Authors' Response to Comments of

An optimization and simulation method for the home service assignment, routing, and scheduling problem with stochastic travel time, service time, and cancellation

Team *The Optimistics*
Daniel Yamín, Daniel Barahona
Advisor: Alfaima L. Solano-Blanco
13th AIMMS-MOPTA Optimization Modeling Competition

RC: ***Reviewers' Comment***, AR: Authors' Response

We are very grateful for the valuable feedback and constructive comments of the committee. Their insightful remarks significantly improved the solution method and the report. This document details the actions we have taken in response to each of the referees' comments. The revised version of the manuscript highlights the substantial changes in blue.

1. Summary of main changes in the revised manuscript

The main changes in the revised manuscript are:

- We add a brief literature review about the home service problem with underlying uncertainty.
- To overcome the limitation of using expected values, we remark that we can add robustness to the optimization model by using percentiles.
- We design most of the figures again for the sake of clarity in the exposition.
- We develop a heuristic method to solve the pricing problem efficiently based on the route-first cluster-second principle.
- We propose a heuristic configuration of the method to solve the H-SARA problem within seconds.
- We present an overview of the method of moments to adapt the proposed ideas to other probability distributions.
- We completely overhaul Section 6 of computational experiments to
  1. have more insight into the performance of the proposed method,
  2. analyze the impact of changing parameters on the solution,
  3. describe in detail some features of the AIMMS-based application.
- Regarding the AIMMS application, we add the following features:
  - We present the appointment times in hour format and thus we add the start time parameter.
  - We provide an Excel data reader to upload the customers' information easily.
  - We allow users to add an estimated probability of cancellation for each customer.

2. Answer to the comments of the referees

RC: ***All the teams used test cases for up to 50 customers. How do their approaches scale to say, 200 customers? Would their approaches work for 1000+ customers also, say, 3000 customers? What changes are needed so that an instance of this size can be solved in a reasonable time?***

AR: Motivated by this comment, we tackled the main computational bottleneck of our solution method, namely, solving the pricing problem. In particular, we propose a heuristic method based on the route-first cluster-second principle to solve the subproblem (see Section 4.3 for details). Thus, we now propose two approaches and Section 4.4 reads

> The proposed ideas can be embedded as a building block in different algorithms. For the purpose of the present study, we propose two algorithm configurations: (i) the Exact Method (EM), in which the pricing problem is solved exactly using the mathematical programming-based method; and (ii) the Heuristic Method (HM), in which we use the route-first cluster-second split mechanism to solve the subproblem. Since the EM solves the pricing problem exactly, its solution to the LP relaxation of the set covering model (4.1)–(4.3) is a dual (i.e., lower) bound on the optimal objective and can be used to compute an optimality gap. Regarding the parameter $t_{\max}$, it allows decision makers to balance between the computational time and the quality of the solution.

Regarding test cases with 50 and 200 customers, the experiments in Section 6.1 conclude that the HM finds good solutions in a reasonable time. To show the scalability of the proposed ideas, we also execute experiments with 1,000 and 3,000 customers using the initial solution (found by the split mechanism) to solve the sizing, assignment, and routing problem, meaning that we skip the route-finding procedure. On average, the method finds solutions for test cases with 1,000 and 3,000 customers in about 110 and 3,000 seconds, respectively.

RC: ***What changes are needed if the "homogeneous teams" assumption is dropped? i.e., different customers have different skill needs, say randomly generate from 3 team types.***

AR: This is a very interesting question which we (coincidentally) discussed in-depth during one group meeting. The proposed method can be easily adapted to tackle the problem in which there is a set T of team types, and customers need to be visited by a certain type according to its needs. In such a case, we would decompose the H-SARA problem and solve it for each type of team (considering only the customers that need to be visited by that specific type). Thus, we would need to solve $|T|$ H-SARA problems independently, which could be efficiently performed through parallel computing.

RC: ***What changes are needed if the customer appointment times are known (removing the Scheduling part of the problem) ?***

AR: Conveniently, in our solution scheme, we decompose the H-SARA problem into two stages. In the first stage, we solve the sizing, assignment, and routing problem employing a column generation-based heuristic. In the second stage, we solve the appointment scheduling problem using a simulation-driven approach. Thus, if we remove the scheduling part of the problem, we would only execute the first stage of our solutions scheme; that is, the column generation-based heuristic.

RC: ***There is a lot of recent literature on stochastic (single and multiple servers) appointment scheduling and stochastic home care related to H-SARA that need to be acknowledged.***

AR: We thank the committee for this insightful comment. We have incorporated into our report a brief review on the home service problem with underlying uncertainty as follows

More recently, the home service problem with underlying uncertainty has gained the attention of researchers. For instance, Chen et al. (2017) tackle the problem by formulating an integer program with chance constraints to cope with uncertainty in durations. Also, Cappanera and Scutellà (2021) address the consistency and demand uncertainty in the home care planning problem. We refer to Zhan et al. (2021) for a complete literature review on the home service routing and appointment scheduling with stochastic times. We highlight that most solution methods are either very complex or unable to find rapidly high-quality solutions for large-scale, real-world applications.

RC: ***We assume that travel times satisfy the triangle inequality. WHY? What happens when this assumption does not hold?***

AR: The triangle inequality should naturally hold in real-world applications since the least costly way to move from one place to another is to go directly without intermediate stops. The importance of the triangle inequality assumption in this study is the approximation ratio of the approximation algorithm. More precisely, for Algorithm 1 to be a 2-approximation algorithm, the triangle inequality must hold. If the assumption does not hold, we would not have information regarding the quality of the solution found by Algorithm 1. We refer to the proof of Theorem 35.2 in Cormen et al. (2009) for details. Interestingly, Theorem 35.3 in Cormen et al. (2009) establishes that, without the triangle inequality, a polynomial-time approximation algorithm for the TSP does not exists unless P $=$ NP.

RC: ***Model (1)-(10) only incorporate expected values of random parameters? This is not enough to model uncertainty and is a limitation of this work.***

AR: Using expected values is very convenient since, by the linearity of expectation, we minimize the expected travel time and overtime costs when solving the sizing, assignment, and routing problem. Also, the expected utility is one of the most used metrics to evaluate decisions with underlying uncertainty (Li et al., 2016) and similar approaches have been used in the literature (e.g., Taş et al., 2013). Finally, we remark that, in our solution scheme, the uncertainty is mainly accounted for in the simulation model when scheduling the appointments. Nevertheless, we acknowledge that model (1)-(10) can lack robustness. In this light, we incorporate the following approach

Note that using expected values is convenient by the linearity of expectation. Note too that the expected utility is the most straightforward metric to evaluate decision-making under uncertainty (Li et al., 2016). Alternatively, to add robustness to the optimization model, one could calculate a high percentile, say 95, of the distribution of travel and service times, and use those values rather than the expected value.

RC: ***You assumed that the distribution service time is exponential. Various other studies show that it is log-normal. Can you approach accommodate other distributions? Would your insights and analysis change? What if the distribution is unknown?***

AR: This is an intriguing question since we initially explained in the report that the approach could be used with other distributions. For the sake of space economy, we removed that part. To answer this question, we added the following paragraph to the manuscript

Using distributions that correctly model the stochastic parameters enhances our solution scheme as it becomes more robust and reliable. Nevertheless, the proposed ideas are not limited to these distributions and can be flexibly adapted to other families. To that end, Appendix B presents an overview of the method of moments in order to estimate the parameters of a given probability distribution. Also, we remark that one can find a well-suited distribution to a parameter by performing a goodness-of-fit test on the historical data set.

In short, any distribution can be used in the approach. For the sake of simplicity and along the lines of related literature, we suggest using the method of moments to estimate the parameters of the distribution. For that purpose, we added Appendix B with an overview of the method. Also, note that through a goodness-of-fit test, a well-suited distribution can be found. Using different distributions changes the solution (i.e., decisions) but does not impact the algorithm's running time, as the sensitivity analysis presented in Section 6.2 shows.

RC: ***I found the attention to the customer cancellation lacking.***

AR: We thank the referees for raising this important source of confusion. We made some additions to the report in order to clarify how our approach accounts for customer cancellation. In Section 2, we incorporate customer cancellations into the model (2.1)–(2.10). In Section 5.1, we characterize customer cancellations through Bernoulli distribution. In Section 5.2, we conduct Monte-Carlo simulations in which we generate random Bernoulli trials to simulate the cancellations. To address this source of confusion, we wrote a part of §5.2 as

> We remark that the simulation model does account for the customers' cancellations. In particular, if the result of the Bernoulli trial is a success for a given customer –meaning that the customer cancels its appointment– the service teams go directly to the location of the next customer on the route. Note that, in this way, appointments are scheduled early when the previous customer on the route has a high chance of canceling. Thus, when a cancellation occurs, the vehicles' idle time is minimized.

We also change the title of the manuscript to emphasize that our approach accounts for customers' cancellations.

RC: ***An interesting approach in route first, cluster second. What alternatives did you explore? (obvious one being cluster first, route second which could be promising as routing problems are typically more difficult to solve than clustering problems).***

AR: We thank the committee for acknowledging the approach taken in the route-first cluster-second. The method works really well in the H-SARA problem and thus, in the revised version, we use it no only to obtain an initial solution but also to solve the pricing problem heuristically. Before using this approach, we first consider solving a bipartite matching problem (which can be done efficiently through the Hungarian algorithm, for example) to match customers close to each other. Next, we tried the k-means clustering algorithm to find groups of customers. However, this approach was insufficient as it was difficult to choose the number of clusters (i.e., solve the sizing problem). Therefore we also tried hierarchical clustering, and more precisely, using a *dendrogram* to cluster customers. Other approaches we experimented with include classical heuristics to solve the routing problem, such as the Clarke and Wright method and the sweep-based algorithm. We believe that extending the route-first cluster-second principle to the H-SARA problem could be a valuable asset to the literature.

RC: ***5.2 Simulation Model: Formalizing this idea, given a value of α –which is specified by the decision maker– we simply set w_i to the α-percentile of the arrival time distribution $p_{\psi_i}(\cdot)$. This approach is very simple and assume that the decision-maker knows how to find the percentile of the arrival time distribution, which implies that the decision-maker can also model the arrival time distribution. How to model this distribution? The arrival time to each customer is a random decision variable that depend on scheduled time, travel time, and other factors.***

AR: As mentioned in the question, the arrival time to each customer is a random variable that depends on several factors meaning that it is the sum of distinct random variables. Since it can be tricky (at least theoretically) to find its probability distribution, we use a more practical approach. Using a Monte-Carlo simulation, we find the empirical probability distribution of arrival times. The simulation model is convenient and straightforward since it circumvents estimating a complicated distribution and is computationally-wise very efficient. As in the AIMMS application, the decision-maker only needs to choose the desired probability of on-time arrival. The Monte-Carlo simulation and the calculation of the percentile are coded in the Python-based program.

RC: ***More results/insights related to the application will improve the report (e.g., decisions vs. different values of cost parameters, performance vs. different values of cost parameters).***

AR: In light of this comment, we added a sensitivity analysis on the costs and the decision-maker profiles (Section 6.2) to the document. In these experiments, we analyze the changes in the solution and the performance of the algorithm when varying the fixed hiring cost, the unit travel time cost, the unit overtime cost, and the desired probability of on-time arrival.

RC: ***5.3 More details on AIMMS tool and how a decision-maker can use it to generate insights.***

AR: We thank the committee for this valuable comment. We have re-written Section 5.3 as follows

> We embedded our solution method into a user-friendly high-level decision support system (DSS) developed in AIMMS. This computerized program can be used in decision-making in the home service industry, allowing decision-makers to solve problems timely, improve their efficiency, and perform other critical tasks related to operations, planning, and management. Conveniently, the DSS is flexible and adaptable to accommodate changes in the environment and the user's decision-making. The application can be used by upper- and mid-level managers to analyze multiple outcomes based on optimization models and analytical techniques.
>
> We discuss some of the essential features of the AIMMS-based application. First, as Figure 7 shows, the user interface of the DSS is intuitive, easy to use, and visually attractive to help managers interact with the tool. Second, the DSS provides a real-world geographical visualization of the solution, facilitating planning and operations-related tasks. Third, the information system presents a cost breakdown so that users gain insight into their operational costs. This feature allows decision-makers to decrease costs, a critical activity in a competitive market. Fourth, the DSS details information regarding the cost, travel time, and visiting customers of each route to help automate the managerial processes. Fifth, the application allows users to perform a what-if analysis on the costs, the desired on-time probability, and the number of customers serviced. This analysis reveals new approaches for the company to increase its competitiveness. Sixth, the DSS outlines a schedule for each customer, including its appointment time and route, to increase the control of the operation. Seventh, the users can specify a maximum running time to balance between the computational time and the quality of the solution. We highlight that this feature allows managers to find solutions within seconds for on-time applications or near-optimal solutions (at the expense of longer computational time) for long-term planning applications. Finally, the information can be uploaded to the application from an external file to facilitate the handling of large data sets. In short, the AIMMS-based DSS helps managers make critical decisions based on analytical methods improving the efficiency and competitiveness of their companies.