Networks and Spatial Economics (2022) 22:33–64
https://doi.org/10.1007/s11067-021-09553-6

A Granular Local Search Matheuristic for a Heterogeneous Fleet Vehicle Routing Problem with Stochastic Travel Times

Ramon Faganello Fachini[1] · **Vinícius Amaral Armentano**[1] · **Franklina Maria Bragion Toledo**[2]

Accepted: 16 August 2021 / Published online: 24 January 2022

Abstract
This paper addresses a multi-attribute variant of the vehicle routing problem which encompasses a heterogeneous fixed fleet, flexible time windows and stochastic travel times. The objective is to minimize the sum of the transportation and the service costs. The former comprises the vehicle fixed costs and route variable costs, and the latter corresponds to the penalty costs for violating customer time windows. The problem is formulated as a two-stage stochastic mixed-integer program with recourse and solved by a granular local search matheuristic. The stochastic travel times are approximated by a finite set of scenarios generated by Burr type XII distribution. Extensive computational tests are performed on 216 benchmark instances, and the advantages of both flexible windows and stochastic travel times are stressed. The experiments show that, compared to a state-of-art mathematical programming solver, the developed matheuristic found better solutions in 81% of the instances within shorter computational times. The proposed solution method also far outperformed an alternative decomposition algorithm based on the augmented Lagrangian relaxation. Furthermore, the flexible time windows yielded overall cost savings for 68% of the instances compared to the solutions obtained for hard time window problems. Finally, explicitly modeling the stochastic travel times provided 66% more feasible solutions than the adoption of a deterministic model with the random parameters fixed at their expected values.

Keywords Vehicle routing problem · Heterogeneous fleet · Flexible time windows · Stochastic travel times · Matheuristic · Granular local search

✉ Ramon Faganello Fachini
ramonfachini@gmail.com

Extended author information available on the last page of the article

1 Introduction

The freight distribution industry is not only essential to modern society but also responsible for significant revenues. For example, Fedex Ground reports revenue of 18.395 billion US dollars stemming from the pickup and delivery of small packages in the United States and Canada in 2018 (Fedex 2018). The vehicle routing problem (VRP) lies at the center of this industry and focuses on providing a minimum-cost service to a set of geographically dispersed customers by employing a limited and capacitated fleet of vehicles that is located at a central depot. Each customer can only be visited once by a single vehicle and all routes must start and finish at the depot (Dantzig and Ramser 1959). Given the VRP relevance, academics have developed a large amount of work targeting this problem and its extensions (Braekers et al. 2016; Golden et al. 2008; Toth and Vigo 2014).

The VRP becomes more realistic when the fleets are heterogeneous, i.e., vehicles differ in their capacities and costs (Taillard 1999). Hoff et al. (2010) analyze industrial aspects of heterogeneous fleets and stress that vehicles usually present different characteristics in practice since fleets are acquired over long periods and automotive technology continuously evolves. Moreover, companies often keep a varied set of vehicle types due to the need for flexibility. For thorough reviews on heterogeneous fleet VRP (HFVRP), see Baldacci et al. (2008) and Koç et al. (2016).

Another important feature of the VRP is the time window constraints that impose a time interval for collecting or delivering of goods at each customer (Desaulniers et al. 2014). The literature classifies these constraints as hard or soft. In the first case, vehicles can arrive early and wait, but the service must start within the time windows. On the other hand, the second case represents a more realistic situation in which the time windows can be violated at penalty costs for early or late customer service. A particular case of the VRP with soft time windows (VRPSTW), which intensifies its practical perspective, is the VRP with flexible time windows (VRPFlexTW). In this problem, introduced by Taş et al. (2014c), early and late service deviations at each customer are bounded by an outer or flexible time window generated with respect to the length of the original or hard time window. The carrier pays a penalty cost only if the arrival time takes place outside the hard time window, but within its deviation bounds. The interested reader is referred to Salani and Battarra (2018) for a recent survey of the VRPSTW.

Both HFVRP and VRPFlexTW assume that all the information necessary to formulate the problems is known and readily available. However, in real-life applications, there are situations where the parameters of such problems have a stochastic nature (Gendreau et al. 2014). In particular, travel times are most subject to uncertainty caused by weather conditions, car accidents and traffic congestion (Gendreau et al. 2016). This uncertainty entails further feasibility conditions and additional costs. Therefore, ignoring the stochastic nature of travel times may result in arbitrarily bad VRP solutions.

Despite the practical relevance of the previously mentioned VRP attributes, namely heterogeneous fleets, flexible time windows and stochastic travel times, no research has simultaneously addressed all of them. The first HFVRP variant that incorporates flexible time windows is proposed by Firouzi et al. (2018), whereas the only stochastic version of such a problem is the mixed fleet stochastic VRP investigated by Teodorović et al. (1995). The latter involves an unlimited fleet of heterogeneous vehicles and customers with stochastic demands.

Apart from these studies, the closest related literature focuses on the VRPSTW under travel time uncertainty. Ando and Taniguchi (2006) suggest a model for this problem where each vehicle can make multiple routes per day within a given scheduling horizon. The objective is to minimize the sum of vehicle fixed costs, operating costs and expected earliness and lateness costs. Triangular distributions are estimated by using real data to describe the stochastic travel times, and a genetic algorithm is devised to handle the problem.

Russell and Urban (2008) deal with a VRPSTW in which travel times are random variables modeled with the Erlang distribution, a special case of the gamma distribution with an integer shape parameter. The authors minimize a weighted average of three objectives, specifically: the number of vehicles employed, the total distance traveled, and the expected earliness and lateness penalties, calculated by closed-form expressions that reflect constant, linear or quadratic costs. The problem is solved in three phases. The first two phases build an initial solution and improve it through tabu search algorithms, respectively, while the third optimizes the waiting time before each customer using a generalized reduced gradient method.

Li et al. (2010) propose two formulations of a VRP variant in which both travel and service times are normally distributed random variables. The first includes hard time windows modeled as chance constraints, whereas the second considers soft time windows and corresponds to a two-stage stochastic program with recourse which designs routes in the first stage aiming to minimize the expected costs of driver's remuneration and late arrivals in the second stage. These models are solved by a tabu search metaheuristic that incorporates a Monte Carlo simulation procedure to estimate the expected values and probabilities of the stochastic parameters. The assumption that travel times follow a normal distribution is also employed by Thompson et al. (2011). The authors present alternative formulations based on stochastic programming and robust optimization of the VRPSTW under travel time uncertainty. Because the former requires expensive numerical integrations, the latter is used in a case study conducted by the authors.

Another VRPSTW with stochastic travel and service times is analyzed by Zhang et al. (2013). They suggest a new stochastic programming model that incorporates service level constraints to ensure a minimum on-time arrival probability at each customer and a hierarchical objective function to minimize three components, namely: vehicle fixed costs, mean total travel time, and the weighted sum of costs stemming from earliness, lateness and route duration excess. An approximation method called α-discrete, which estimates the vehicle arrival time distributions at customers, is embedded in an iterated tabu search algorithm to solve the problem. Experiments were conducted with travel and service times following lognormal and normal distributions, respectively.

The more recent studies on VRPSTW with stochastic travel times are from Taş and colleagues (Taş et al. 2013, 2014a, b). Taş et al. (2013) develop a three-phase method. An initialization algorithm obtains a starting feasible solution in the first phase. A tabu search metaheuristic improves such a solution in the second phase, and a post-optimization procedure is called in the third phase to adjust the departure times of vehicles from the depot to minimize penalties due to time window violations. The same problem is exactly solved by Taş et al. (2014b) through a branch-and-price solution approach. Taş et al. (2014a) further extended the VRPSTW to incorporate time-dependent and stochastic travel times. They adapt the three-phase method given in Taş et al. (2013) and implement an adaptive large neighborhood search metaheuristic to tackle such a problem. In these three works, travel times are assumed to be gamma distributed.

This paper deals with the industrially relevant variant of VRP called heterogeneous fixed fleet VRP with flexible time windows and stochastic travel times (HFFVRP-FlexTW-STT). The objective is to minimize the sum of the transportation costs and service costs. The transportation costs comprise the vehicle fixed costs and route variable costs, while service costs correspond to the penalty costs for violating customer time windows. These two costs provide an easy way of exploring the trade-offs between the expenses of the carrier company and the customer service reliability as pointed out by Taş et al. (2013).

The main contributions of this paper are fivefold: i) to the best of our knowledge, the proposition of the first mathematical formulation for the HFFVRP-FlexTW-STT; ii) the development of a two-stage stochastic mixed-integer program with recourse, where the assignment of customers to vehicles make up the first stage, and recourse decisions are made in the second stage to find minimum-cost routes for vehicles according to observed travel times; iii) the suggestion of a scenario generation procedure which describes stochastic travel times using the Burr type XII distribution (Burr 1942), which has been shown to represent variations in day-to-day travel times better than other distributions such as lognormal, gamma, Weibull and normal (Susilawati et al. 2013; Taylor 2017); iv) the design of a novel granular local search matheuristic to solve the problem; v) extensive computational experimentation on benchmark instances and the assessment of benefits gained by flexible windows and stochastic travel times.

The computational experiments show that our matheuristic outperforms a state-of-art mathematical programming solver in terms of solution quality and runtime. The proposed solution method also overcomes an alternative heuristic algorithm based on the augmented Lagrangian relaxation. Furthermore, the results underline the advantages of both flexible windows and stochastic travel times. The former drastically reduce the solution's overall cost, whereas the latter are critical for their feasibility in a stochastic environment.

The remainder of the article is organized as follows. Section 2 introduces the problem description. Section 3 describes the granular local search matheuristic. Computational results are reported in Sect. 4, and conclusions are outlined in Sect. 5.

2 Problem Description

The HFFVRP-FlexTW-STT is defined on a complete graph $G = (\mathcal{N}, \mathcal{E})$ where the node set $\mathcal{N} = \{0,1,..,n\}$ consists of the depot $\{0\}$ and the set of customers $\mathcal{C} = \mathcal{N}\backslash\{0\}$, while the set $\mathcal{E} = \{(i,j) : i,j \in \mathcal{N}, i \neq j\}$ represents the edges between the nodes. A heterogeneous fleet of vehicles is positioned at the depot in order to supply the customers. The set $\mathcal{M} = \{1,...,m\}$ represents the m distinct types of vehicles. The set $\mathcal{M}_r = \left\{1,...,m_r\right\}$ represents the vehicles available at the depot of type $r \in \mathcal{M}$, each one having capacity Q_r and an associated fixed cost f_r. For each edge $(i,j) \in \mathcal{E}$ and for each vehicle type $r \in \mathcal{M}$, a deterministic travel cost $c_{ij}^r = c_{ji}^r$ is given for which we assume that the triangular inequality holds. Furthermore, let $\hat{t}_{ij}^r, (i,j) \in \mathcal{E}, r \in \mathcal{M}$, be stochastic travel times following a known probability distribution.

Each node $i \in \mathcal{N}$ presents a demand q_i, a service time W_i, a time window $\left[e_i, l_i\right]$ and associated fractions f_i^e and f_i^l used to set the early and late service deviation bounds, respectively. The time window $\left[e_0, l_0\right]$ at the depot represents the planning horizon, and $q_0 = W_0 = 0$ by definition. Following Taş et al. (2014c), we define flexible time windows $\left[e'_i, l'_i\right]$ for each $i \in \mathcal{N}$ with limits expressed by $e'_i = \max\left\{e_i - f_i^e\left(l_i - e_i\right), 0\right\}$ and $l'_i = \min\left\{l_i + f_i^l\left(l_i - e_i\right), l_0 + f_0^l\left(l_0 - e_0\right)\right\}$. Servicing a customer within $\left[e'_i, e_i\right]$ is penalized by δ_α for one unit of earliness, while servicing a customer within $\left[l_i, l'_i\right]$ is penalized by δ_β for one unit of lateness.

Define the binary variables $y_i^{rk}, i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r$, such that, if the demand of node $i \in \mathcal{N}$ is satisfied by vehicle $r^k, r \in \mathcal{M}, k \in \mathcal{M}_r$, and $y_i^{rk} = 0$, otherwise. The decisions regarding the assignment of customers to available vehicles can formulated as follows:

$$min \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} f_r y_0^{rk} + E_\epsilon \mathcal{Q}(\mathbf{y},\mathbf{t}(\omega)) \tag{1}$$

$$\sum_{i\in\mathcal{C}} q_i y_i^{rk} \leq Q_r y_0^k \quad r \in \mathcal{M},\ k \in \mathcal{M}_r \tag{2}$$

$$\sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} y_i^{rk} = 1 \quad i \in \mathcal{C} \tag{3}$$

$$\sum_{k\in\mathcal{M}_r} y_0^{rk} \leq m_r \quad r \in \mathcal{M} \tag{4}$$

$$y_i^{rk} \in \{0,1\} \quad i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r \tag{5}$$

where ε is a random vector that defines the travel time distributions, $\omega \in \Omega$ is an outcome in sample space Ω of random events, and $\mathcal{Q}(\mathbf{y},\mathbf{t}(\omega))$ is equal to the costs of routes that vehicles will follow to service their assigned customers given realized travel times $\mathbf{t}(\omega)$. The objective function (1) minimizes the carrier's overall cost, including vehicle fixed costs $\sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} f_r y_0^{rk}$ as well as expected traveling costs and penalty costs for violating customer time windows $E_\epsilon \mathcal{Q}(\mathbf{y},\mathbf{t}(\omega))$. Constraints (2) are the vehicle capacity constraints, while constraints (3) require the assignment of a single vehicle to each customer. Constraints (4) and (5) limit the number of vehicles that leave the depot for each type r and impose the integrality requirements on assignment variables, respectively.

Model (1) – (5) is a two-stage stochastic mixed-integer program with recourse (Birge and Louveaux 2011) based on the cluster-first and route-second approach (Fisher and Jaikumar 1981). In the first stage, customers are grouped in clusters and assigned to vehicles before the stochastic travel times being revealed. After these times are realized, recourse decisions related to the route of each vehicle are made in the second stage. The second stage recourse function can be formulated as a set of independent routing problems over the customers assigned to each vehicle, presenting the special structure of the traveling salesman problem with flexible time windows (TSPFlexTW) (Fachini and Armentano 2020a):

$$\mathcal{Q}(\mathbf{y},\mathbf{t}(\omega)) = \min \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}, j\neq i} \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} c^r_{ij} x^{rk}_{ij} + \sum_{i\in\mathcal{N}} \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} \left(\delta_\alpha \alpha^{rk}_i + \delta_\beta \beta^{rk}_i\right) \tag{6}$$

$$\sum_{i\in\mathcal{N}} x^{rk}_{ij} = y^{rk}_j \quad j\in\mathcal{N}, j\neq i, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{7}$$

$$\sum_{j\in\mathcal{N}} x^{rk}_{ij} = y^{rk}_i \quad i\in\mathcal{N}, i\neq j, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{8}$$

$$w^{rk}_i + W_i + t^r_{ij}(\omega) - [M^r_{ij}(\omega)(1 - x^{rk}_{ij})] \leq w^{rk}_j \; i,j\in\mathcal{N}, i\neq j, j\neq 0, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{9}$$

$$e'_i \leq w^{rk}_i \leq l'_i \quad i\in\mathcal{N}, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{10}$$

$$\alpha^{rk}_i \geq e_i - w^{rk}_i \quad i\in\mathcal{N}, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{11}$$

$$\beta^{rk}_i \geq w^{rk}_i - l_i \quad i\in\mathcal{N}, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{12}$$

$$\left(w^{rk}_i, \alpha^{rk}_i, \beta^{rk}_i\right) \in \mathbb{R}_+, \left(x^{rk}_{ij}, y^{rk}_i\right) \in \{0,1\} \; i,j\in\mathcal{N}, i\neq j, r\in\mathcal{M}, k\in\mathcal{M}_r \tag{13}$$

where x^{rk}_{ij} is a binary variable such that $x^{rk}_{ij} = 1$, if vehicle $r\in\mathcal{M}, k\in\mathcal{M}_r$ travels directly from node $i\in\mathcal{N}$ to node $j\in\mathcal{N}, i\neq j$, otherwise, $x^{rk}_{ij} = 0$, and continuous variables $\left(w^{rk}_i, \alpha^{rk}_i, \beta^{rk}_i\right)$ denote service start time, earliness and lateness at node $i\in\mathcal{N}$ when serviced by vehicle $r^k, r\in\mathcal{M}, k\in\mathcal{M}_r$, respectively. The objective function (6) minimizes the sum of traveling costs, called route variable costs, and service penalty costs. Constraints (7) and (8) guarantee that every node is visited exactly once when $y^{rk}_j = 1$ and $y^{rk}_i = 1$. Constraints (9), in which $M^r_{ij}(\omega) = \max\left\{l'_i - e'_j + t^r_{ij}(\omega) + W_i, 0\right\}$ (see Desrosiers et al. (1995)), ensure that the service start time at a customer must be greater or equal to the sum of the service start time with the service time and the traveling time of its immediate predecessor. Constraints (10) impose the flexible time windows at nodes, while constraints (11) and (12) connect their service start times with the service earliness and lateness, respectively. Constraints (13) indicate the domain of the variables.

Let $\mathcal{S} \subseteq \Omega$ define a set of scenarios for the random event obtained by a sampling procedure. One can approximate formulation (1) – (13) through the following multi-scenario deterministic model (MSDM):

$$MSDM = \min \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} f_r y_0^{rk} + \sum_{s\in\mathcal{S}} Pr(s) \Bigg(\sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}, j\neq i} \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} c_{ij}^r x_{ij}^{rks} + \sum_{i\in\mathcal{N}} \sum_{r\in\mathcal{M}} \sum_{k\in\mathcal{M}_r} \left(\delta_\alpha \alpha_i^{rks} + \delta_\beta \beta_i^{rks}\right) \Bigg) \tag{14}$$

$$Constraints\ (2) - (4)$$

$$\sum_{i\in\mathcal{N}} x_{ij}^{rks} = y_j^{rk} \quad j \in \mathcal{N}, j \neq i, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{15}$$

$$\sum_{j\in\mathcal{N}} x_{ij}^{rks} = y_i^{rk} \quad i \in \mathcal{N}, i \neq j, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{16}$$

$$w_i^{rks} + W_i + t_{ij}^{rs} - [M_{ij}^{rs}(1 - x_{ij}^{rks})] \leq w_j^{rks} i, j \in \mathcal{N}, i \neq j, j \neq 0, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{17}$$

$$e_i' \leq w_i^{rks} \leq l_i' \quad i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{18}$$

$$\alpha_i^{rks} \geq e_i - w_i^{rks} \quad i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{19}$$

$$\beta_i^{rks} \geq w_i^{rks} - l_i \quad i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{20}$$

$$\left(w_i^{rks}, \alpha_i^{rks}, \beta_i^{rks}\right) \in \mathbb{R}_+, \left(y_i^{rk}, x_{ij}^{rks}\right) \in \{0, 1\}\, i, j \in \mathcal{N}, i \neq j, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S} \tag{21}$$

where $Pr(s)$ defines the probability of scenario $s \in \mathcal{S}$ and $M_{ij}^{rs} = \max \left\{ l'_i - e'_j + t_{ij}^{rs} + W_i, 0 \right\}$. Note that the second stage variables $\left(x_{ij}^{rks}, w_i^{rks}, \alpha_i^{rks}, \beta_i^{rks}\right)$ and the, now deterministic, travel times t_{ij}^{rs} are scenario specific. MSDM is a large-scale mixed-integer programming (MIP) model whose solution provides a customer-vehicle assignment $\mathbf{y}$ that minimizes vehicle fixed costs as well as expected route variable costs and penalty costs.

3 Granular Local Search Matheuristic for the HFFVRP-FlexTW-STT

We propose a solution method for the HFFVRP-FlexTW-STT based on the local search matheuristic devised by Agra et al. (2016, 2018) to handle stochastic production and inventory routing problems. Following the local branching concept of Fischetti and Lodi (2003), such a procedure iteratively tries to improve the best current solution for a stochastic MIP (SMIP) by searching a neighborhood induced by the addition in SMIP of general linear inequalities called local branching constraints.

The main framework of the matheuristic of Agra et al. (2016, 2018) is summarized in Fig. 1. Initially, a simplified version of the SMIP is solved by considering a single scenario in which all stochastic parameters are set to their expected values (line 1). Once a starting solution has been obtained, a local branching constraint is added to the SMIP with all scenarios (line 3), and the resulting model is solved by a MIP solver within a time limit of LS_time seconds (line 4). If a better solution is found, the current solution is updated (line 5) and the procedure is repeated, otherwise, the local search stops.

The algorithm developed in this section modifies the one of Agra et al. (2016, 2018) through two significant adaptations: (i) additional steps are performed to generate the starting solution; and (ii) the local search moves are restricted to a list of promising neighbors using the reduction technique of granular neighborhoods (Toth and Vigo 2003). These adaptations are included in our heuristic to address difficulties of HFFVRP-FlexTW-STT not present in problems studied by Agra et al. (2016, 2018), namely the absence of complete or relatively complete recourse and the presence of second stage discrete variables. In the following sections, the local branching constraints are defined first, followed by the description of the aforementioned adaptations.

3.1 Local Branching Constraints

Assume that $\bar{\mathbf{y}}$ represents the values of the first stage variables $\mathbf{y}$ in a given reference solution of MSDM and, for each node $i \in \mathcal{N}$, let $\bar{S}_i = \left\{(r,k) \in \mathcal{B} : \bar{y}_i^{rk} = 1\right\}$ denote the binary support of $\bar{\mathbf{y}}$. We define the following local branching constraint for the HFFVRP-FlexTW-STT:

$$\Delta\left(\mathbf{y}, \bar{\mathbf{y}}\right) \leq \kappa \tag{22}$$

where $\Delta(\mathbf{y}, \bar{\mathbf{y}})$ is the Hamming distance function expressed by

Algorithm Local search matheuristic

1 Solve a simplified SMIP considering a single scenario in which all stochastic parameters are set to their expected values
2 **Repeat**
3 Add a local branching constraint to the SMIP with all scenarios
4 Solve the resulting model using a MIP solver for LS_time seconds
5 Update the current solution
6 **Until** no improvement in the objective function is observed

Fig. 1 Local search matheuristic

$$\Delta(\mathbf{y}, \bar{\mathbf{y}}) = \sum_{(r,k)\in\bar{S}_i} \sum_{i\in N} \left(1 - y_i^{rk}\right) + \sum_{(r,k)\in\mathcal{B}\setminus\bar{S}_i} \sum_{i\in N} \left(y_i^{rk}\right) \tag{23}$$

and κ is an integer positive parameter that defines the neighborhood size. The latter should be calibrated to yield a neighborhood sufficiently small to be explored within reasonable computing time, but large enough to improve the best current solution.

3.2 Starting Solution

As in Agra et al. (2016, 2018), a starting solution is generated by solving the model MSDM considering a single scenario s^* in which the stochastic travel times $t_{ij}^{rs^*}$ are set to their expected values $E\left(\hat{t}_{ij}^{r}\right)$. The resulting problem is a deterministic heterogeneous fixed fleet VRP with flexible time windows (HFFVRP-FlexTW), which is solved by the large neighborhood search (LNS) matheuristic proposed by Fachini and Armentano (2020b). In this method, a feasible solution is obtained by an extension of the Solomon's I1 heuristic (Solomon 1987), and further improved by the alternated application of destruction and reconstruction procedures. The former comprises four customer removal operators devised by Prescott-Gagnon et al. (2009), whereas the latter rebuilds ruined solutions using a logic-based Benders decomposition (LBBD) method (Hooker and Ottosson 2003). Because LNS was originally designed for the deterministic heterogeneous fixed fleet VRP with hard time windows (HFFVRPTW), the following adjustments are made to take into account the flexible time windows:

- The hard time windows are replaced by the flexible time windows in all required calculations;
- The LBBD subproblems, which present the special structure of TSPFlexTW, are solved by an extension of the heuristic dynamic programming algorithm of Balas and Simonetti (2001) that incorporates a label-correcting strategy called EL, as suggested by Fachini and Armentano (2020a).

LNS stops as soon as it reaches *LNS_time* seconds of CPU time, generating initial reference values $\bar{\mathbf{y}}$ for variables $\mathbf{y}$. However, these values do not ensure second stage feasibility because our problem does not have complete or relatively complete recourse (Birge and Louveaux 2011). Therefore, before local search starts, an auxiliary routine is performed to update vector $\bar{\mathbf{y}}$ with values that guarantee the feasibility of all second stage problems. The idea is to move from the first stage solution found for the single scenario s^* to a nearby one, which is feasible for all scenarios. This routine exploits a proximity function similar to that used by Fischetti and Monaci (2014) in the heuristic proximity search (PS). More specifically, the objective function (14) is replaced by

$$\min \Delta(\mathbf{y}, \bar{\mathbf{y}}) \tag{24}$$

where $\Delta(\mathbf{y}, \bar{\mathbf{y}})$ is the Hamming distance defined in (23), and $\bar{\mathbf{y}}$ is the first stage solution vector generated by LNS matheuristic. The resulting formulation, called LNS model (LNSM), is given by objective function (24) and MSDM constraints. Such model is then solved by a MIP solver within a time limit *PS_time*. With the completion of this step, the reference vector $\bar{\mathbf{y}}$ is updated with the best feasible solution found by the solver and original objective function (14) is restored. Since LNSM takes into account every HFFVRP-FlexTW-STT scenario, the new values obtained for vector $\bar{\mathbf{y}}$ ensure the feasibility of all second stage problems.

3.3 Granular Neighborhood

Granular neighborhoods were proposed by Toth and Vigo (2003) in the context of a tabu search algorithm for the VRP. Their purpose is to decrease the computational effort of local search through a candidate list strategy (Glover 1997). The key feature of this technique is to replace the original complete graph $G = (\mathcal{N}, \mathcal{E})$ by a new sparse one, denoted by $G' = \left(\mathcal{N}, \mathcal{E}'\right)$ with $\left|\mathcal{E}'\right| \ll |\mathcal{E}|$, which induces neighborhoods that can be quickly examined without affecting the quality of solutions found. The restricted edge set $\mathcal{E}'$ should contain only promising or "short" edges whose costs do not exceed a granularity threshold value.

In our implementation, the granularity threshold value ϑ is given by

$$\vartheta = \phi \frac{VC^*}{\left(n + \sum_{r \in \mathcal{M}} m_r\right)}$$

where VC^* represents the variable costs of routes in the solution found by LNS for scenario s^*, ϕ is a sparcification parameter that controls the size of granular neighborhood, n is the number of customers and $\sum_{r \in \mathcal{M}} m_r$ is the total of heterogeneous vehicles. Therefore, the restricted edge set can be defined as $\mathcal{E}' = \left\{(i,j) \in \mathcal{E}, \max_{r \in \mathcal{M}} \left\{c_{ij}^r\right\} \leq \vartheta\right\} \cup \mathcal{I}$, where $\mathcal{I}$ is a set of important edges, i.e., those connected to the depot and those belonging to the best solution found by LNS for s^*.

Since HFFVRP-FlexTW-STT presents second stage discrete variables, the solution of model (14) – (21) is time-consuming even with the addition of local branching constraints (22). Thus, to accelerate our local search matheuristic, the above-mentioned granular neighborhood discards the high-cost or "long" edges $(i,j) \in \mathcal{E} \backslash \mathcal{E}'$ of MSDM, producing a restricted MIP formulation with a smaller number of constraints and variables. More specifically, the removal of an edge $(i,j) \in \mathcal{E} \backslash \mathcal{E}'$ has the following implications for the model: (i) exclusion of variables x_{ij}^{rks} related to the removed edge for all $r \in \mathcal{M}, k \in \mathcal{M}_k, s \in \mathcal{S}$; and (ii) exclusion of constraints (17) associated with the removed edge.

Following previous studies (Branchini et al. 2009; Toth and Vigo 2003; Schneider et al. 2017), the cardinality of set $\mathcal{E}'$ is dynamically adjusted to diversify the search procedure. For this purpose, we do not fix the value of ϕ. Instead, we first set $\phi = \phi_{step}$ and, whenever a local optimum is found, we make $\phi \leftarrow \phi + \phi_{step}$ until such a parameter reaches the maximum value $\phi_{\max}$.

3.4 Algorithm Implementation

Figure 2 shows the pseudo-code of the granular local search matheuristic proposed for the HFFVRP-FlexTW-STT. Initially, MSDM is solved considering the single scenario s^*, in which $t_{ij}^{rs^*} = E\left(\hat{t}_{ij}^{r}\right)$ (lines 1 – 2). Parameter ϕ is initialized (line 3), enabling the calculation of granularity threshold ϑ (line 4) and the construction of a restricted edge set $\mathcal{E}'$ (line 5). As soon as these steps have been carried out, the multi-scenario version of MSDM is built by taking into account the sparse graph G' (line 6), and the auxiliary routine described in Sect. 3.2 is triggered to generate a reference solution feasible for all scenarios (lines 7 – 10). The granular local search

Algorithm granular local search matheuristic for the HFFVRP-FlexTW-STT

// Solution of MSDM considering single scenario s^*
1 Generate scenario s^* such that $t_{ij}^{rs^*} \leftarrow E\left(\hat{t}_{ij}^{r}\right)$
2 Solve the resulting HFFVRP-FlexTW by applying LNS matheuristic for LNS_time seconds and obtain $(VC^*, \bar{\mathbf{y}})$
// Construction of sparse graph $G' = (\mathcal{N}, \mathcal{E}')$
3 $\phi \leftarrow \phi_{\text{step}}$
4 $\vartheta \leftarrow \phi VC^* \Big/ \left(n + \sum_{r \in \mathcal{M}} m_r\right)$
5 $\mathcal{E}' \leftarrow \left\{(i,j) \in \mathcal{E}, \max_{r \in \mathcal{M}} \left\{c_{ij}^{r}\right\} \le \vartheta\right\} \cup \mathcal{I}$
6 Build MSDM by taking into account the sparse graph $G' = (\mathcal{N}, \mathcal{E}')$
// Generation of a starting feasible solution
7 Replace objective function (14) by expression (24) in order to generate LNSM
8 Solve LNSM by using a MIP solver for PS_time seconds
9 Update $\bar{\mathbf{y}}$ with the best solution found by the MIP solver
10 Restore original objective function (14), i.e., MSDM
// Granular local search procedure
11 **While** $\phi \le \phi_{\max}$ **do**
12 **Repeat**
13 Add constraint (22) to MSDM
14 Solve (14), (2) – (4), (15) – (22) using a MIP solver for LS_time seconds
15 Update the current reference solution
16 **Until** no improvement in the objective function is observed
17 *// Sparse graph update*
18 $\phi \leftarrow \phi + \phi_{\text{step}}$
19 **If** $\phi \le \phi_{\max}$ **then**
20 $\vartheta \leftarrow \phi CV^* \Big/ \left(n + \sum_{r \in \mathcal{M}} m_r\right)$
21 $\mathcal{E}' \leftarrow \left\{(i,j) \in \mathcal{E}, \max_{r \in \mathcal{M}} \left\{c_{ij}^{r}\right\} \le \vartheta\right\} \cup \mathcal{I}$
22 Update model MSDM
23 **End If**
24 **End While**
25 Output the minimum-cost solution found for the HFFVRP-FlexTW-STT

Fig. 2 Granular local search matheuristic for the HFFVRP-FlexTW-STT procedure (lines 11 – 24) is analogous to that one presented in Fig. 1. The main difference is that the search does not stop when a local optimum is reached. Instead, the value of ϕ is increased (line 18), the granularity threshold ϑ is recalculated (line 20), and both the restricted edge set $\mathcal{E}'$ and the model MSDM are updated (lines 21 – 22). Once a local optimum has been found by using the maximum value allowed for the sparcification parameter, i.e., $\phi = \phi_{max}$, the algorithm stops and outputs the minimum-cost solution found (line 25).

4 Computational Experiments

This section describes the computational results of the proposed algorithm that was coded in C# and run on a CPU Intel® Core™ i7-4790 CPU at 3,6 GHz with 16 GB RAM. As MIP solver, we used the Gurobi v.6.05 software. We start by presenting the benchmark instances and the parameters used within our algorithm in Sects. 4.1 and 4.2, respectively. In Sect. 4.3, we compare the proposed matheuristic with the direct application of Gurobi v.6.05 to solve the HFFVRP-FlexTW-STT model. Section 4.4 evaluates the solutions obtained by the granular local search with respect to those obtained by an alternative decomposition algorithm based on the augmented Lagrangian relaxation. To highlight the benefits gained by flexible time windows, in Sect. 4.5, we assess the results found by LNS matheuristic for the HFFVRP-FlexTW associated with scenario s^* against corresponding HFFVRPTW solutions. Finally, in Sect. 4.6, we analyze the advantage of solving the suggested stochastic model rather than its deterministic counterpart.

4.1 Instance Generation

Our data set was derived from 216 benchmark instances proposed by Fachini and Armentano (2020b) for the HFFVRPTW. Each instance combines: customer parameters derived from Solomon's instances (Solomon 1987), which comprise both short scheduling horizon classes R1, C1 and RC1 and long scheduling horizon classes R2, C2 and RC2, vehicle parameters derived from Liu and Shen (1999b) subclasses A, B and C, and fixed fleets obtained by Liu and Shen (1999a) for the corresponding fleet size and mix vehicle routing problem with time windows. There are problems with different time window densities $TW_{density}$, i.e., percentage of customers that have time windows, and various time window widths $TW_{width} = l_i - e_i$, $i \in \mathcal{C}$. Moreover, the data set includes instances with sizes n = 25, 50 and 100. The name of each one specifies its attributes of class, identification number in class, number of customers and subclass, e.g., R104.25A is the fourth instance of class R1, with 25 customers and vehicles of subclass A. For further details, see Fachini and Armentano (2020b).

We adapt each instance to the HFFVRP-FlexTW-STT context by means of two modifications. First, as in Taş et al. (2014c), we set costs $\left(\delta_\alpha, \delta_\beta\right)$ to $(0.50, 1.00)$ and fractions $\left(f_i^e, f_i^l\right)$ to $(0.10, 0.10)$ for all $i \in \mathcal{N}$. Second, we change the deterministic travel time t_{ij}^r of each edge $(i,j) \in \mathcal{E}$ to a random variable $\hat{t}_{ij}^r$ following the Burr type XII probability distribution (Burr 1942), which has been shown to properly represent variations in day-to-day travel times (Susilawati et al. 2013; Taylor 2017). The cumulative distribution function of the 3-parameter version of Burr type XII distribution (Tadikamalla 1980) is

$$F\left(x, \Theta, \gamma_1, \gamma_2\right) = 1 - \left(1 + \left(\frac{x}{\Theta}\right)^{\gamma_1}\right)^{-\gamma_2}$$

where $x \geq 0$ is a random variable, $\Theta > 0$ is a scaling parameter and $\gamma_1 > 0$ and $\gamma_2 > 0$ are shape parameters.

Random variables $\hat{t}^r_{ij}$ are approximated by a set $\mathcal{S}$ with 20 scenarios, each one with an associated probability $Pr(s) = 1/20$. For this purpose, we set $E\left(\hat{t}^r_{ij}\right) = t^r_{ij}$ and chose values $\gamma_1 = 2$ and $\gamma_2 = 1$ for shape parameters, as suggested by Gómez et al. (2016). In this way, the scaling parameter is given by $\Theta = 0.6366 t^r_{ij}$ since

$$E\left(\hat{t}^r_{ij}\right) = \frac{\Theta \gamma_2 \Gamma\left(\gamma_2 - \frac{1}{\gamma_1}\right)\Gamma\left(\frac{1}{\gamma_1} + 1\right)}{\Gamma\left(\gamma_2 + 1\right)}$$

Based on this parameter configuration, the following sampling procedure is adopted. For each triplet (i, j, r) of each scenario $s \in \mathcal{S}$, we first draw a random number U from a uniform distribution in the interval (0,1]. Then, the value of travel time t^{rs}_{ij} can be obtained by setting $t^{rs}_{ij} \leftarrow F^{-1}(U, \Theta, \gamma_1, \gamma_2)$, which gives

$$t^{rs}_{ij} \leftarrow \Theta \sqrt[\gamma_1]{(1 - U)^{-1/\gamma_2} - 1}$$

Because the employed probability distribution has very high variance and skewness, the sampling procedure may output infeasible instances. It can even generate scenarios for which no feasible solution exists at all. Since no reasonable SMIP can be constructed in such a circumstance, we modify model (14) – (21) by allowing feasible solutions with unserved customers at a given penalty cost ρ_u. More specifically, constraints (3) are replaced by

$$\sum_{r \in \mathcal{M}} \sum_{k \in \mathcal{M}_r} y^{rk}_i + u_i = 1 \quad i \in \mathcal{C} \tag{25}$$

where u_i is a binary variable such that $u_i = 1$, if customer $i \in \mathcal{C}$ is left unserved, $u_i = 0$, otherwise. Furthermore, a new term $\sum_{i \in \mathcal{N} \setminus \{0\}} \rho_u u_i$ is added to objective function (14) that becomes

$$\min \sum_{r \in \mathcal{M}} \sum_{k \in \mathcal{M}_r} f_r y^{rk}_0 + \sum_{i \in \mathcal{C}} \rho_u u_i$$
$$+ \sum_{\omega \in \Omega} pr(\omega) \left(\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}, j \neq i} \sum_{r \in \mathcal{M}} \sum_{k \in \mathcal{M}_r} c^r_{ij} x^{rk\omega}_{ij} + \sum_{i \in \mathcal{N}} \sum_{r \in \mathcal{M}} \sum_{k \in \mathcal{M}_r} \left(\delta_\alpha \alpha^{rk\omega}_i + \delta_\beta \beta^{rk\omega}_i\right) \right) \tag{26}$$

Therefore, we use the modified MSDM (M-MSDM), given by (26), (2), (4), (15) – (21), (25), in the computational experiments of next sections. Following Wang and Lin (2017), we set $p_u = 200$. The full data set is publicly available at https://sites.

google.com/view/hffvrp-flextw-stt/home. The instances themselves consist of a possible contribution of this paper is in the form of a freely available benchmark for the HFFVRP-FlexTW-STT.

4.2 Parameter Setting

To tune the parameters employed in the granular local search matheuristic, we perform a number of preliminary tests involving a subset of 36 representative instances. This subset comprises instances R104, C103, RC101, R202, C204 and RC204, with sizes $n=25$, 50 and 100, and vehicles of subclasses A and C.

The parameter setting occurred in two steps. In order to calibrate the LNS matheuristic, we started by only considering the expected value of scenario s^* associated with each instance. As shown in Table 1, the LNS parameters non-sensitive to the instance characteristics were tested in different ranges and the best configuration of values was chosen. On the other hand, the LNS problem-dependent parameters are given by mathematical expressions imported from Fachini and Armentano (2020b). Such expressions are presented in Table 2. For a full description of LNS control parameters, see Fachini and Armentano (2020b).

In a second step, we carried out test runs including all 20 scenarios generated for each instance to calibrate the overall procedure of granular local search. After a sensitivity analysis of intuitively selected value combinations, the following setting was adopted for the remaining parameters:

- $(LNS\_time, PS\_time, LS\_time, \kappa, \phi_{step}, \phi_{\max}) = (1800, 800, 450, 4, 2.5, 5)$, for instances with sizes $n = 25$ and $n = 50$;
- $(LNS\_time, PS\_time, LS\_time, \kappa, \phi_{step}, \phi_{\max}) = (2400, 1200, 600, 2, 1.25, 5)$, for instances with size $n = 100$.

Table 1 LNS non-sensitive parameters

Parameter	Range of tested values	Step size	Selected value	
			General case	Special cases[a]
ψ_1	[0.30, 0.70]	0.01	0.55	0.40
ψ_2	[0.30, 0.70]	0.01	0.45	0.60
η	[0.00, 1.00]	0.10	0.40	0.40
λ	[0.00, 1.00]	0.50	0.00	0.00
α_1	[0.00, 1.00]	0.10	0.00	0.30
α_2	[0.00, 1.00]	0.10	1.00	0.70
μ	[0.00, 1.00]	0.50	0.00	0.50
D	[30.00, 40.00]	5.00	35.00	35.00
K	[8.00, 14.00]	1.00	12.00	12.00
q	[10.00, 30.00]	5.00	15.00	15.00

[a]Instances of classes R2, C2 and RC2 with $TW_{density} \leq 75\%$

Table 2 LNS problem-dependent parameters

Parameter	Description	Expression
θ	Parameter used to build LNS starting feasible solution	$-TW_{density}{}^2 + 0.9TW_{density} + 0.6$
ξ	Number of customers removed per LNS iteration	$\max\left\{2, \left\lceil 4\ln\left(100TW_{density} \Big/ \left(n + \sum_{r\in\mathcal{M}} m_r + TW_{width}\right)\right)\right\rceil + 9\right\}$

We note that, in our implementation, the stopping criterion of LNS is given by the parameter *LNS_time*, tuned in the second step.

4.3 Comparison with a State-of-art MIP Solver

Since no other approach has been proposed for the HFFVRP-FlexTW-STT in the literature, we present a comparative analysis of results obtained by our granular local search matheuristic with those generated by the direct application of Gurobi v.6.05 state-of-art MIP solver to tackle this problem. For all instances, we imposed a CPU time limit of 10,800 s (3 h). Table 3 summarizes the mean results for each subset of 4 instances with the same size and belonging to the same class and subclass by using the following notation: "#Opt" (number of instances for which the optimal solution was found); "#Best" (number of instances for which the algorithm in question found the best upper bound); "Mean Gap%" (mean optimality gap $100(UB - LB)/UB$, where UB and LB are, respectively, the upper and lower bounds on the HFFVRP-FlexTW-STT optimal cost); and "Mean CPU" (mean time in seconds). A bar "-" indicates that the optimality gap is unavailable because no valid lower bound was generated for the instances in question. We remark that the quality assessment of the solutions obtained by our matheuristic was based on the lower bounds returned by the MIP solver. The interested reader is referred to Section A of the Supplementary Electronic Material for the detailed results obtained by both methods.

The MIP solver found proven optimal solutions for 12 (5%) instances with up to 50 customers and returned suboptimal solutions for 204 (95%) instances. Regarding such suboptimal solutions, the optimality gaps range from 2.93% to 95.63% in 180 (84%) instances and are unknown in 24 (11%) instances for which the linear relaxation of M-MSDM remained unsolved by the MIP solver after 10,800 s. In this last subset of 24 instances, Gurobi v.6.05 heuristically found an initial feasible solution before starting its presolve routine. The inherent difficulty of solving some M-MSDM linear relaxations at the root node can be accounted for the high dimensionality of the addressed problems, e.g., instance R101.100A presents 4 994 049 constraints, 4 850 524 binary variables and 145 440 continuous variables for $|\mathcal{S}| = 20$ scenarios.

The granular local search matheuristic found feasible solutions for all 216 instances by expending a mean computational time of 5387 s. With respect to the 192 instances for which the MIP solver has provided lower bounds, 186 (97%) of

Table 3 MIP solver versus granular local search matheuristic

n	Class	Subclass	Gurobi v.6.05 MIP solver				Granular local search matheuristic			
			#Opt	#Best	Mean Gap%	Mean CPU	#Opt[a]	#Best	Mean Gap%[a]	Mean CPU
25	R1	A (4)	1	0	55.17	8107	1	3	16.10	1794
		B (4)	1	1	54.76	8109	1	2	29.49	2095
		C (4)	1	1	54.83	8111	0	3	29.97	2553
	C1	A (4)	1	1	10.21	8131	0	3	8.30	1409
		B (4)	1	0	19.69	8113	1	3	14.01	1767
		C (4)	1	0	42.99	8113	1	3	16.73	1679
	RC1	A (4)	0	0	33.68	10,800	0	4	17.57	1852
		B (4)	0	0	54.60	10,800	0	4	25.64	2402
		C (4)	0	0	54.00	10,800	0	4	31.78	2195
	R2	A (4)	0	1	42.99	10,800	0	3	39.32	3688
		B (4)	0	1	46.50	10,800	0	3	28.37	3255
		C (4)	0	1	32.99	10,800	0	3	25.42	3523
	C2	A (4)	1	3	28.84	8180	0	1	32.05	2536
		B (4)	1	2	43.33	8213	0	2	33.66	2778
		C (4)	1	2	41.40	8123	0	2	38.64	3604
	RC2	A (4)	0	1	42.03	10,800	0	3	28.86	3025
		B (4)	0	1	51.06	10,800	0	3	30.79	2752
		C (4)	0	1	36.25	10,800	0	2	31.58	3275
50	R1	A (4)	1	1	59.01	8823	0	3	28.96	3250
		B (4)	1	0	67.07	9236	1	3	38.95	3756
		C (4)	1	0	68.28	9306	1	3	41.37	3746
	C1	A (4)	0	1	45.25	10,800	0	3	13.08	3385
		B (4)	0	1	52.96	10,800	0	3	22.73	2580
		C (4)	0	1	56.11	10,800	0	3	23.83	4028

Table 3 (continued)

n	Class	Subclass	Gurobi v.6.05 MIP solver				Granular local search matheuristic			
			#Opt	#Best	Mean Gap%	Mean CPU	#Opt[a]	#Best	Mean Gap%[a]	Mean CPU
	RC1	A (4)	0	0	71.00	10,800	0	4	27.90	3483
		B (4)	0	0	87.27	10,800	0	4	43.38	3484
		C (4)	0	0	88.98	10,800	0	4	45.44	4871
	R2	A (4)	0	0	71.04	10,800	0	4	27.68	6737
		B (4)	0	0	84.82	10,800	0	4	35.63	5633
		C (4)	0	0	82.42	10,800	0	4	34.41	7072
	C2	A (4)	0	4	26.63	10,800	0	0	41.07	6442
		B (4)	0	3	58.39	10,800	0	1	57.65	5549
		C (4)	0	3	50.51	10,800	0	1	59.64	5105
	RC2	A (4)	0	1	58.75	10,800	0	3	37.52	6628
		B (4)	0	1	74.95	10,800	0	3	45.12	5230
		C (4)	0	0	80.93	10,800	0	4	46.53	5676
100	R1	A (4)	0	0	-	10,800	0	4	-	10,800
		B (4)	0	0	-	10,800	0	4	-	10,800
		C (4)	0	0	-	10,800	0	4	-	10,800
	C1	A (4)	0	1	46.91	10,800	0	3	15.07	8451
		B (4)	0	0	71.78	10,800	0	4	24.05	8276
		C (4)	0	1	62.80	10,800	0	3	30.46	8455
	RC1	A (4)	0	0	79.49	10,800	0	4	71.69	10,800
		B (4)	0	0	92.43	10,800	0	4	87.99	10,800
		C (4)	0	0	94.05	10,800	0	4	88.04	10,800
	R2	A (4)	0	0	83.90	10,800	0	4	76.09	8293

Table 3 (continued)

n	Class	Subclass	Gurobi v.6.05 MIP solver				Granular local search matheuristic			
			#Opt	#Best	Mean Gap%	Mean CPU	#Opt[a]	#Best	Mean Gap%[a]	Mean CPU
		B (4)	0	0	93.37	10,800	0	4	90.14	8962
		C (4)	0	0	94.58	10,800	0	4	92.50	8688
	C2	A (4)	0	0	50.53	10,800	0	4	37.57	6170
		B (4)	0	0	75.77	10,800	0	4	48.11	5428
		C (4)	0	0	77.75	10,800	0	4	50.27	4968
	RC2	A (4)	0	0	79.57	10,800	0	4	59.31	9446
		B (4)	0	0	92.59	10,800	0	4	77.85	8105
		C (4)	0	0	94.22	10,800	0	4	81.17	8024
	All (216)		12	34	59.89	10,262	6	175	39.09	5387

[a]Based on lower bounds returned by Gurobi v.6.05 MIP solver

these solutions are suboptimal with gaps ranging from 0.13% to 95.09%, and 6 (3%) correspond to optimal solutions with up to 50 customers.

The outcomes in Table 3 reveals that our matheuristic found better upper bounds for 175 (81%) instances, while the MIP solver generated better solutions for 34 (16%) instances, with 7 (3%) ties. Such a comparison also points out that the advantage of the granular local search matheuristic over the MIP solver increases with the instance size. As an example, 105 (73%) out of the 144 instances with sizes $n = 25$ and $n = 50$ were better solved by our algorithm, while the same method was superior for 70 (98%) out of the remaining 72 instances with size $n = 100$.

Another interesting trend can be seen in Fig. 3, which plots mean optimality gaps returned by both algorithms for each group of 36 instances of the same class, i.e., R1, C1, RC1, R2, C2 or RC2. What stands out in this chart is the propensity of the suggested matheuristic to provide greater gap reductions when compared to the MIP solver for instances with short scheduling horizons, i.e., those belonging to classes R1, C1 and RC1. This result suggests that longer scheduling horizons may entail additional complexity to the HFFVRP-FlexTW and, therefore, impact negatively on the performance of granular local search matheuristic.

Taken together, these results underline the advantage of the granular local search matheuristic over the MIP solver, which was outperformed both in terms of CPU time and solution quality for most of instances. Such an advantage is intensified for instances with short scheduling horizons

4.4 Comparison with an Alternative Heuristic Method

Given the lack of benchmarks, in a second experiment to evaluate the granular local search performance, we have implemented a heuristic version of the Progressive Hedging Algorithm (PHA) (Rockafellar and Wets 1991) to the HFFVRP-FlexTW-STT.

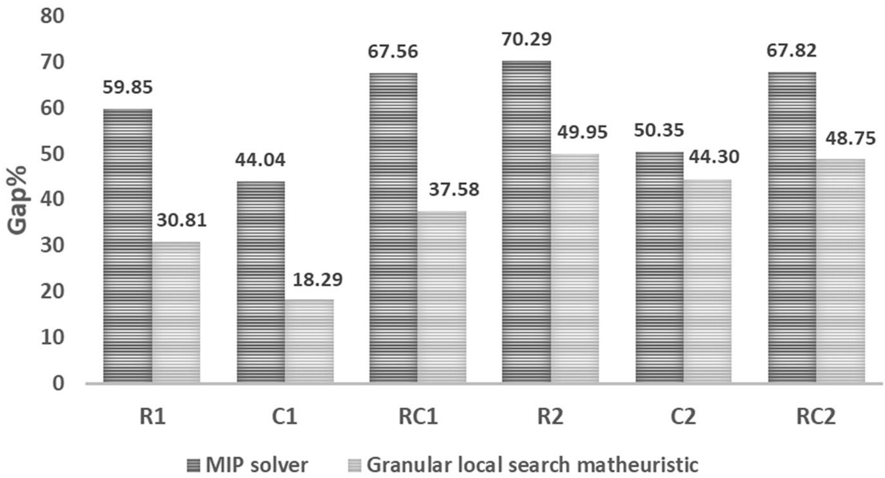

Fig. 3 Effect of classes on MIP solver and granular local search matheuristic

The PHA applies a scenario decomposition technique, in which the augmented Lagrangian relaxation is used to partition a multi-scenario problem into individual scenario subproblems. For this purpose, we create a copy $y_i^{rks} \in \{0,1\}, i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S}$ of variables $\boldsymbol{y}$ for each scenario $s \in \mathcal{S}$. Such variables, together with the addition to the problem of "nonanticipativity constraints"

$$y_i^{rks} = y_i^{rkh} i \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s, h \in \mathcal{S}, s \neq h$$

ensure that the mathematical model decisions are not tailored for each particular scenario but reflect the complete original problem. Since the number of these constraints may be large, they are replaced by

$$y_i^{rks} = \bar{y}_i^{rk} \, i, \in \mathcal{N}, r \in \mathcal{M}, k \in \mathcal{M}_r, s \in \mathcal{S},$$

which are then relaxed using an augmented Lagrangian strategy.

The main idea of PHA is to iteratively solve $|\mathcal{S}|$ deterministic subproblems until the algorithm converges by obtaining a consensus between all the optimized scenarios, i.e., if the gap between all variables y_i^{rks} f and the consensus parameter $\bar{y}_i^{rk}$ falls below a given threshold value, e.g., $\epsilon = 0.0001$. After each algorithmic iteration there are two parameter vectors called *Lagrangian multiplier* and *penalty parameter* that are updated by enabling the convergence of $\boldsymbol{y}$ variables to the values of consensus parameters. For more details on the implementation of similar approaches, see Crainic et al. (2011, 2016) Lamghari and Dimitrakopoulos (2016) and Quddus et al. (2017).

Our PHA implementation adopts the same parameters to initialize and update the *Lagrangian multiplier* and the *penalty parameter* as in Crainic et al. (2011). We considered $\epsilon = 0.0001$ as the threshold value for consensus evaluation. For each iteration, the deterministic scenarios are solved by the LNS matheuristic of Fachini and Armentano (2020b) under a time limit of $\left\lceil 125 \ln \left(n + \sum_{r \in \mathcal{M}} m_r \right) - 400 \right\rceil$ seconds.

For the algorithmic comparison, we selected a subset of 9 representative instances from our test bed with diverse characteristics. These instances were solved by PHA under the same CPU time limit of 10,800 s (3 h), which had been also imposed to the granular local search matheuristic. Since PHA might not converge within this time limit, we adopted the following two step strategy: (i) PHA runs at most for 10,200 s; (ii) the variables y_i^{rks} for which a consensus has been obtained among the scenarios within this time limit are fixed in their obtained values, and the resultant partially fixed model (26), (2), (4), (15) – (21), (25) is then solved by Gurobi v.6.05 in the remaining processing time.

Table 4 compares, for each instance, the solutions generated by both methods. Column 1 shows the instance name. Columns 2 and 3 show, respectively, the results of the granular local search matheuristic and of the PHA. The last column refer to the percentage differences $\Delta\%$ between the total costs obtained by granular local search matheuristic (GLSM) (TC_{GLSM}) and those obtained by PHA (TC_{PHA}), i.e., $\Delta\% = 100\left(TC_{PHA} - TC_{GLSM}\right) / TC_{GLSM}$. A bar '-' indicates that no feasible solution was found for the pair algorithm/instance in question. The bottom of the table reports average results

As can be seen from Table 4, the advantage of the granular local search matheuristic becomes clear. Such method found better solutions for all instances, yielding percentage differences $\Delta\%$ in the interval from 7% to 100%, 53% on average. For 3 out of 9 instances (33%) the PHA algorithm could not even find a feasible solution within the 10,800 s time limit. Such results reinforce the effectiveness of the proposed matheuristic.

4.5 Effect of Accounting Flexible Time Windows

To evaluate the benefits obtained when flexible time windows are considered, in this section, we conduct a numerical comparison between the solutions generated by LNS matheuristic for the deterministic HFFVRP-FlexTW associated with scenario s^*, and for the corresponding HFFVRPTW. To address the latter problem, we used the same setting of parameters as in Fachini and Armentano (2020b) and, therefore, the CPU time limit of LNS matheuristic was given by expression $\left\lceil 1295 \ln\left(n + \sum_{r\in\mathcal{M}} m_r\right) - 3400\right\rceil$ in both experiments.

Table 5 summarizes the results obtained. Columns 1 – 3 show the sizes, classes and subclasses of the instances. The next three columns refer to the solutions generated for the HFFVRPTW, which encompasses their total cost ("TC"), the corresponding vehicle fixed costs ("FC") and route variable costs ("VC"). The following four columns present the results obtained when solving the HFFVRP-FlexTW related to scenario s^*, including penalty costs for violating customer time windows ("PC"). Column 11 displays the percentage difference $\Delta_1\%$ between the total costs of HFFVRPTW and HFFVRP-FlexTW, i.e., $100\left(TC_{HFFVRPTW} - TC_{HFFVRP-FlexTW}\right)/TC_{HFFVRP-FlexTW}$. Columns 12 and 13 show, respectively, the percentage differences $\Delta_2\%$ and $\Delta_3\%$ between fixed and variables costs obtained for such VRP variants, i.e.,$100\left(FC_{HFFVRPTW} - FC_{HFFVRP-FlexTW}\right)/FC_{HFFVRP-FlexTW}$ and $100\left(VC_{HFFVRPTW} - VC_{HFFVRP-FlexTW}\right)/VC_{HFFVRP-FlexTW}$. Each entry in this table corresponds to mean results for each subset of four instances with the same size and belonging to the same class and subclass. Detailed results, for each instance, can be found in Section B of the Supplementary Electronic Material.

Considering the 54 subsets of instances analyzed in Table 5, flexible time windows leaded to lower total costs for 40 (74%), lower vehicle fixed costs for 23 (43%)

Table 4 Granular local search matheuristic versus PHA – total costs and mean percentage differences

Instance	TC_{GLSM}	TC_{PHA}	Δ %
R101.25A	2688.13	3116.08	16%
C101.25A	2891.05	3492.35	21%
RC102.25A	1811.73	2106.16	16%
R101.50A	5839.40	6234.13	7%
C101.50A	5360.63	7662.31	43%
RC102.50A	3351.73	-	100%
R101.100A	10,790.56	-	100%
C101.100A	11,240.70	19,670.71	75%
RC102.100A	12,338.07	-	100%
All (9)	6256.89	7046.96	53%

Table 5 HFFVRPTW versus HFFVRP-FlexTW – costs and mean percentage differences

n	Class	Subclass	HFFVRPTW			HFFVRP-FlexTW				Percentage differences		
			TC	FC	VC	TC	FC	VC	PC	$\Delta_1\%$	$\Delta_2\%$	$\Delta_3\%$
25	R1	A (4)	1147.09	605.00	542.09	1141.24	605.00	536.24	0.23	0.54	0.00	1.16
		B (4)	651.45	127.00	524.45	649.34	130.00	519.34	0.65	0.39	-2.05	0.98
		C (4)	586.66	64.00	522.66	584.34	65.00	519.34	0.65	0.41	-1.27	0.60
	C1	A (4)	1871.63	1600.00	271.63	1873.72	1600.00	273.72	0.00	-0.10	0.00	-0.95
		B (4)	599.24	320.00	279.24	593.72	320.00	273.72	0.00	0.92	0.00	2.50
		C (4)	433.72	160.00	273.72	431.63	160.00	271.63	0.00	0.52	0.00	0.99
	RC1	A (4)	1593.97	1095.00	498.97	1582.45	1080.00	502.45	0.99	0.73	1.43	-0.65
		B (4)	711.15	231.00	480.15	697.56	235.50	462.06	1.56	1.92	-1.72	3.91
		C (4)	592.70	123.75	468.95	565.11	120.00	445.11	1.56	4.99	3.12	5.42
	R2	A (4)	1517.92	1075.00	442.92	1502.27	1075.00	427.27	4.49	1.10	0.00	3.67
		B (4)	661.41	225.00	436.41	619.44	202.50	416.94	1.03	7.61	16.67	4.57
		C (4)	552.69	118.75	433.94	534.08	107.50	426.58	1.45	3.49	12.50	2.14
	C2	A (4)	3003.86	2775.00	228.86	3004.68	2775.00	229.68	0.00	-0.01	0.00	-0.33
		B (4)	685.29	455.00	230.29	785.41	555.00	230.41	0.00	-8.61	-10.64	-0.08
		C (4)	505.77	277.50	228.27	507.91	277.50	230.41	0.00	-0.43	0.00	-0.93
	RC2	A (4)	1828.00	1387.50	440.50	1956.08	1525.00	431.08	1.02	-4.75	-6.25	2.21
		B (4)	732.26	312.50	419.76	737.54	312.50	425.04	1.02	-0.57	0.00	-1.11
		C (4)	555.30	152.50	402.80	556.33	147.50	408.83	1.69	-0.09	0.22	-0.73
50	R1	A (4)	2277.15	1285.00	992.15	2292.20	1325.00	967.20	1.05	-0.60	-2.99	2.97
		B (4)	1209.73	271.00	938.73	1195.93	270.00	925.93	0.56	1.09	0.46	1.25
		C (4)	1073.43	139.75	933.68	1067.86	138.75	929.11	0.58	0.60	0.73	0.58
	C1	A (4)	3569.23	3025.00	544.23	3565.26	3025.00	540.26	0.00	0.12	0.00	0.64
		B (4)	1158.74	605.00	553.74	1144.05	605.00	539.05	0.00	1.28	0.00	2.43
		C (4)	842.65	302.50	540.15	851.85	302.50	549.35	0.00	-1.00	0.00	-1.42

Table 5 (continued)

n	Class	Subclass	HFFVRPTW			HFFVRP-FlexTW				Percentage differences		
			TC	FC	VC	TC	FC	VC	PC	Δ_1%	Δ_2%	Δ_3%
	RC1	A (4)	2984.78	1912.50	1072.28	2934.42	1920.00	1014.42	2.46	1.69	-0.30	5.84
		B (4)	1373.11	406.50	966.61	1353.09	411.00	942.09	1.39	1.44	-1.04	2.51
		C (4)	1190.81	208.50	982.31	1144.40	201.75	942.65	1.22	3.95	3.44	4.00
	R2	A (4)	2672.59	1975.00	697.59	2498.81	1800.00	698.81	7.91	7.69	11.81	-0.25
		B (4)	1082.72	337.50	745.22	1037.38	337.50	699.88	6.77	4.37	0.00	6.54
		C (4)	896.53	180.00	716.53	857.00	180.00	677.00	8.47	4.86	0.00	6.04
	C2	A (4)	5232.44	4775.00	457.44	5183.88	4775.00	408.88	0.00	2.09	1.25	13.06
		B (4)	1428.50	1005.00	423.50	1365.69	905.00	460.69	0.00	5.31	12.50	-7.44
		C (4)	967.39	527.50	439.89	870.43	477.50	392.93	0.00	10.81	11.25	11.73
	RC2	A (4)	3370.88	2600.00	770.88	3108.84	2325.00	783.84	4.32	8.51	11.04	-1.68
		B (4)	1290.49	482.50	807.99	1297.43	527.50	769.93	7.82	-0.53	-6.40	6.05
		C (4)	1019.40	277.50	741.90	999.90	226.25	773.65	4.42	2.34	18.97	-3.51
100	R1	A (4)	4393.22	2662.50	1730.72	4324.80	2647.50	1677.30	1.45	1.59	0.56	3.20
		B (4)	2246.53	534.00	1712.53	2230.66	536.50	1694.16	1.84	0.75	-0.48	1.17
		C (4)	1971.93	270.25	1701.68	1936.85	267.00	1669.85	1.11	1.84	1.20	1.94
	C1	A (4)	7589.45	6275.00	1314.45	7612.50	6275.00	1337.50	0.00	-0.32	0.00	-1.52
		B (4)	2608.99	1255.00	1353.99	2605.29	1270.00	1335.29	0.00	0.13	-0.90	0.34
		C (4)	2012.82	627.50	1385.32	1933.76	627.50	1306.26	0.00	3.82	0.00	5.27
	RC1	A (4)	5491.59	3442.50	2049.09	5302.08	3382.50	1919.58	7.17	3.73	1.76	7.62
		B (4)	2701.81	691.50	2010.31	2583.68	682.50	1901.18	7.39	4.76	1.32	6.09
		C (4)	2293.06	345.75	1947.31	2245.26	342.75	1902.51	10.48	2.52	0.86	2.88
	R2	A (4)	3571.79	2312.50	1259.29	3477.61	2312.50	1165.11	15.97	2.79	0.00	8.23
		B (4)	1676.90	462.50	1214.40	1606.96	462.50	1144.46	17.95	4.14	0.00	5.70

Table 5 (continued)

n	Class	Subclass	HFFVRPTW			HFFVRP-FlexTW				Percentage differences		
			TC	FC	VC	TC	FC	VC	PC	Δ_1%	Δ_2%	Δ_3%
		C (4)	1478.96	231.25	1247.71	1433.06	231.25	1201.81	14.99	3.24	0.00	3.82
	C2	A (4)	6853.84	6025.00	828.84	6894.40	6025.00	869.40	7.26	-0.39	0.00	-3.80
		B (4)	2044.87	1205.00	839.87	2062.70	1205.00	857.70	8.68	-0.19	0.00	-0.73
		C (4)	1448.39	602.50	845.89	1477.55	602.50	875.05	7.72	-1.37	0.00	-2.36
	RC2	A (4)	5024.88	3400.00	1624.88	4917.03	3400.00	1517.03	16.30	2.12	0.00	7.37
		B (4)	2313.64	680.00	1633.64	2156.29	680.00	1476.29	14.08	7.38	0.00	11.10
		C (4)	1927.42	340.00	1587.42	1830.64	340.00	1490.64	14.79	5.35	0.00	6.77
All (216)			2028.16	1162.65	865.51	1994.86	1154.37	840.49	3.90	1.93	1.43	2.51

and lower route variables costs for 38 (70%) subsets. On the other hand, higher total costs were found for 14 (26%) subsets and solutions with higher fixed and variables costs were obtained for 16 (30%) subsets as a counterpart of the flexibility. Ties occurred for 15 (27%) subsets with reference to vehicle fixed costs. The bottom of this table shows that lower costs "TC", "FC" and "VC", and therefore, positive mean percentage differences $\Delta_1\%$, $\Delta_2\%$ and $\Delta_3\%$ were obtained for the whole set of instances.

In Table 6, we provide a further assessment of the gains stemming from the flexible time windows. This table presents the number of instances and its percentage relative to the total of 216 instances for which positive (">0"), null ("$=0$") or negative ("<0") percentage differences were obtained. It also displays the maximum and the minimum values of $\Delta_1\%$, $\Delta_2\%$ and $\Delta_3\%$. We have positive percentage differences $\Delta_1\%$ (68%) and $\Delta_3\%$ (61%) for most of instances, pointing out that HFFVRP-FlexTW yields fewer total costs and route variable costs than HFFVRPTW. In contrast, null values prevail in the data set for $\Delta_2\%$ (51%), which evince a reduced gain in terms of vehicle fixed costs. For each of the examined percentage differences, negative values were obtained for approximately 20% of the instances.

From Tables 5 and 6, it is possible to attest that, in general, the use of flexible time windows is beneficial and promotes relevant total cost savings. Moreover, the values found for percentage differences $\Delta_2\%$ and $\Delta_3\%$ outline that such savings mostly arise from reductions in the route variable costs.

4.6 Effect of Accounting Travel Time Uncertainty

In this section, we evaluate the advantage of explicitly modeling travel time uncertainty through the HFFVRP-FlexTW-STT model compared to solving its deterministic counterpart, i.e., HFFVRP-FlexTW. For this purpose, we use the well-known stochastic programming measure called the value of the stochastic solution (VSS) (Birge 1982; Birge and Louveaux 2011).

We first state a problem called expected value (EV), a deterministic approximation of the proposed recourse problem (RP) in which only the expected value scenario s^* is considered:

Table 6 HFFVRPTW versus HFFVRP-FlexTW – summary of percentage differences

Percentage difference	>0		$=0$		<0		Maximum value	Minimum value
	#	%	#	%	#	%		
$\Delta_1\%$	147	68	16	7	53	25	23.64	-33,57
$\Delta_2\%$	65	30	110	51	41	19	50.00	-42.55
$\Delta_3\%$	132	61	34	16	50	23	33.47	-31.78

$$EV = \min \sum_{r \in \mathcal{M}} \sum_{k \in \mathcal{M}_r} f_r y_0^{rk} + E_\varepsilon \left[\mathcal{Q}\left(\mathbf{y}, E\left(\hat{t}_{ij}^r \right) \right) \right]$$

$$Constraints\ (2) - (5)$$

where $E\left(\hat{t}_{ij}^r\right)$ denotes the expectation of stochastic travel times, i.e., $t_{ij}^{rs^*}$. Then, we define the expected value of using the EV solution (EEV) as the solution value of model (26), (2), (4), (15) – (21), (25) with the first stage decision variables fixed at the optimal values $\bar{\mathbf{y}}$ obtained by solving EV, i.e.,

$$EEV = Objective\ (26)$$

$$Constraints\ (2), (4), (15) - (21), (25)$$

$$\boldsymbol{y} = \bar{\boldsymbol{y}}$$

The VSS can be written as follows:

$$VSS = EEV - RP \tag{27}$$

Whenever this measure presents large positive values, there is a potential benefit from solving RP over EV (Birge and Louveaux 2011).

We selected instance RC201.25A to illustrate the VSS calculation. For such an instance, it is possible to solve EV to proven optimality using a MIP solver. This software returns the optimal cost EV = 2080.69 and, from Table A.2 of the Supplementary Electronic Material, we have that RP = 2620.33 is the best solution found by our granular local search matheuristic for RP.

Turning now to EEV computation, when the first stage decision variables $\mathbf{y}$ are fixed at the optimal values obtained for EV, i.e., $\mathbf{y} = \mathbf{y}_{EV}$, no feasible solution exists for second stage decision variables $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Therefore, $EEV = +\infty$ (see Birge (1982)) and, from (27), we have $VSS \rightarrow +\infty$. It means that travel time uncertainty really matters in our context. Figure 4 depicts clusters of customers $\bar{\mathcal{H}}_1^{\mathrm{EV}} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 13, 17\}$ and $\bar{\mathcal{H}}_2^{\mathrm{EV}} = \{0, 9, 10, 11, 12, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25\}$, associated with solution vector $\mathbf{y}_{EV}$, and clusters $\bar{\mathcal{H}}_1^{\mathrm{RP}} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 17\}$ and $\bar{\mathcal{H}}_2^{\mathrm{RP}} = \{0, 11, 12, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25\}$, related to $\mathbf{y}$ values obtained for RP. Despite pairs $\left(\bar{\mathcal{H}}_1^{\mathrm{EV}}, \bar{\mathcal{H}}_1^{\mathrm{RP}}\right)$ and $\left(\bar{\mathcal{H}}_2^{\mathrm{EV}}, \bar{\mathcal{H}}_2^{\mathrm{RP}}\right)$ present many customers in common, the assumption of deterministic travel times leads to HFFVRP-FlexTW-STT infeasibility.

To compute the VSS measure to a broader set of instances, we have proceeded as follows. First, we have calculated EEV by solving model (26), (2), (4), (15) – (21), (25) with its first stage variables $\mathbf{y}$ fixed at their corresponding values returned by LNS matheuristic of Fachini and Armentano (2020b) for the deterministic HFFVRP-FlexTW associated with scenario $\bar{\mathbf{y}}$. In this step, the resultant problems were solved by Gurobi v.6.05 within a CPU time limit of 10,800 s (3 h). Second, we have assumed the RP value as the best solution found by our granular local search

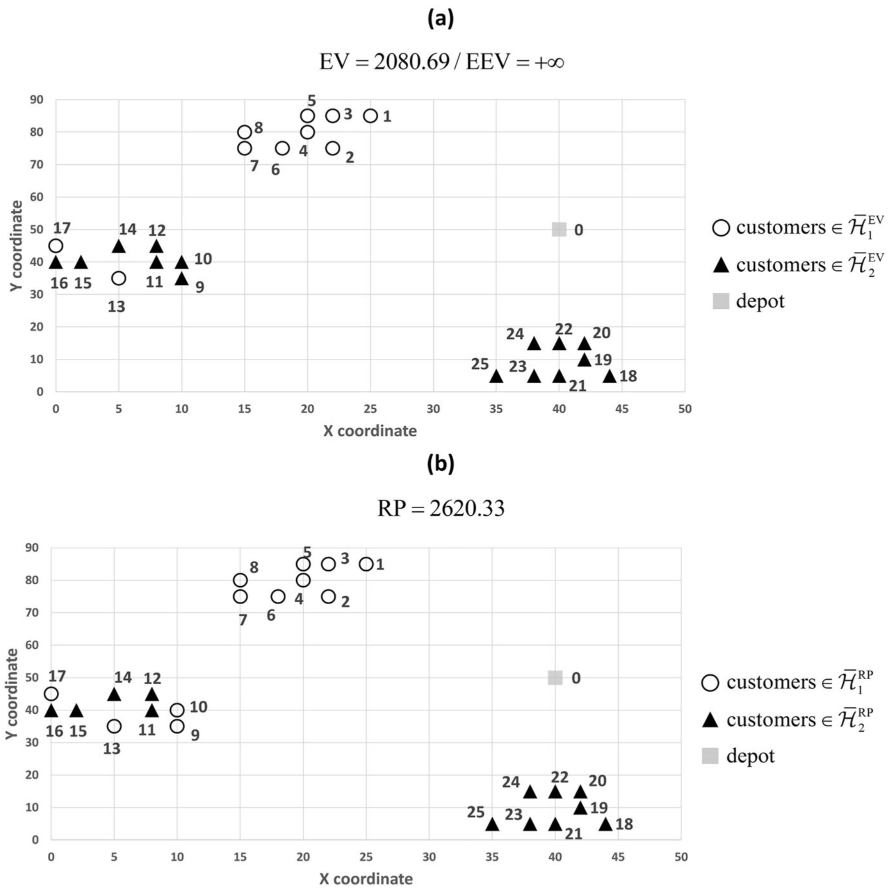

Fig. 4 Clusters of customers obtained for EV and RP – instance RC201.25A

matheuristic to model (26), (2), (4), (15) – (21), (25) under the same time limit. Third, we calculate VSS as the difference EEV and RP. Given the inherent difficulty of solving the HFFVRP-FlexTW-STT by using Gurobi v.6.05 for instances with $n=100$, we limited the scope of VSS calculations to the 144 instances with sizes $n=25$ and $n=50$. Table 7 displays aggregated results grouped by instance size, subclass and feasibility of EEV problem by using the following notation: ‘‘#Instances’ (number of instances belonging to the group), and “VSS %” (VSS mean percentages, i.e., $s*$). Detailed results, for each instance, can be found in Section C of the Supplementary Electronic Material.

We note from Table 7 that it clearly pays off to use the stochastic solution rather than the deterministic mean value solution. The latter resulted in infeasible problems regarding the second stage decision variables $100(VSS)/RP$ for 95 out of 144 instances (66%), which entails $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. By contrast, as a counterpart of the increased complexity of solving RP rather than of successively tackling EV and EEV, we found

Table 7 Value of the stochastic solution

n	Subclass	EEV feasibility	#Instances	VSS%[a]
25	A	feasible	10	-6.95
		infeasible	14	$+\infty$
	B	feasible	10	-8.73
		infeasible	14	$+\infty$
	C	feasible	11	-9.67
		infeasible	13	$+\infty$
50	A	feasible	6	-15.53
		infeasible	18	$+\infty$
	B	feasible	5	-7.17
		infeasible	19	$+\infty$
	C	feasible	7	-15.96
		infeasible	17	$+\infty$
	All	feasible	49	-10.28
		infeasible	95	$+\infty$

[a]Negative values of VSS% are possible because the adopted solutions of RP stems from a heuristic approach

negative mean VSS% values for instances with feasible EEV problems. Such negative values benefit the deterministic mean solution instead of the stochastic one. Regarding this result, there are two important remarks: (i) it is only possible to obtain negative values of VSS% because the adopted solutions of RP stems from a heuristic approach and therefore are not necessarily optimal, otherwise such metric would be always non-negative since RP takes into account the minimization of the expected costs over all scenarios, including $VSS \rightarrow +\infty$; and (ii) despite the negative mean VSS% results generated for these instances, their absolute values range from -6.95% to -15.96%, demonstrating that even in such few situations the losses associated with the increased complexity of the stochastic model are low.

From the examination of Table 7, a last conclusion can be drawn. The advantage of the stochastic solution over the deterministic mean value solution increases with the instance size. For example, 41 (57%) out of the 72 instances with size s^* resulted in infeasible EEV problems, while this number increases to 54 (75%) out of the remaining 72 instances with size $n = 25$. This trend reinforces the practical importance of explicitly modeling the stochastic travel times, since real-life VRP problems tend to be large scaled.

5 Conclusions

This paper has introduced a new variant of the VRP that includes a heterogeneous fixed fleet of vehicles, flexible time windows and stochastic travel times. We developed a two-stage stochastic mixed-integer model with recourse to this problem. Customers' assignment to vehicles makes up the first stage, while recourse decisions are made in the second stage to find minimum-cost routes for vehicles given realized travel times. The objective is to minimize the sum of transportation costs and service costs. The former comprise the vehicle fixed costs and route variable costs, and the latter correspond to the penalty costs for violating customer time windows.

We developed a scenario generation algorithm that describes stochastic travel times using the Burr type XII distribution (Burr 1942), and an effective granular local search matheuristic was devised to tackle the problem. Extensive computational results obtained for 216 benchmark instances attest the effectiveness of our approach. It outperformed Gurobi v.6.05 MIP solver, finding better solutions for 175 (81%) instances and consuming a mean CPU time around 48% lower. The proposed matheuristic also far outperformed an alternative decomposition algorithm based on the augmented Lagrangian relaxation.

The advantages of both flexible windows and stochastic travel times have been also assessed. Flexible time windows yielded overall cost savings for 68% of the instances compared to the solutions obtained for hard time window problems. Furthermore, an in-depth analysis of the well-known stochastic programming measure VSS illustrated the potential benefits of considering stochastic travel times over solving an approximated deterministic model. Explicitly modeling the stochastic travel times showed to be a critical issue, since the adoption of a deterministic problem with the random parameters fixed at their expected values resulted in infeasible solutions for 66% of the tested instances.

Future research may develop of an exact method for the HFFVRP-FlexTW-STT. Another research avenue involves applying the algorithmic techniques suggested here to other stochastic VRP variants, which have received much less attention from academics than their deterministic counterparts. Finally, other interesting research direction would be to integrate the proposed method into the location and routing problem with stochastic travel times, since it has been proven that solving the location and routing problem is superior to solving the facility location problem and VRP separately (Salhi and Rand 1989).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11067-021-09553-6.

Acknowledgements This research was funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil, grants 141064/2015-3 and 305577/2014-0) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil, grants 2016/01860-1 and 2016/06566-4). We are grateful for the valuable suggestions made by the editor and referees.

References

Agra A, Christiansen M, Hvattum LM, Rodrigues F (2016) A MIP Based Local Search Heuristic for a Stochastic Maritime Inventory Routing Problem. In: Paias A, Ruthmair M, Voß S (eds) Proceedings of the 7th International Conference on Computational Logistics (ICCL 2016). Springer, Cham 18–34

Agra A, Requejo C, Rodrigues F (2018) An adjustable sample average approximation algorithm for the stochastic production-inventory-routing problem. Networks 72(1):5–24

Ando N, Taniguchi E (2006) Travel time reliability in vehicle routing and scheduling with time windows. Netw Spat Econ 6(3–4):293–311

Balas E, Simonetti N (2001) Linear time dynamic-programming algorithms for new classes of restricted TSPs: A computational study. INFORMS J Comput 13(1):56–75
Baldacci R, Battarra M, Vigo D (2008) Routing a heterogeneous fleet of vehicles. In: Golden BL, Raghavan S, Wasil EA (eds) The Vehicle Routing Problem: Latest Advances and New Challenges. Springer, New York, pp 3–27
Crainic TG, Fu X, Gendreau M, Rei W, Wallace SW (2011) Progressive hedging-based metaheuristics for stochastic network design. Networks 58(2):114–124
Crainic TG, Gobbato L, Perboli G, Rei W (2016) Logistics capacity planning: A stochastic bin packing formulation and a progressive hedging meta-heuristic. Eur J Oper Res 253:404–417
Birge JR (1982) The value of the stochastic solution in stochastic linear programs with fixed recourse. Math Program 24(1):314–325
Birge JR, Louveaux F (2011) Introduction to Stochastic Programming, 2nd edn. Springer, New York
Braekers K, Ramaekers K, Van Nieuwenhuyse I (2016) The vehicle routing problem: State of the art classification and review. Comput Ind Eng 99:300–313
Branchini RM, Armentano VA, Løkketangen A (2009) Adaptive granular local search heuristic for a dynamic vehicle routing problem. Comput Oper Res 36(11):2955–2968
Burr IW (1942) Cumulative Frequency Functions. Ann Math. Stat 13(2):215–232
Dantzig GB, Ramser JH (1959) The truck dispatching problem. Manage Sci 6(1):80–91
Desaulniers G, Madsen OB, Røpke S (2014) The vehicle routing problem with time windows. In: Toth P, Vigo D (eds) Vehicle Routing: Problems, Methods, and Applications, 2nd edn. SIAM, Philadelphia, pp 119-159
Desrosiers J, Dumas Y, Solomon MM, Soumis F (1995) Time constrained routing and scheduling. In: Ball MO, Magnanti TL, Monma CL, Nemhauser FL (eds) Handbooks in Operations Research and Management Science, vol 8. Network Routing. Elsevier, Amsterdam, pp 35–139
Fachini RF, Armentano VA (2020a) Exact and heuristic dynamic programming algorithms for the traveling salesman problem with flexible time windows. Optim Lett 14:579–609
Fachini RF, Armentano VA (2020b) Logic-based Benders decomposition for the heterogeneous fixed fleet vehicle routing problem with time windows. Comput Ind Eng 148:106641
Fedex (2018) Fedex Annual Report 2018. http://s1.q4cdn.com/714383399/files/oar/2018/AnnualReport2018/assets/pdf/FedEx-Annual-Report-2018.pdf. Accessed 05 September 2018
Firouzi M, Jouzdani J, Shirouyehzad H, Paydar MM (2018) A rich heterogeneous fleet vehicle routing problem with flexible time windows: a case study of dairy supply chain. International Journal of Logistics Systems and Management 30(3):386–405
Fischetti M, Lodi A (2003) Local branching. Math. Program 98(1–3):23–47
Fischetti M, Monaci M (2014) Proximity search for 0–1 mixed-integer convex programming. J Heuristics 20(6):709–731
Fisher ML, Jaikumar R (1981) A generalized assignment heuristic for vehicle routing. Networks 11(2):109–124
Gendreau M, Jabali O, Rei W (2014) Stochastic vehicle routing problems. In: Toth P, Vigo D (eds) Vehicle Routing: Problems, Methods, and Applications, 2nd edn. SIAM, Philadelphia, pp 213–239
Gendreau M, Jabali O, Rei W (2016) 50th Anniversary Invited Article – Future Research Directions in Stochastic Vehicle Routing. Transport Sci 50(4):1163–1173
Glover F, Laguna M (1997) Tabu Search. Kluwer, Massachusetts
Golden BL, Raghavan S, Wasil EA (eds) (2008) The Vehicle Routing Problem: Latest Advances and New Challenges. Springer, New York
Gómez A, Mariño R, Akhavan-Tabatabaei R, Medaglia AL, Mendoza JE (2016) On modeling stochastic travel and service times in vehicle routing. Transport Sci 50(2):627–641
Hoff A, Andersson H, Christiansen M, Hasle G, Løkketangen A (2010) Industrial aspects and literature survey: Fleet composition and routing. Comput Oper Res 37(12):2041–2061
Hooker JN, Ottosson G (2003) Logic-Benders decomposition. Math. Program 96(1):33–60
Koç C, Bektas T, Jabali O, Laporte G (2016) Thirty years of heterogeneous vehicle routing. Eur J Oper Res 249(1):1–21
Lamghari A, Dimitrakopoulos R (2016) Progressive hedging applied as a metaheuristic to schedule production in open-pit mines accounting for reserve uncertainty. Eur J Oper Res 253(3):843–855
Li X, Tian P, Leung SC (2010) Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. Int J Prod Econ 125(1):137–145
Liu FH, Shen SY (1999a) A method for vehicle routing problem with multiple vehicle types and times windows. Proc Natl Sci Counc Repub China, Part A: Phys Sci Eng 23(4):526–536

Liu FH, Shen SY (1999b) The fleet size and mix routing problem with time windows. J Oper Res Soc 50(7):721–732

Prescott-Gagnon E, Desaulniers G, Rousseau LM (2009) A Branch-and-Price-Based Large Neighborhood Search Algorithm for the Vehicle Routing Problem with Time Windows. Networks 54(4):190–204

Quddus MA, Hossain NUI, Mohammad M, Jaradat RM, Roni MS (2017) Sustainable network design for multi-purpose pellet processing depots under biomass supply uncertainty. Comput Ind Eng 110:462–483

Rockafellar RT, Wets RJB (1991) Scenarios and Policy Aggregation in Optimization under Uncertainty. Math Oper Res 16(1):119–147

Russell RA, Urban TL (2008) Vehicle routing with soft time windows and erlang travel times. J Oper Res Soc 59(9):1220–1228

Salani M, Battarra M (2018) The opportunity cost of time window violations. EURO J Transp Logist 7(4):343–361

Salhi S, Rand GK (1989) The effect of ignoring routes when locating depots. Eur J Oper Res 39(2):150–156

Schneider M, Schwahn F, Vigo D (2017) Designing granular solution methods for routing problems with time windows. Eur J Oper Res 263(2):493–509

Solomon MM (1987) Algorithms for the vehicle routing and scheduling with time windows constraints. Oper Res 35(2):254–265

Susilawati S, Taylor MAP, Somenahalli SVC (2013) Distributions of travel time variability on urban roads. J Adv Transport 47(8):720–736

Tadikamalla PR (1980) A Look at the Burr and Related Distributions. Int Stat Rev 48(3):337–344

Taillard ÉD (1999) A heuristic column generation method for the heterogeneous fleet VRP. RAIRO 33(1):1–14

Taş D, Dellaert N, Van Woensel T, de Kok AG (2013) Vehicle routing problem with stochastic travel times including soft time windows and service costs. Comput Oper Res 40(1):214–224

Taş D, Dellaert N, Van Woensel T, de Kok AG (2014a) The time-dependent vehicle routing problem with soft time windows and stochastic travel times. Transport Res C-Emer 48:66–83

Taş D, Gendreau M, Dellaert N, Van Woensel T, de Kok AG (2014b) Vehicle routing with soft time windows and stochastic travel times: A column generation and branch-and-price solution approach. Eur J Oper Res 236(3):789–799

Taş D, Jabali O, Van Woensel T (2014c) A vehicle routing problem with flexible time windows. Comput Oper Res 52:39–54

Taylor MAP (2017) Fosgerau's travel time reliability ratio and the Burr distribution. Transport Res B-Meth 97:50–63

Teodorović D, Krcmar-Nozic E, Pavkovic G (1995) The mixed fleet stochastic vehicle routing problem. Transport Plan Techn 19(1):31–43

Thompson R, Taniguchi E, Yamada T (2011) Estimating the benefits of considering travel time variability in urban distribution. Transp Res Record 2238(1):86–96

Toth P, Vigo D (2003) The granular tabu search and its application to the vehicle-routing problem. INFORMS J Comput 15(4):333–346

Toth P, Vigo D (eds) (2014) Vehicle Routing: Problems, Methods, and Applications, 2nd edn. SIAM, Philadelphia

Wang Z, Lin WH (2017) Incorporating travel time uncertainty into the design of service regions for delivery/pickup problems with time windows. Exp Syst Appl 72(15):207–220

Zhang J, Lam WH, Chen BY (2013) A stochastic vehicle routing problem with travel time uncertainty: trade-off between cost and customer service. Netw Spat Econ 13(4):471–496

Authors and Affiliations

Ramon Faganello Fachini[1] · **Vinícius Amaral Armentano**[1] ·
Franklina Maria Bragion Toledo[2]

Vinícius Amaral Armentano
vinicius@densis.fee.unicamp.br

Franklina Maria Bragion Toledo
fran@icmc.usp.br

1 Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, SP 13083-852 Campinas, Brazil

2 Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, SP 13560-970 São Carlos, Brazil