 4 destroy-operators and 3 repair-operators, which are selected in each iteration of the LNS by a roulette wheel mechanism.

The destroy-operators are:

- **Random node removal**: q customer nodes from both echelons are removed at random
- **Worst node removal**: The q customer nodes with the highest removal savings are removed regardless the echelon they belong to
- **Random route removal**: One route from any echelon is removed at random
- **Max waiting removal**: The route with the longest waiting time at satellites is removed regardless the echelon it belongs to

The parameter $q$ is assumed as a random number with $q = U[\lfloor N_s/8 \rfloor, \lfloor N_s/4 \rfloor]$. These values are based on considerations that too few nodes removed at a single LNS step do not diversify the solution enough to escape local optima. On the other hand, too many nodes removed at a single LNS step cause a lot of additional computational time without contributing much to the solution quality.

The repair-operators are:

- **Greedy insertion**: One node after the other is inserted at its current best position; the order of insertion is chosen at random
- **Best insertion**: The best insertion position for each node is determined and the node with the lowest insertion cost is inserted at its best position
- **2-regret insertion**: The 3 best insertion positions are determined for each node and then the regret values between the best and the third best insertion position are calculated. The node with the highest regret value is inserted at its best insertion position

When using the repair operators, first-echelon and second-echelon customers can only be inserted in the respective echelon, whereas for ‘grey zone’ customers insertion positions in routes of both echelons are checked. After applying one destroy-operator and one repair-operator, the positions of the satellites are checked and adapted as necessary trying to get a feasible solution $s'$ with respect to the synchronization requirement. For this step a two-stage procedure as described in Section 5.1 is used. If $s'$ is a feasible solution the local search operators are applied to further improve $s'$, otherwise $s'$ is discarded.

If $s'$ is feasible and the objective function value $f(s')$ is better than the current best one $f(s^*)$, $s'$ is accepted in any case. If the objective function value of the new solution is worse, it is accepted based on the Metropolis criterion often used in simulated annealing, with an acceptance probability $p = e^{-\frac{f(s') - f(s^*)}{T}}$ (Aarts, Korst, & Michiels, 2014). $T$ represents the temperature which cools down by $T_{t+1} = \alpha T_t$, and $0 < \alpha < 1$ whenever a solution $s'$ is accepted. Whenever a solution $s'$ is accepted, it is used as a new starting solution $s = s'$ and the next iteration of LNS starts. This is repeated till the maximum number of LNS iterations $t_{\text{max}}$ is reached.

5.3. Heuristic rectangle/cuboid splitting

The HRS by Matl et al. (2019), which has already been mentioned in Section 2, has the advantage that it can be used with any heuristic, which is appropriate to address the problem at hand for one objective. In addition, this method allows to update the Pareto front with respect to intermediate solutions which are later dominated by new-found ones. HRS is based on the Box algorithm by Hamacher, Pedersen, and Ruzika (2007). It starts with solving the minimization problem at hand for objective $f_1$ without any constraint on $f_2$, which yields the upper left black square shown in Step 1 of Fig. 4. Then the other extreme value to the problem is determined by $(f_{\text{max}}^{f_1}, f_{\text{min}}^{f_2})$, which denotes the maximum value of $f_1$ and the minimum value of $f_2$. Now a rectangle is formed out of these two extreme values, represented by the grey rectangle in Step 1 of Fig. 4.

This rectangle can then be split by choosing the $\epsilon$-value as the value which halves the rectangle into a temporarily feasible half (the lower half in Step 1 of Fig. 4) and a temporarily infeasible half (the upper half in Step 1 of Fig. 4). The solution to the corresponding sub-problem with the additional constraint on the maximum value on $f_2$ represented by the chosen $\epsilon$-value (see red diamond in Step 1 of Fig. 4) allows to discard areas which are dominated by the solutions found or should not contain Pareto-optimal solutions anymore (see remaining grey rectangles in Step 2 of Fig. 4, which represent areas where Pareto-optimal solutions may still be found).

In each iteration the largest remaining rectangle is chosen to be split and the temporarily feasible (lower) half is tested for a Pareto-optimal solution and if one is found the rectangles are updated as described before (see Steps 3 + 4 of Fig. 4). This procedure is repeated till a termination criterion (e.g., the maximum number of sub-problems solved) is reached.

If no solution can be found to a sub-problem, the respective area is completely discarded (see Steps 5 + 6 of Fig. 4). If a solution is found which dominates a previous solution (this may happen if a heuristic is used to find solutions to the problem), the dominated one is removed from the set of Pareto-optimal solutions (see Steps 7 + 8 of Fig. 4). The light-grey rectangles in Step 8 of Fig. 4, which would have been added to the remaining rectangles by the original Box algorithm, are not included in the HRS, because such an enlargement could cause a kind of sub-cycling and so increase the run-time significantly.

For further details we refer the interested reader to Matl et al. (2019).

An advantage of HRS is that the algorithm quickly converges to a representative approximation of the complete set of Pareto-optimal solutions. Additionally, the difficulty of determining the value of the $\epsilon$-parameter is eliminated, because it is simply set to the value that halves the largest remaining rectangle.
In this paper, we extend the principle of HRS to be able to find a set of Pareto-optimal solutions $P(S_1)$ to a problem with three objectives. This yields a heuristic cuboid splitting (HCS).

The start is similar to the HRS. The solution $s'$ of the single-objective problem without any constraints for the other two objectives yields the upper, front, right point $(f_{1}^{\min}, f_{2}^{\max}, f_{3}^{\max})$ in the left part of Fig. 5. The lower, back, left point can be set to trivial values: It is assumed simply as $(M, 0, 0)$ (see Algorithm 1 lines 4–6). This means the maximum value for the main objective is assumed as a big enough constant, whereas the minimum values for the other two objectives are assumed to be 0. The such defined initial cuboid $r^*$ is halved in two directions (see right part of Fig. 5).

This splitting yields one feasible cuboid $r^*$ (the dark grey shaded cuboid in the left part of Fig. 6) and three infeasible ones (the light grey shaded cuboids in the left part of Fig. 6). Then the temporarily feasible cuboid $r''$ is checked for a feasible solution (see Algorithm 1 lines 7–11). If a feasible solution $s''$ is found, this solution is added to the Pareto front $P$ and the cuboids in $R$ are updated, i.e., cuboids which are dominated by the solution found are removed as well as the cuboid which would dominate this solution, as it is assumed that it is an efficient solution. Remaining parts are merged if they form a regular cuboid again (as shown in the right part of Fig. 6 and in lines 15–23 of Algorithm 1).

The procedure continues with the largest remaining cuboid $r'$ (see left part of Fig. 7 and Algorithm 1 line 9). If no feasible solution $s''$ can be found (see Algorithm 1 lines 12–14), the respective cuboid $r''$ is removed from the set of feasible cuboids $R$ (see right part of Fig. 7).

If a solution $s''$, which is later found in the solution procedure, dominates a previous potentially Pareto-optimal solution $s_i$, this solution is removed from the set and the set of cuboids is updated accordingly (see Algorithm 1 lines 17–21). The algorithm stops when the stopping condition is met (see Algorithm 1 line 24).

This can be assumed as a maximum number of iterations, a computational time limit or also a requirement on the minimum size of the largest remaining feasible region which has not been searched yet.

To conclude, the HCS works quite similarly to the above described HRS. Only the updating procedure of the remaining cuboids is a bit more demanding.

5.4. Calculation of GHG emissions

GHG emissions are assumed to be distance-based and have a direct component caused by burning fossil fuel and an indirect one caused by providing the required fuel or energy. Hence, also second-echelon vehicles cause a small amount of indirect GHG emissions, as they are assumed to be electrically assisted emission-free vehicles. For the calculation of GHG emissions of first-echelon

\begin{algorithm}
\caption{Heuristic cuboid splitting}
\begin{algorithmic}[1]
\State $P(S_1) \leftarrow \emptyset$ \Comment{\textit{i.e.} empty set of non-dominated solutions $\mathcal{P}$}
\State $R(p_i/q_i) \leftarrow \emptyset$ \Comment{\textit{i.e.} empty set of cuboids spanned by opposite edge points $\mathcal{R}$}
\State $i \leftarrow 0$
\State $s' \leftarrow \text{LNS}$ \Comment{\textit{i.e.} initial best solution by LNS for first objective without constraints on other objectives $\mathcal{S}$}
\State Add $s'$ to $P$
\State $r^* = (f_i(s')/(M, 0, 0))$
\State Add $r^*$ to $R$
\Repeat
\State $r' \leftarrow \text{max}(R)$ \Comment{\textit{i.e.} select largest cuboid in $R$ $\mathcal{R}$}
\State $r'' \leftarrow \text{feasible half} \triangleright \text{intermediate feasible part of cuboid $\mathcal{R}$}$
\State $s'' \leftarrow \text{LNS in $r''$} \triangleright \text{find best non-dominated solution in selected cuboid by LNS $\mathcal{S}$}$
\If{ $s'' = \emptyset$} \Comment{\textit{i.e.} no appropriate solution is found $\mathcal{S}$}
\State Remove $r'$ from $R$
\EndIf
\If{$s''$ is feasible}
\State Add $s''$ to $P$
\For{each $s_i \in P$}
\If{$s_i$ dominates $s_i$}
\State Remove $s_i$ from $P$
\EndIf
\EndFor
\EndIf
\Until{$i = \text{maxit}$} \Comment{\textit{i.e.} stopping condition is met $\mathcal{S}$}
\State return $P$
\end{algorithmic}
\end{algorithm}
Fig. 5. On the left: visualization of starting point \((f_{1}^{\text{min}}, f_{2}^{\text{min}}, f_{3}^{\text{min}})\) - black dot) and diagonally opposite extreme point \((M, 0, 0)\) spanning initial cuboid \(r^*\).
On the right: splitting cuboid \(r^*\) into 4 parts by halving it in both directions.

Fig. 6. On the left: intermediately feasible (dark grey shaded) cuboid \(r''\) and infeasible (light grey shaded) cuboids obtained by splitting initial cuboid into halves in two directions.
On the right: best feasible solution \(s''\) (red dot) to the sub-problem; determines which parts of remaining cuboids in \(R\) cannot contain non-dominated solutions anymore \(\rightarrow\) removed from set of cuboids; cuboids together forming a regular cuboid are combined (update of \(R\)).
*(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)*

Fig. 7. On the left: continuation of HCS procedure.
On the right: a cuboid \(r''\) is removed completely from \(R\) if no feasible solution can be found for this cuboid.
vehicles, the average fuel consumption is multiplied with the CO2-factor of the fuel used. Similarly, for the second-echelon vehicles the average electric power consumption is multiplied with the CO2-factor of the overall fuel mix. Data used for the calculations in this paper, are listed in Table 4 in the Appendix.

To reflect the increase in emissions caused by very low speeds (e.g., searching for a parking space, continuing the route after serving a customer) for every visit of a first-echelon customer, the amount of GHG emissions of a FEV is increased by an average amount of emissions of 0.5 (Barth & Boriboonsomsin, 2009).

5.5. Calculation of disturbance

In this paper, disturbance is assumed to be related to the number of affected people along a traversed path. Therefore, this objective can only be considered if appropriate data is available (in our case this holds only for the realistic test instance of Vienna, named vienna (see Section 6.1 for further details of this instance)). The calculation of disturbance along a path, which we suggest in this paper, can be seen as distance-based but weighted by the respective population density of each city district the respective path crosses (see different population density in different districts of Vienna in Fig. 13 in the Appendix).

In addition, selected points of interest (POIs) – in this paper, we consider schools and hospitals as such points – increase the number of people affected because delivery operations are mainly done during the daytime when large numbers of people are present at the above-mentioned POIs (see the locations of schools and hospitals in Vienna displayed in Fig. 13 in the Appendix).

For each single shortest path (based on the minimum distance for each echelon) between two nodes the number of people affected is calculated by multiplying the distance of each part of the path in a certain city district with the respective population density of this area. For each POI along this path an additional number of people affected is added (see Fig. 14 in the Appendix).

Finally, the total number of people affected along this path is multiplied by a certain impact factor for each echelon. This is due to the fact that the main disturbance factor – noise – is above all caused by FEVs, whereas SEVs hardly cause disturbance. Therefore, the impact factor of FEVs is assumed to be significantly higher (in our tests we assume a 30 times higher impact factor for FEVs) than that of SEVs. This assumption is based on the maximum noise level of light duty vehicles of currently 74 dB(A) (Parliament, 2014) and considerations that e-cargo bikes, for which no detailed information could be found, cause noise comparable to a talk in low voice of about 30 dB(A). As an increase of 10 dB(A) is equivalent to a doubling of loudness and we want to clearly express the difference in disturbance for the two echelons, we assume for our calculations that first-echelon vehicles have a 30 times higher negative impact than second-echelon vehicles.

6. Computational results

The algorithm was implemented in C/C++ and tested under Linux Ubuntu 16.04 LTS running on a virtual machine (using two processors and 2 gigabyte memory) on a host Intel(R) Core(TM) i5-3320 M CPU @ 2.60 gigahertz 4 gigabyte RAM.

6.1. Test instances

For our tests we consider three types of instances. First, we use a set of adapted Solomon instances (C101, C201, R101, R201, RC101, RC201) introduced in Anderluh et al. (2017). In these instances the original Solomon-depot is used as a second-echelon depot which is surrounded by a rectangle of 8 satellites. At the border of the area considered a first-echelon depot is added. This set is used for testing the quality of LNS (see Section 6.3) compared to the GRASP-PR method introduced in Anderluh et al. (2017).

Second, we consider the test instance vienna based on realistic data of Vienna and already used in Anderluh et al. (2017). Customer locations are based on the location of 100 randomly selected pharmacies. 18 satellites are chosen at appropriate places around the city center. The first-echelon depot represents the depot of one of the pharmacy wholesalers in Vienna and the second-echelon depot is assumed at an appropriate place in the first district of Vienna. Distances are based on the real street network, which in some cases yields different distances for FEVs and SEVs as they may use different types of streets. Travel times for FEVs (which are assumed to be vans) are based on floating car data provided by Taxi 31300 in cooperation with the Austrian Institute of Technology, whereas travel times for SEVs (which are assumed to be cargo bikes) are derived based on the distance and an average speed factor.

Third, we generate a set of test instances where the relative location of the city center surrounded by potential satellites within the city and its location with respect to the depot of the first echelon vary. In our tests we focus on three different settings (S). Setting A represents a city with the city center in the middle

Table 2

<table>
<thead>
<tr>
<th>instance</th>
<th>GRASP+PR</th>
<th>LNS</th>
<th>imp. LNS in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>st. dev.</td>
<td>st. dev. in %</td>
</tr>
<tr>
<td>C101</td>
<td>2289.41</td>
<td>33.93</td>
<td>1.48%</td>
</tr>
<tr>
<td>C201</td>
<td>1675.33</td>
<td>5.96</td>
<td>0.36%</td>
</tr>
<tr>
<td>R101</td>
<td>1016.42</td>
<td>3.50</td>
<td>0.34%</td>
</tr>
<tr>
<td>R201</td>
<td>712.77</td>
<td>1.57</td>
<td>0.22%</td>
</tr>
<tr>
<td>RC101</td>
<td>958.79</td>
<td>2.30</td>
<td>0.24%</td>
</tr>
<tr>
<td>RC201</td>
<td>730.95</td>
<td>3.61</td>
<td>0.49%</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>0.52%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3

Comparison of average total cost for solutions with preassigned customers compared to solutions with a ‘grey zone’

<table>
<thead>
<tr>
<th>instance</th>
<th>preassigned</th>
<th>‘grey zone’</th>
<th>improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>C101</td>
<td>2278.40</td>
<td>2125.58</td>
<td>-6.66%</td>
</tr>
<tr>
<td>C201</td>
<td>1608.60</td>
<td>1602.61</td>
<td>-0.37%</td>
</tr>
<tr>
<td>R101</td>
<td>971.75</td>
<td>859.49</td>
<td>-11.55%</td>
</tr>
<tr>
<td>R201</td>
<td>686.04</td>
<td>672.75</td>
<td>-1.94%</td>
</tr>
<tr>
<td>RC101</td>
<td>924.54</td>
<td>839.79</td>
<td>-9.17%</td>
</tr>
<tr>
<td>RC201</td>
<td>711.13</td>
<td>690.06</td>
<td>-2.96%</td>
</tr>
<tr>
<td>vienna</td>
<td>890.70</td>
<td>830.68</td>
<td>-6.74%</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td>-5.44%</td>
</tr>
</tbody>
</table>
of the city area and the first-echelon depot on the outskirts of the city. Settings B and C represent cities each with the city center at one border of the city (for example a seaside city) and the first-echelon depot either opposite on the other side of the city (setting B) or diagonally opposite (setting C). A graphical representation of these settings can be seen in Fig. 15 in the Appendix.

Customers are located (l) based on the principle used in the Solomon instances. Thus, we have either randomly located customers (r), clustered ones (c) or a combination of the former two (rc). The number of customers (n) is set to 100 and we generate 2 instances (z) of each type with \( z = 2 \) having a 50% increased second-echelon vehicle's capacity. This yields a total of 18 artificial test instances with the above-explained naming convention Snlz with \( S = \{ A, B, C \} \), \( n = 100 \), \( l = \{ r, c, rc \} \) and \( z = \{ 1, 2 \} \).

The demand of each customer is obtained by sampling a uniform distribution \( (d_i = U[1.8] \text{ for } i \in V_c) \) and the service time of each customer is obtained similarly as \( \lambda_i = U[1.20] \text{ for } i \in V_c \). The service time of satellites is set to \( \lambda_i = 10 \text{ for } i \in V_s \), whereas the service times of the first-echelon and second-echelon depots are assumed to be \( \lambda_i = 20 \) and \( \lambda_i = 0 \), respectively.

The capacity of a first-echelon vehicle is set to \( Q^1 = 100 \) and the capacity of a SEV is assumed as \( Q^2 = 16 \) in all instances with setting \( z = 1 \) and instance vienna. In all instances with setting \( z = 2 \), the SEV's capacity is set to \( Q^2 = 24 \).

Details of all Snlz-instances generated and instance vienna with respect to the number of first-echelon customers, second-echelon customers and 'grey zone' customers as well as the respective demand can be found in Table 5 in the Appendix.
6.2. Parameter setting

The parameters for LNS are determined based on pretests with the adapted Solomon instances described in Section 6.1 on the single-objective problem with cost as objective. The parameter values tested are based on all combinations with parameter $\alpha = \{0.99, 0.95, 0.90\}$, parameter $\beta = \{20, 30, 40\}$ and parameter $\gamma = \{20, 30, 40\}$.

Fig. 16 in the Appendix shows average total cost versus average computational time for all parameter combinations. The very left combination of parameters (shown in Fig. 16 in the Appendix) yields minimum total cost at a rather short computational time. Therefore, we use these parameter values for our computational tests. For the starting temperature $T_0 = \beta|V_c|$, the parameter $\beta$ is set to $\beta = 20$. The parameter $\alpha$ is set to $\alpha = 0.90$ and for the maximum number of iterations $t_{i_{\text{max}}}$ we assume $t_{i_{\text{max}}} = \gamma|V_c|$ with $\gamma = 30$.

6.3. Quality of LNS

Pretests with the six adapted Solomon instances and five runs for each instance regarding the performance of the LNS-operators summarized in Table 6 in the Appendix show that the two random removal operators are significantly more often involved in finding a new best solution, with the random node removal being clearly the one most often involved in finding a new best solution. In contrast to that, the three repair-operators behave rather similarly in this respect.

Considering computational time, our pretests show that the node-based destroy-operators require approximately twice the
computational time than the route-based ones. Regarding the repair-operators it becomes obvious that the greedy one is the fastest and the other two need approximately three times the computational time than the former one.

The quality of our LNS method is then tested in comparison with the GRASP+PR method used in Anderluh et al. (2017). We compare results of the single-objective problem without a ‘grey zone’. The GRASP method, which has already been described in Section 5.1, is used to build a starting pool of solutions. Then in the PR-step, two solutions out of this pool are chosen and the first (initial) solution is transformed step by step into the second (guiding) solution. After each step the intermediate solution is checked if it is fit to be included into the pool of solutions. The interested reader is referred to Anderluh et al. (2017) for further details.

Computational results for the Solomon instances show that LNS can improve the solution quality compared to the one by GRASP+PR by 3.15% on average. Detailed results are represented in Table 2, with the far right column showing the relative improvement in mean objective values of LNS compared to GRASP+PR. Also the variation on obtained results expressed by standard deviation can be reduced by our LNS (see respective columns in Table 2 for detailed results).

6.4. Impact of the ‘grey zone’

The impact of the ‘grey zone’ in comparison to fully preassigned customers is tested on the adapted Solomon instances for a single objective. Results in Table 3 point out an improvement in to-
The results in Table 3 underline the improvement in the solution quality by the usage of a 'grey zone' when considering total cost as a single objective.

Figs. 8–11, described in detail in Section 6.5, show the impact of a 'grey zone' instead of preassigned customers also for two objectives. If this solution is along the Pareto front, it is represented by a green triangle, otherwise it is represented by a red diamond.

Our tests show that only for 6 out of the 20 calculated Pareto fronts the solution found when all customers are preassigned to either the first or the second echelon (as shown in Fig. 2) is nondominated (instances A100cl1, A100r1 and A100rc2 in Fig. 8; instances B100r2 and B100rc2 in Fig. 9; instance C100r2 in Fig. 10). Thus, especially for instances in which customers are not or only partly clustered, the complete pre-assignment yields rather nondominated solutions. The more clustered the customers are, the worse is the solution obtained by complete pre-assignment of customers. For instance vienna, this solution is rather far away from the Pareto front (see Fig. 11).

To conclude, especially for instances with clustered customers (see Figs. 8–11), using a ‘grey zone’ can improve the solution quality compared to solutions found for the same instances with complete pre-assignment of all customers.

6.5. Impact of the city layout

For testing the impact of the city layout in a multi-objective setting we focus on total cost as the first objective of the optimization problem and GHG emissions or disturbance as the second objective (details on the calculation of GHG emissions and disturbance factors can be found in Sections 5.4 and 5.5, respectively).

As the social objective can only be realistically assessed if data for population density per district as well as points of interest are available, this objective is only considered for instance vienna (conducted tests with randomly generated disturbance values for the artificial test instances provided no meaningful insights).

The Pareto fronts of all test instances are shown in Figs. 8–11. Data labels in the figures give the number of customers served by a second-echelon vehicle and the number of synchronized transshipment meetings required for the respective solution. For better comparability of the results all objective values are normalized.

In general, for all instances ‘100’2, non-dominated solutions with more customers served by second-echelon vehicles can be found. This is due to the fact that for these instances the capacity of the second-echelon vehicles is assumed to be 50% increased compared to the second-echelon vehicles’ capacity in instances of type ‘100’1.

A special case is instance B100r1 (see middle left part of Fig. 9), for which no Pareto front could be found in our tests. This could be due to the fact that the reduction in GHG emissions for serving more customers by second-echelon vehicles is compensated for by additional necessary synchronized transshipment meetings with first-echelon vehicles. Thus, for this instance there seems to be no trade-off between the economic and the environmental objectives. Regarding the city layout, it can be seen from Figs. 8–10 that in general settings A and B are more appropriate for testing a ‘grey zone’ and a multi-objective setup than setting C. This effect is increased when customers are clustered.
A comparison of the Pareto front of the economic and the environmental objectives (see left part of Fig. 11) with the Pareto front of the economic and the social objectives of instance vienna (see right part of Fig. 11) shows significantly fewer non-dominated solutions for the latter one. So, not every Pareto-optimal solution with respect to the economic and the environmental objectives is also efficient when considering the social objective.

The interested reader is referred to the tables available as ‘Supplementary Materials’ for detailed results of all Pareto-optimal solutions.

6.6. Pareto surface for three objectives

When focusing on all three objectives – the economic, the environmental and the social one – the resulting set of non-dominated solutions forms a Pareto surface. The approximated Pareto-set of solutions for instance vienna obtained by LNS embedded in the heuristic cuboid splitting described in Section 5.3 is shown in Fig. 12. When considering all three objectives, the non-dominated solutions found for instance vienna (see Fig. 12) underline the impact of considering all three objectives at once. The Pareto-set forms a kind of a ‘bowl’, such that solutions which are very cost-efficient cause a rather high amount of emissions but are nevertheless good regarding disturbance. On the other side of the ‘bowl’ (upper right point in Fig. 12), the solution found is very good with respect to emissions, but costly and causing a high amount of disturbance. Solutions with lowest disturbance show an average performance regarding cost and emissions. Hence, this Pareto surface provides decision support in all three aspects of sustainability.

7. Conclusion

In this paper, we deal with the MO2eVRPSynGZ. The assignment to echelons of customers located in the ‘grey zone’, which is the zone between the inner-city center and the area around it, is part of the model and is done in the solution procedure of the optimization problem.

In the model we focus on the economic objective expressed in costs as first objective. Additional objectives of citizens and municipal authorities are included by considering negative external effects of freight transport such as GHG emissions as environmental objective and disturbance of citizens, caused by noise and congestion, as social objective.

Solutions for the formulated multi-objective optimization problem are found by LNS embedded in an \( \epsilon \)-constraint method – the heuristic rectangle splitting. The latter method is extended in this paper to a heuristic cuboid splitting which is able to deal with three objectives.

Computational results for artificial test instances and a realistic instance of Vienna show that the quality of the provided LNS can improve results significantly compared to GRASP+PR, the application of the ‘grey zone’ concept can improve the solution quality of the single-objective problem and also of the approximated Pareto-set of solutions in the multi-objective case, especially if customers are clustered. Furthermore, our results emphasize that the city layout has to be considered in the decision process if the provided freight distribution scheme should be established, because especially for cities in which the city center is located in the middle of the city this scheme turned out to be advantageous. Finally, the economic, the environmental and the social objectives can be considered at the same time by the provided heuristic cuboid splitting, which yields a Pareto surface and underlines the trade-off between all three objectives.

Because of lacking data, disturbance as additional objective to economic costs and GHG emissions was considered only for the realistic instance of Vienna. Therefore, future work has to focus on additional tests with all three objectives and, hence, realistic test instances for other cities with data about population density as well as points of interest are required to further confirm the findings of this paper.

Acknowledgments

Part of this work was funded by the Austrian Research Promotion Agency as part of the JPI CONCOORD project (FP7 project No. 839739). Furthermore, we gratefully acknowledge the travel time data provided by Taxi 31300, as well as the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through its Discovery Grants program.

Appendix

Table 4 Data for the calculation of GHG emissions.

<table>
<thead>
<tr>
<th></th>
<th>First echelon (Light duty vehicle [3.5t])</th>
<th>Second echelon (eBullitt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average fuel</td>
<td>0.085 liter per kilometer</td>
<td>10.45 watt-hour per kilometer</td>
</tr>
<tr>
<td>consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO2-factor</td>
<td>0.52 (status 2017)</td>
<td>5 gram per kilometer</td>
</tr>
<tr>
<td>CO2-equivalent</td>
<td>279.4 gram per kilometer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Umweltbundesamt (n. d.)³</td>
<td></td>
</tr>
</tbody>
</table>

⁴ Provided numbers are subject to regular updates, previous data is available on request.

Table 5 Instance characteristics.

| Instance | | \( \sum_{i} d_{Vi} \in V_{i} \) | | \( \sum_{i} d_{Vi} \in V_{i} \) | | \( \sum_{i} d_{Vi} \in V_{i} \) | | \( \sum_{i} d_{Vi} \in V_{i} \) |
|----------|----------------|----------------|----------------|----------------|----------------|----------------|
| A100c1   | 31             | 150            | 65             | 270            | 4              | 21             |
| A100c2   | 39             | 190            | 56             | 264            | 5              | 31             |
| A100r1   | 14             | 61             | 81             | 388            | 5              | 23             |
| A100r2   | 22             | 97             | 73             | 359            | 5              | 13             |
| A100rc1  | 24             | 111            | 73             | 311            | 3              | 13             |
| A100rc2  | 17             | 71             | 74             | 340            | 9              | 46             |
| B100c1   | 31             | 126            | 60             | 260            | 9              | 31             |
| B100c2   | 37             | 178            | 55             | 248            | 8              | 42             |
| B100c1   | 18             | 82             | 73             | 339            | 4              | 17             |
| B100c2   | 18             | 79             | 80             | 367            | 2              | 14             |
| C100c1   | 21             | 101            | 71             | 317            | 8              | 32             |
| C100c2   | 18             | 95             | 76             | 325            | 6              | 28             |
| vienna   | 54             | 222            | 41             | 190            | 5              | 10             |

Table 6 Performance of LNS operators.

<table>
<thead>
<tr>
<th>Operator</th>
<th>New best solution found</th>
<th>Average computational time for execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>random node removal</td>
<td>54.79%</td>
<td>0.00091 second</td>
</tr>
<tr>
<td>worst node removal</td>
<td>8.68%</td>
<td>0.00099 second</td>
</tr>
<tr>
<td>random route removal</td>
<td>27.68%</td>
<td>0.00046 second</td>
</tr>
<tr>
<td>max waiting removal</td>
<td>8.85%</td>
<td>0.00052 second</td>
</tr>
<tr>
<td>greedy insertion</td>
<td>36.13%</td>
<td>0.037 second</td>
</tr>
<tr>
<td>best insertion</td>
<td>30.25%</td>
<td>0.099 second</td>
</tr>
<tr>
<td>2-regret insertion</td>
<td>33.62%</td>
<td>0.106 second</td>
</tr>
</tbody>
</table>
Fig. 13. Population density of city districts provided by the municipality of Vienna, location of POIs considered (schools and hospitals) and nodes of instance vienna.

Fig. 14. Basis for calculation of disturbance of a certain path of instance vienna; path is divided into segments based on city districts.
**Supplementary material**

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2019.07.049.

**References**


