ABSTRACT
In modern society, sustainable transportation practices in smart cities are becoming increasingly important for both companies and citizens. These practices constitute a global trend, which affects multiple sectors resulting in relevant socio-economic and environmental challenges. Moreover, uncertainty plays a crucial role in transport activities; for instance, travel time may be affected by road work, the weather, or accidents, among others. This paper addresses a rich extension of the capacitated vehicle routing problem, which considers sustainability indicators (i.e., economic, environmental and social impacts) and stochastic traveling times. A simheuristic approach integrating Monte Carlo simulation into a multi-start metaheuristic is proposed to solve it. A computational experiment is carried out to validate our approach, and analyze the trade-off between sustainability dimensions and the effect of stochasticity on the solutions.

1 INTRODUCTION
Nowadays, the growing public concern for the environment preservation and social welfare is leading to more sustainable cities. In this sense, smart cities are places that have implemented information and communication technologies for getting an optimal transport system, considering economic, environmental, and social aspects in urban zones. For instance, most companies are starting to design and apply smart strategies to control environmental impacts. While a number of studies tackle the sustainability issues from an environmental perspective, the sustainability also involves social and economic factors (McKinnon et al. 2015). In addition, governments create monetary instruments such as road pricing and fines related to emissions excess and traffic noise, among others. Thus, a route cost varies according to the considered country or region, and the type of vehicle and road, to mention some examples.

In this context, we define the capacitated vehicle routing problem with stochastic traveling times (CVRP-ST). Traditionally, the goal of the CVRP is to design routing plans to serve a set of customers from one depot minimizing the traveling distance. A high correlation among traveling times and distances is typically assumed, which is unrealistic in urban routing (Figure 1). Recently, this problem has been enriched with goals related to fuel consumption. However, there are other negative factors that may be reduced by an efficient distribution planning. In fact, there is a lack of works focused on sustainability indicators (Eshtehadi, Fathian, and Demir 2017). Here, we consider the three cost dimensions: i.e., economic, environmental, and social dimensions. Moreover, authors tend to work on deterministic problems, but this assumption is too demanding for routing problems, where there is a wide range of elements with unpredictable but potentially significant effects on traveling times such as road works, the weather, accidents, etc. The high number of agents interacting also adds stochasticity to the traffic flow. Here, we model traveling times as random variables following specific probability distributions, either theoretical or empirical ones.
To reduce the existing gap in the literature, we present a simheuristic algorithm that provides high-quality solutions in a few seconds. This approach provides robust solutions against variations in traveling times, i.e., under any urban conditions the robust solution is a good solution that ensures a minimal impact. Being a rich VRP (Cáceres-Cruz et al. 2014; Miranda and Conceição 2016), the CVRP-ST is \( \text{NP}\)-hard, which means that heuristics/metaheuristics are needed to solve real-size instances in a reasonable computing time. However, these approaches are typically designed for deterministic combinatorial optimization problems (COPs). For this reason, we apply simheuristics (Juan et al. 2015), which integrate simulation into metaheuristics to solve stochastic COPs. In particular, our approach integrates Monte Carlo simulation (MCS) into a multi-start metaheuristic (Martí, Resende, and Ribeiro 2013). While the metaheuristic searches for promising solutions, the MCS component assesses them. This assessment is done by simulating a number of scenarios, where each scenario is defined by the generation of one value per stochastic variable relying on the corresponding probability distribution. Then, performance measures are computed such as expected total cost or total cost variance. Since MCS may be time-consuming, only a few promising solutions are assessed and a relatively small number of scenarios are simulated. Additionally, simulation is used to perform a risk analysis considering the top best solutions found during the execution of the algorithm. For example, the best solution in terms of expected total cost may be different, in terms of variance or a given percentile, from the second best solution. This analysis is based on accurate measures (i.e., using a higher number of simulation runs) and suitable visual techniques to easily show the behavior of the best solutions. If realistic restrictions based on time are imposed (e.g., limiting the number of hours that a driver may drive without resting), reliability analysis may be carried out with simulation. In other words, simulation allows us to estimate the probability of violating this restriction, since the stochasticity makes it impossible to guarantee that it will be satisfied in all possible scenarios. There are several works employing a simheuristic approach for addressing a routing problem (see, e.g., Cáceres-Cruz et al. 2012; González-Martin et al. 2012; Muñoz-Villamizar et al. 2013). However, to the best of our knowledge this is the first work considering sustainable indicators and stochastic traveling times. A computational
experiment is performed to illustrate and test our approach, and quantify the effect of stochasticity on the solutions’ performance. The paper is organized as follows. While section 2 reviews related works, section 3 defines the problem addressed. Section 4 presents the proposed approach. Sections 5 and 6 describe the computational experiment and the results, respectively. Finally, a few conclusions are drawn in section 7.

2 BACKGROUND

According to Eshtehadi, Fathian, and Demir (2017), the number of works related to the green VRP (G-VRP) has dramatically increased during the last years. Most studies focus on theoretical scenarios and evaluate emissions or fuel consumption without considering the social dimension. Erdogan and Miller-Hooks (2012) introduce the G-VRP, which extends the classical VRP by including the minimization of fuel consumption to reduce operational costs, which also minimizes the CO$_2$ emissions. Demir, Bektas, and Laporte (2011) propose a model to estimate the pollutants released based on distance traveled, and vehicles’ weight and speed. In the same lines, Xiao et al. (2012) estimate fuel consumption, which varies according to the distance traveled and the load carried. Later, Zhang et al. (2015) and Kuo (2010) solve the VRP minimizing fuel consumption cost and emissions. Faulin et al. (2011); Liu et al. (2014) and Zhang et al. (2015) solve the fuel consumption VRP (FCVRP) focusing on loadings, which constitute an essential variable determining the fuel consumption and the level of emissions. Xiao and Konak (2015); Demir, Bektas, and Laporte (2011) and Kuo (2010) solve the green heterogeneous VRP (G-HVRP) taking into account the influence of traffic congestion, road gradient, speed variations and distance traveled on the route efficiency. Scott, Urquhart, and Hart (2010) analyze the emissions and fuel consumption for internal combustion engine vehicles considering asymmetric costs, which are defined by the road gradient and the vehicle’s loading. In addition to fuel consumption, other negative factors are considered when designing urban routing. For instance, Jabbarpour, Noor, and Khokhar (2015), Meng, Liu, and Wang (2012), and Uchida (2014) study congestion issues. Delucchi and McCubbin (2010) and Nash (2003) estimate an economic factor to quantify the accident risk for pedestrian and vehicles according to speed variations.

Regarding simheuristics, this approach is becoming increasingly popular. For instance, an algorithm for solving the arc routing problem with stochastic demands is proposed by González-Martin et al. (2012). The approach combines MCS with the RandSHARP algorithm, which makes use of an adapted version of the Clarke and Wright Savings (CWS) heuristic (Clarke and Wright 1964) integrating biased-randomization techniques (Juan et al. 2011). A review on these techniques and simheuristics to address vehicle and arc routing problems is presented in Gonzalez-Martin et al. (2014). Cáceres-Cruz et al. (2012) introduce an algorithm for addressing the single-period inventory routing problem with stochastic demands, relying on the multi-start metaheuristic. Juan et al. (2014) analyze a similar problem considering stock-outs. Muñoz-Villamizar et al. (2013) consider the integrated location and routing problem in urban logistics. Recently, Guimarans, Dominguez, and Juan (2016) present a hybrid sim heuristic algorithm that combines biased-randomized routing and packing heuristics within a multi-start framework for solving the two-dimensional VRP, where customers’ demands are composed by sets of non-stackable items. Moreover, it is important to highlight the potential of simheuristics to help decision-makers to consider objectives of environmental and social welfare. Simheuristics allows the quick generation and assessment of multiple promising solutions. It allows companies to better address real and challenging transportation problems which appear in any smart city.

3 PROBLEM DESCRIPTION

The CVRP-ST may be defined by a directed graph $G = (N, A)$. $N = 0 \cup N_c$ is the set of nodes, where 0 corresponds to the depot, and $N_c = \{1, 2, ..., n\}$ is the subset of customers. $A = \{(i, j) : i, j \in N, i \neq j\}$ is the set of arcs that connect all nodes in $N$. Each customer $i$ has a known positive demand $q_i$. There is a set $K$ of $k$ homogeneous vehicles with a capacity of $Q$. Each route starts and ends at the depot, and all customers’ demands must be satisfied. Each arc $(i, j)$ is characterized by a traveling distance ($d_{ij}$) and a traveling
time \((T_{ij})\). Times are assumed to be random variables, since they depend on external and unpredictable factors. The total traveling time of a route is limited to \(T_{lim}\), which represent the maximum time a driver can drive without resting. This constitutes a hard constraint in a deterministic environment. However, this condition cannot be guaranteed in the stochastic environment. In this case, a penalization \(t'\) in terms of time is added if the constraint is not fulfilled. For a given solution, each arc has a cost \((c_{ij})\) that represents the impacts of traveling through it considering the following dimensions:

- **Economic dimension.** It includes two indicators: the total traveling time, which affects the amount paid for driver wages and vehicle fixed costs, and the total traveling distance, which is monetized relying on the oil price.
- **Environmental dimension.** Kuo (2010) and Zhang et al. (2015) introduce a model that considers CO\(_2\) emissions as an impact related to fuel consumption. Here, we estimate the fuel consumption and the emissions as suggested in these works.
- **Social dimension.** Related indicators are very subjective, because most of the negative impacts generate intangible effects on people. These effects are difficult to measure and depend on the perspective analyzed, which leads to diverse results and practical implications (Navarro et al. 2016; Demir et al. 2015; McKinnon et al. 2015; Anand et al. 2012). We follow the approach of Delucchi and McCubbin (2010), which monetizes the accident risk for pedestrian and vehicles. This risk depends on the distance and the vehicle’s loading on that arc.

All in all, the objective is defined as a multi-criteria function considering the expected traveling time cost, traveling distance cost, the environmental cost and the social cost.

4 SOLVING APPROACH

A simheuristic methodology integrating MCS into a multi-start metaheuristic is presented, which relies on the biased randomization version of the CWS (BR-CWS) heuristic (Juan et al. 2011) to generate solutions. Our approach (Figure 2) relies on two facts: (i) the CVRP-ST can be seen as the CVRP when the random times have zero variance; and (ii) efficient but simple algorithms exist for the CVRP. The multi-start metaheuristic generates a number of solutions for a given amount of time and returns the best one.

The first step is transforming the CVRP-ST instance into a CVRP one replacing the random variable \(T_{ij}\) by its mean value \(t_{ij} = E[T_{ij}]\). Then, an initial solution \((\text{initSol})\) is created. A solution consists of a set of routes, in which each route is represented as a sequence of customers to visit. The performance of this solution in the stochastic environment is assessed with MCS following these steps: (i) simulate a scenario in which a value is generated per random variable according to the associated probability distribution; (ii) compute the total cost; and (iii) repeat the two previous steps until having simulated \(n_{Sim}\) scenarios, and compute the expected total cost. Step (ii) adds the penalization to a route’s cost if the restriction of duration is violated. Afterwards, a solution \((\text{bestDetSol})\) is created to store the ‘best deterministic solution’ (i.e., that providing the lowest total cost considering the deterministic instance). In addition, a list of \(l\) solutions \((\text{bestStochSolList})\) is created to include the ‘best stochastic solutions’ (i.e., those with the lowest expected total cost). Initially, \(\text{initSol}\) is copied into \(\text{bestDetSol}\) and \(\text{bestStochSolList}\).

Then, a loop with a stopping criterion based on the elapsed computational time is started. First, a solution \((\text{newSol})\) is created and an acceptance criterion is employed to decide whether it is classified as promising. If \(\text{newSol}\) is not promising then it is discarded, and another iteration starts. Oppositely, if \(\text{newSol}\) is promising, MCS is applied to assess it. \(\text{bestDetSol}\) is replaced by \(\text{newSol}\) if the latter presents a lower total cost, and \(\text{bestStochSolList}\) is updated according to the expected total cost. Once the loop is ended, the solutions in \(\text{bestStochSolList}\) are assessed again performing \(n_{Sim}\) simulations runs. While \(n_{Sim}\) is a relatively low value to obtain rough estimates of the expected total costs in a short amount of time, \(n_{Sim}\) is a greater value that provides more accurate estimates for the best stochastic solutions found.

Finally, a risk analysis is performed, where the solutions are compared not only based on expected total
costs, but also on other measures such as variance, quartiles, etc. Additionally, the reliability of a solution, defined as one minus the probability of violating the duration limit of at least one route, can be estimated by computing the proportion of scenarios where that happens.

In order to classify a solution as promising or not, a variable \( rpd \) is computed, which measures the relative difference percentage between the total cost of \( best\text{DetSol} \) and \( new\text{Sol} \). If there is an improvement (i.e., \( rpd < 0 \)), \( new\text{Sol} \) is considered a promising solution. Otherwise, \( new\text{Sol} \) is also declared promising with a probability of \( \exp(-rpd) \). This acceptance criterion aims to avoid entrapment at local optimum.

As discussed in Hatami, Ruiz, and Andrés-Romano (2015), it is similar to a simulated annealing criterion, but simpler and without parameters.

![Diagram of the proposed approach for the CVRP-ST with sustainability indicators using BR-CWS and MCS.](image)

The generation of solutions is based on the BR-CWS heuristic. Biased randomization allows the randomization of deterministic and iterative heuristics, and aims to obtain a high number of promising solutions. In particular, the choice of one element from a list is done by assigning a probability to each element, being this probability correlated with a measure of preference. This concept relies on the fact that
the best step in the short term is not necessarily the best one in the long term. A geometric distribution is used, which assigns a probability of selection to each edge according to its position inside the sorted list of savings. Finally, to consider sustainability indicators, the classical distance-based savings of the CWS are replaced by ‘rich savings’ including all costs.

5 COMPUTATIONAL EXPERIMENTS

Our algorithm, represented in Figure 3, has been implemented in Java and run on a personal computer with 8 GB of RAM and an Intel Core i7 of 1.8 GHz. The following CVRP benchmark instances, which may be found in the CVRPLIB library (http://vrp.atd-lab.inf.puc-rio.br/index.php/en/), are used for the experiment: E-n23-k3, E-n33-k4, E-n51-k5 and E-n60-k9 proposed by Christofides and Eilon 1969, and A-n60-k9 (Augerat et al. 1995). The numbers appended to ‘n’ and ‘k’ represent the number of nodes (i.e., customers and the depot) and vehicles, respectively.

Firstly, we validate our approach (identified by RCTFJ) by comparing the deterministic solutions obtained by minimizing total distance with the best known solutions (BKS). For each instance, the information included in Table 1 is the following: instance name, distances of the BKS and our solution, gap and maximum computing time. According to the results, our approach reaches the optimal solution in one instance, and presents a gap lower than 1.5%, on average.

Secondly, we adapted the instances to consider traveling times. Following the approach of Feng, Zhang, and Jia (2017), traveling speeds are assumed to be a mixture of truncated normal distributions (to select only positive values). Table 2 describes the different distributions employed and their weights. Additionally, an smoothing procedure is applied, which defines the final speed $v_{ij}$ as $\alpha \cdot \hat{v}_{ij} + (1 - \alpha) \cdot \mu$, where $\mu$ refers to the mean velocity and $\alpha$ represents a weight set to 0.8. Having the distances and the speeds, it is immediate to obtain the traveling times.

In order to introduce stochasticity, each traveling time $t_{ij}$ is replaced by a random variable $T_{ij}$ which is assumed to follow a lognormal distribution through location parameter, $\mu_{ij}$ and scale parameter $\sigma_{ij}$ with $E[T_{ij}] = t_{ij}$ and $Var[T_{ij}] = k \cdot t_{ij}$, such that:

Table 1: Validation of our approach in a deterministic environment considering distances.

<table>
<thead>
<tr>
<th>Instances</th>
<th>BKS</th>
<th>RCTJ</th>
<th>Gap (%)</th>
<th>Max. time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-n23-k3</td>
<td>569</td>
<td>569</td>
<td>0.00</td>
<td>100</td>
</tr>
<tr>
<td>E-n33-k4</td>
<td>835</td>
<td>838</td>
<td>0.36</td>
<td>100</td>
</tr>
<tr>
<td>E-n51-k5</td>
<td>521</td>
<td>534</td>
<td>2.50</td>
<td>100</td>
</tr>
<tr>
<td>A-n60-k9</td>
<td>1354</td>
<td>1368</td>
<td>1.03</td>
<td>300</td>
</tr>
<tr>
<td>E-n76-k10</td>
<td>830</td>
<td>847</td>
<td>2.05</td>
<td>300</td>
</tr>
</tbody>
</table>

Table 2: Description of the parameters used to generate speeds (km/h) in different instances.

<table>
<thead>
<tr>
<th>Distributions</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Weights</td>
<td>20%</td>
<td>40%</td>
<td>40%</td>
</tr>
</tbody>
</table>
Figure 3: Scheme of our algorithm.
\[ \mu_{ij} = \ln(E[T_{ij}]) - \frac{1}{2} \cdot \ln \left( 1 + \frac{\text{Var}[T_{ij}]}{E[T_{ij}]^2} \right) \]

\[ \sigma_{ij} = \sqrt{\ln \left( 1 + \frac{\text{Var}[T_{ij}]}{E[T_{ij}]^2} \right)} \]

It is expected that as \( k \) converges to zero the results from the stochastic version converge to those obtained in the deterministic scenario. Notice that the assumption regarding the time distribution is needed only to generate data. In a real-life application, information and communication technologies would enable the use of real data. 3 values are used in the experiments to study different levels of stochasticity: 0.25, 0.35, and 0.45. The penalization \( t' \) is set to 4. The limit of traveling time per route ranges between 6 and 9, depending on the instance.

The algorithm parameters are specified as follows: (i) the geometric distribution parameter \( \alpha \) is set to 0.3; (ii) the number of simulation runs during the loop and the refinement are set to 200 and 2000, respectively; and (iii) the maximum time is limited to 60 seconds. The number of simulations were experimentally chosen, considering the trade-off between computational time (since MCS tends to be time-consuming) and accuracy of the measures of performance for each solution (i.e., expected total cost and reliability). According to Talbi (2009), it is a good practice when working with randomized algorithms to run several independent executions. Indeed, it may lead to significantly better solutions. We perform 5 executions, each one with a different seed.

6 ANALYSIS OF RESULTS

First, the weight of each component in the expected total cost is computed for the solution provided by the simheuristic considering the lowest level of stochasticity. The results for each instance are shown in Figure 4. It can be observed that the weights are very different, which may be due to the different scales of the instances’ characteristics. On average, the distances and the emissions represent the highest weights. The cost of the social dimension, which basically depends on the demands and distances, presents the highest variability.

![Figure 4: Composition of the best stochastic solutions.](image)
Then, the need of a simheuristic approach is assessed by comparing the performance of the best deterministic solutions (i.e., those minimizing the total cost) and the best stochastic solutions (i.e., those minimizing the expected total cost). Table 3 gathers the results for each instance with the highest level of stochasticity. In preliminary experiments, it was noted that the differences were negligible for low levels of stochasticity. Obviously, the expected total costs are significantly higher than the total costs, since they include the penalizations. It is concluded that the reliabilities tend to be very high, i.e., the solutions found are reliable. In most cases the best deterministic and stochastic solutions are the same, except for the instances E-n51-k5 and A-n60-k9. In these cases, the best deterministic solution shows a lower total cost, while the best stochastic one shows a lower expected total cost.

Table 3: Comparison of best deterministic and stochastic solutions considering a high level of stochasticity.

<table>
<thead>
<tr>
<th>Instances</th>
<th>Best deterministic solution</th>
<th></th>
<th>Best stochastic solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total cost</td>
<td>Expected TC</td>
<td>Reliability</td>
</tr>
<tr>
<td>E-n23-k3</td>
<td>651.05</td>
<td>662.82</td>
<td>0.9024</td>
</tr>
<tr>
<td>E-n33-k4</td>
<td>2067.25</td>
<td>2072.14</td>
<td>0.9701</td>
</tr>
<tr>
<td>E-n51-k5</td>
<td>368.75</td>
<td>371.06</td>
<td>0.9860</td>
</tr>
<tr>
<td>A-n60-k9</td>
<td>2136.05</td>
<td>2179.34</td>
<td>0.9277</td>
</tr>
<tr>
<td>E-n76-k10</td>
<td>576.24</td>
<td>579.09</td>
<td>0.9939</td>
</tr>
</tbody>
</table>

7 CONCLUSIONS

The concern for environmental and social welfare has been growing during the last two decades encouraging the transformation of cities into more sustainable places. As a consequence, many decision-makers conceive that technology and optimization techniques help to reach efficient logistic processes. In this context, we have addressed the capacitated vehicle routing problem with stochastic traveling times. The aim is to design routes that minimize the negative impacts measured by a few sustainability indicators. We propose a simheuristic approach to address this problem, which integrates Monte Carlo simulation into a multi-start metaheuristic. This procedures integration allows solving routing problem considering a random behavior to simulate different traveling times. One of the main contributions of our methodology is that it allows to consider personalized guidelines to get a balanced integration of the sustainability criteria. In that way, the simheuristic contributes to reduce negative impacts caused by transport activities. Another important contribution is that our approach can be used with any probability distribution to estimate traveling times, which means that our simheuristic allows us to test promising solutions in realistic scenarios.

According to our results, the proposed approach is able to provide high-quality solutions in short computational times, while coping with the stochasticity of traveling times. In addition, it has been observed that ignoring this stochasticity leads to worse solutions even in environments with a low level of uncertainty. Regarding sustainability indicators, it has been observed that the expected traveling time costs and the distance costs represent a high proportion of the expected total costs.

Several lines of future research stem from this work. For instance, our approach may be extended to consider asymmetric costs and stochastic demands, which are realistic characteristics. Another interesting line is to introduce an heterogeneous fleet of vehicles. In addition, sensibility analysis may be performed to study how the costs related to each indicator affect the solutions.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness and FEDER (TRA2013-48180-C3-P, TRA2015-71883-REDT), the Erasmus+ program (2016-IES01-KA108-023465),
and the Ibero-American Program for Science and Technology for Development (CYTED2014-515RT0489). We also thank the support of the UPNA and UOC doctoral programmes.

REFERENCES


AUTHOR BIOGRAPHIES

LORENA S. REYES-RUBIANO is a PhD student at the Public University of Navarra. She holds a MSc in Operations Management. Her work is related to the application of operations research and simulation techniques to solve combinatorial optimization problems in logistics and transportation. Her email address is lorena.reyes@unavarra.es.

LAURA CALVET is a PhD student at the IN3 - Open University of Catalonia. She holds a MSc in Statistics and Operations Research, a BSc in Economics and a BSc in Applied Statistics. Her work is related to the combination of statistical learning and simheuristics for solving complex combinatorial optimization problems under uncertainty. Her website address is https://lauracalvetl.wordpress.com and her email address is lcalvet@uoc.edu.

JAVIER FAULIN is a Full Professor of Statistics and Operations Research at the Public University of Navarre (Pamplona, Spain). He holds a Doctoral Degree in Economics and Business Administration from the University of Navarra. His research interests include transportation and logistics, vehicle routing problems and simulation modelling and analysis, particularly in the practical resolution of logistics and delivery problems of companies. His work is also related to the evaluation of the environmental impact of freight transportation and his email address is javier.faulin@unavarra.es.

ANGEL A. JUAN is Associate Professor in the Computer Science Dept. at the Open University of Catalonia, and Principal Investigator at IN3. He holds a PhD in Industrial Engineering and a MSc in Mathematics. His research interests include applications of randomized algorithms and simheuristics in logistics, transportation, and finance. He has published over 50 JCR-indexed articles in these fields. His website address is http://ajuanp.wordpress.com and his email address is ajuanp@uoc.edu.