

Soeffker, Ninja; Ulmer, Marlin W.; Mattfeld, Dirk C.

Article — Published Version

Balancing resources for dynamic vehicle routing with stochastic customer requests

OR Spectrum

Provided in Cooperation with:

Springer Nature

Suggested Citation: Soeffker, Ninja; Ulmer, Marlin W.; Mattfeld, Dirk C. (2024) : Balancing resources for dynamic vehicle routing with stochastic customer requests, OR Spectrum, ISSN 1436-6304, Springer, Berlin, Heidelberg, Vol. 46, Iss. 2, pp. 331-373, <https://doi.org/10.1007/s00291-024-00747-1>

This Version is available at:

<https://hdl.handle.net/10419/313828>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



Balancing resources for dynamic vehicle routing with stochastic customer requests

Ninja Soeffker¹ · Marlin W. Ulmer² · Dirk C. Mattfeld³

Received: 23 May 2022 / Accepted: 10 January 2024 / Published online: 15 February 2024
© The Author(s) 2024

Abstract

We consider a service provider performing pre-planned service for initially known customers with a fleet of vehicles, e.g., parcel delivery. During execution, new dynamic service requests occur, e.g., for parcel pickup. The goal of the service provider is to serve as many dynamic requests as possible while ensuring service of all initial customers. The allocation of initial services impacts the potential of serving dynamic requests. An allocation aiming on a time-efficient initial routing leads to minimal overall workload regarding the initial solution but may congest some vehicles that are unable to serve additional requests along their routes. An even workload division is less efficient but grants all vehicles flexibility for additional services. In this paper, we investigate the balance between efficiency and flexibility. For the initial customers, we modify a routing algorithm to allow a shift between efficient initial routing and evenly balanced workloads. For effective dynamic decision making with respect to the dynamic requests, we present value function approximations with different feature sets capturing vehicle workload in different levels of detail. We show that sacrificing some initial routing efficiency in favor of a balanced vehicle workload is a key factor for a flexible integration of later customer requests that leads to an average improvement of 10.75%. Further, we show when explicitly depicting heterogeneity in the vehicle workload by features of the value function approximation provides benefits and that the best choice of features leads to an average improvement of 5.71% compared to the worst feature choice.

Keywords Dynamic vehicle routing · Same-day service · Approximate dynamic programming · Value function approximation

1 Introduction

In the time of growing customer expectations and service orientation (United Parcel Service of America, Inc 2018), many service providers offer customers service at their homes, preferably even on the day of the service request. Examples for such

Extended author information available on the last page of the article

services are the same-day pickup of parcels or technicians performing instant repair services at customers' homes. In many cases, these spontaneous requests are integrated in the "standard" routes of vehicles performing pre-planned services. For example, parcel pickups are integrated in conventional parcel delivery routes.

While historically, such standard routes were optimized for efficiency, the combination of routes for initial customers and integration of new customers poses new challenges for companies. On the one hand, optimizing initial routes for efficiency may free resources such as the working time of vehicle drivers to serve more spontaneous dynamic requests. On the other hand, it may also congest some vehicles which are then inflexible and unable to add any close by requests to their initial route. In particular, we consider two extremes of finding a solution to the initial routing problem in the beginning of the day. The first one is an efficient solution to the static initial routing problem. Such a solution might be desirable if the initial customers require different services or driver qualifications and should be consolidated as much as possible. The second one is finding a solution in which the routing may appear less efficient, but provides more flexibility to integrate new requests. We refer to the first extreme as being efficient in terms of the initial routing, this solution may distribute the customers very unevenly to the fleet leading to routes of unequal lengths. We then refer to the second extreme as being balanced in terms of the initial routing as the goal is to obtain routes consisting of similar numbers of customers.

In this research, we investigate this balance between efficiency and flexibility for a dynamic vehicle routing problem with stochastic requests. For the initial routing, we present a modified savings algorithm allowing a controlled shift between balanced and efficient resource use. The modified savings algorithm can be used to find both extremes as well as intermediate solutions for the initial routing. For the dynamic assignments leading to a routing update, we rely on value function approximation (VFA), a method of approximate dynamic programming (Powell 2011) and reinforcement learning (Barto 1998) that learns the value of being in a certain (post-decision) state or of taking a particular decision. In our case, the VFA frequently simulates the process offline, stores post-decision state values and updates them after each simulation run. In the next simulation run, the new values are used for decision making. The trained VFA can then be applied instantly in the online execution to respond to new requests. Since the number of post-decision states is typically vast for complex problems of realistic size, VFAs usually aggregate them to a set of features. In the case of performing such an aggregation, selecting features is challenging as they need to ensure effective learning. Both too many and not enough features may impede the learning process. A particular new challenge of our problem is the consideration of resource balance. To address and analyze this balance, we present three new feature sets tailored to the specific needs of the problem. The first one considers the available resources of the vehicles individually, the second one considers them in a condensed way, and the last one includes the resource balance directly. In a comprehensive study, we show that our VFA outperforms benchmark heuristics significantly. We also show that even though balancing initial resources may be less efficient, the gained flexibility is very valuable when new customers request service. Furthermore, we highlight that depicting the resource balance or imbalance explicitly by means of the chosen features becomes particularly important for cases where

the fleet's resources are less balanced. In some cases, the service provider might be limited in choosing the initial routing flexibly. This could be the case if, for example, the initial customers are scheduled for complex services that require special equipment or a level of qualification that not all drivers possess. In such cases, it is helpful to choose the solution approach parameters in the best manner according to such restrictions.

Our work makes the following contributions. We are among the first to shed light on the interplay between initial routing and dynamic rerouting highlighting the tradeoff between efficiency and flexibility. We propose a solution approach that consists of one part for the initial routing and another part for the dynamic decisions. While both ideas have been considered individually in literature, both the combination as well as the impact on the solution quality have not been investigated. The developed approach for the initial routing can be parametrized to allow for either an efficient or a flexible initial routing. For the dynamic decisions, we develop a value function approximation methodology with several new feature combinations that differ in how they are able to depict the balance of the resources among the fleet. Further, we show which feature combinations are beneficial depending on the initial routing as well as depending on problem characteristics. We not only contribute to the literature by means of the proposed solution approach, but also analyze the interaction between the initial routing and the chosen features and show several important problem and methodology-specific insights.

This article is structured as follows. We review related literature on the problem setting and the existent research gap in Sect. 2 before describing and modeling the problem setting considered here in Sect. 3. The chosen value function approximation along with the proposed routing and aggregation choices are explained in Sect. 4. Sect. 5 then depicts the experimental setup and presents results. This article is then concluded in Sect. 6.

2 Related literature

In this section, we first provide a literature review on stochastic dynamic vehicle routing problems related to our problem setting. To this end, we characterize the literature based on different aspects regarding the initial routing, the acceptance mechanism of dynamic requests, and anticipation. In addition, we analyze the literature in terms of the considered application and dynamism before summarizing the research gap that our work aims to fill.

The literature on dynamic vehicle routing problems is rather vast, but the research interest is still growing due to the technological possibilities to track vehicles, goods, and information as well as increasing communication possibilities. For a general overview on dynamic vehicle routing problems, the interested reader is referred to Pillac et al. (2013) and Psaraftis et al. (2016), for a more specific overview on dynamic and stochastic vehicle routing problems, the reader is referred to Ritzinger et al. (2016) and Soeffker et al. (2022). In the recent years, literature on dynamic vehicle routing has been extended by also considering customer preferences which enables to provide different service options to customers. While we

consider the simplest version of demand management by accepting or rejecting customer requests, the reader interested in more sophisticated approaches is referred to Fleckenstein et al. (2023) for a survey on demand management and vehicle routing.

While there are different ways to categorize literature on stochastic dynamic vehicle routing, there are three aspects of particular interest for this article. The first interesting aspect is how the initial routing for the initially known customers is conducted. The second aspect refers to whether the acceptance mechanism is selective or non-selective about the dynamic customer requests. The third aspect refers to the level of anticipation of the applied solution approach. The first and third aspect refer to whether the solution approach is anticipatory with regard to the initial and the dynamic customers, respectively. The second aspect rather refers to a problem characteristic as it describes whether customers have to be served if they can be served. In the following, the literature is discussed with regard to these aspects.

The first aspect relevant for categorizing literature concerns how initial customers are allocated to the vehicles before the start of the time horizon and before the dynamic customers arrive. Most early articles on dynamic vehicle routing problems, like Kilby et al. (1998), Montemanni et al. (2005), or Hanshar and Ombuki-Berman (2007), solve static versions of the problem in specific time points. That is, all currently known customers are included in a classical VRP using objective functions like minimum travel distance or time. If the share of initial customers is small in comparison to the time capacity available, such a procedure tends to create only a small number of initial routes which are rather long while other routes are empty. We refer to such approaches as being not anticipatory in their initial routing decision. In contrast to considering only currently available information, articles like Bent and Van Hentenryck (2004) or Zhang et al. (2023) use scenarios that include artificial customer requests that serve as a possible realization of the future thus being anticipatory in the initial routing. Although the artificial requests are removed from the solutions, the routes are already designed in a way that incorporating potential future requests is enabled. Such initial routes are more likely to be rather balanced in terms of the resource usage. For a more detailed overview, how the initial routing is created in the literature, the reader is referred to Zhang et al. (2023).

In this article, our contribution to the literature with regard to the initial routing is twofold. On the one hand, we propose an approach that uses a parameter to control the balance of the initial resources. On the other hand, we also explicitly conduct analyses on the impact of the initial routing on the solution quality as well as on the interplay with the dynamic decision making.

The second aspect concerns whether the problem definition allows the decision about acceptances of new customer requests. We differentiate selective and non-selective acceptance mechanisms, that is, approaches that accept all requests that can be accepted without violating constraints and approaches that may also decide to reject a customer request despite feasibility. From a customer perspective, it seems more reasonable to serve customers on a first-come-first-serve basis. That would lead to accepting customers if their request can be served without violating any constraints like work hours or vehicle ranges. However, some customers might require more resources to be served, for example because they are located far away from other customer locations. In this case, a service provider

may end up serving more customers when a selective acceptance mechanism is applied. While some customers may benefit from this, this is usually to the disadvantage of those customers in “inconvenient” locations (Soeffker et al. 2017). Both approaches therefore have advantages and disadvantages.

The third aspect for differentiation of literature is the level of anticipation of the applied solution approach with respect to the dynamic decisions. Non-anticipatory approaches only use the current information as a basis for decision making. In contrast, anticipatory approaches try to incorporate information about potential future developments in the external information, about future decisions, or a combination of both.

Some examples of non-selective acceptance mechanisms in combination with non-anticipatory approaches are provided in the following. Gendreau et al. (1999) suggest a parallel tabu search that improves a set of current plans until a new request is revealed. If the new request can be accommodated in any of the plans, the request is accepted and all plans not complying with the insertion are deleted. Only if none of the plans can be modified to include the new request, it is rejected. The focus therefore is on the optimization of the routing plans as the acceptance decision is based on feasibility. For a dynamic dial-a-ride problem, Attanasio et al. (2004) also suggest to accept requests if they can be inserted feasibly. Upon insertion of the new request, the plan is then improved by means of parallel tabu search. Ichoua et al. (2000) extend the idea of a parallel tabu search by allowing a diversion of the route. These approaches have to accept requests if they can and then focus on improving the routing plans for all known customers. The disadvantage is that they are not anticipatory and may accept customer requests at all cost, substantially obstructing the vehicles’ routes and eventually leading to a limited number of services overall.

In the literature using anticipatory approaches with non-selective acceptance mechanisms, a variety of different concepts is proposed. Bent and Van Hentenryck (2004) suggest an approach where a set of plans is maintained and customer requests are accepted if any of them can feasibly include the new request. However, artificial customers are sampled and the routing first includes these customers that are then removed. The plan actually carried out is the one that is most similar to the other ones leading to a focus on flexible routing plans. For a similar problem, Bent and Van Hentenryck (2007) extend the sampling idea by Bent and Van Hentenryck (2004) by a waiting and a relocation heuristic. Ichoua et al. (2006) extend the approach of Gendreau et al. (1999) by a waiting heuristic that considers information about the chances for future customer requests in certain parts of the service area. Thomas (2007) propose heuristics based on analytically derived properties of the optimal solution for a problem similar to the one discussed here. In the heuristics proposed, knowledge about future requests is assumed, and in each decision the maximum number of customers that requested service and can be inserted feasibly is accepted. The same problem is discussed in Ghiani et al. (2012) where two approaches are compared that are both anticipatory, but accept as many customer requests in a decision point as possible while not violating the time limit. For the emerging field of same-day delivery, Voccia et al. (2019) combine a sampling strategy with a waiting strategy. While the approaches aim to create a routing that

anticipates future customers, they are applied to problems that do not allow a selection of customers which potentially leads to fewer customers that can be visited.

The last category are approaches that are anticipatory with selective acceptance mechanisms like the one we are proposing as it provides the most flexibility for decision making. In Angelelli et al. (2009), a multi-period pick-up problem is considered where some customers can be rejected for same-day service, but will be served on the following day. Azi et al. (2012) consider a delivery problem where customers can be accepted or rejected for service. The decision about acceptance is based on scenarios like in Bent and Van Hentenryck (2004). In Klapp et al. (2018), a dynamic dispatch waves problem is considered where requests are not rejected immediately, but can simply not be served which is a special case of a selective approach. Ulmer et al. (2018) and Soeffker et al. (2019) consider a one-vehicle problem where customers can be accepted or rejected for same-day service, Ulmer et al. (2018) extend this problem to a multi-period problem where customers are rejected for same-day service and then have to be served on the next day. Angelelli et al. (2021) consider a similar problem, but acceptances and rejections for their stochastic and dynamic orienteering problem are handled on the day before the actual conduction of the route. The problem described in this article is adapted from Ulmer et al. (2018) and Soeffker et al. (2019) by extending the problem setting to a fleet of vehicles where the conduction of the route takes place while new requests for the same day arrive. For a same-day delivery problem setting, Klein and Steinhardt (2023) combine a sampling strategy and a demand management approach. Also for same-day delivery applications, Bozanta et al. (2022) and Chen et al. (2022) propose to use reinforcement learning approaches and analyze the impact of different features considered in their solution approaches. While these articles consider problems in which a rejection is allowed and that apply anticipatory solution approaches, most of them consider a single vehicle, some of them do not consider initial customers, or they apply a non-anticipatory routing of those initial customers.

Zhang et al. (2023) consider a closely related problem setting to ours and we highlight the differences in the following. To this end, we draw on the classifications proposed in Fleckenstein et al. (2023) and Soeffker et al. (2022). In the classification of Fleckenstein et al. (2023), the booking and service horizon overlap in both settings, the constraints consider a fleet in both cases, and the processing happens in real-time for both cases. However, Zhang et al. (2023) consider the objective function to maximize the number of served requests while our objective is to maximize revenue generated by accepting requests. When comparing the considered problem settings based on the source of uncertainty as described in Soeffker et al. (2022), both problem settings consider stochastic customer requests, which is an uncertainty in the demand. However, the characteristics of these requests differ in the two problem settings. The set of feasible locations is limited to a given node set in Zhang et al. (2023) while it is not in our problem setting. On the other hand, the requests in Zhang et al. (2023) come with varying service times which is not the case in our problem setting. All customers generate the same contribution towards the objective function in Zhang et al. (2023), while customer requests vary in the associated reward in our problem setting. In addition, to approach their problem setting, Zhang et al. (2023) design both an anticipatory initial routing approach and an anticipatory dynamic decision approach. However, they apply another

solution approach where considering different feature combinations or analyzing the impact of features on the solution quality or their interaction with the initial routing is not applicable. Our work thus contributes to the literature by filling this research gap.

In Table 1, we summarize the described literature according to relevant problem characteristics, the anticipation in the considered approach, and the conducted analysis. While most early literature did not allow anticipatory rejections of customers, later research did incorporate this aspect. However, to allow for anticipatory dynamic decision making, most of the recent literature focuses on a single vehicle. The initial routing was not considered in an anticipatory manner in most literature. In recent years, there has been a trend towards considering same-day delivery applications in research. While most of these articles now consider a fleet and apply anticipatory solution approaches, they consider problems without initial customers. They therefore do not have to find an initial routing and they cannot analyze the interplay between the initial routing and the dynamic decision making.

In addition to the classification regarding problem, solution approach, and analysis aspects, we also provide an overview of the degree of dynamism (dod, Larsen et al. 2002) associated with typical applications considered in the literature. The dod describes the share of dynamic requests in the total number of requests. In Fig. 1, we provide a similar literature list as in Table 1 and indicate the application field along with the degree of dynamism considered in the computational studies. There are two large groups of applications: courier/express/parcel providers and same-day delivery (including here e-groceries and food delivery). In addition to these main groups, there is also a large body of literature not focusing very much on a particular application and one article with a dial-a-ride application. It can be observed that there has been a recent trend to focus on same-day delivery applications which typically do not include any known customer requests at all. Apart from same-day delivery applications, most articles consider degrees of dynamism between 0.5 and 0.75.

Based on the literature review presented above with the resulting literature gap, we consider in this article a routing problem with a fleet, with initial customers, with the option of anticipatory rejections. As a solution approach, we propose an anticipatory value function approximation and we propose an approach that can be anticipatory with regard to the initial routing. While a closely related problem setting has been solved with an anticipatory solution approach in the literature, we extend the literature by analyzing the interplay between the initial and the dynamic decision making and proposing and evaluating different feature combinations for the value function approximation. Based on typical applications of the considered problem setting, we further compare the performance of the value function approximation for different degrees of dynamism.

3 Dynamic vehicle routing problem with stochastic customer requests

In this section, we first describe the considered problem setting in Sect. 3.1. After the problem description, we formally model the problem as a Markov decision process model in Sect. 3.2.

Table 1 Classification of literature with respect to considered problem characteristics, anticipation within the applied solution approach, and conducted analysis

	Problem			Anticip.		Analysis	
	Fleet	Anticipatory rejection	Initial customers	Initial routing	Dynamic decisions	Interplay initial & dynamic	Feature combinations
Kilby et al. (1998)	✓		✓				
Gendreau et al. (1999)	✓		✓				
Ichoua et al. (2000)	✓		✓				
Attanasio et al. (2004)	✓		✓				
Bent and Van Hentenryck (2004)	✓		✓	✓	✓		
Montemanni et al. (2005)	✓		✓				
Hvattum et al. (2006)	✓		✓	✓	✓		
Ichoua et al. (2006)	✓				✓		
Bent and Van Hentenryck (2007)			✓	✓	✓		
Hanshar and Ombuki-Berman (2007)	✓		✓				
Thomas (2007)		✓	✓		✓		
Angellelli et al. (2009)	✓	✓	✓		✓		
Azi et al. (2012)	✓	✓	✓		✓		
Ghiani et al. (2012)		✓	✓	✓	✓		
Klapp et al. (2018)		✓			✓		
Ulmer et al. (2018)		✓	✓		✓		
Ulmer et al. (2018)		✓	✓		✓		
Soeffker et al. (2019)		✓	✓		✓		
Voccia et al. (2019)	✓				✓		
Angellelli et al. (2021)		✓			✓		
Basso et al. (2022)				✓	✓		
Bozanta et al. (2022)	✓	✓	✓		✓		✓
Chen et al. (2022)	✓	✓			✓		✓

Table 1 (continued)

	Problem	Anticip.			Analysis	
		Anticipatory rejection	Initial customers	Initial routing	Dynamic decisions	Interplay initial & dynamic
	Fleet					Feature combinations
Giménez-Palacios et al. (2022)	✓		✓			
Zhang et al. (2023)	✓	✓	✓	✓	✓	
Klein and Steinhardt (2023)	✓	✓			✓	
This work	✓	✓	✓	✓	✓	✓

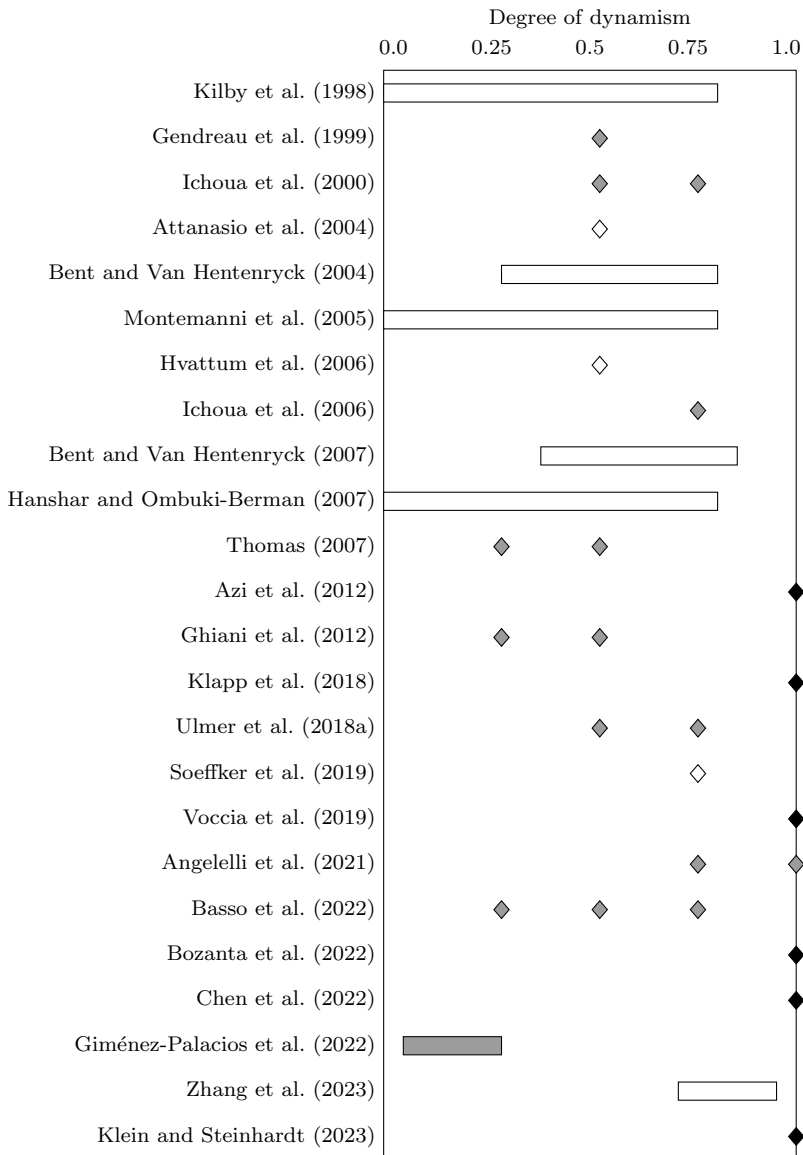


Fig. 1 Overview of applied degrees of dynamism, color of filling indicates application (black: same-day delivery/e-groceries/food delivery, gray: pickup in courier/express/parcel applications, white: dial-a-ride or no application indicated)

3.1 Problem description

In the considered dynamic vehicle routing problem with stochastic customer requests (DVRPSR), drivers of a service provider navigate a fleet of vehicles from a depot through a given area to serve customer requests during a service horizon

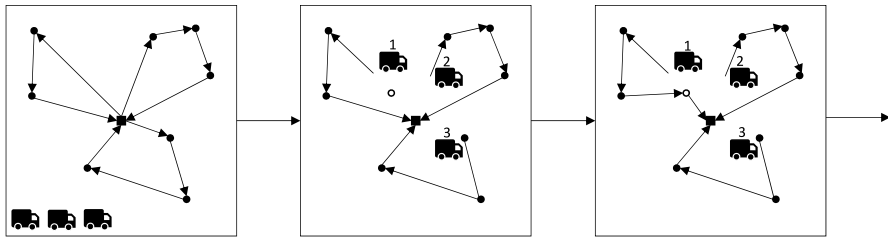


Fig. 2 Example of beginning of service horizon: Examples for initial routing (left panel), new customer request (middle panel), acceptance and inclusion of the new customer request (right panel)

and to return to the depot until the end of the service horizon. Both driving times and service times at customer locations are assumed to be known and deterministic. Some of the service requests are known in advance (early request customers) and for these initial customers, an initial routing is created. Other requests only occur during the service horizon (late request customers), these are dynamically accepted and included into a route (or the request for same-day service is rejected).

The temporal and spatial distributions of the service requests are known. Late request customers have a value that can be interpreted as the revenue for the service provider for serving the customer on the same day. When a new customer requests service, the service provider is informed about the location and the associated revenue of the request. Then, the service provider decides whether to accept or reject the request and which vehicle is assigned to serve the customer request. We assume that the acceptance or rejection and assignment decisions are fixed once they are made. That is, accepted requests have to be served. The objective is to maximize the expected revenue accumulated over the day.

In Fig. 2, a small example of the beginning of the time horizon is depicted. In the left part of the figure, an exemplary initial routing is depicted where three routes are created for the three vehicles. All eight initial customers are included in the initial routes. In the middle part of the figure, the service horizon is ongoing. The vehicles are already conducting their routes, and a new customer requested service. This new customer request is depicted as a small white circle. In the right part of the figure, the new customer request was accepted and included into the first vehicle's route.

3.2 Markov decision process model

In this section, we describe the given problem setting by means of a formal model.

Initial routing. For the given problem, each early request customer $c \in C^e$ is assigned to a vehicle route θ_0^v of a vehicle $v \in F$ before the service horizon starts. For the fleet of vehicles F , the set of routes is denoted by Θ_0 . The routes need to comply with several classical VRP constraints that are described briefly in the following. All vehicle routes start and end at the depot, the number of routes cannot exceed the number of available vehicles. Each customer $c \in C^e$ has to be visited by exactly one route. All routes have to end at the depot before or at the end of the time horizon T .

Overview. Due to the dynamic nature of the DVRPSR, the dynamic decision process described in Sect. 3.1 is modeled as a finite Markov decision process (MDP). The MDP contains sequences of states, decisions, post-decision states, and transitions. In states s_k in decision points $k = 1, \dots, K$, a decision x_k from a set of possible decisions $X_k(s_k)$ is chosen. Post-decision states s_k^x describe the state immediately following the combination of state s_k and decision x_k and thus do not cover any effects other than those caused by the decision. The concept of post-decision states therefore allows to split the deterministic effects due to the decision on the one side and the effects caused by the stochastic process on the other side. After the deterministic post-decision state, transitions induced by stochastic realizations ω_{k+1} lead to the next decision point $k + 1$ resulting in a new decision state s_{k+1} . This continues until the final state s_K is reached.

States. States s_k in the MDP contain all information that is necessary to describe the decision situation sufficiently. For the described problem, s_k for $k = 1, \dots, K - 1$ therefore contains information about the time t_k where $t = 0, \dots, T$ in basic units of \bar{t} , e.g. minutes or seconds, in the service horizon T , the planned routes of all vehicles Θ_k , and information about the new customer request c_k from the set C^l of late request customers. This information about c_k consists of a location l_k and revenue u_k associated with the request.

Initial state. Notably, the initial state s_0 and the final state s_K are exceptions as they do not contain new customer requests and therefore no decision options. The initial state s_0 contains the time $t_0 = 0$ and the route set Θ_0 after the routing of the early request customers is planned. The remaining resources after the initial state are therefore not the same on each day and in each initial state.

Decisions and rewards. With the start of the service horizon, all service vehicles follow their routes θ^v and those with a non-empty route leave the depot to visit customers. A decision point k is triggered by a new service request from a late request customer $c_k \in C^l$. The service provider is informed about the location l_k and the associated revenue u_k for the request. The dispatcher of the service provider then has to choose a decision $x_k \in X_k(s_k)$. Such a decision determines whether the customer request will be served in the same service horizon or whether it is rejected. In the case of an acceptance, the request has to be included in exactly one route $\theta^v \in \Theta$. The decision options $X_k(s_k)$ can contain options in which the customer request is rejected, that is, the set of customers to be served is unchanged, as well as options in which the set of customers includes all previous customers and the new customer request. For both the rejection and the acceptance of the request, different routing options may be available as long as the vehicles can visit and serve all customer requests and return to the depot before the end of the service horizon. Also, we assume that upon starting the service at one customer, the next customer in the route will already be informed about the upcoming service, the new request can therefore not be inserted before that informed customer. The set of routes after the implementation of the decision x_k and thus before the occurrence of new customer requests and before the conduction of the routes is denoted as Θ_k^x . Associated with each decision option is a reward $R(s_k, x_k)$ that describes the immediate contribution towards the objective function. Since the objective is to maximize the expected overall revenue of the service provider, the reward $R(s_k, x_k)$ is the revenue u_k of the customer,

so $R(s_k, x_k) = u_k$, in case of an acceptance or zero, so $R(s_k, x_k) = 0$, if the customer request is rejected.

Post-decision state and transition. The deterministic post-decision state (PDS) consists of a time $t_k^x = t_k$ and the updated set of routes Θ_k^x . After the PDS, a stochastic transition leads to the next decision point $k + 1$ that occurs upon the request of a new customer request $c_{k+1} \in C^l$. Until this new decision point, the vehicles follow their assigned routes θ^v and customers whose service has been completed are removed from the routes. The final decision point s_K occurs when the service horizon is over, that is, when $t_K = T$ and all service vehicles returned to the depot.

Objective function. The objective function is to maximize the sum of revenue accumulated throughout the service horizon through the acceptance of late request customers. A solution to an MDP is a policy π that describes which decision $x_k \in X_k(s_k)$ to choose in $s_k \in S$. The optimal policy π^* is the one maximizing the objective function, that is,

$$\pi^* = \arg \max_{\pi \in \Pi} \mathbb{E} \left[\sum_{k=0}^K R(s_k, X_k^{\pi^*}(s_k)) \mid s_0 \right]. \quad (1)$$

This is equivalent to choosing in each decision point k the decision that maximizes the sum of the immediate and the expected future reward assuming that the optimal policy π^* is followed afterward (Bellman's Equation, Bellman 1957). This means that

$$X_k^{\pi^*}(s_k) = \arg \max_{x_k \in X(s_k)} \left\{ R(s_k, x_k) + \mathbb{E} \left[\sum_{j=k+1}^K R(s_j, X_j^{\pi^*}(s_j)) \mid (s_k, x_k) \right] \right\}. \quad (2)$$

The second part, the expected future rewards following the combination of state s_k and decision x_k , can also be denoted as the value of the post-decision state $V(s_k^x)$.

4 Solution approach

In the problem considered in this article, there are two different types of decisions about the usage of the available resources. The first type arises before the service vehicles leave the depot. The customers that are already known before the time horizon need to be assigned to vehicles and an initial routing needs to be determined for each vehicle. This initial routing impacts the remainder of the decision horizon as it defines how parts of the resources are used. Then, throughout the time horizon, new customers request service. When such a request occurs, the service provider needs to determine whether service will be offered to this customer or not. In case of a service offering, the routing of the vehicles needs to be adapted to incorporate the new customer which has an impact on the remaining resources.

The problem setting therefore requires a careful balancing of resources throughout the time horizon. If the initial routing is constructed in a way that minimizes travel distances, in the extreme case, some vehicles might be idling at the depot

while others might not have capacities to integrate any future requests. We therefore explicitly incorporate the balancing of resources within the initial routing part of the solution approach. In the solution approach, we propose a parameter that defines whether we require a rather efficient or a rather balanced initial routing.

After creating the initial routes, we further assume that keeping track of the resource balance might be crucial for the dynamic decision making of the dynamic part of the problem. Solving the dynamic vehicle routing problem with stochastic customer requests optimally would require to compute the value of each post-decision state s_k^x (or state-decision pair (s_k, x_k)). Unfortunately, the problem suffers from the “curses of dimensionality” (Powell 2011). The *transition space* depicts all possibilities for transitioning from one state-decision pair to a next state and contains information about the probabilities for any of these transitions. In the described problem setting, new customer requests can occur anywhere in the service area and at any point in time meaning that the number of possible next states is infinite due to the continuous distributions. Thus, the transition space is too large to compute the value of states. In addition, the *state space* (or post-decision state space as in this case) is too large to determine values for all of them individually for the same reason of customer requests being possible anywhere in the service area. However, because states and post-decision states contain not only locations of customers, but entire routes, the curse of the (post-decision) state space is even more complex. Finally, the *decision space* is too large to evaluate all possible options in a decision point. While the decision space contains a finite number of options, it contains for both the acceptance and the rejection of the new request an entire solution space of the underlying routing problem. In summary, it is not possible to recursively compute the values of all post-decision states and, thus, the optimal policy.

Therefore, we propose a value function approximation that uses offline simulation to approximate the values for all post-decision states. However, since the post-decision state space is too large, we propose three feature aggregation choices to summarize important information about the post-decision state. In particular, we apply features that allow measuring the resource use itself to support efficient resource usage. However, to investigate the importance of representing the resource balance, the three different feature choices vary from depicting the balance of the fleet not at all over depicting the fleet’s balance generally to depicting the balance by representing each vehicle individually. Further, due to the remaining size of the post-decision state space, we also use a partitioning approach to only represent relevant areas of the post-decision state space. In the classification of Soeffker et al. (2022), the solution approach VFA falls in the category prescriptive and internal use of information model. In the classification of Fleckenstein et al. (2023), the feasibility check of decision options is route-based, the cost estimation is predictive, and the demand control considers accept and reject.

In the remainder of this section, we describe our solution approach in more detail. First, an overview on the solution approach is provided in Sect. 4.1. Then, we describe the construction of the initial routes in Sect. 4.2 and explain the initial routing parameter that defines the balance of the initial routing. Further, we describe the components addressing the three curses of dimensionality in Sects. 4.3, 4.4, and 4.5. The proposed solution approach explicitly incorporates

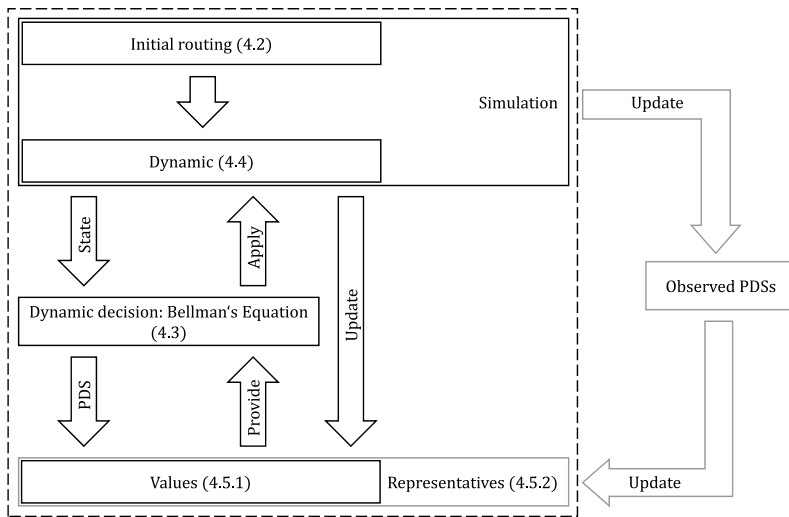


Fig. 3 Schematic representation of solution approach that simulates the entire horizon, draws on values stored in representatives, updates the values, and contains a component that updates the representatives, adapted from Soeffker et al. (2019)

the initial routing. Further, it extends solution approaches from the literature from a single-vehicle case to a multi-vehicle case by considering more decision options and proposing three feature aggregation choices.

4.1 Solution approach overview

To provide an overview of the overall solution approach, we depict it in Fig. 3. In the following sections, we motivate and explain the individual parts in detail. Here, we briefly describe the process depicted in Fig. 3.

The solution approach repeatedly simulates the entire decision horizon which includes the initial routing and the dynamic decision making (upper part of the figure). The initial routing is described in detail in Sect. 4.2, the dynamic decision making in Sect. 4.4. In a decision point of the dynamic decision making, represented by “State”, a decision according to Bellman’s Equation (middle part and described in Sect. 4.3) is chosen. This happens based on the immediate rewards and the expected values of the resulting post-decision states (“PDS”) that are stored in representatives (bottom part). After a decision is chosen and applied, the simulation continues until the end of the horizon. Then, the information about the actually obtained rewards during the simulated decision horizon following all of the observed post-decision states is used to update the values in the representatives, which is described in Sect. 4.5.1. The applied adaptive state space partitioning approach (ASSP, Soeffker et al. (2019), represented by the right part of Fig. 3) collects observed post-decision states over several simulation runs and

uses them in specific intervals to update the representatives that are used to store values (right part of the figure and described in Sect. 4.5.2).

4.2 Initial routing

The initial routing is a classical VRP except for its objective function and the fact that it sets the resources for the remaining dynamic process. Due to the dynamic nature of the problem, it is not necessarily desirable to find the optimal solution in terms of the route lengths. This is because suboptimal routes are longer which means that the insertion of new requests can be easier in later decision points. One classical approach for the VRP is the Clarke Wright Savings algorithm (Clarke and Wright 1964) that creates individual routes for each customer and then merges routes if time capacities are not violated and the overall routes' length can be reduced. However, for the problem at hand, not all customers are initially present. The functionality of the Savings algorithm would lead to an imbalanced allocation of the fleet's time resources with some vehicles fully utilized for the entire day and others (nearly) empty.

In that case, only one vehicle route may spread over the service area while the other vehicles are waiting in the depot. From an intuition perspective, this seems not desirable as the vehicles in the depot spend their resources with waiting instead of serving customers while other vehicle routes might be able to include further customer requests. In such a case, a new request could be included in the longest route for efficiency reasons while the other routes remain empty and the vehicles remain in the depot. As an alternative, a rather equal allocation of customers to vehicles could be applied in which the vehicles' resources are rather balanced. For this, we could alternatively apply a fully balanced Savings algorithm where we compute a maximum number of customer requests per route as

$$z = \left\lceil \frac{|C^e|}{|F|} \right\rceil \quad (3)$$

with $|C^e|$ being the number of initial customers to be visited and $|F|$ the number of vehicles. The fully balanced Savings algorithm is then not allowed to merge routes if the resulting route exceeds this number. Notably, in rare cases, for example with high demand or when merging routes is not efficient, this can lead to more routes than vehicles. If this happens, we enforce to merge the shortest routes. If at least one of the resulting routes exceeds the time horizon, the instance is considered to be infeasible. Thus, the dynamic decision making is skipped and a reward of 0 is obtained over the entire day. Since the limitation of merging routes restricts the Savings algorithm, the resulting routes might be in total longer which might seem less efficient. To allow a compromise between the classical Savings and the fully balanced Savings, we propose a modified Savings algorithm that uses a tuning parameter m between 0 (classical Savings) and 1 (rather balanced initial routes) and then compute the maximum number of customers per route as

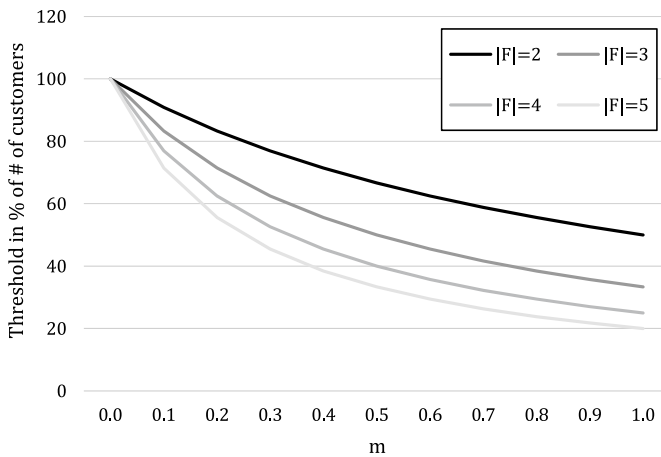


Fig. 4 Theoretical threshold for modified Savings (in percentage of number of customers)

$$z = \left\lceil \frac{|C^e|}{1 + m \cdot (|F| - 1)} \right\rceil. \quad (4)$$

For $m = 0$, this results in no threshold, that is, the Savings algorithm can merge all routes. For $m = 1$, this results in the previous equation and approximately the same number of customers per route. In Fig. 4, we show the theoretical impact of this parameter on the maximum number of customers per route as a share of the total number of customers to be included.

It can be seen that the threshold is not linearly dependent of the parameter. However, by choosing different values for parameter m , all thresholds between potentially allowing all customers in one route and requiring a rather balanced routing can be achieved. In the computational study, we evaluate the impact of the initial routing by means of parameter m .

4.3 Decision space

In the given problem setting, decisions contain two parts: the acceptance and the routing. Theoretically, the decision space considers many options for the acceptance of the new customer request and also many options for the rejection of the new request. Technically, this includes not only all options to insert the new customer into the current routes, but also all modifications of the current routes except for those that change the assignment of the already fixed customers. To avoid this curse of dimensionality, we reduce the size of the decision space by not modifying the routing in case of a rejection and by relying on a heuristic procedure for the routing in case of an acceptance.

In this work, we apply a cheapest insertion heuristic as described in Rosenkrantz et al. (1974) and check the position and the resulting post-decision state of an insertion in the cheapest insertion in each vehicle's route. Therefore, for

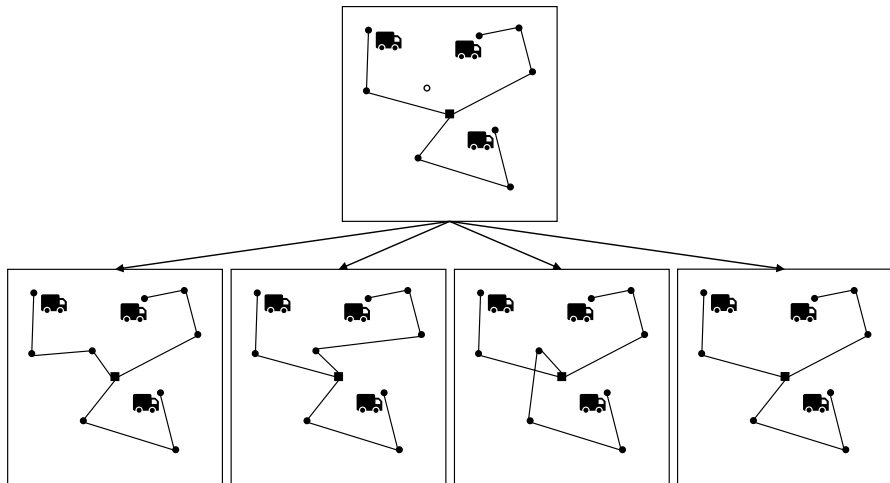


Fig. 5 Example for a decision situation (upper part) and four different decision options: Acceptance of request with inclusion of new request in upper left route (bottom, left figure), in upper right route (bottom, second from left figure), in bottom route (bottom, second from right figure), and rejection of request (bottom, right figure)

m vehicles, $m + 1$ decision options are available. Figure 5 depicts an exemplary set of options for a new customer request. In the first three options, the customer request is included in one of the existing routes. In the last option, it is not accepted and therefore not part of one of the routes.

4.4 Transition space

As customers may request service at any point in the given service area and at any point in time, the transition space is too large making it impossible to find the optimal policy π^* by means of dynamic programming. Value function approximation, a method of approximate dynamic programming (ADP, Powell 2011) approximates the optimal policy and the values of (post-decision) states. Due to the challenging size of the transition space, it is helpful to separate the effects of the decision and the stochastic realization that both affect the transition between a state and the next state. Therefore, we work with a model that includes a post-decision state that captures the effects of the decision, but not yet the effects of the stochastic realization. VFA then approximates the optimal policy by replacing the actual value of a state or, as in our case, of a post-decision state $V(s_k^x)$ with an approximation $\hat{V}(s_k^x)$. In this work, we apply a value function approximation relying on offline simulation that aims to find reliable approximations of the values of the post-decision states. We apply a non-parametric value function

approximation, a very generic approach that does not expect the relationship between post-decision state features and value of the post-decision state to follow a certain functional form.¹ Such a non-parametric VFA can be organized in the form of a lookup table, but it can also be more flexible. If the values $V(s_k^x)$ are replaced by $\hat{V}(s_k^x)$, the approximated Bellman's Equation then reads as follows:

$$X_k^{\pi^*}(S_k) = \arg \max_{x \in X(S_k)} \{R(S_k, x) + \hat{V}(s_k^x)\}. \quad (5)$$

4.5 State space

For an approximation of each PDS's value, the post-decision state space is too large. To overcome this problem, there are two measures taken in this work. The first one is to aggregate information of the original MDP in order to reduce the size of the post-decision state space. After this, the question of an efficient state space partitioning arises as some areas of the space may be more important for the decision making than others. In the following Sects. 4.5.1 and 4.5.2, both measures are described in more detail.

4.5.1 Aggregation

In this section, we discuss potential options to aggregate the information of the original MDP to obtain an aggregated post-decision state space. As described earlier, the post-decision state in the original MDP contains information about the time and the routes of all vehicles. Using this for a non-parametric value function approximation is not practicable as the number of customers is not limited and customers can be located anywhere leading to an infinite post-decision space. In previous work that considered only one vehicle (Ulmer et al. 2018), time and slack, the remaining budget a vehicle has after returning to the depot and before the end of the horizon, were discussed to be appropriate choices as time is the resource to be managed in this problem. Here, we extend this line of reasoning and propose and compare three different aggregation levels that all are based on information about time. The three aggregation options depict the fleet resource balancing in different degrees of detail. In Fig. 6, one post-decision state is shown that contains information about all routes of the vehicles. In the following, we describe the three different options to aggregate information based on temporal information that are depicted in Fig. 6.

The first, most straightforward, approach is to use the time and one dimension per vehicle describing the slack for each vehicle. This aggregation approach, depicted in the left subfigure of Fig. 6 is the most detailed one as it contains explicit information about each vehicle in the fleet. The second approach, shown in the middle subfigure of Fig. 6, is the least detailed as it computes the mean slack over all vehicles

¹ We present results for a related problem setting with a single vehicle in Appendix A.3 to demonstrate why we are not applying a parametric VFA assuming a linear relationship between the features and the value.

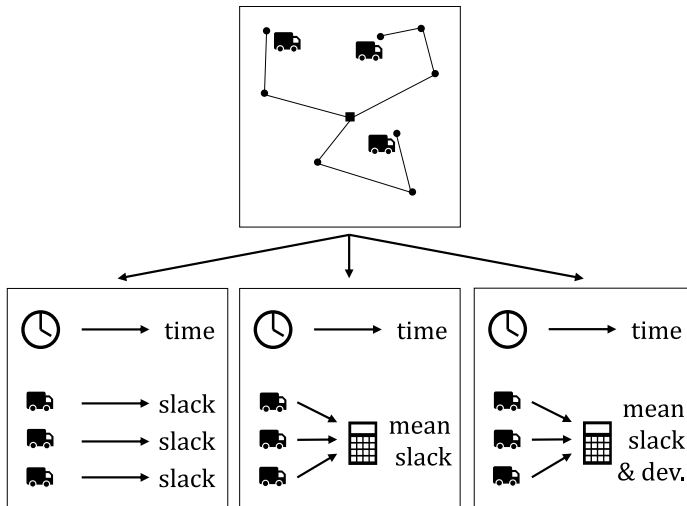


Fig. 6 Original post-decision state (top) and different aggregation options for an exemplary post-decision state: time and individual slacks (bottom, left), time and mean slack (bottom, middle), time and combination of mean slack and deviation (bottom, right)

and uses therefore only one aggregated feature for the whole fleet which is a rather implicit form of information. The third approach, which is depicted in the right sub-figure of Fig. 6, also uses time and mean slack, but adds a third dimension which is the standard deviation from the mean slack. This option therefore incorporates not only the information about the mean, but also information about the spread within the fleet.

4.5.2 Partitioning

After aggregating the post-decision state space of the original MDP to a smaller one, the resulting aggregated post-decision state space is still quite large. For the sake of simplicity, we refer to the aggregated feature vectors as post-decision states although they do not contain the original information. For a problem with three vehicles, a time horizon of 480 min and a discretization to minutes, the aggregated post-decision state space would consist of $480 \cdot 480 = 230,400$ aggregated post-decision states with an aggregation to time and mean slack, but of $5.3 \cdot 10^{10}$ aggregated post-decision states with an aggregation to time and individual slacks. If each PDS would be evaluated separately, this would require visiting each PDS multiple times during the approximation process and a high computational effort. Further, in many problems, some areas of the PDS space are not even accessible by a state-decision combination due to logical reasons. Such areas will never be visited and their representation in the post-decision state space is therefore not helpful. Also, depending on the problem and the PDS space, some PDSs may be similar enough to treat them equally and evaluate them as if they were the same. Especially because of the computational problems, it is very common to partition the state space in some form

to reduce the number of values that have to be approximated. The most traditional and intuitive approaches are approaches that partition the state space in static partitions of the same size, usually like a look-up table with certain interval sizes (Powell 2011). The larger these interval sizes are, the lower the computational burden, but the higher the chance to treat PDSs equally that are rather dissimilar. However, small interval sizes not only come with high computational efforts, but also with the chance of insufficient approximations if partitions are not visited often enough. One other partitioning approach is to combine different approximation architectures with different levels of detail (weighted look-up tables, Powell 2011). Another option is to use the knowledge from the simulation that is used to approximate the values to find a partitioning iteratively (Lee and Lau 2004; Ikonen et al. 2016; Ulmer et al. 2018; Agussurja et al. 2019; Soeffker et al. 2019). For the partitioning of the post-decision state space, we apply an adaptive state space partitioning approach (ASSP, Soeffker et al. 2019). ASSP aims to find those areas of the post-decision state space that are relevant for the decision making in an adaptive manner. In Soeffker et al. (2019), it was applied to a simplified problem with a two-dimensional state space and achieved better results than a static partitioning approach of the same size. ASSP is based on the idea to collect the aggregated PDSs that are gathered during the simulation anyway to create the partitioning which is based on representatives. To this end, the simulation is started with initialized partitions (where each partition consists of the area around a representative) and decision making is conducted based on this initialization. After each simulation run, the representatives' values are updated with the newly gathered information and the observed aggregated PDSs are stored. After a certain number of simulation runs, the stored PDSs are clustered to form a set of representatives that are now based on the problem-specific observations. These representatives replace the initial ones, their values are initialized and the process of learning and collecting PDSs starts again. After a set number of iterations, the representatives are set and only learning is conducted. In Soeffker et al. (2019), the decision making is conducted by applying a modified Bellman's Equation that contains, next to the immediate reward and the expected future rewards, an accuracy correction term. This term is calculated as a multiplication of the accuracy correction factor (ACT) with the distance between the PDS to be evaluated and the corresponding representative used for the approximation.

5 Computational study

In this section, we present the results of the computational study. In particular, we demonstrate that a more balanced initial routing is beneficial for the overall objective of the considered problem setting and we identify a strong connection between the initial routing and the feature selection for the aggregation of the solution approach. Further, we show that the best feature selection also depends on the degree of dynamism. To highlight the benefits of the proposed VFA approach, we also show that anticipatory decision making is beneficial and we highlight that the chosen partitioning approach has a substantial impact on the

solution quality. To this end, we describe in Sect. 5.1 the instance settings of the problem and in Sect. 5.2 the parameter tuning of the approaches we apply as well as benchmarks. We then state the results of the computational study in Sect. 5.3.

5.1 Instances

In the computational study, we consider the problem from Sect. 3 and generate instances in the following way. A fleet of $|F| = 3$ vehicles serves customer requests in a service area with a size of $20 \text{ km} \times 20 \text{ km}$. Distances in this area are Euclidean and the speed of the vehicles is 25.0 km/h (approx. 15.5 mph). The service is conducted during a time horizon of $T = 480 \text{ min}$, customers can request service throughout the whole time horizon, and the service time at a customer is $t^s = 15 \text{ min}$. Information using time units is discretized to minutes. For one time horizon, which is both the service horizon as well as the request horizon, a total of 100 customers is expected to request service either in advance or during the day. To investigate different business concepts with different shares of initial and dynamic customers, we vary the percentage of dynamic requests. The degree of dynamism (dod, Larsen et al. 2002), describing the share of dynamic requests in the overall requests, is set to 0.5, 0.75, and 0.9 to investigate the solutions for different instance characteristics. Based on our analysis summarized in Fig. 1, 0.5 and 0.75 are the most typical choices in literature. Additionally, to account for the trend towards very dynamic business models, a dod of 0.9 still contains some initial customers, but a very high share of dynamic customers. The spatial distribution of customers (both early and late request) is assumed to be uniform over the service area, the arrival rate of customers does not change over the time horizon. The rewards assigned to the late requests are distributed following a normal distribution with a mean of 5 (monetary units, units are omitted for convenience) and a standard deviation of 2.

5.2 Parameters and benchmark approaches

In this section, the tuning parameters for the applied approach (Sect. 5.2.1) as well as the benchmark approaches (Sect. 5.2.2) are summarized.

5.2.1 Parameters of the proposed VFA approach

In general, the VFA consists of three phases: one phase to find the representatives of the state space partitioning, one phase to approximate the values of the representatives, and one phase to evaluate the policy induced by the values. The search for the state space partitioning happens in an iterative procedure (see Soeffker et al. 2019), we here chose three iterations of 300 simulation runs each. During these 300 runs, PDSs are collected and after collection, the partitioning for the next iteration (or the approximation phase) is found by means of a clustering. As in Soeffker et al. (2019), we choose an k -medoids approach (Kaufman and Rousseeuw 1987) with the number of clusters (denoted as the parameter k) set to 2000. K -medoids is a

common clustering approach that creates k groups of elements that are defined over their medoid that has to be an element itself. We choose k -medoids as it is not only common and intuitive, but also allows to set the number of partitions in advance. Note that the number k is the same for all aggregation choices despite their different number of post-decision state space dimensions for a better comparability under the assumption that we cannot or do not want to create a rather large number of partitions. Using the same number of representatives spread over a larger number of dimensions means that the representatives are, in tendency, further apart. However, the additional information may outweigh this loss in detail. The distance function is Euclidean, this applies both to the clustering as well as to determining the closest representative in the PDS space. When new representatives are set, their values are initialized with a high value to enforce exploration. In the first iteration, the values are not updated, the resulting acceptance policy is therefore purely myopic (thus accepting requests if they can be feasibly inserted). After the first clustering, however, values of the representative are updated to be an average over all observed values that are associated with this representative. For the ACT factor p used in the partitioning approach by Soeffker et al. (2019), we use $p = 0$ as tests for a dod of 0.75 showed that this performed best in most cases (see Appendix A.2). The approximation phase consists of 100,000 simulation runs and the found policy that is induced by the values of the representatives is evaluated in 10,000 simulation runs. Throughout the computational study, time is discretized to minutes. Each parameter combination experiment was conducted five times and the reported results are the averages over those five results.

For the initial routing component, we use the parameter m to vary the initial routing between a classical Savings approach ($m = 0$) and a rather balanced initial routing ($m = 1$).

5.2.2 Benchmark approaches and their parameters

While the objective of this work is to analyze the impact of the initial routing and the aggregation of the post-decision state space, we use two benchmarks to compare the achieved results to. The first benchmark approach (“myopic”) is a myopic policy that accepts customer requests if a feasible insertion into any of the vehicle routes is possible. If the request is accepted, the customer is inserted in the cheapest insertion position that can be found over all routes. This benchmark is used to show the value of anticipation. In the results, the myopic approach with a classical Savings approach ($m = 0$) will be denoted as “benchmark” that we use as a baseline to compare the other approaches to. The second benchmark (“LT”) is a value function approximation without the partitioning approach ASSP (Soeffker et al. 2019). Instead of an adaptive partitioning, the post-decision state spaces are a priori and statically partitioned in lookup-table-like entries where the entries have the same size. The feature combinations for the post-decision state space and the number of partitions used are the same as for the ASSP approach. This means that the same number of partitions is created for LT and ASSP. The LT representatives are created statically all over the post-decision state space whereas the ASSP representatives are created dynamically depending on the post-decision states observed during the simulation making it the

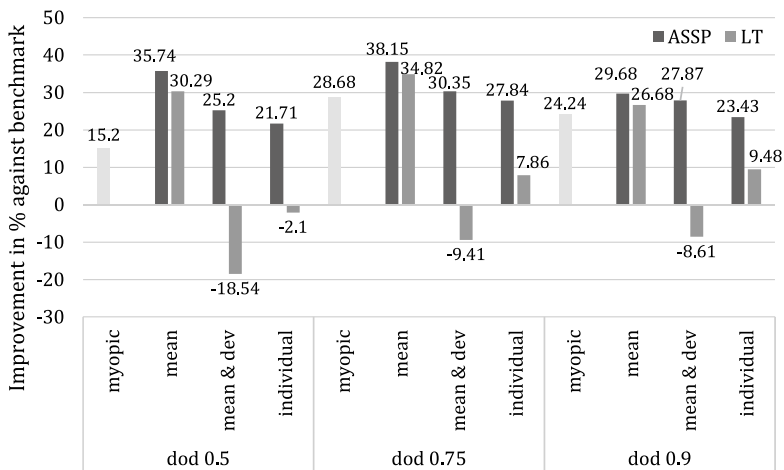


Fig. 7 Overview on improvement of solution quality for all proposed approaches in comparison to benchmark (myopic approach with $m = 0$), for $\text{dod} = 0.5, 0.75, 0.9$

more flexible approach. This benchmark helps to obtain an intuition of the solution quality of the partitioning approach. The parameters for the LT benchmark are the same as for ASSP. That is, the partitions are set in the beginning of the learning, the representatives' values are updated during 100,000 simulations as average over the observed values that are associated with this representative. After this approximation phase, the resulting policy is evaluated in 10,000 simulation runs.

5.3 Results

We report the results of the computational study with a focus on different aspects. We start by reporting the best results for the myopic benchmark approach as well as for the different feature choices for both VFA approaches ASSP and LT in Sect. 5.3.1. Afterward, we show the impact of the initial routing in Sect. 5.3.2. Finally, we present insights into the impact of the feature choice on the solution quality and how this relates to the initial routing choice in Sect. 5.3.3. While summarized results are presented in this section, the detailed results can be found in Appendix A.1.

5.3.1 Overview

In Fig. 7, we report the improvement of the best solution qualities in comparison with the benchmark (myopic with $m = 0$) for a degree of dynamism of 0.5, 0.75, and 0.9. The best solution quality means that we report the results of the best initial routing factor m for the three approaches “myopic”, “ASSP”, and “LT”. For ASSP and LT, we also differentiate the results for the different feature choices.

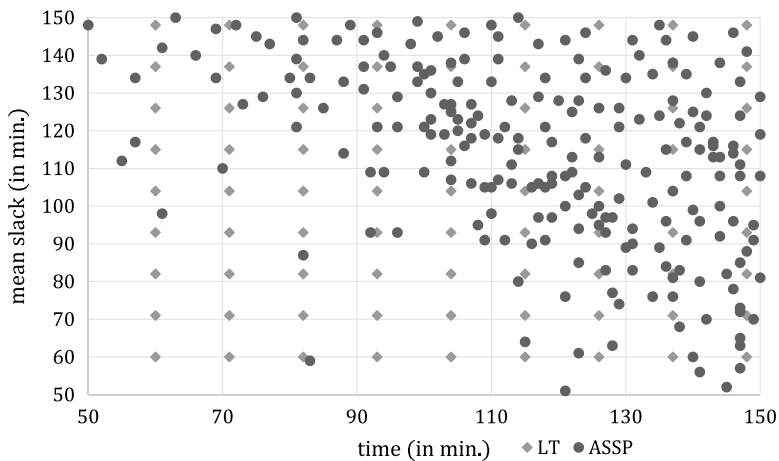


Fig. 8 Excerpt of representatives for LT (medium gray) and ASSP (dark gray) for an aggregation to the features time (x -axis) and mean slack (y -axis) for dod of 0.75

Three main observations can be drawn from Fig. 7. The first observation is that the myopic approach with the best value for m outperforms the myopic approach with $m = 0$. We will provide more information on this observation in the remainder of this section, but this already indicates that a balanced initial routing is preferable in comparison to a seemingly efficient initial routing. The second observation is that the solution quality obtained by the myopic benchmark can be outperformed by a VFA, but is not necessarily outperformed by all VFA options. As we discuss later, the feature choice as well as the approximation architecture of the VFA play an essential role in the success of the value approximation. Further, for all three feature choices, the proposed approach ASSP performs better than the VFA benchmark approach LT. On average over all degrees of dynamism, ASSP provides an improvement in solution quality of 21.70% with the largest improvement being 26.29% for a dod of 0.5 and the smallest improvement being 18.54% for a dod of 0.9. This indicates that the partitioning approach can have a severe impact on the solution quality. To explain this finding, Fig. 8 shows an excerpt of the created representatives for an aggregation to the features time and mean slack for LT in medium gray and ASSP in dark gray for a dod of 0.75. It can be seen that the LT representatives follow a static pattern and the ASSP representatives are created in a rather different pattern. Due to the observed post-decision states from the simulation, ASSP created almost no representatives in the lower left corner of the figure which represents a time of 50 to 90 min and a remaining mean slack of 50 to 80 min. That is, this area did not seem to play an important role in the observed post-decision states. Instead, the ASSP representatives are located more densely in the upper right corner of the figure where they can thus represent differences better than LT.

In the following sections, we will provide results for different initial routing options and provide more insights into the impact of the feature choice.

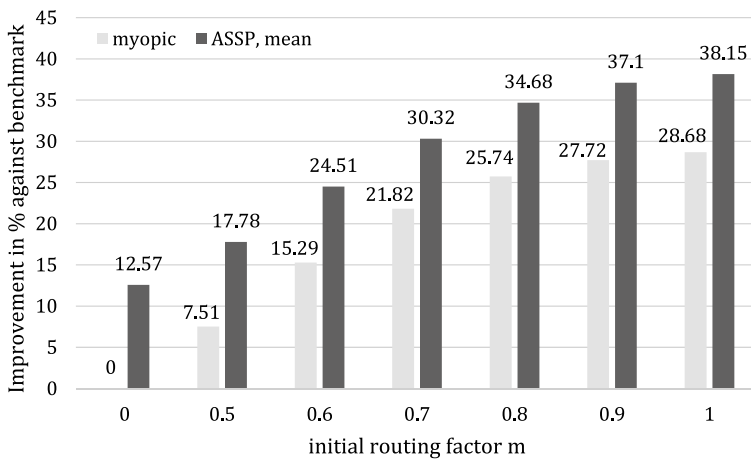


Fig. 9 Impact of initial routing factor on improvement of solution quality for myopic and ASSP for a dod of 0.75

5.3.2 Impact of initial routing

For determining the impact of the initial routing, we first vary the initial routing factor m for the myopic benchmark as there is no interference with the feature choice for this approach for a dod of 0.75. Additionally, we vary the initial routing factor m for ASSP with the features time and mean slack as this showed the best performance for the dod of 0.75. Here, we test $m = \{0, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ as preliminary tests showed that the solution qualities are increasing from $m = 0$ over $m = 0.5$ until $m = 1.0$. As the myopic approach with $m = 0$ is referred to as the benchmark, the improvement of solution qualities obtained for these settings is depicted in Fig. 9.

It can be observed that the solution quality increases with an increasing initial routing factor m . This indicates that it is beneficial to provide rather balanced initial routes (represented by higher values of m). Intuitively, this makes sense as some of the resource time is “wasted” when a vehicle is not included in the initial routing and therefore waits at the depot. For the myopic approach, new customer requests are also inserted into the route where it causes the smallest detour. If one route covers large parts of the service area and other routes never leave the depot, then it is also very likely that new requests are inserted into the route that is already long, leaving the remaining vehicles idling in the depot. This then even increases the effect of “wasting resources” which will be demonstrated in Fig. 10.

To show the impact of the initial routing factor m on the balance of the vehicle resources, Fig. 10 depicts an example of the slack captured in state space representatives from the initial state, that is, in $t = 0$ for a dod of 0.75. For this purpose, two examples of the experiment runs were chosen. While both are for ASSP with individual slacks, one has a rather imbalanced initial routing with $m = 0.0$ and the other has a rather balanced initial routing, that is, $m = 1.0$.

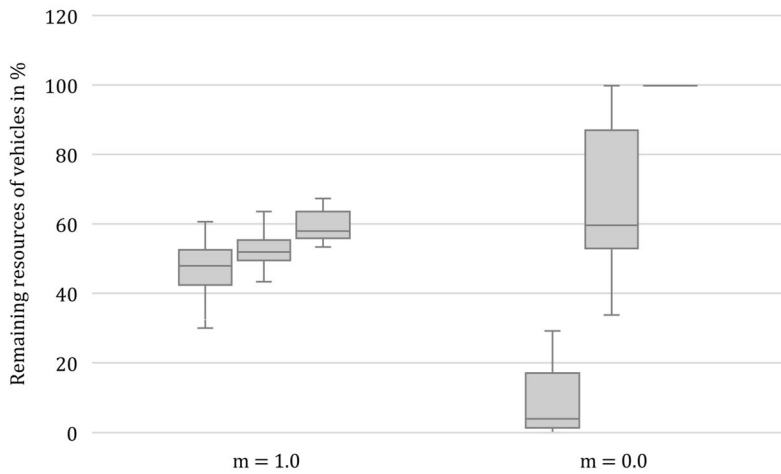


Fig. 10 Example for remaining vehicle resources after the initial routing for dod of 0.75, proposed approach with individual slacks and $m = 1.0$ (left) and $m = 0.0$ (right)

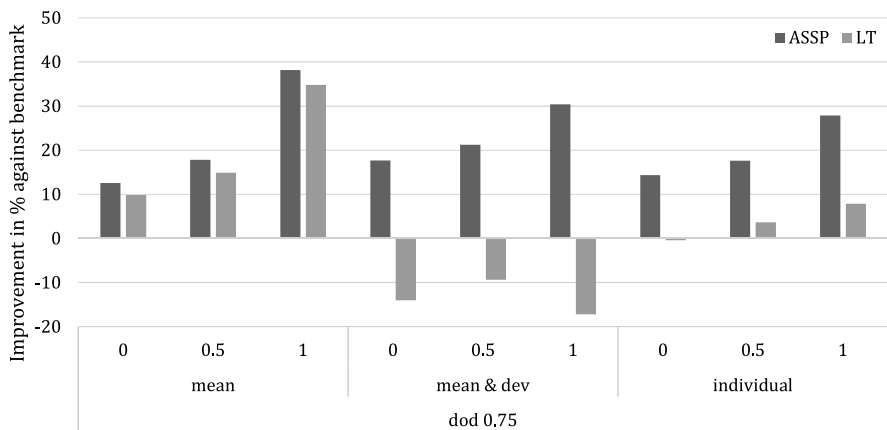


Fig. 11 Impact of initial routing factor on improvement of solution quality for LT (medium gray) and ASSP (dark gray) for a dod of 0.75

Figure 10 depicts the remaining vehicle slack after the initial routing for all three vehicles. While the information describes the obtained representatives and therefore a summary of all observed situations after the initial routing, it still provides some insight. It can be seen in Fig. 10 that using $m = 1.0$ leads to a rather balanced utilization of the vehicle resources as the three slacks are rather similar. In contrast, using $m = 0.0$ leads to a situation where one slack covers the entire time horizon. That is, one vehicle is not used meaning that it will remain in the depot until customers are assigned to it. As time of the service horizon passes by, these resources are then “wasted” as they are not spent on serving customers.

After observing the effect of the initial routing on the myopic benchmark, we also present results depicting the impact of the initial routing on the solution quality for the VFA benchmark as well as for the proposed approach.

In Fig. 11, we present the impact of the initial routing on the improvement of the solution quality for the two VFA approaches LT and ASSP for the different feature choices in comparison to the benchmark of the myopic approach with $m = 0.0$. Figure 11 depicts the results for a dod of 0.75, but the general findings for a dod of 0.5 and 0.9 are similar (see Appendix A.1). The solution quality improvement for the VFA benchmark LT is depicted in medium gray, the solution quality improvement for the proposed approach ASSP is depicted in dark gray.

In general, we can observe that a balanced routing is beneficial except for cases where decision making is obstructed anyway. For LT with features describing time and the mean slack over the fleet (“mean”), it can be seen that a rather balanced initial routing, represented by $m = 1.0$, results in the highest improvement of solution quality. The same increasing tendency with an increasing m can be seen when considering LT with the features time and the individual slacks of the fleet vehicles (“individual”). For the features time, mean slack, and slack deviation (“mean & dev”), an initial routing factor of $m = 0.5$ results in a better solution quality than $m = 0.0$ and $m = 1.0$ for LT. However, the solution quality of this approach with this feature choice performs considerably poor overall as it is even outperformed by the myopic approach with the worst initial routing choice. A solution quality worse than the myopic benchmark indicates that the decision making is impeded by LT (as the decisions made by the myopic approach could have been chosen also). In this case, if the initial routes are rather balanced, the slack deviation is rather small for most observed post-decision states. At the same time, the lookup-table-like approach covers the entire post-decision state space and is thus not able to focus on the differences between the observations.

For our proposed approach ASSP, we see a similar impact of the initial routing on the solution quality improvement for all three aggregation options. That is, for an increasing initial routing factor m , we observe an increasing solution quality meaning that rather balanced initial routes are beneficial for the solution quality. Over all degrees of dynamism, the average improvement of ASSP when applying an initial routing factor of $m = 1.0$ rather than $m = 0.0$ is 10.75% with the smallest improvement being 3.32% and the largest being 22.73%. As for the myopic approach, this can be explained by the “waste of resources” that can occur when vehicles stay in the depot. However, in contrast to LT, ASSP outperforms the myopic benchmark with $m = 0.0$ highlighting that an appropriate partitioning approach is necessary for a VFA approach.

Summarizing, it can be seen that more balanced initial routes (associated with higher values of m) are beneficial compared to seemingly efficient initial routes. That is, if the service provider has the free choice which initial routing to choose, a more balanced initial routing should be preferred. In some cases, for example due to staff qualifications or equipment requirements, it may be beneficial to apply a more efficient initial routing. In the following, we therefore also discuss other initial routing choices.

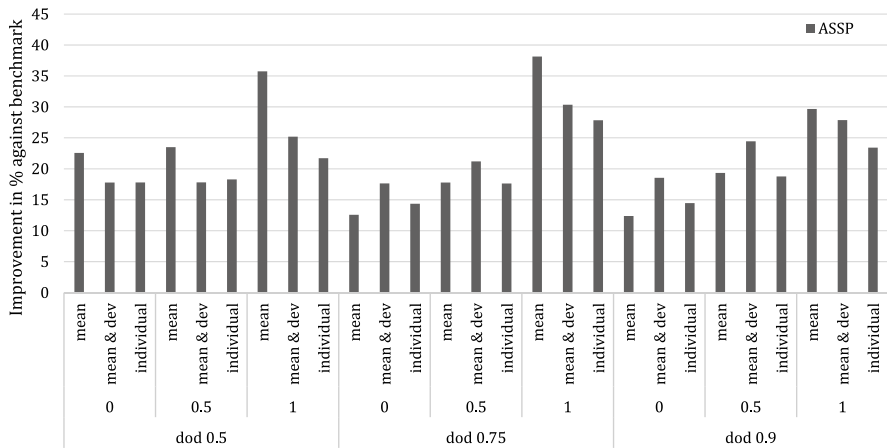


Fig. 12 Impact of feature choice on improvement of solution quality for ASSP (dark gray) for all values of *dod*

5.3.3 Impact of feature choice and relationship with initial routing

After depicting the impact of the initial routing on the solution quality, we now focus on the impact of the feature choice on the solution quality of the proposed approach ASSP. The figures presented here show the improvement of the solution quality in comparison to the myopic approach with $m = 0$. In Appendix A.1, an additional figure shows the impact of the initial routing on the solution quality of not only ASSP, but also the benchmark approaches myopic and LT.

In Fig. 12, the improvement of the solution quality is presented for ASSP (dark gray). The results are shown for the three *dod* values of 0.5, 0.75, and 0.9 and then presented for the different feature combinations and initial routing factors.

For the proposed ASSP, the best feature choice depends on the initial routing factor m and the degree of dynamism. For a *dod* of 0.5, where only a small fraction of the overall driver resources are unplanned, there is not much flexibility in decision making and it is always the best choice to use the features time and mean slack. For a *dod* of 0.75 or 0.9, there is a larger share of dynamic customers and therefore more of flexibility in decision making. Here, for $m = 0.0$ and $m = 0.5$, both indicating possibly rather efficient and less balanced initial routes, the results show that it is beneficial to not only consider time and the mean slack, but additionally the slack deviation. Notably, the number of partitions is the same for all feature choices which means that spreading the partitions over three dimensions for “mean & dev” provides a benefit over spreading it over only two dimensions as in “mean”. This means that the additional information outweighs the decreasing level of detail that can be provided by the representatives and their state space coverage. It can also be noted that all aggregation choices outperform not only the myopic benchmark with $m = 0$, but also the myopic approach with the same value of m . When the initial routes are rather balanced, that is, $m = 1.0$, it is sufficient to focus on time and the mean slack. Here, the additional information

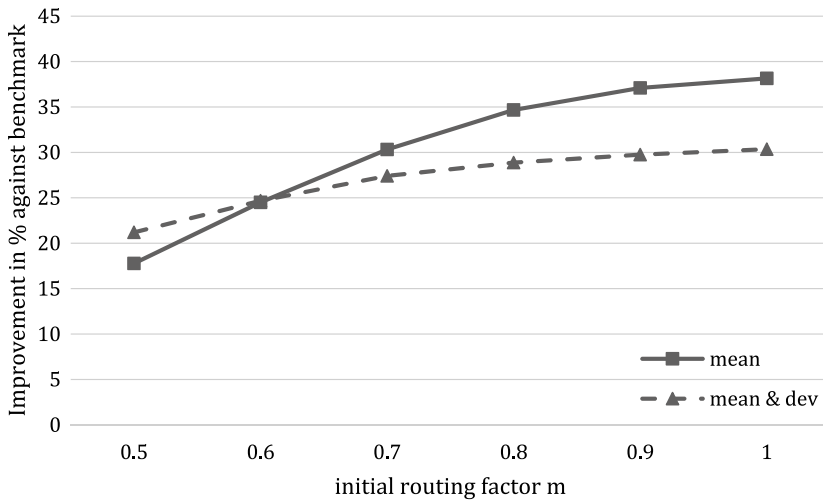


Fig. 13 Impact of initial routing factor m on solution quality for ASSP for dod of 0.75 and feature choices “mean” and “mean & dev”

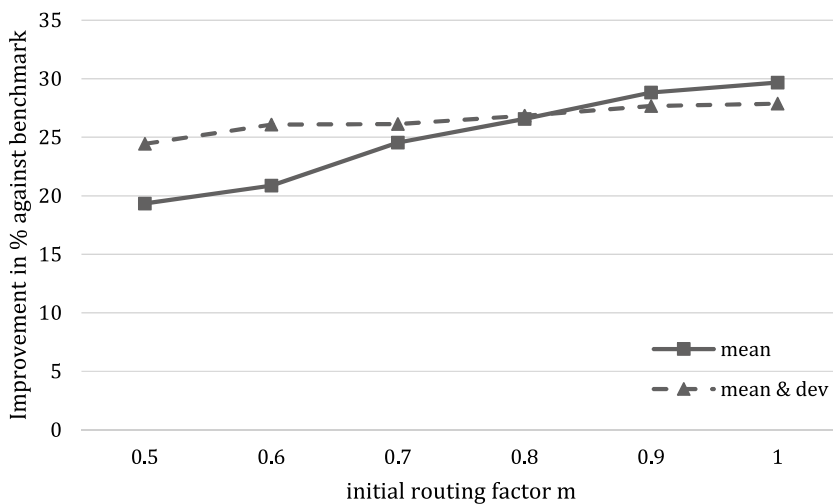


Fig. 14 Impact of initial routing factor m on solution quality for ASSP for dod of 0.9 and feature choices “mean” and “mean & dev”

provided by the deviation is not helpful which is probably also because the rather balanced initial routes lead to less deviation of the individual slacks as it can be seen in Fig. 10. Over all degrees of dynamism, the average improvement of the best performing feature choice compared to the worst performing feature choice when using the same initial routing is 5.71% underlining the impact the best feature choice has on the solution quality.

For ASSP, we can summarize that for a dod of 0.5, “mean” always performs better than “mean & dev”. However, for a dod of 0.75 and 0.9, “mean & dev” performs best until an initial routing factor m of 0.5 and is outperformed by “mean” for $m = 1.0$. Thus, we test intermediate values for m of 0.6, 0.7, 0.8, 0.9 to determine more closely when “mean” starts to be the best feature choice. Figures 13 and 14 depict the impact of the initial routing factor m on the solution quality for ASSP with the features time and mean slack as well as time, mean slack, and deviation for a dod of 0.75 and 0.9, respectively.

For a dod of 0.75, “mean” performs better than “mean & dev” for values of m starting at 0.7. For the dod of 0.9, “mean” only starts to perform better than “mean & dev” at $m = 0.9$. Summarizing, it can be observed that for higher degrees of dynamism, the feature choice “mean & dev” stays relevant for higher values of the initial routing factor. For smaller degrees of dynamism, the additional information becomes less important already for less balanced routes.

For the LT (see Appendix A.1), while the initial routing has an impact on the solution quality, the aggregation to time and mean slack provides the best solution quality for all values of dod and initial routing factor m . The second best solution quality is obtained by an aggregation to time and all individual slacks. The aggregation to time, mean slack, and the deviation is outperformed by the myopic approach because the deviation dimension is covered entirely by LT although only small parts of this dimension have to be covered in detail.

Summarizing, several main conclusions can be drawn from the results. In addition to the earlier insights about the value of anticipation and the benefit of balanced initial routes, it can be also seen that the best feature choice depends on the instance specifics and also the initial routing. This is especially important if the service provider chooses the initial routing first with efficiency of the initial routes, for example due to driver or equipment reasons. Balancing the initial routes leads to average improvements for the proposed approach of 10.75% compared to unbalanced initial routes while the best feature choice leads to average improvements for the proposed approach of 5.71%.

6 Conclusion and future work

In this article, we propose to solve the dynamic vehicle routing problem with stochastic customer requests (DVRPSR) by means of a solution approach that combines an initial routing approach with a value function approximation approach. The initial routing approach contains a parameter that defines whether a balanced usage of fleet resources has to be achieved in the initial routing. The proposed value function approximation applies a partitioning approach from literature and we contribute to the literature by proposing three different sets of features that capture the fleet resource balancing over the time horizon in a different level of detail.

In this work, we analyze the interplay between the initial routing and the dynamic decision making as well as the design of the solution approach. In our computational study, we show that solving the DVRPSR with the proposed approach outperforms both a myopic benchmark as well as a value function approximation with a simpler

partitioning approach. We show that the initial routing, which constitutes the first decision of the problem, has a large impact on the solution quality as it defines the initial usage of resources. For both a myopic benchmark as well as our proposed approach, results show that it is beneficial to use a rather balanced initial routing that avoids “wasting resources”. For the proposed approach, the average improvement of the rather balanced initial routing is 10.75% compared to a rather efficient initial routing with individual improvements of up to 22.73%. For the myopic approach, an improvement of over 28% was obtained when balancing the initial routes in comparison to applying a rather efficient initial routing. For the value function approximation with a simple partitioning approach, decision making is severely impacted by the partitioning approach leading to both a poor solution quality in general, sometimes even up to 9% worse than an imbalanced myopic approach. On average, the proposed partitioning approach leads to an improvement of 21.70% compared to the simple partitioning approach. Further, the value function approximation with the simple partitioning approach leads to a situation where a rather balanced initial routing is not the most successful initial routing strategy. These results along with the obtained solution quality underline the need for a partitioning approach that adapts to the problem specifics. We also evaluate the impact of three feature aggregation choices that are adapted from literature for this problem. In particular, we show that additional information can be beneficial for the solution quality if the initial routes are not very balanced to depict the heterogeneity in the resource usage, and that this additional information is not necessary for a rather balanced initial routing. Summarizing, our results showed that the best feature aggregation choice depends both on the initial routing as well as on instance characteristics like the degree of dynamism and that the best feature choice leads to average improvements of 5.71% for the proposed value function approximation.

In the future, it should be investigated how decision making is changed if customers have to be accepted if they can be feasibly inserted. Also, with increasing computational and methodological advances, it would be interesting to see the impact of more detailed features or more considered routing options on the solution quality. Once the general relationship between the features and the stored values are understood clearly and it can be evaluated which features have an impact on the values, neural networks might also be used as more sophisticated and flexible value function approximation architectures which may further increase the solution quality.

Appendix

In the appendix, we provide further information on the results. In Section A.1, we present the detailed results for the experiments. Sect. A.2 contains results on the impact of the ASSP factor p on the results and Sect. A.3 provides results for a parametric value function approximation applied to the single vehicle case to show that ASSP outperforms such a parametric value function approximation.

Table 2 Solution quality of myopic approach for different values of the initial routing factor m and for a dod of 0.5

m	v1	v2	v3	v4	v5	avg
0.0	23.78	23.79	23.76	23.68	23.70	23.74
0.5	23.88	24.00	23.83	23.89	23.80	23.88
1.0	27.42	27.33	27.38	27.28	27.35	27.35

Table 3 Solution quality of myopic approach for different values of the initial routing factor m and for a dod of 0.75

m	v1	v2	v3	v4	v5	avg
0.0	34.99	35.09	34.93	35.08	34.99	35.02
0.5	37.58	37.69	37.72	37.66	37.59	37.65
0.6	40.29	40.38	40.38	40.37	40.44	40.37
0.7	42.74	42.71	42.63	42.61	42.60	42.66
0.8	43.95	44.04	44.05	44.08	44.05	44.03
0.9	44.71	44.70	44.75	44.74	44.72	44.72
1.0	45.02	45.01	45.08	45.13	45.08	45.06

Table 4 Solution quality of myopic approach for different values of the initial routing factor m and for a dod of 0.9

m	v1	v2	v3	v4	v5	avg
0.0	38.41	38.48	38.42	38.45	38.48	38.45
0.5	42.54	42.56	42.55	42.61	42.50	42.55
1.0	47.76	47.82	47.81	47.73	47.73	47.77

Table 5 Solution quality of LT for the three feature choices and different values of the initial routing factor m and a dod of 0.5

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	27.97	28.09	28.27	27.94	28.02	28.06
Mean, 0.5	28.17	28.34	28.16	28.01	28.30	28.20
Mean, 1.0	30.96	31.11	30.91	30.78	30.90	30.93
Mean & dev, 0.0	19.07	19.22	19.07	19.08	19.10	19.11
Mean & dev, 0.5	19.41	18.99	19.60	19.63	19.07	19.34
Mean & dev, 1.0	19.33	18.54	18.85	19.06	18.23	18.80
Individual, 0.0	22.36	22.16	22.57	22.36	22.05	22.30
Individual, 0.5	22.15	22.30	22.42	22.35	22.36	22.31
Individual, 1.0	23.37	23.20	23.27	23.16	23.21	23.24

Detailed results

In this section, we report the individual solution quality in terms of the percentage of accumulated rewards of the experiments. To this end, we present the results of the myopic approach with different values of the initial routing factor m in Tables 2, 3, and 4 for a dod of 0.5, 0.75, and 0.9, respectively. We present the results of the VFA

Table 6 Solution quality of LT for the three feature choices and different values of the initial routing factor m and a dod of 0.75

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	38.45	38.39	38.54	38.47	38.42	38.46
Mean, 0.5	40.13	40.34	40.34	40.03	40.24	40.22
Mean, 1.0	47.82	47.05	47.22	46.97	47.01	47.21
Mean & dev, 0.0	30.02	30.51	30.16	29.84	29.93	30.09
Mean & dev, 0.5	31.87	31.78	31.42	31.54	32.00	31.72
Mean & dev, 1.0	28.78	27.49	29.21	29.74	29.72	28.99
Individual, 0.0	34.75	35.10	34.78	34.85	34.87	34.87
Individual, 0.5	36.52	36.18	36.05	36.28	36.43	36.29
Individual, 1.0	37.46	38.03	37.63	38.19	37.54	37.77

Table 7 Solution quality of LT for the three feature choices and different values of the initial routing factor m and a dod of 0.9

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	41.73	41.82	41.86	42.32	41.68	41.88
Mean, 0.5	44.56	44.79	45.17	44.57	44.32	44.68
Mean, 1.0	48.47	48.85	48.71	49.04	48.47	48.71
Mean & dev, 0.0	33.61	33.36	33.18	33.42	33.49	33.41
Mean & dev, 0.5	35.14	35.05	35.13	35.13	35.24	35.14
Mean & dev, 1.0	32.53	32.22	32.32	33.08	34.21	32.87
Individual, 0.0	40.51	38.95	38.84	39.58	40.40	39.65
Individual, 0.5	40.58	40.74	40.35	41.02	39.89	40.52
Individual, 1.0	42.36	41.81	42.15	41.49	42.67	42.10

Table 8 Solution quality of ASSP for the three feature choices and different values of the initial routing factor m and a dod of 0.5

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	29.32	29.03	28.93	29.04	29.18	29.10
Mean, 0.5	29.21	29.63	29.16	29.26	29.24	29.32
Mean, 1.0	32.19	32.10	32.08	32.39	32.38	32.27
Mean & dev, 0.0	27.97	28.09	27.97	27.91	27.88	27.97
Mean & dev, 0.5	27.73	28.06	28.22	27.90	27.94	27.97
Mean & dev, 1.0	29.48	29.72	29.90	29.77	29.75	29.72
Individual, 0.0	27.84	27.88	28.13	28.08	27.91	27.97
Individual, 0.5	28.01	28.37	27.87	28.05	28.13	28.09
Individual, 1.0	28.87	28.99	28.59	29.21	28.81	28.89

benchmark approach “LT” for the initial routing factor $m = 0, 0.5, 1$ and for the three different feature choices in Tables 5, 6, and 7 for a dod of 0.5, 0.75, and 0.9, respectively. Last, we present the results of the proposed approach “ASSP” for the initial

Table 9 Solution quality of ASSP for the three feature choices and different values of the initial routing factor m and a dod of 0.75

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	39.36	39.45	39.39	39.42	39.47	39.42
Mean, 0.5	41.12	41.36	41.21	41.19	41.34	41.24
Mean, 0.6	43.58	43.51	43.57	43.60	43.74	43.60
Mean, 0.7	45.72	45.52	45.60	45.60	45.72	45.63
Mean, 0.8	47.19	47.21	46.94	47.25	47.22	47.16
Mean, 0.9	48.25	47.69	48.03	48.08	48.00	48.01
Mean, 1.0	48.39	48.31	48.55	48.41	48.24	48.38
Mean & dev, 0.0	41.44	41.21	40.92	41.44	40.98	41.20
Mean & dev, 0.5	42.49	42.42	42.61	42.29	42.29	42.44
Mean & dev, 0.6	43.56	43.66	44.09	43.57	43.41	43.66
Mean & dev, 0.7	44.31	44.96	44.52	44.63	44.66	44.62
Mean & dev, 0.8	44.98	45.44	44.71	45.14	45.37	45.13
Mean & dev, 0.9	45.58	45.38	45.33	45.52	45.37	45.44
Mean & dev, 1.0	45.42	45.82	45.68	45.84	45.47	45.65
Individual, 0.0	39.51	40.64	40.06	39.92	40.10	40.05
Individual, 0.5	40.93	41.22	41.30	41.09	41.40	41.19
Individual, 1.0	45.01	44.73	44.81	44.62	44.67	44.77

Table 10 Solution quality of ASSP for the three feature choices and different values of the initial routing factor m and a dod of 0.9

Feature choice, m	v1	v2	v3	v4	v5	avg
Mean, 0.0	42.89	43.33	43.14	43.09	43.63	43.21
Mean, 0.5	45.69	45.98	46.03	45.94	45.80	45.89
Mean, 0.6	46.50	46.46	46.55	46.50	46.36	46.48
Mean, 0.7	47.78	47.88	48.08	47.80	47.89	47.89
Mean, 0.8	48.69	48.70	48.47	48.75	48.73	48.67
Mean, 0.9	49.54	49.48	49.54	49.50	49.59	49.53
Mean, 1.0	49.80	49.87	49.95	49.97	49.70	49.86
Mean & dev, 0.0	45.49	45.91	45.55	45.22	45.74	45.58
Mean & dev, 0.5	47.58	47.89	48.63	47.26	47.87	47.85
Mean & dev, 0.6	48.35	48.35	49.20	48.01	48.50	48.48
Mean & dev, 0.7	47.75	48.77	48.77	48.41	48.80	48.50
Mean & dev, 0.8	48.61	48.99	49.16	48.64	48.48	48.78
Mean & dev, 0.9	49.27	49.22	49.13	48.97	48.85	49.09
Mean & dev, 1.0	49.19	48.81	49.55	49.07	49.22	49.17
Individual, 0.0	43.95	44.37	44.21	43.53	43.99	44.01
Individual, 0.5	45.54	45.38	45.44	45.98	45.98	45.66
Individual, 1.0	47.46	47.33	47.53	47.32	47.65	47.46

routing factor $m = 0, 0.5, 1$ and for the three different feature choices in Tables 8, 9, and 10 for a dod of 0.5, 0.75, and 0.9, respectively.

For all experimental settings, five independent runs were conducted and the values used for the calculation of the improvements reported in the main body of

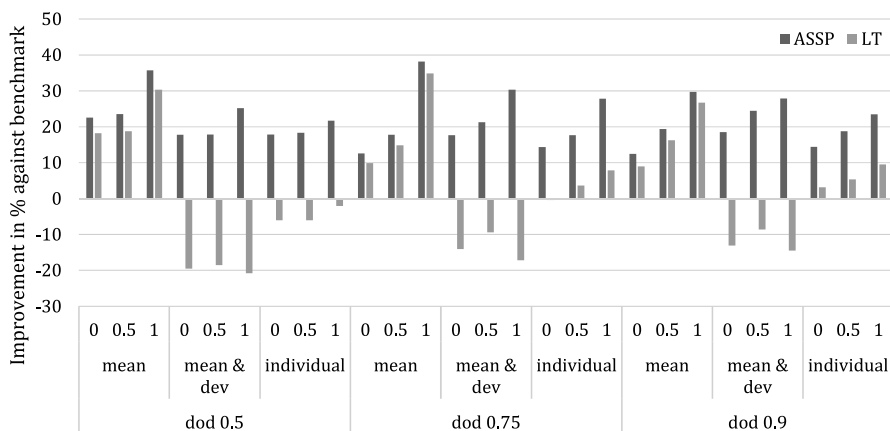


Fig. 15 Impact of initial routing factor on improvement of solution quality for LT (medium gray) and ASSP (dark gray) for all choices of dod

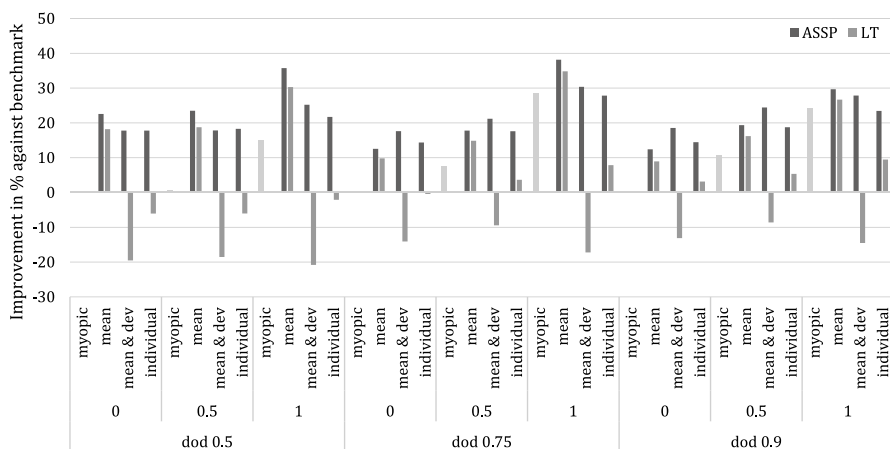


Fig. 16 Impact of feature choice on improvement of solution quality for myopic (light gray), LT (medium gray), and ASSP (dark gray) for all values of dod

the article are the averages. Here, the individual results (denoted by “v1”..“v5”) are reported along with the averages (denoted by “avg”).

In addition to Fig. 11 that showed the impact of the initial routing only for a dod of 0.75, in Fig. 15, we also show the impact of the initial routing on the improvement of the solution quality for both LT (in medium gray) and ASSP (in dark gray) for a dod of 0.5, 0.75, and 0.9.

In addition to Fig. 12, that showed the impact of the feature choice on the improvement of the solution quality only for the proposed approach ASSP, Fig. 16 depicts the impact of the feature choice on the improvement of the solution quality for myopic, LT, and ASSP.

Table 11 Impact of balancing the initial routes: Improvement in % for ASSP per feature choice and dod

	dod 0.5	dod 0.75	dod 0.9
Mean	10.74	22.73	15.38
Mean & dev	6.29	10.80	7.86
Individual	3.32	11.79	7.84

Table 12 Impact of the feature choice: Improvement in % for ASSP per initial routing factor m and dod

	dod 0.5	dod 0.75	dod 0.9
$m = 0.0$	4.08	4.52	5.48
$m = 0.5$	4.82	3.04	4.79
$m = 1.0$	11.53	8.07	5.06

Table 13 Improvement in % for ASSP over the benchmark LT per feature choice, dod, and initial routing factor m

	dod 0.5			dod 0.75			dod 0.9		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
Mean	3.71	3.99	4.19	2.50	2.55	2.47	3.18	2.70	2.36
Mean & dev	46.36	44.64	58.09	36.92	33.79	57.46	36.42	36.16	49.56
Individual	25.41	25.87	24.32	14.84	13.49	18.52	10.99	12.70	12.74

While the initial routing has an impact on the solution quality for LT, the general structure is the same for all values of dod and all initial routing factor values. That is, for all values of m , the best solution quality is obtained by an aggregation to time and the mean slack over all vehicles. For all values of m , the second best solution quality is achieved by considering time and all individual slacks. The worst solution quality is obtained for an aggregation to time, the mean slack, and the deviation. As discussed, the inflexible structure of the lookup-table-like partitioning approach probably has difficulties with the deviation feature which is particularly obvious for the rather balanced initial routes where the deviation of the slacks is small. Notably, it can be seen that LT with the aggregation to time and mean slack outperforms both the myopic benchmark with $m = 0$ as well as the myopic approach with the same value of m .

To summarize the performance of the proposed approach ASSP with regard to the impact of balancing the initial routes and the impact of the best feature choice, we provide an overview in the following Tables 11 and 12.

For the impact of balancing the initial routes, we compare for each degree of dynamism and feature combination the solution quality of ASSP for the balanced initial routing factor $m = 1.0$ with the rather unbalanced initial routing factor $m = 0.0$. The improvements are listed in Table 11, the average improvement over all degrees of dynamism and feature choices is 10.75%.

Table 14 Solution quality of the proposed approach for aggregation to time and mean slack, different values of the initial routing factor m , and different values of the ASSP factor p

m, p	v1	v2	v3	v4	v5	avg
0.0, 0.0	39.36	39.45	39.39	39.42	39.47	39.42
0.0, 0.05	39.52	39.50	39.54	39.57	39.20	39.47
0.0, 0.1	39.68	39.52	39.71	39.76	39.64	39.66
0.0, 0.2	39.48	39.58	39.67	39.52	39.54	39.56
0.0, 0.3	39.69	39.48	39.33	39.60	39.54	39.53
0.0, 0.4	39.38	39.62	39.25	39.36	39.35	39.39
0.0, 0.5	39.36	39.45	39.28	39.25	39.31	39.33
0.5, 0.0	41.12	41.36	41.21	41.19	41.34	41.24
0.5, 0.05	41.07	41.26	41.09	41.03	41.29	41.15
0.5, 0.1	41.18	41.01	41.04	41.24	41.43	41.18
0.5, 0.2	41.11	41.13	41.17	40.96	41.35	41.14
0.5, 0.3	41.35	41.33	40.95	40.92	41.25	41.16
0.5, 0.4	40.97	41.02	40.95	41.12	41.17	41.05
0.5, 0.5	40.85	41.01	40.77	40.96	41.06	40.93
1.0, 0.0	48.39	48.31	48.55	48.41	48.24	48.38
1.0, 0.05	47.30	47.50	47.26	46.98	47.16	47.24
1.0, 0.1	47.47	47.40	47.34	47.51	47.18	47.38
1.0, 0.2	47.43	47.56	47.41	47.36	47.34	47.42
1.0, 0.3	47.25	47.72	47.20	47.13	47.17	47.29
1.0, 0.4	47.57	47.08	47.43	47.27	47.30	47.33
1.0, 0.5	47.19	47.23	47.10	47.19	47.20	47.18

For the impact of the feature choice, we compare for each degree of dynamism and initial routing factor m the solution quality of ASSP for the best feature choice with the solution quality obtained with the worst feature choice. The improvements are listed in Table 12, the average improvement over all degrees of dynamism and initial routing factor values is 5.71%.

Further, we also summarize the algorithmic performance of the proposed approach ASSP relative to the benchmark “LT”, the value function approximation with a simpler partitioning approach, in Table 13. Over all combinations, the average improvement is 21.70%.

Impact of ASSP factor p on results for dod of 0.75

In Soeffker et al. (2019), the solution approach ASSP contains the ACT factor p that modifies the Bellman’s Equation to incorporate the uncertainty of the value approximation. When a PDS has to be evaluated, the value stored in the closest representative is used. However, when the PDS is far away from the representative, this approximation might not be too well. Thus, the value of the representative is reduced depending on the ACT factor p which is multiplied with the distance between PDS and representative. In this section, we report the individual solution quality in terms of the percentage of accumulated rewards of

Table 15 Solution quality of the proposed approach for aggregation to time, mean slack, and deviation, different values of the initial routing factor m , and different values of the ASSP factor p

m, p	v1	v2	v3	v4	v5	avg
0.0, 0.0	41.44	41.21	40.92	41.44	40.98	41.20
0.0, 0.05	40.78	41.27	40.96	41.03	41.06	41.02
0.0, 0.1	40.82	41.11	40.98	40.64	40.96	40.90
0.0, 0.2	41.03	40.80	41.20	41.13	40.91	41.01
0.0, 0.3	41.03	40.72	40.58	40.99	40.57	40.47
0.0, 0.4	40.38	40.36	40.32	40.62	40.24	40.38
0.0, 0.5	40.27	40.36	40.09	40.03	40.31	40.21
0.5, 0.0	42.49	42.42	42.61	42.29	42.29	42.44
0.5, 0.05	41.58	41.18	41.67	41.81	41.68	41.58
0.5, 0.1	41.39	41.48	41.18	41.64	41.58	41.45
0.5, 0.2	41.49	41.70	41.63	41.75	41.66	41.64
0.5, 0.3	40.89	41.16	41.52	41.17	41.99	41.34
0.5, 0.4	41.68	41.24	41.96	41.42	41.65	41.59
0.5, 0.5	41.32	41.13	41.29	41.16	41.40	41.26
1.0, 0.0	45.42	45.82	45.68	45.84	45.47	45.65
1.0, 0.05	44.94	44.83	44.77	44.89	44.90	44.86
1.0, 0.1	45.29	44.73	44.55	45.12	45.32	45.00
1.0, 0.2	45.13	44.52	45.12	45.00	44.88	44.93
1.0, 0.3	45.27	44.79	44.84	45.56	45.34	45.16
1.0, 0.4	44.69	44.79	44.99	44.85	44.99	44.86
1.0, 0.5	45.06	44.46	44.77	44.82	44.78	44.78

the experiments for varying values of the ACT factor p to show why a value of $p = 0.0$ was chosen for the remainder of the experiments. In particular, we present the results of the proposed approach “ASSP” for the initial routing factor $m = 0, 0.5, 1$, the three different feature choices, and different values of the ASSP-factor p in Tables 14, 15, and 16. For all experimental settings, the α was set to 0.75. Further, five independent runs were conducted and the values used for the calculation of the improvements reported in the main body of the article are the averages. Here, the individual results (denoted by “v1”..“v5”) are reported along with the averages (denoted by “avg”). The best value of the ACT factor is highlighted in bold.

Comparison to linear value function approximation

In this section, we apply a parametric linear value function approximation in comparison to ASSP, LT, and myopic. To provide a clear comparison without the interference of several vehicles and the heterogeneous customer rewards, we report results for the problem setting considered in Soeffker et al. (2019). The problem setting differs from the one considered here in the following. There is only one vehicle instead of a fleet and customer acceptances are made upon the arrival at a customer

Table 16 Solution quality of the proposed approach for aggregation to time and individual slacks, different values of the initial routing factor m , and different values of the ASSP factor p

m, p	v1	v2	v3	v4	v5	avg
0.0, 0.0	39.51	40.64	40.06	39.92	40.10	40.05
0.0, 0.05	39.66	39.60	40.00	40.02	40.03	39.86
0.0, 0.1	39.94	39.99	39.94	39.89	40.10	39.97
0.0, 0.2	39.57	39.77	39.44	40.07	40.00	39.77
0.0, 0.3	38.74	38.63	39.12	39.14	39.38	39.00
0.0, 0.4	38.00	38.25	38.21	38.93	37.77	38.23
0.0, 0.5	37.18	37.52	37.60	37.67	37.14	37.42
0.5, 0.0	40.93	41.22	41.30	41.09	41.40	41.19
0.5, 0.05	41.04	40.60	40.91	40.45	40.85	40.77
0.5, 0.1	40.66	40.60	40.53	40.73	40.68	40.64
0.5, 0.2	40.51	41.13	41.04	40.71	40.48	40.77
0.5, 0.3	40.46	40.34	40.06	40.26	40.43	40.31
0.5, 0.4	39.81	40.17	40.33	39.74	39.86	39.99
0.5, 0.5	38.74	39.79	39.32	39.36	39.38	39.32
1.0, 0.0	45.01	44.73	44.81	44.62	44.67	44.77
1.0, 0.05	44.23	44.10	44.07	44.11	44.48	44.20
1.0, 0.1	44.34	43.96	44.96	44.48	44.71	44.49
1.0, 0.2	44.44	44.94	44.94	44.59	44.42	44.67
1.0, 0.3	44.72	44.77	44.49	44.52	44.82	44.66
1.0, 0.4	44.45	44.57	44.54	44.57	45.04	44.63
1.0, 0.5	44.41	44.45	44.09	44.83	43.48	44.25

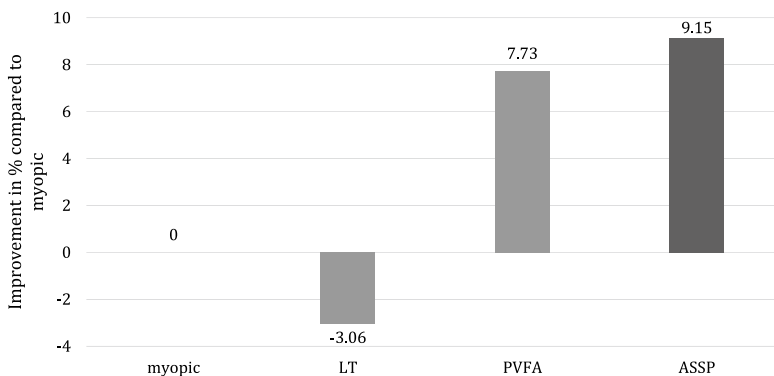


Fig. 17 Improvement of solution quality in comparison to myopic approach for the proposed approach ASSP and the two benchmarks LT and PVFA, adapted from Soeffker et al. (2019) with additional results

location. Therefore, decisions may be made on sets of requests rather than on individual ones. Further, all customers provide the same immediate reward.

In Soeffker et al. (2019), results were reported for a myopic benchmark approach and for two value function approximations where one uses a lookup-table like representation (“LT”) and the other one applies ASSP (“ASSP”). Both of those

approaches use 500 representatives and use the features time and slack. Here, we now extend those results by also providing results for a parametric value function approximation using multiple linear regression (“PVFA”) to motivate why this approach was not studied for the fleet case of this article. For those experiments, a total of 60 customers was expected with a degree of dynamism of 0.75 leading to an expected number of 45 dynamic customers. The service time is assumed to be 5 min.

In Fig. 17, the improvement of solution quality compared to the myopic approach is depicted for the three approaches. It can be seen that LT leads to a solution quality 3.06% worse than the myopic approach. PVFA and ASSP yield a solution quality better than myopic with ASSP outperforming PVFA. In this single vehicle setting, the dependency between the value and the considered features resembles a linear function in certain parts of the state space, but not in all. Due to the superior performance of ASSP, a parametric value function approximation with a multiple linear regression was not considered in this article.

Acknowledgements Marlin Ulmer’s work is funded by the DFG Emmy Noether Programme, project 444657906. We gratefully acknowledge their support. We thank the two anonymous reviewers and the editors of the special issue. Their comments helped to significantly improve the article.

Funding Open access funding provided by University of Vienna.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agussurja L, Cheng SF, Lau HC (2019) A state aggregation approach for stochastic multiperiod last-mile ride-sharing problems. *Transp Sci* 53(1):148–166
- Angelesli E, Bianchessi N, Mansini R, Speranza MG (2009) Short term strategies for a dynamic multi-period routing problem. *Transp Res Part C Emerg Technol* 17(2):106–119
- Angelesli E, Archetti C, Filippi C, Vindigni M (2021) A dynamic and probabilistic orienteering problem. *Comput Oper Res* 136:105454
- Attanasio A, Cordeau JF, Ghiani G, Laporte G (2004) Parallel tabu search heuristics for the dynamic multi-vehicle dial-a-ride problem. *Parallel Comput* 30(3):377–387
- Azi N, Gendreau M, Potvin JY (2012) A dynamic vehicle routing problem with multiple delivery routes. *Ann Oper Res* 199(1):103–112
- Barto AG (1998) Reinforcement learning: an introduction. MIT Press

- Basso R, Kulcsár B, Sanchez-Diaz I, Qu X (2022) Dynamic stochastic electric vehicle routing with safe reinforcement learning. *Transp Res Part E Logist Transp Rev* 157:102496
- Bellman R (1957) *Dynamic programming*, 1st edn. Princeton University Press, Princeton, NJ, USA
- Bent RW, Van Hentenryck P (2004) Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Oper Res* 52(6):977–987
- Bent R, Van Hentenryck P (2007) Waiting and relocation strategies in online stochastic vehicle routing. In: *IJCAI*, pp 1816–1821
- Bozanta A, Cevik M, Kavaklioglu C, Kavuk EM, Tosun A, Sonuc SB, Duranel A, Basar A (2022) Courier routing and assignment for food delivery service using reinforcement learning. *Comput Ind Eng* 164:107871
- Chen X, Ulmer MW, Thomas BW (2022) Deep q-learning for same-day delivery with vehicles and drones. *Eur J Oper Res* 298(3):939–952
- Clarke G, Wright JW (1964) Scheduling of vehicles from a central depot to a number of delivery points. *Oper Res* 12(4):568–581
- Fleckenstein D, Klein R, Steinhardt C (2023) Recent advances in integrating demand management and vehicle routing: a methodological review. *Eur J Oper Res* 306(2):499–518
- Gendreau M, Guertin F, Potvin JY, Taillard E (1999) Parallel tabu search for real-time vehicle routing and dispatching. *Transp Sci* 33(4):381–390
- Ghiani G, Manni E, Thomas BW (2012) A comparison of anticipatory algorithms for the dynamic and stochastic traveling salesman problem. *Transp Sci* 46(3):374–387
- Giménez-Palacios I, Parreño F, Álvarez-Valdés R, Paquay C, Oliveira BB, Carravilla MA, Oliveira JF (2022) First-mile logistics parcel pickup: vehicle routing with packing constraints under disruption. *Transp Res Part E Logist Transp Rev* 164:102812
- Hanshar FT, Ombuki-Berman BM (2007) Dynamic vehicle routing using genetic algorithms. *Appl Intell* 27(1):89–99
- Hvattum LM, Løkketangen A, Laporte G (2006) Solving a dynamic and stochastic vehicle routing problem with a sample scenario hedging heuristic. *Transp Sci* 40(4):421–438
- Ichoua S, Gendreau M, Potvin JY (2000) Diversion issues in real-time vehicle dispatching. *Transp Sci* 34(4):426–438
- Ichoua S, Gendreau M, Potvin JY (2006) Exploiting knowledge about future demands for real-time vehicle dispatching. *Transp Sci* 40(2):211–225
- Ikonen E, Selek I, Najim K (2016) Process control using finite Markov chains with iterative clustering. *Comput Chem Eng* 93:293–308
- Kaufman L, Rousseeuw PJ (1987) Clustering by means of medoids. In: Dodge Y (ed) *First international conference on statistical data analysis based on the L1-norm and related methods*
- Kilby P, Prosser P, Shaw P (1998) *Dynamic VRPS: a study of scenarios*. University of Strathclyde Technical Report 1(11)
- Klapp MA, Erera AL, Toriello A (2018) The one-dimensional dynamic dispatch waves problem. *Transp Sci* 52(2):402–415
- Klein V, Steinhardt C (2023) Dynamic demand management and online tour planning for same-day delivery. *Eur J Oper Res* 307(2):860–886
- Larsen A, Madsen OB, Solomon MM (2002) Partially dynamic vehicle routing-models and algorithms. *J Oper Res Soc* 53(6):637–646
- Lee IS, Lau HY (2004) Adaptive state space partitioning for reinforcement learning. *Eng Appl Artif Intell* 17(6):577–588
- Montemanni R, Gambardella LM, Rizzoli AE, Donati AV (2005) Ant colony system for a dynamic vehicle routing problem. *J Comb Optim* 10(4):327–343
- Pillac V, Gendreau M, Guéret C, Medaglia AL (2013) A review of dynamic vehicle routing problems. *Eur J Oper Res* 225(1):1–11
- Powell WB (2011) *Approximate dynamic programming: solving the curses of dimensionality*, Wiley series in probability and statistics, vol 842. Wiley, New York
- Psarafitis HN, Wen M, Kontovas CA (2016) Dynamic vehicle routing problems: three decades and counting. *Networks* 67(1):3–31
- Ritzinger U, Puchinger J, Hartl RF (2016) A survey on dynamic and stochastic vehicle routing problems. *Int J Prod Res* 54(1):215–231
- Rosenkrantz DJ, Stearns RE, Lewis PM (1974) Approximate algorithms for the traveling salesperson problem. In: *switching and automata theory, 1974., IEEE conference record of 15th annual symposium on*, IEEE, pp 33–42

- Soeffker N, Ulmer MW, Mattfeld DC (2019) Adaptive state space partitioning for dynamic decision processes. *Bus Inf Syst Eng* 61(3):261–275
- Soeffker N, Ulmer MW, Mattfeld DC (2022) Stochastic dynamic vehicle routing in the light of prescriptive analytics: a review. *Eur J Oper Res* 298(3):801–820
- Soeffker N, Ulmer MW, Mattfeld DC (2017) On fairness aspects of customer acceptance mechanisms in dynamic vehicle routing. In: Large RO, Kramer N, Radig AK, Schäfer M, Sulzbach A (eds) *Proceedings of Logistikmanagement 2017*, pp 17–24
- Thomas BW (2007) Waiting strategies for anticipating service requests from known customer locations. *Transp Sci* 41(3):319–331
- Ulmer MW, Mattfeld DC, Köster F (2018) Budgeting time for dynamic vehicle routing with stochastic customer requests. *Transp Sci* 52(1):20–37
- Ulmer MW, Soeffker N, Mattfeld DC (2018) Value function approximation for dynamic multi-period vehicle routing. *Eur J Oper Res* 269(3):883–899
- United Parcel Service of America, Inc (2018) UPS Pulse of the Online Shopper™ Study - Global Study Executive Summary. <https://www.ups.com/assets/resources/media/knowledge-center/ups-pulse-of-the-online-shopper.PDF>. [Online; accessed 16-December-2021]
- Voccia SA, Campbell AM, Thomas BW (2019) The same-day delivery problem for online purchases. *Transp Sci* 53(1):167–184
- Zhang J, Luo K, Florio AM, Van Woensel T (2023) Solving large-scale dynamic vehicle routing problems with stochastic requests. *Eur J Oper Res* 306(2):596–614

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ninja Soeffker¹  · Marlin W. Ulmer² · Dirk C. Mattfeld³

✉ Ninja Soeffker
ninja.soeffker@univie.ac.at

Marlin W. Ulmer
marlin.ulmer@ovgu.de

Dirk C. Mattfeld
d.mattfeld@tu-braunschweig.de

¹ Department of Business Decisions and Analytics, University of Vienna, Vienna, Austria

² Chair of Management Science, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany

³ Decision Support Group, Technische Universität Braunschweig, Braunschweig, Germany