

UNIVERSITÀ DELLA CALABRIA



UNIVERSITA' DELLA CALABRIA
Dipartimento di Matematica e Informatica

Dottorato di Ricerca in
Matematica e Informatica
(Ricerca Operativa)

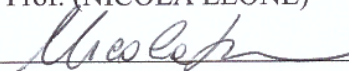
CICLO
XXXI

TITOLO TESI

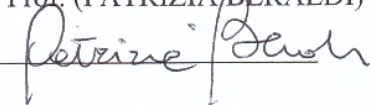
**Latency-based routing problems under uncertainty with some applications in
health care systems: models and methodologies**

Settore Scientifico Disciplinare MAT/09

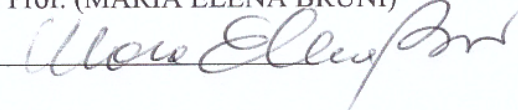
Coordinatore: Ch.mo Prof. (NICOLA LEONE)

Firma 

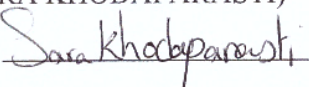
Supervisore/Tutor: Ch.mo Prof. (PATRIZIA BERARDI)

Firma 

Ch.mo Prof. (MARIA ELENA BRUNI)

Firma 

Dottorando: Dott./ssa (SARA KHODAPARASTI)

Firma 

To Maria Elena Bruni

Acknowledgements

I would like to express my deepest sense of indebtedness and profound gratitude to my thesis supervisors Prof. P. Beraldi and Prof. M.E. Bruni for their great inspiration, constant encouragement, practical and invaluable advice.

Abstract

In this thesis, we address some latency-based vehicle routing problems under the uncertainty of travel and service times. We review the contributions in the literature and present more efficient mathematical models enhanced by a prototype metaheuristic approach providing near-optimal solutions in low computational time. We also implement our contribution on different applications in disaster management and scheduling sector. As another contribution, we also address the equity in strategic and tactical problems arising in emergency medical service and primary health care sector. This is justified by the close connection of equity and latency, as two important performance measures, both related to the customer-centricity concept.

Contents

1	Preface	1
2	The risk-averse traveling repairman problem with profits	5
2.1	Introduction	5
2.2	The risk-averse K -TRPP under uncertain travel times	11
2.3	Mathematical formulation with position dependent variables	14
2.3.1	The iterated greedy heuristic	19
2.3.2	Computational results	22
2.4	Mathematical formulation defined on a layered graph	25
2.4.1	The hybrid reactive greedy randomized adaptive search heuristic	28
2.4.2	Computational results	37
2.4.3	Results for the P-instances	39
2.4.4	Results for the E-instances	42
2.4.5	Large size instances	46
2.4.6	Deterministic K -TRPP	47
2.5	Conclusions and future research directions	49
3	Uncertain latency-based routing problem with service level constraint: an application to disaster relief	50
3.1	Introduction and motivation	50
3.2	Disaster relief response and logistics planning: literature review	53
3.2.1	Deterministic models for humanitarian relief response	54
3.2.2	Stochastic models for humanitarian relief response . .	63
3.2.3	Simulation, GIS, and prediction methods	67
3.2.4	Findings and potential gaps	68

3.3	A latency-based vehicle routing problem with service level constraint under uncertainty	71
3.3.1	The Mathematical Model	74
3.3.2	The case study	77
3.3.3	Performance of the model	80
3.3.3.1	The impact of the risk	80
3.3.3.2	The fleet size	82
3.3.3.3	The service level	84
3.3.3.4	Model validation	85
3.3.3.5	The analysis under a disruption scenario: the occurrence of a tsunami	86
3.3.3.6	The analysis under a disruption scenario: the failure of the depot	88
3.3.3.7	The Monte Carlo simulation	91
3.3.4	The heuristic approach	93
3.3.5	Conclusions and future research directions	98
4	Asymmetric latency-based vehicle routing problems with uncertain travel and service times	99
4.1	The generalized robust model	100
4.2	The metaheuristic approach	103
4.2.1	Removal heuristics	107
4.2.2	Insertion heuristics	108
4.2.3	Adaptive mechanism	110
4.2.4	The acceptance criterion	111
4.2.5	The VND-based heuristic	111
4.2.6	Local search procedure	112
4.3	Application in the multi-machine scheduling context	113
4.3.1	Application description	113
4.3.2	Computational experiments	114
4.3.3	Proposed mathematical model versus the traditional order acceptance model	118
4.4	Conclusions and future research directions	120

5	Incorporating equity into emergency medical service strategic planning	121
5.1	Locating Emergency Medical Services: literature review and new challenges	121
5.1.0.1	Incorporating equity	123
5.1.0.2	Incorporating uncertainty	128
5.2	Balancing Efficiency and Equity in location-allocation models with an application to strategic EMS design	136
5.2.1	Introduction	137
5.2.2	The use of DEA in facility location models	138
5.2.3	Designing an equitable and efficient system	139
5.2.3.1	The EMS case	142
5.2.4	The exact solution approach	144
5.2.5	Computational results	147
5.2.6	Conclusions and future research directions	154
6	Location-allocation models in the primary health care sector	156
6.1	Enhancing Community based Health Programs in Iran: a multi-objective location-allocation model	157
6.1.1	Introduction	157
6.1.2	Literature review	159
6.1.2.1	Location-allocation models	159
6.1.2.2	Multi-objective models in health care sector	160
6.1.3	Problem description and formulation	163
6.1.4	The mathematical model	165
6.1.5	The fuzzy goal programming approach	169
6.1.6	Case study	172
6.1.6.1	Presentation of the results	175
6.1.6.2	The current system	176
6.1.6.3	Case A	177
6.1.6.4	Case B	179
6.1.6.5	Case C	181
6.1.7	Discussion	183
6.1.8	Conclusions and future research directions	183

6.2	A Multi-period Location–Allocation Model for Nursing Home Network Planning Under Uncertainty	184
6.2.1	Introduction	184
6.2.2	Literature Review	187
6.2.3	Problem description and mathematical formulation . .	192
6.2.3.1	The multi–period probabilistic location-allocation model	194
6.2.3.2	The deterministic equivalent formulation . .	197
6.2.4	Case Study	200
6.2.4.1	Case study description and input data	200
6.2.4.2	Results and findings	202
6.2.4.3	Current system evaluation	206
6.2.4.4	Probabilistic versus deterministic and time- invariant model	207
6.2.4.5	Monte Carlo simulation	209
6.2.5	Conclusions and future research directions	210
A	Appendix	212
A.0.1	Open facilities for Cases A, B, and C	212
A.0.2	Linearization	213
A.0.3	Coordinate transformation	214
A.0.4	Location coordinates of population centers and facilities	215
	Bibliography	218

List of Figures

2.1	Toy example	9
2.2	Optimal paths	9
2.3	Toy example revisited	16
2.4	Optimal paths in the risk-averse model	17
2.5	Frequency histograms: Risk-neutral model	18
2.6	Frequency histograms: Risk-averse model	18
3.1	The frequency histogram of selected papers in journals	54
3.2	Decision layers in humanitarian relief network	55
3.3	Haiti hearthquake, Encyclopaedia Britannica, Inc.	78
3.4	51 communal sections with potential damage higher than moderate 78	
3.5	Classification of the affected areas based on population and the severity level	79
3.6	<i>EAT</i> versus <i>STDAT</i>	81
3.7	Tsunami event-NOAA Center for Tsunami Research	86
3.8	The broken links under the tsunami scenario	87
3.9	Two numerical solutions	88
3.10	Disrupted links under the depot failure scenario	89
3.11	Two numerical solutions	90
3.12	Two numerical solutions	92
5.1	Emergency Care Pathway and related problems	123
5.2	Risk level α vs. within- and between-group Gini indices	150
5.3	The trade-off between efficiency and equity criteria in Pareto front	153
5.4	The percentage of trade-off: efficiency versus equity criteria in Pareto front	153

6.1	The system structure	164
6.2	9 Municipal zones of Shiraz city	173
6.3	Spatial distribution of population centers and existing and potential locations of facilities	174
6.4	Spatial distribution of optimal locations for case A	178
6.5	Spatial distribution of optimal locations for the case B	181
6.6	Spatial distribution of optimal locations for the case C	182
6.7	Spatial distribution of demand nodes and the locations of nursing homes in Shiraz city	201
6.8	Spatial distribution of optimal nursing homes' locations during each period	203
6.9	Coverage versus distance threshold	205
6.10	Stochastic time-invariant demand model versus the proposed model 209	

List of Tables

2.1	Table of notation	15
2.2	Simulation results	19
2.3	Results for the P-instances	23
2.4	Results for the E-instances	24
2.5	LBA versus BA and FBA for Pn55k7	39
2.6	Results for the P-instances, $\lambda = 0.1$	40
2.7	Results for the P-instances, $\lambda = 0.5$	42
2.8	Results for the P-instances, $\lambda = 0.9$	43
2.9	Upper bound performance for the P-instances	44
2.10	Results for the E-instances, $\lambda = 0.1$	45
2.11	Results for the E-instances, $\lambda = 0.5$	45
2.13	Upper bound performance for the E-instances	45
2.12	Results for the E-instances, $\lambda = 0.9$	46
2.14	Results for the CMT-instances	47
2.15	Results for the deterministic P-instances	48
2.16	Results for the deterministic E-instances	49
3.1	Classification of literature on deterministic models for humanitarian routing	59
3.2	Classification of literature on stochastic models for humanitarian routing	66
3.3	Sensitivity of the solution with respect to λ	81
3.4	Sensitivity of solution with respect to fleet size k	83
3.5	Sensitivity analysis as function of Γ	84
3.6	Real-road network versus the Euclidean-based network	85
3.7	Simulation results	93
3.8	Results for larger cases: proposed heuristic versus SCIP	97

4.1	Heuristic settings	114
4.2	Results for the instances with $n = 10$ nodes and $k = 1$	116
4.3	Results for the instances with $n = 15$ nodes and $k = 1$	117
4.4	Results for the instances with $n = 25$ nodes and $k = 2$	118
4.5	Comparing the results of two models	120
5.1	E^3MS in comparison with pure efficiency and equity models	151
5.2	The results for the Pareto front with $\alpha = 0.05$ and $RTT = 10$ minutes	151
6.1	Input and output parameters for CCs	174
6.2	The DEA score of CCs	175
6.3	The optimal results for case A	177
6.4	The optimal results for case B	179
6.5	The optimal results for the case B with priority structure	180
6.6	The optimal results for case C	182
6.7	Estimated elderly population	202
6.8	Case study inputs	202
6.9	The optimal nursing homes' sites	203
6.10	Classifying covered nodes based on the relative improvement in accessibility criterion	204
6.11	Sensitivity analysis with respect to α, ρ_0	205
6.12	The optimal nursing homes' sites for the deterministic model	208
6.13	The reliability level of the deterministic model	208
6.14	The probability of violation in the Monte Carlo simulation	210
A.1	Optimal locations for Scenario A	212
A.2	Optimal locations for Scenario B	212
A.3	Optimal locations for Scenario C	213
A.4	Location coordinates of population centers at zones 1-3	215
A.5	Location coordinates of population centers at zones 4-9	215
A.6	The coordinates of potential facility sites	216
A.7	Distance traveled by covered demands	217

List of Algorithms

1	Scheme of the heuristic	20
2	Adaptive local search	22
3	General structure of the RGRASP	29
4	GRASP(α): The construction phase	32
5	Scheme of the local search	33
6	Scheme of the HILS	37
7	The proposed heuristic	94
8	Initial solution	96
9	Perturbation (s, δ)	96
10	Hybrid ALNS-VND heuristic	106
11	The VND-based heuristic	112
12	The exact ϵ -constraint method	146
13	The exact fuzzy goal programming method	171

Chapter 1

Preface

This thesis is mainly focused on latency-based vehicle routing problems under uncertainty. A wide range of applications arising in different domains are inherently customer-centric posing customer's satisfaction, typically measured in terms of responsiveness, as a primary goal. For example, in the humanitarian logistics, that represents our elective application domain, the arrival time plays a crucial role in saving people's lives and reducing the mortality rate. While the scientific literature on traditional vehicle routing problems, both in the deterministic and stochastic variants, is rich and consolidated, very few papers analyze latency-based routing problems under uncertainty.

In Chapter 2, we first introduce the traveling repairman problem with profits (TRPP). In simple words, the TRPP is a variant of the well-known traveling salesman problem in which a repairman is supposed to visit a subset of nodes of a graph in order to maximize the total profit which is a decreasing function of arrival times (or latencies) at the customers. In Section 2.2, we present a risk-averse approach to handle the uncertainty of travel times in the TRPP via a mean-risk model and run a Monte Carlo simulation experiment to show how neglecting the stochasticity of travel times may result in unstable solutions. As our first contribution, we propose a position-dependent mathematical formulation which extends the traditional single-server model proposed by Dewilde ([87]) for the TRPP to the multi-vehicle case. This contribution is also enhanced by a prototype metaheuristic approach which provides good solutions in reasonable computational time. This contribution was presented in the *EURO Conference on Advances in Freight Transportation and Logistics (2018)* that was later published in the *Transportation Research Procedia* journal [92].

The TRPP position-dependent mathematical model is not the best model in terms of computational tractability, especially when commercial solvers are used to solve the problem. This motivated us to present in Section 2.4 a more efficient mathematical model built on a layered graph based on the same idea presented in [194]. We then proposed a metaheuristic approach, including three different algorithms which can be found in Subsection 2.4.1. Some parts of the latter contribution were presented in the *2nd European Conference on Stochastic Optimization (2017)* as well as in the *EURO/ALIO conference (2018)* and was later published in *Electronic Notes in Discrete Mathematics* journal [54]. The extended contribution was submitted to the *Computers & Operations Research* journal and is currently under second revision.

In Chapter 3, we address a latency-based vehicle routing problem with service level constraint considering again the uncertainty of travel times. The problem is closely related to the multi-vehicle TRPP, and, in fact, can be regarded as a variant of that problem in which the objective is to minimize the total latency and profits are instead considered into a constraint enforcing the achievement of a minimum service level. This contribution is enhanced by a simple iterative local search heuristic which was presented in the *46th Annual Conference of the Italian Operational Research Society (2016)*. Later, we also implemented the model for a real case study on the Haiti earthquake 2010 to assess the validity of the proposed approach as a decision-support tool addressing routing decision-makings arising in the disaster relief phase. The initial results of this contribution were presented in the *ORAHs conference (2017)*. We also submitted an extended version of this contribution, including a detailed discussion on the managerial insights and enriched by a variable neighborhood descent heuristic to the *European Journal of Operational Research*. The paper is currently under revision [39].

Chapter 4 presents an extension of the problem addressed in Chapter 3, where travel times are asymmetric and uncertain and uncertain service times are also considered. Specifically, we adopt a robust distributionally optimization approach to deal with the uncertainty of the input parameters varying over an interval set. From the solution viewpoint, we also propose a metaheuristic algorithm hybridizing the large neighborhood search with a variable neighborhood descent heuristic. The proposed mathematical model

as well as the algorithm are then applied to the selective multi-machine scheduling problem with uncertain processing and setup times. The problem, arising in the manufacturing sector, is indeed closely related to the vehicle routing problem. We also show how the proposed model, in the deterministic context, outperforms the existing model in the literature.

In addition to the contributions on routing problems, we have also investigated how the equity concept can be incorporated into strategic and tactical decision-makings arising in the health care sector. In effect, equity and latency can be both considered as customer-centricity performance measures. In Chapter 5, we first present a comprehensive review on strategic location and tactical allocation decisions in the emergency medical service (EMS) sector along with the potential gaps and challenges ahead ([21]). The Section 5.1 has been selected from a comprehensive review paper we published in the *Computers & Operations Research* journal [21].

Next, in Section 6.2 we propose a novel bi-objective location-allocation model balancing efficiency and equity criteria in the EMSs. We also present an exact solution approach to generate the Pareto-efficient solutions. A paper describing the application of the proposed model to a real case study was published in the *Optimization Letters* journal [140].

In Chapter 6, we focus on location-allocation problems arising in the primary health care sector. In Section 6.1, we address a novel location-allocation problem arising in a hierarchical public health care system that provide consulting services and addiction treatment cares to the residents. We also propose an exact fuzzy goal programming approach to solve the problem and test the model on a real case study for the city of Shiraz, in Iran. Our contribution provides interesting managerial insights on the performance of the current system and how it can be improved under different scenarios. This contribution was published in the *Health care Management Science* journal [141]. In Section 6.2, we deal with the optimal location and allocation of nursing homes in order to optimize the efficiency of the system while improving the accessibility over a multi-period planning horizon. The model aims at determining the optimal location of new nursing homes to be added into the system, at the beginning of each decision period, when new resources are provided. To deal with the ambiguity of demand changing over the long-term planning horizon, we propose a probabilistic chance constraint

approach in which the uncertainty of the demands is taken into account. Alike the former contributions, the model is tested on a real case study. This final contribution was published in the *Operations Research for Health Care* journal [[139](#)].

Chapter 2

The risk-averse traveling repairman problem with profits

2.1 Introduction

The traveling repairman problem (TRP) is a variant of the well-known traveling salesman problem in which a repairman is supposed to visit all the nodes of a graph exactly once minimizing the the sum of the arrival times (or latencies) at the customers. Besides being at the heart of important customer-centric vehicle routing applications, where some quality criterion regarding the customer's satisfaction must be pursued, this problem also appears in a variety of other contexts, such as smart grid maintenance, manufacturing, data retrieval networks in computer networks, home delivery and machine scheduling with sequence-dependent processing times.

The TRP (also known in the literature as minimum latency problem (MLP)) has been extensively studied by a large number of researchers who proposed several exact and non-exact approaches. Lucena [156] and Bianco et al. [45] proposed early exact enumerative algorithms, in which lower bounds are derived using a Lagrangian relaxation. Fischetti et al. [103] proposed an enumerative algorithm that makes use of lower bounds obtained from a linear integer programming formulation. Van Eijl [257], Méndez-Díaz et al. [172], and Ezzine et al. [97] developed mixed integer programming formulations with various families of valid inequalities. Bigras et al. [46] suggested a number of integer programming formulations as well as a branch-and-bound algorithm. Some approximation algorithms are also known for the MLP. The first one was suggested by Blum et al. [49] with an approx-

imation factor of 144. The current best approximation algorithm for the MLP is due to Chaudhuri et al. [72] for general metric spaces and to Archer and Blasiak [20] for the case where an edge-weighted tree is considered.

Up to this date, few metaheuristics are available for the TRP. Salehipour et al. [218] first proposed a simple composite algorithm based on a GRASP, improved with a variable neighborhood search procedure. In [178], Mladenović et al. presented a general variable neighborhood search metaheuristic enhanced with a move evaluation procedure facilitating the update of the incumbent solution. Silva et al. [235] presented a composite multi-start metaheuristic approach consisting of a GRASP and a randomized variable neighborhood descent algorithm for the construction and improvement phases, respectively. They also adopted an efficient move evaluation procedure to speed up the search over different neighborhoods.

Recently, some interesting problem variants of the TRP/MLP have been proposed. A direct generalization of the TRP that includes identical vehicles ([98]), is the K -traveling repairman problem. In this routing problem a set of customers should be serviced by a fleet of K vehicles with the aim of collecting a profit, which is a monotonically decreasing function of the arrival time of the vehicle at the customer. The Cumulative VRP (CumVRP) generalizes the TRP, by considering a demand associated to each node. The CumVRP can be further regarded as a special case of the weighted vehicle routing problem proposed by Zhang et al. [271]. Later on, the presence of multiple depots was included in the model by Zhang et al. [272]. Kara et al. [137] considered a flow-based formulation of the Cumulative Capacitated VRP (CCVRP), where the objective function to be minimized is not the sum of arrival times, but rather the sum of arrival times multiplied by the demand of the node. Ngueveu et al. presented in [188] a memetic heuristic aimed at visiting a set of customers with a homogenous capacitated vehicle fleet. The proposed metaheuristic seeks appropriate upper bounds while the lower bounds are obtained by exploiting the properties of the problem. In [210], Ribeiro and Laporte presented an adaptive large neighborhood metaheuristic, applying different repair and destroy procedures adopted from the literature. In a recent paper, Sze et al. [245] presented a hybrid metaheuristic algorithm for the CCVRP, in which a two-stage adaptive variable neighborhood search algorithm that incorporates large neighborhood search

as a diversification strategy is designed. This study also considers the min-max objective, adapting appropriately the proposed algorithm.

A recent interesting variant of the problem is the traveling repairman problem with profits (TRPP) introduced by Dewilde et al. [87]. The problem belongs to the class of vehicle routing problems with profits, a flourishing literature stream that has attracted the attention of the operations research community in the last ten years. It is worthwhile remarking that some discrepancies in problem definition and taxonomy might be encountered in the literature. Generally speaking, a profit is associated with each customer and only a subset of customers to be served is chosen; the decision is made on the basis of an objective function that might include the collected profit and/or the travel cost. Depending on the type of the objective function chosen, different names can be found in the literature. When the objective function is the maximization of the total profit collected, and a maximum duration constraint or a capacity constraint (or both) are imposed on vehicle routes, the problems are usually referred to as problem with profits. If the objective function is the minimization of the total traveling cost and a constraint on the minimum amount of profit to be collected is imposed, we categorize the problem as prize-collecting. If the objective function is the maximization of the difference between the total collected profit and the traveling cost, then the routing problem is cast as a profitable problem. It is worth noting that, notwithstanding the objective function in the TRPP is the collected profit minus the total arrival time, the problem is commonly referred in the literature as a problem with profits.

In this problem, a revenue is obtained the first time a location is visited and the visit of each location is optional, for a single vehicle problem. More formally, let us consider a set of customers $\bar{V} = \{1, \dots, n\}$ indexed by i, j and a pre-specified central depot (denoted with 0) where a homogenous fleet of K vehicles is located. Let $V = \bar{V} \cup \{0\}$ and $G = (V, E)$ ($E = \{(ij) \subseteq V \times V\}$) be a complete undirected graph. We denote by p_i the revenue associated to each customer $i \in \bar{V}$ and we assume that the profit collected by visiting a customer decreases with the arrival time, expressed as a function of the edge travel times, denoted by d (in line with [87]). In particular, the profit assumes the maximum value (the revenue value p_i) when the arrival time at node i (t_i) is zero. The objective is to identify a set of feasible vehicle

paths (each starting at the depot and visiting a different subset of nodes) that maximizes the collected profit expressed as a decreasing function of the arrival times. In particular, the total profit associated to a given solution (set of vehicle paths) is

$$\sum_{k=1}^K \sum_{i \in V^k} (p_i - t_i) \quad (2.1)$$

where we denote by V^k , the set of nodes visited by the vehicle k . For a given solution, composed by K disjoint paths $\pi^k = [0, l_{[1]}, \dots, l_{[L^k]}]$, $k = 1, \dots, K$ defined by an ordered set of links indexed by l (here L^k denotes the length of path of the vehicle k), the total profit can be rewritten, for each vehicle, k , as

$$\sum_{i \in V^k} p_i - t_i = \sum_{q=1}^{L^k} p[q] - (L^k - q + 1)(d_{[q]}) \quad (2.2)$$

where q is the position of the link $\ell \in \pi^k$ assuming value in the interval $[1, \dots, L^k]$. The notation $[\cdot]$ denotes the position of the link in the path. In fact, equation (2.8) sums the difference between the profit of the q^{th} visited node after the depot and the number of times the edge preceding that node is counted in the total latency, including the total latency of node q and all the subsequent nodes visited after.

This highlights the similarities with one of the variants of the time-dependent routing problem considered by Picard and Queyranne [199]. In that problem, the travel cost function between two cities not only depends on the distance, but also on the position of the arc in the tour. It is worthwhile noticing that this is also the distinguishing characteristic of the TRPP and its counterpart without profit (TRP), since the arrival time, or latency, at a given node, depends on the position of the node in the path.

In what follows, we present a fictitious small example shown in Figure 2.1, including five potential nodes to be visited by a fleet of two homogeneous vehicles. Over each edge is reported the corresponding travel time, whereas the value under each node represents the revenue collected while visiting the node. The optimal paths (obtained by solving the model presented in [87]) are shown in Figure 2.2.

The total profit gained by visiting nodes 2, 3, 4, and 5 is 20, calculated as the sum of the profit over the path $0 - 2 - 3$ ($19 - (2 \times (3)) + 1 \times (3)) = 10$), and over the path $0 - 4 - 5$ ($20 - (2 \times (3.5)) + 1 \times (3)) = 10$).

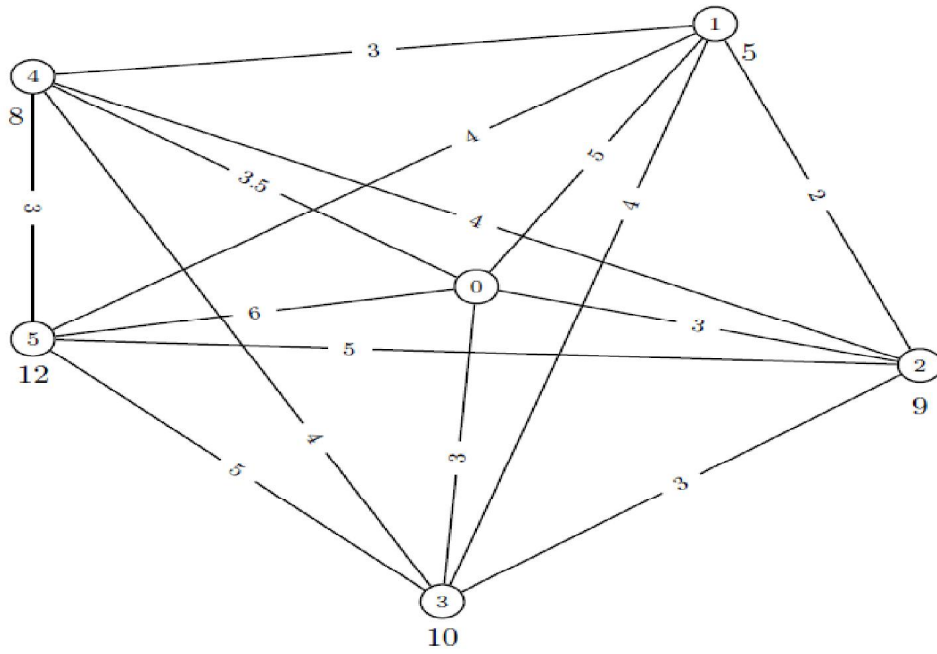


Figure 2.1: Toy example

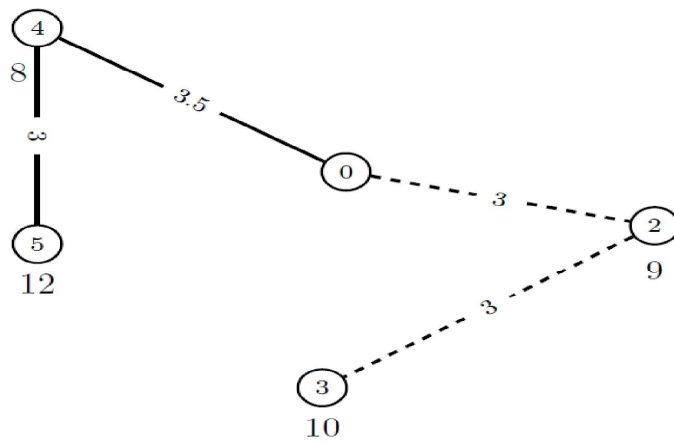


Figure 2.2: Optimal paths

In the extant literature, it is usually assumed that the repairman travels at a constant speed, which is not a realistic assumption in many cases. The speed and, consequently, the travel time might vary depending on different factors such as, but not limited to, the fluctuations in weather condition, traffic congestion, and vehicle breakdowns. The issue of incorporating the uncertainty becomes particularly important in customer-centric vehicle routing problems (and hence both in TRPs and TRPPs), where the attention is directed towards customers related measures. In disaster relief operations and humanitarian logistics operations, for instance, the arrival time has a crucial role in saving people’s lives and might significantly contribute to decrease the rate of mortality.

To the best of our knowledge, notwithstanding there is a vast literature on the incorporation of stochasticity for vehicle routing problems ([40, 56, 108]), a few contributions exist on profit-based routing latency problems at the presence of uncertainty. In [256], the a priori traveling repairman problem is defined as a stochastic problem with recourse, where the tour should minimize the second-stage cost defined as the sum of expected latencies arising from the revealed set of uncertain vertices to be visited.

In this Chapter, we present our contributions in the stochastic routing literature focusing on the K -TRPP. We propose a formulation explicitly addressing uncertainty in the arrival times. We study the problem under the assumption that the probability distribution of the random travel times is not completely specified and that only the mean and the variance of the travel times is known. We propose, in particular, a risk-averse stochastic version of the TRPP under uncertain travel times modeled through a mean-risk approach. This broadens the applicability of our models, that can be applied to tackle the solution of different real-life scenarios, hampered by the lack of a universally acceptable distribution for modeling travel and service times. To the best of our knowledge, no existing study has yet investigated profit-based routing latency problems at the presence of uncertainty.

The contribution of the Chapter is twofold. Firstly, we extend the mathematical model proposed in [87] to address the problem in a stochastic setting, where the uncertainty affects the travel times. By adopting a risk-averse approach, we present a mean-risk model that can solve small to medium size instances to optimality. We present a validation of the model through a

Monte Carlo simulation analysis. Moreover, a heuristic approach, easy to implement and very effective, is also proposed. Secondly, a more efficient formulation of the problem is presented and the first heuristic extended and enhanced by introducing a move evaluation procedure which speeds up the computation. Moreover, in order to assess the quality of the solutions obtained by the heuristic method, we introduce three different upper bounding procedures.

2.2 The risk-averse K -TRPP under uncertain travel times

In this section, we address the problem under a risk-averse perspective. In this case, the decision-maker is willing to trade-off some profit against more stable solutions in terms of variability. There are different ways to capture and model risk. The expected utility theory ([262]), that has been prevalently used in economics, uses concave utility functions. In the late nineties, an axiomatic approach to risk has been proposed ([147, 212]), introducing the theory of coherent risk measures. Here, we account for risk by controlling the standard deviation. This is an intuitive risk measure that can be widely communicated by the experts and easily computed, since it can be applied whenever the first and the second moments of the random travel times are known, regardless the specific distribution function we consider. This assumption is reasonable since it is unlikely that the full distribution of travel times, if ever available, is known. In order to represent different risk attitudes, from the risk-neutral to the more risk-averse one, we adopt a mean-risk approach, striking the balance between two conflicting objectives, i.e. the expected profit maximization and the risk minimization, following the classical mean-risk framework proposed by Markowitz in [163], in the portfolio optimization theory.

We assume that the travel time d_l of each edge l of the network is represented by a random variable with mean $\mathbb{E}(\tilde{d}_l)$ and variance $VAR(\tilde{d}_l)$. For ease of exposition, we consider travel times are independent, which might be not the case, especially in emergency situations. We should mention that, even though the following derivation is presented for the uncorrelated case, the approach proposed is general and can be applied also when the travel

times are dependent (see [38] for more details). When the travel times are considered random, the arrival time of each vehicle at generic node i is itself a random variable (denoted with \tilde{t}_i).

A risk measure can be incorporated in our problem, by combining the mean travel time with some measure of the dispersion. The standard deviation is a very intuitive measure of variability, which is also related to the value-at-risk objective [41, 42, 43] and that can be used whenever the first and the second moments of the distribution function of the travel times are known [40].

The mean-risk function associated to a set of routes π^k , $k = 1, \dots, K$ under uncertain travel times is then

$$\lambda \mathbb{E} \left[\sum_{k=1}^K \sum_{i \in V^k} (p_i - \tilde{t}_i) \right] + (1 - \lambda) \sqrt{\text{VAR} \left[\sum_{k=1}^K \sum_{i \in V^k} (p_i - \tilde{t}_i) \right]}. \quad (2.3)$$

Since the decision-maker is risk-averse, the problem does not merely entail the maximization of the expected profit, but it must also consider the travel time variability. The weighting parameter $\lambda \in [0, 1)$ reflects different risk-averse preferences of the decision maker and plays the role of the trade-off weight in mean-risk models [152]. By decreasing its value more weight is put on the non-linear part of the objective function, reflecting a risk-averse behavior of the decision maker.

We note that even though the arrival times are dependent, it turns out that the total variance can be computed independently for each vehicle as the paths cover disjoint subsets of nodes.

In particular, the arrival time at each node is the sum of the travel times associated to the links $l \in \pi_i^k$ i.e. belonging to the subpath connecting the depot to the node i .

Hence,

$$\mathbb{E}(\tilde{t}_i) = \mathbb{E} \left[\sum_{l \in \pi_i^k} \tilde{d}_l \right] = \sum_{l \in \pi_i^k} \mathbb{E}(\tilde{d}_l) \quad (2.4)$$

Under the assumption that there is not any common edge traversed by two

different vehicles, we have

$$\begin{aligned} \text{VAR} \left[\sum_{k=1}^K \sum_{i \in V^k} (p_i - \tilde{t}_i) \right] &= \sum_{k=1}^K \text{VAR} \left[\sum_{i \in V^k} \tilde{t}_i \right] = \\ &= \sum_{k=1}^K \left(\sum_{i \in V^k} \text{VAR}(\tilde{t}_i) + \sum_{i \in V^k} \sum_{\substack{j \in V^k \\ i \neq j}} \text{COV}(\tilde{t}_i, \tilde{t}_j) \right). \end{aligned} \quad (2.5)$$

Here,

$$\text{VAR}(\tilde{t}_i) = \text{VAR} \left(\sum_{l \in \pi_i^k} \tilde{d}_l \right) = \sum_{l \in \pi_i^k} \text{VAR}(\tilde{d}_l) \quad (2.6)$$

and the covariance between \tilde{t}_i and \tilde{t}_j can be evaluated as follows:

$$\begin{aligned} \text{COV}(\tilde{t}_i, \tilde{t}_j) &= \mathbb{E}[\tilde{t}_i \tilde{t}_j] - \mathbb{E}[\tilde{t}_i] \mathbb{E}[\tilde{t}_j] = \mathbb{E} \left[\sum_{l \in \pi_i^k} \tilde{d}_l \times \sum_{l' \in \pi_j^k} \tilde{d}_{l'} \right] - \mathbb{E} \left[\sum_{l \in \pi_i^k} \tilde{d}_l \right] \mathbb{E} \left[\sum_{l' \in \pi_j^k} \tilde{d}_{l'} \right] \\ &= \mathbb{E} \left[\sum_{l \in \pi_i^k \cap \pi_j^k} \tilde{d}_l^2 + \sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} \tilde{d}_l \tilde{d}_{l'} \right] - \left[\sum_{l \in \pi_i^k} \mathbb{E}(\tilde{d}_l) \right] \left[\sum_{l' \in \pi_j^k} \mathbb{E}(\tilde{d}_{l'}) \right] = \mathbb{E} \left[\sum_{l \in \pi_i^k \cap \pi_j^k} \tilde{d}_l^2 \right] + \\ &+ \mathbb{E} \left[\sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} \tilde{d}_l \tilde{d}_{l'} \right] - \sum_{l \in \pi_i^k \cap \pi_j^k} \mathbb{E}^2(\tilde{d}_l) - \sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} \mathbb{E}(\tilde{d}_l) \mathbb{E}(\tilde{d}_{l'}) = \\ &= \sum_{l \in \pi_i^k \cap \pi_j^k} \mathbb{E}(\tilde{d}_l^2) - \sum_{l \in \pi_i^k \cap \pi_j^k} \mathbb{E}^2(\tilde{d}_l) + \sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} \mathbb{E}(\tilde{d}_l \tilde{d}_{l'}) - \sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} \mathbb{E}(\tilde{d}_l) \mathbb{E}(\tilde{d}_{l'}) = \\ &= \sum_{l \in \pi_i^k \cap \pi_j^k} [\mathbb{E}(\tilde{d}_l^2) - \mathbb{E}^2(\tilde{d}_l)] + \sum_{l \in \pi_i^k, \notin \pi_j^k} \sum_{l' \in \pi_j^k, \notin \pi_i^k} [\mathbb{E}(\tilde{d}_l \tilde{d}_{l'}) - \mathbb{E}(\tilde{d}_l) \mathbb{E}(\tilde{d}_{l'})] = \sum_{l \in \pi_i^k \cap \pi_j^k} \text{VAR}(\tilde{d}_l) \end{aligned} \quad (2.7)$$

since the second term is the covariance between the links l and l' which is zero, for hypothesis.

Proposition 2.2.1. Let q be the position of the link $\ell \in \pi^k$ assuming value in the interval $[1, \dots, L^k]$. For each vehicle, $k = 1, \dots, K$

$$\text{VAR} \left[\sum_{i \in V^k} \tilde{t}_i \right] = \sum_{q=1}^{L^k} (L^k - q + 1)^2 \text{VAR}(\tilde{d}_{[q]}) \quad (2.8)$$

Proof.

$$\begin{aligned} \text{VAR} \left[\sum_{i \in V^k} \tilde{t}_i \right] &= \sum_{i \in V^k} \text{VAR}(\tilde{t}_i) + \sum_{i \in V^k} \sum_{j \in V^k, i \neq j} \text{COV}(\tilde{t}_i, \tilde{t}_j) = \\ &= \sum_{i \in V^k} \sum_{l \in \pi_i^k} \text{VAR}(\tilde{d}_l) + \sum_{i \in V^k} \sum_{j \in V^k, i \neq j} \sum_{l \in \pi_i^k \cap \pi_j^k} \text{VAR}(\tilde{d}_l) \end{aligned} \quad (2.9)$$

In the first term, we count $L^k - q + 1$ times the variance $VAR(\tilde{d}_{[q]})$ for each $q = 1, \dots, L^k$.

As far as the second term is concerned, we observe that we count $VAR(\tilde{d}_l)$ a number of times equal to

$$2 \binom{L^k - q + 1}{2} = \frac{(L^k - q + 1)!}{(L^k - q + 1 - 2)!2!} \quad (2.10)$$

since it is considered for each non-ordered pair of nodes (ij) . But

$$\frac{(L^k - q + 1)!}{(L^k - q + 1 - 2)!2!} = \frac{(L^k - q + 1)(L^k - q + 1 - 1)(L^k - q + 1 - 2)!}{(L^k - q + 1 - 2)!2!} = \frac{(L^k - q + 1)(L^k - q)}{2} \quad (2.11)$$

Then

$$\begin{aligned} VAR \left[\sum_{i \in V^k} \tilde{t}_i \right] &= \sum_{q=1}^{L^k} VAR(\tilde{d}_{[q]}) \left[(L^k - q + 1) + 2 \frac{(L^k - q + 1)(L^k - q)}{2} \right] = \\ &= \sum_{q=1}^{L^k} VAR(\tilde{d}_{[q]}) [(L^k - q + 1) + (L^k - q + 1)(L^k - q)] = \\ &= \sum_{q=1}^{L^k} VAR(\tilde{d}_{[q]}) [L^k - q + 1](1 + (L^k - q))] = \sum_{q=1}^{L^k} (L^k - q + 1)^2 VAR(\tilde{d}_l). \end{aligned} \quad (2.12)$$

□

2.3 Mathematical formulation with position dependent variables

As highlighted in the previous Section, the position of each edge in the path should be considered for a correct definition of the total latency. Dewilde et al., in [87], presented a mathematical formulation based on position-dependent variables for the deterministic TRPP with a single vehicle. The integer linear programming formulation proposed requires as an input parameter the number of visited customers, or equivalently the path length. In order to obtain the solution, all the possible path lengths should be considered and the model should be separately solved many times. Hence, this solution framework can be considerably time-consuming for real applications. In what follows, we will present our mean-risk model for the TRPP, as an extension of the Dewilde's model to the multi-vehicle case. The notation used in the model is reported here for the sake of clarity.

Table 2.1: Table of notation

Sets and input parameters:	
$\bar{V} = \{1, \dots, n\}$	set of nodes by i, j
V	$\bar{V} \cup \{0\}$
$\{1, \dots, K\}$	set of paths (vehicles) indexed by k
g	possible path length $g \in \{1, \dots, n - K + 1\}$
$\{1, \dots, g\}$	set of positions of traversed links indexed by q
p_i	profit collected at node i
μ_{ij}	expected travel time of the link (ij)
σ_{ij}^2	travel time variance of the link (ij)
Decision Variables:	
$x_{ijq}^{g,k}$	$= \begin{cases} 1, & \text{if } (ij) \text{ is the } q^{\text{th}} \text{ link over the path of the vehicle } k \text{ with length } g \\ 0, & \text{otherwise} \end{cases}$

Considering the above notation, the mathematical formulation based on position-dependent variables is as follows:

$$\max : \sum_{k=1}^K \sum_{g=1}^{n-K+1} \sum_{q=1}^g \sum_{i \in V} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \lambda [p_j - (g+1-q)\mu_{ij}] x_{ijq}^{g,k} - (1-\lambda) \sqrt{\sum_{k=1}^K \sum_{g=1}^{n-K+1} \sum_{q=1}^g \sum_{i \in \bar{V}} \sum_{\substack{j \in V \\ j \neq i}} (g+1-q)^2 \sigma_{ij}^2 x_{ijq}^{g,k}}, \quad (2.13)$$

$$\sum_{g=1}^{n-K+1} \sum_{j \in V} x_{0j1}^{g,k} = 1, \quad k = 1, \dots, K \quad (2.14)$$

$$\sum_{\substack{i \in \bar{V} \\ i \neq j}} x_{ijq}^{g,k} - \sum_{\substack{i \in V \\ i \neq j}} x_{ji, q+1}^{g,k} = 0, \quad j \in V, g = 2, \dots, n - K + 1 \\ q = 1, \dots, g - 1, k = 1, \dots, K \quad (2.15)$$

$$\sum_{g=q}^{n-K+1} \sum_{i \in \bar{V}} \sum_{\substack{j \in V \\ j \neq i}} x_{jiq}^{g,k} \leq 1, \quad q = 1, \dots, g, \quad k = 1, \dots, K \quad (2.16)$$

$$\sum_{k=1}^K \sum_{g=1}^{n-K+1} \sum_{q=2}^g \sum_{j \in V} x_{0jq}^{g,k} = 0, \quad (2.17)$$

$$\sum_{k=1}^K \sum_{g=1}^{n-K+1} \sum_{q=1}^g \sum_{\substack{i \in \bar{V} \\ j \neq i}} x_{ijq}^{g,k} \leq 1, \quad j \in V \quad (2.18)$$

$$x_{ijq}^{g,k} \in \{0, 1\}, \quad i \in \bar{V}, \quad j \in V, \quad g = 1, \dots, n - K + 1, \\ q = 1, \dots, g, \quad k = 1, \dots, K \quad (2.19)$$

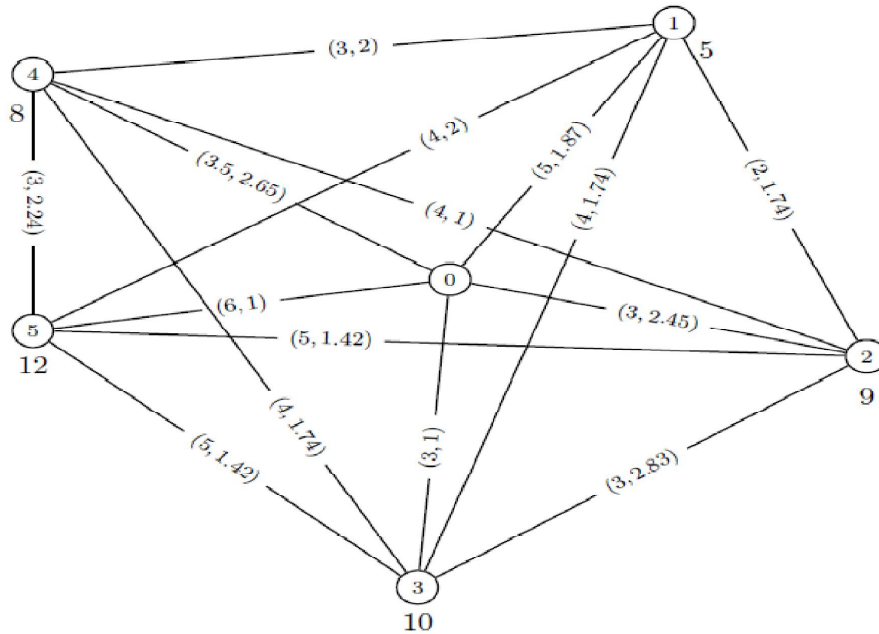


Figure 2.3: Toy example revisited

The objective (2.13) maximizes the total collected revenue expressed in terms of the linear combination of the expected value and risk. These two complementary measures help the decision-maker to investigate promising and competitive solutions through the trade-off between the associated expected value and risk. It also benefits from the fact that a better use of historical information on uncertain parameters is made, finally, resulting in a better and more accurate model. The set of constraints in (2.14) ensures the departure of all vehicles from the depot. The connectivity constraints are reported in (2.15). Constraints (2.16) guarantee that each position over each path is assigned to at most one link. The set of constraints in (2.17) establishes that no link starting from the depot can be used in any other position except the first one, over each feasible path. The set of constraints in (2.18) requires that each affected area is visited at most once over all paths. The set of constraints in (2.19) expresses the binary nature of variables.

The optimal solution of this model, applied to the toy example 2.1 and considering the case in which the decision-maker is concerned about risk ($\lambda = 0.1$), is shown in Figure 2.4. Figure 2.3 shows the network, where the ordered pair over each edge represents the expected travel time over the edge and its standard deviation, and the value under each node represents

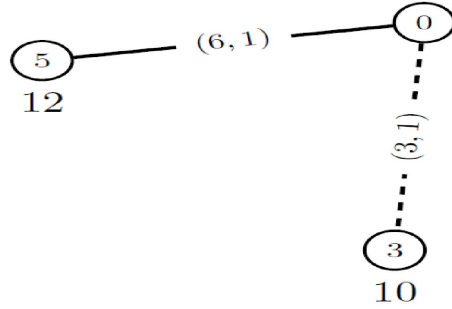


Figure 2.4: Optimal paths in the risk-averse model

the revenue collected while visiting the node.

The total expected profit of visiting nodes 3 and 5 is $12 - (6) + 10 - (3) = 13$ and the total standard deviation is $\sqrt{1 + 1} = 1.41$, which is much lower than the standard deviation of the deterministic risk-neutral solution (calculated as

$$\sqrt{(2^2 \times (6) + 1^2 \times (8)) + (2^2 \times (7.02) + 1^2 \times (5.01))} = 8.06$$

It is important to evaluate the reliability of the optimal solution resulted from the proposed approach. To this end, we run a Monte Carlo simulation, as one of the most common uncertainty assessment tools, to investigate the efficiency of the proposed risk-averse model compared to the risk-neutral one.

In particular, we have run a Monte Carlo simulation, including 50,000 different scenarios generated using the lognormal distribution, to represent different possible travel times values over each edge (ij) . The total expected profit under each scenario is expressed as the total difference between the collected revenue at a node and the realized arrival time to reach that node. Figures 2.5 and 2.6 report the frequency histograms of the simulated profits for the risk-neutral and the risk-averse solutions, respectively.

The simulation results show that the solution of the risk-neutral model is highly unstable with high variations in the total profit confirmed by the long tail in Figure 2.5.

We should note that the risk-neutral model ignores the variance completely and it is reasonable to expect high profit spread. Interestingly, the negative profit values (on the left of the vertical line) in Figure 2.5 underline the

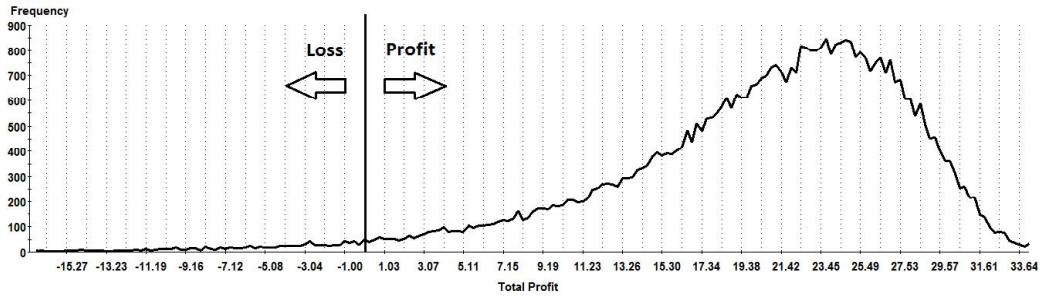


Figure 2.5: Frequency histograms: Risk-neutral model

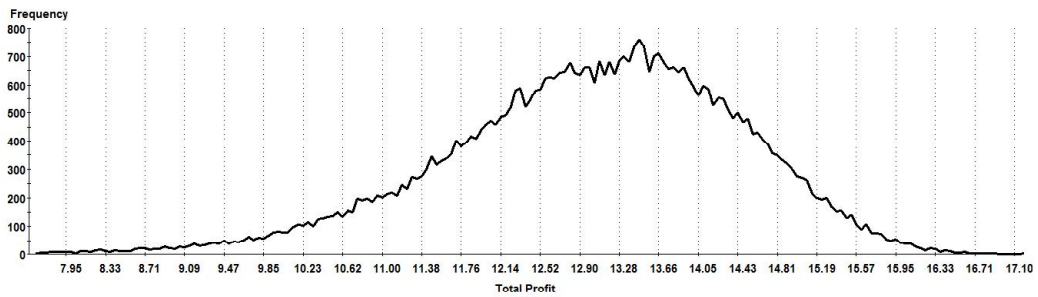


Figure 2.6: Frequency histograms: Risk-averse model

probability of experiencing loss. This occurrence is avoided by adopting a more risk-averse approach (see Figure 2.6).

	<i>Mean</i>	<i>Median</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
Risk-averse Model	12.99	13.11	17.33	2.34	1.42
Risk-neutral Model	20.01	21.58	35.70	-104.13	8.07

Table 2.2: Simulation results

Table 2.2 reports the arithmetic mean, the median, the max and min values, as well as the standard deviation (SD) of the simulation results. In general, we should expect that, in terms of the total expected profit, the risk-neutral case provides better results than the risk-averse model, (supported by the higher *Mean*, *Median*, and the *Max* values reported for the risk-neutral model). The long range $Max - Min = 139.8$ as well as the high standard deviation ($SD = 8.07$) show the unstable behavior of the risk-neutral model, compared with the risk-averse one with range of 14.99 and SD of 1.42.

2.3.1 The iterated greedy heuristic

In the following we present a heuristic approach to efficiently tackle the solution of the problem (2.13)–(2.19).

The heuristic approach relies on an iterated greedy method that alternates between constructive and destructive phases. The greedy constructive method builds a solution R^{curr} involving a subset S of the node set \bar{V} . Visited nodes in R^{curr} are selected from $U \setminus TBL$ where U and TBL are, respectively, the set of unvisited and the set of temporary forbidden nodes. Additionally, an adaptive local search phase is applied to R^{curr} and a new solution \hat{R} is obtained. Then, during the destructive phase, a percentage of the visited nodes in S ($10\%n$) are removed randomly from the current solution and put in temporary blacklist TBL . The set of visited nodes S is updated and the constructive phase is applied again to rebuild the solution. In fact, at each iteration, the set of visited nodes S in R^{curr} is partially destroyed over the **Destroy procedure** which is then repaired and updated

over the **Construction procedure** in the next iteration. The method iterates this pattern until a given number of iteration It_{max} is reached. The best solution R^* and the best value of the objective function $\mathcal{O}\mathcal{F}(R^*)$ are stored and returned at the end of the algorithm.

The pseudocode of the iterated greedy heuristic is presented in Algorithm 1.

Algorithm 1: Scheme of the heuristic

```

1 Initialization:
    $S \leftarrow \emptyset, U \leftarrow \bar{V}, TBL \leftarrow \emptyset; R^*, \hat{R}, R^{curr} \leftarrow null; \mathcal{O}\mathcal{F}(\cdot) \leftarrow null$ 
2 for ( $It = 1$  to  $It_{max}$ ) do
3    $R^{curr} \leftarrow null$ 
4    $(S, R^{curr}) \leftarrow \mathbf{Construction\ procedure}(U \setminus TBL)$ 
5    $\hat{R} \leftarrow \mathbf{Adaptive\ local\ search}(R^{curr})$ 
6   if  $\mathcal{O}\mathcal{F}(\hat{R}) > \mathcal{O}\mathcal{F}(R^*)$  then
7      $R^* \leftarrow \hat{R}$ 
8      $\mathcal{O}\mathcal{F}(R^*) \leftarrow \mathcal{O}\mathcal{F}(\hat{R})$ 
9   end
10  Destroy procedure( $S$ )
11   $TBL \leftarrow \emptyset$ 
12   $t \leftarrow 1$ 
13  while ( $t \leq 10\%n$ ) do
14    Select randomly a node  $i^* \in S$ 
15     $TBL \leftarrow TBL \cup \{i^*\}$ 
16     $t++$ 
17  end
18   $S \leftarrow S \setminus TBL$ 
19 end
20 return  $R^*, \mathcal{O}\mathcal{F}(R^*)$ 

```

In the following, we discuss the main steps of the proposed heuristic in more detail.

During the constructive phase, the **Construction Procedure** is called to build an initial solution. Nodes $i \in S$ are sorted in ascending order with respect to the following criterion $\frac{\lambda\mu_{0i} + (1-\lambda)\sqrt{\sigma_{0i}^2}}{\lambda p_i}$ which accounts for both the distance from the depot and the revenue. One seed node is then inserted into each vehicle route, following a greedy ordering criterion. Then, each unvisited node is inserted one at a time, in the best position in the best route (on the basis of the increment of the profit value).

There are many different alternatives that can be considered for the local search. Among them, we opted for a self-adaptive mechanism implemented in the **Adaptive local search** procedure. The idea is to randomly select one neighborhood in the set of neighborhoods, accordingly to the associated selection probability. The set of neighborhoods $\{1, \dots, \mathcal{N}\}$ is the following;

1. Intra-route exchange operation: exchanges the position of a pair of non-adjacent nodes over a path.
2. 2-opt operation: deletes two non-adjacent edges along the path and add two other edges such that the direction of middle unchanged edges in the new path are reversed.
3. Or-opt operation: deletes any triple of non-adjacent edges within the path and reconnects them by adding three new edges such that the order of middle unchanged edges over the path is preserved.
4. Inter-route exchange operation: exchanges a pair of nodes belonging to two different paths.
5. Delete-insert operation: deletes a node from a path and adds it to the other one.

Within the *warm-up* period, a roulette wheel mechanism is applied to determine which neighborhood to explore. The selection probability is the same for all the neighborhoods and set to $\frac{1}{\mathcal{N}}$. After the *warm-up* period, the neighborhood selection is performed using the self-adaptive mechanism considering the success and the failure of the neighborhoods in the past. In particular, the selection probability of the neighborhoods is updated every N_a iterations using the following formula:

$$prob_{\nu} = \frac{\bar{s}_{\nu}}{\sum_{\nu=1}^{\mathcal{N}} \bar{s}_{\nu}} \text{ where } \bar{s}_{\nu} = \frac{s_{\nu}}{s_{\nu} + f_{\nu}} + \epsilon.$$

Here, s_{ν} and f_{ν} count the number of times the neighborhood ν was successful or unsuccessful, respectively, and ϵ is a small value added to provide all the neighborhoods (even the unsuccessful ones) with a chance. When the local search is not able to improve the solution within a given number of iterations, it is stopped and the destruction phase is applied again. The proposed **Destroy procedure** generates a subset of randomly selected nodes in the current solution to be banned from being present in the solution of next

iteration. This kind of diversification mechanism allows us to extensively search the solution space in order to find near-optimal solutions. A scheme of the local search heuristic is shown in Algorithm 2.

Algorithm 2: Adaptive local search

```

1 Input: The current solution  $R^{curr}$ , Maximum number of iterations
   allowed without improvement  $IMP_{max}$ 
2 Initialization:  $prob_\nu \leftarrow \frac{1}{N}, \forall \nu = 1, \dots, V; \hat{R} \leftarrow null, imp \leftarrow 0$ 
3 for  $t = 1$  to  $T$  do
4   | if ( $t > Warm\text{-}up\ Period$  and  $t \% N_a = 0$ ) then
5   |   | Update-Selection-Probability
6   | end
7   | Select a Neighborhood  $\hat{\nu}$  according to the probabilities
   |  $prob_\nu, \forall \nu = 1, \dots, N$ 
8   | Explore the Neighborhood  $\hat{\nu}$  around  $R^{curr}$  and return a solution  $\hat{R}$ 
9   | if ( $\mathcal{OF}(R^{curr}) \geq \mathcal{OF}(\hat{R})$ ) then
10  |   |  $imp++$ 
11  | end
12  | if ( $imp \% IMP_{max} == 0$ ) then
13  |   | return  $\hat{R}$ 
14  | end
15 end
16 return  $\hat{R}$ 

```

2.3.2 Computational results

We report the performance of the proposed heuristic approach on different test cases. The code was implemented in C++ and the experiments performed on an Intel® Core™ i7 2.90 GHz, with 8.0 GB of RAM memory, running under Windows operating system. In order to have an idea of the quality of the solution obtained, the corresponding mathematical model was solved using SCIP. We have tested the heuristic on two sets of instances, including the P -instances and the E -instances used as benchmark in routing problems [194]. The number of customers (vehicles) in the data sets vary from 15 – 75 (2 – 15) and 21 – 75 (3 – 14), respectively. The expected travel time over each link (ij) is set to the Euclidean distance between node i and j and its variance is computed as $\lceil (r\mathbb{E}(\tilde{d}_{(ij)}))^2 \rceil$, where r is a random number uniformly distributed in the interval $[0.1, 0.32]$. In order to randomly generate revenue values, we have slightly modified the

Table 2.3: Results for the P-instances

Instance	n	K	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.9$		
			Iterated greedy	SCIP		Iterated greedy	SCIP		Iterated greedy	SCIP	
			$\Delta Time$	ΔGap	Opt	$\Delta Time$	ΔGap	Opt	$\Delta Time$	ΔGap	Opt
Pn16k8	15	5	0.56	0.09	0.00	1.01	0	0.00	1.02	0.12	0.00
Pn19k2	18	2	0.05	0.06	0.00	1.18	0	0.00	1.19	0	0.00
Pn20k2	19	2	0.05	1.13	0.00	0.75	2.61	0.00	1.43	0	0.00
Pn21k2	20	2	0.03	3.65	0.00	0.86	0.76	0.00	1.12	0.71	0.00
Pn22k2	21	2	0.79	3.09	0.00	0.81	0.27	0.00	1.43	2.01	0.00
Pn22k8	21	8	0.35	0.84	0.00	0.46	0	0.00	0.56	0.33	0.00
Pn23k8	22	8	0.74	1.25	0.00	0.92	0.33	0.00	1.16	0.15	0.00
Pn40k5	39	5	0.03	0.39	7.87	0.15	0.53	0.00	0.84	1.17	0.00
Pn44k5	44	5	0.06	1.27	5.35	0.06	1.28	0.56	0.59	1.17	0.00
Pn50k7	49	7	0.07	1.78	10.19	0.07	1.27	0.75	0.81	0.86	0.00
Pn50k8	49	8	0.06	1.04	9.49	0.40	0.93	0.00	1.13	1.32	0.00
Pn50k10	49	10	0.26	1.41	0.00	0.06	0.4	0.14	3.01	0.43	0.00
Pn51k10	50	10	0.06	0.75	2.15	1.91	0.48	0.00	1.66	0.32	0.00
Pn55k7	54	7	0.09	4.06	9.36	0.09	1.87	0.66	0.81	2.77	0.00
Pn55k8	54	8	0.08	2.94	10.03	0.09	2.17	0.53	1.02	1.62	0.00
Pn55k10	54	10	1.15	3.81	0.00	0.97	2.82	0.00	2.18	2	0.00
Pn55k15	54	15	0.02	2.82	7.82	1.32	2.47	0.00	1.25	2.15	0.00
Pn60k10	59	10	0.58	0.92	0.00	0.11	0.72	0.42	2.48	0.8	0.00
Pn65k10	64	10	0.13	0.74	8.02	0.15	1.94	0.45	0.65	1.34	0.00
Pn70k10	69	10	0.18	0.62	7.58	0.20	1.32	0.46	0.95	1.21	0.00
Pn76k4	75	4	0.60	-	∞	0.59	-56.23	58.49	0.58	-107.59	110.28
Pn76k5	75	5	0.51	-	∞	0.54	-	∞	0.54	-	∞
average			0.29	1.63	4.343	0.58	-1.62	2.97	1.20	-4.15	5.25

formula proposed in [87] in order to account for the stochastic case by considering $p_i \sim U((l, u))$ where $l = \min_{i=1}^n \left(\mathbb{E}(\tilde{d}_{0i}) - \alpha \sqrt{VAR(\tilde{d}_{0i})} \right)$ and $u = \frac{n}{2} \max_{i=1}^n \left(\mathbb{E}(\tilde{d}_{0i}) + \beta \sqrt{VAR(\tilde{d}_{0i})} \right)$. Here α, β are deviation parameters within interval $[0, 1]$ which are determined by the decision maker. Obviously, the profit assigned to the depot is zero.

The number of iterations It_{max} and T have been set to 20 and 1200, respectively. The *warm - up period* has been set to 100 iterations and the local search is stopped after experiencing 20 inner iterations without any improvement. Also, the trade-off parameter λ is taken from the set $\{0.1, 0.5, 0.9\}$ and the probabilities are updated every $N_a = 10$ iterations. Tables 2.3 and 2.4 summarize the obtained results.

The performance of the heuristic has been evaluated by comparing the solution with the one obtained by SCIP within a time limit of one hour. Columns 1, 2, and 3 refer to the name of instances, the number of nodes and the number of vehicles, respectively. Then, for each value of λ the speed up

Table 2.4: Results for the E-instances

Instance	n	K	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.9$		
			Iterated greedy	SCIP		Iterated greedy	SCIP		Iterated greedy	SCIP	
			$\Delta Time$	ΔGap	Opt	$\Delta Time$	ΔGap	Opt	$\Delta Time$	ΔGap	Opt
En22k4	21	4	2	4.03	0.00	4.44	0.45	0.00	4.32	2.49	0.00
En23k3	22	3	3.93	6.04	0.00	5.72	2.84	0.00	4.43	3.76	0.00
En30k3	29	3	0.07	-1.6	17.45	0.16	0.57	0.00	2.42	1.49	0.00
En33k4	32	4	0.1	-0.63	14.04	0.12	0.63	0.88	3.4	0.84	0.00
En51k5	50	5	0.38	-0.5	6.61	0.41	0.51	0.56	2.88	0.5	0.00
En76k7	75	7	1.63	-	∞	1.6	0	1.11	1.61	0.51	0.38
En76k8	75	8	1.46	-1.77	5.42	1.47	0	0.47	1.47	0.19	0.31
En76k10	75	10	1.31	-0.13	2.56	1.28	0.34	0.26	3.08	0.33	0.00
En76k14	75	14	0.97	0.21	1.64	0.99	0.07	0.18	1.05	0.31	0.00
average			1.32	0.71	5.96	1.8	0.6	0.38	2.74	1.16	0.08

(in percentage) in the solution time (evaluated as $\Delta Time = \frac{CPU_{heu}}{CPU_{SCIP}} \times 100$) and the percentage gap of the heuristic solution with respect to the solution provided by SCIP (evaluated as $\Delta Gap = \frac{OF_{SCIP} - OF_{heu}}{OF_{SCIP}} \times 100$) are reported together with the percentage optimality gap (Opt) of the SCIP solution.

By looking at the results in Tables 2.3-2.4, we observe that the heuristic provides quite satisfying solutions with the average gap limited to 1.63% for the P -instances. Moreover, we observe that for the most challenging instances of $Pn76k4$ and $Pn76k5$, either SCIP could not provide any feasible solution (verified by Opt of ∞) or provided low quality feasible solutions compared with the heuristic solutions (verified by the negative ΔGap values). In terms of solution time, the average speed-up ($\Delta Time$) is around 0.29% for the most complex case with $\lambda = 0.1$. The decrease in the value of λ , reflects a risk-averse behavior of the decision maker and exacerbates the complexity of the problem, since more weight is put on the non-linear part of the objective function.

In the case of the E -instances, the average gap is limited to 1.16% and SCIP was not able to find any feasible solution for the most challenging instance $En76k7$ with $\lambda = 0.1$, whereas for five out of eight remaining instances the heuristic outperformed SCIP in terms of solution quality (see column 5 of Table 2.4). In what follows we also discuss about other findings which are not directly reported in the Tables. For instance, the proposed heuristic provides satisfying solutions with an average gap (evaluated over all the λ values) limited to 0.83 for the E -instances and to -1.43% for the P -instances, respectively. In addition, the proposed heuristic outperforms

SCIP in terms of solution time, which is less than 20 seconds and on average around 4 seconds, for the P -instances. The average solution time for the E -instances is around 24 seconds with a maximum value of less than 59 seconds for the instance $En76k7$.

The next Section will present a new model for the same problem which is computationally more efficient and, then, we will develop a series of tailored heuristics.

2.4 Mathematical formulation defined on a layered graph

In this section, we present a new mathematical model which is computationally more efficient than the the previous model. The main idea of the new model is borrowed from the multi-layered based structure presented in [194]. In this work, Guillen et al. ([194]) presented an efficient new formulation, defined on a multi-level network, for the deterministic K -traveling repairman problem without profits enhanced by an iterative greedy metaheuristic. Let us define a set of levels $L = \{1, \dots, r, \dots, N + 1\}$, where $N = n - K + 1$ represents an upper bound on the number of demand nodes served in a path. In any level, a copy of nodes is present, amended also with depot in levels from 2 to N . The $N + 1^{th}$ level is composed of a copy of the depot. In this network, each tour is represented by a path that ends in a first level node and starts in a copy of the depot in some level. The level number represent the position of the node in the path, in such a way that nodes at level 1 are the last, nodes at level two are the second to last, at level three are the third to last and so on. Two paths cannot visit the same node, nor in the same level neither in different levels. Using the multilevel network, the following decision variables are defined. Corresponding to each pair of demand node $i \in \bar{V}$ ($\bar{V} = V \setminus \{0\}$) and visiting level $r \in L$, a binary variable x_i^r is defined taking the value 1 iff node i is visited at level r , and 0 otherwise. In a similar way, the binary variable y_{ij}^r is assigned to each edge (ij) such that $i \in V$, $j \in \bar{V}$ and takes the value 1 if edge (ij) is used to link node i in level $r + 1$ to node j in level r , otherwise its value is set to 0. Defining $\bar{L} = L \setminus \{N, N + 1\}$, the mathematical model can be formulated as follows.

$$\max: z = \lambda \left(\sum_{j \in \bar{V}} \sum_{r \in L} (p_j - r \mu_{0j}) y_{0j}^r + \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \sum_{r \in \bar{L}} (p_j - r \mu_{ij}) y_{ij}^r \right) - (1 - \lambda) \sqrt{\sum_{j \in \bar{V}} \sum_{r \in L} r^2 \sigma_{0j}^2 y_{0j}^r + \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \sum_{r \in \bar{L}} r^2 \sigma_{ij}^2 y_{ij}^r}, \quad (2.20)$$

$$\sum_{r \in L} x_i^r \leq 1, \quad i \in \bar{V} \quad (2.21)$$

$$\sum_{i \in \bar{V}} x_i^1 = K, \quad (2.22)$$

$$\sum_{r \in L} \sum_{j \in \bar{V}} y_{0j}^r = K, \quad (2.23)$$

$$\sum_{\substack{j \in \bar{V} \\ j \neq i}} y_{ij}^r = x_i^{r+1}, \quad i \in \bar{V}, r \in \bar{L} \quad (2.24)$$

$$y_{0j}^r + \sum_{\substack{i \in \bar{V} \\ i \neq j}} y_{ij}^r = x_j^r, \quad j \in \bar{V}, r \in \bar{L} \quad (2.25)$$

$$y_{0j}^N = x_j^N, \quad j \in \bar{V} \quad (2.26)$$

$$x_i^r \in \{0, 1\}, \quad i \in \bar{V}, r \in L \quad (2.27)$$

$$y_{ij}^r \geq 0, \quad i \in V, j \in \bar{V}, r \in \bar{L}. \quad (2.28)$$

The objective function (2.20) maximizes the total profit collected by visiting a subset of nodes which is expressed as the difference between the revenue and the arrival time of the visited nodes.

Constraints (2.21) require that each demand node is visited at most in one level. Constraints (2.22) ensure that each vehicle is assigned to exactly one demand node at the end of its tour (level 1). Constraints (2.23) impose the dispatch of exactly K vehicles from the depot. The constraints in (2.24)-(2.26) are the connectivity constraints and show the relation between the binary variables x_i^r and y_{ij}^r . Constraints (2.24) require that any node i visited at the upper level $r + 1$ should be connected to exactly one upcoming visited node (let say j) by traversing edge (ij) at the lower level r . The constraints (2.25) impose that any node j visited at level r should be linked to exactly one recently visited node (let say i) by traversing edge (ij) or linked directly to the depot by traversing edge $(0j)$ at the same level. Constraints (2.26) require that each node visited at the highest level should be the first visited node over the path which is connected to the depot by traversing edge $(0j)$

at the same level. The constraints in (2.27)-(2.28) express the nature of variables.

Upper bounds

In what follows, we present three upper bounds based on those proposed for the CumVRP by Ngueveu et al. in [188]. The idea is to sort the customers in descending profit order and the edges in ascending expected travel time and variance order. Let \bar{p}_i be the i^{th} highest revenue assigned to a node and $\bar{\mu}_e$ and $\bar{\sigma}_e^2$ be the e^{th} lowest travel time and the e^{th} lowest variance among all the edges, respectively.

To derive a first upper bound, we first assume that l customers are directly served from the depot. We notice that the number of visited nodes $l = K, \dots, n$, (since all the K vehicles are used and each vehicle serves at least one customer) is not known a priori and hence, different upper bound values ub_l^1 are separately evaluated. Considering the closest edges incident to the depot (in this case the index used will be e'), the upper bound is calculated as follows.

$$ub_l^1 = \lambda \left[\sum_{i=1}^l \bar{p}_i - \sum_{e'=1}^l \bar{\mu}_{e'} \right] - (1 - \lambda) \sqrt{\sum_{e'=1}^l \bar{\sigma}_{e'}^2} \quad (2.29)$$

The final upper bound is $UB^1 = \max_{l=K}^n ub_l^1$.

In the second upper bound, the position of each edge in the path is considered for the evaluation of the arrival time. In fact, the travel time of the first edge appears in the evaluation of the arrival times of all the nodes serviced in the path. For instance with $K = 1$ and $l = 5$ (five nodes serviced) the arrival time of the edge in the first position ($e = 1$) is counted $l + K - e$ times. When we consider K vehicles, the average number of times each edge is counted in a balanced solution, servicing l customers, is $\lceil \frac{K+l-e-(l \bmod K)}{K} \rceil$, where the term $\frac{K+l-e}{K}$ accounts for the number of times the edge in position e is counted, and the term $\frac{(l \bmod K)}{K}$ is for balancing the vehicle paths. The second upper bound can be computed as

$$ub_l^2 = \lambda \left[\sum_{i=1}^l \bar{p}_i - \sum_{e=1}^l \left\lceil \frac{K+l-e-(l \bmod K)}{K} \right\rceil \bar{\mu}_e \right] - (1 - \lambda) \sqrt{\sum_{e=1}^l \left(\left\lceil \frac{K+l-e-(l \bmod K)}{K} \right\rceil \right)^2 \bar{\sigma}_e^2} \quad (2.30)$$

and $UB^2 = \max_{l=K}^n ub_l^2$.

The third upper bound is computed by differentiating the set of edges incident to the depot (e') from those which are not incident to the depot (e'').

$$\begin{aligned}
ub_l^3 = & \lambda \sum_{i=1}^l \bar{p}_i - \lambda \left(\sum_{e'=1}^K \left\lceil \frac{K+l-e'-(l \bmod K)}{K} \right\rceil \bar{\mu}_{e'} + \sum_{e''=1}^{l-K} \left\lceil \frac{l-e''-(l \bmod K)}{K} \right\rceil \bar{\mu}_{e''} \right) - \\
& (1-\lambda) \sqrt{\sum_{e'=1}^K \left(\left\lceil \frac{K+l-e'-(l \bmod K)}{K} \right\rceil \right)^2 \bar{\sigma}_{e'}^2 + \sum_{e''=1}^{l-K} \left(\left\lceil \frac{l-e''-(l \bmod K)}{K} \right\rceil \right)^2 \bar{\sigma}_{e''}^2}
\end{aligned} \tag{2.31}$$

and $UB^3 = \max_{l=K}^n ub_l^3$. These upper bounds will be used to evaluate the quality of the solutions generated by the heuristic method that we shall present in the next section.

2.4.1 The hybrid reactive greedy randomized adaptive search heuristic

In this section, we present a hybrid heuristic, based on an iterative multi-start algorithm called greedy randomized adaptive search heuristic (GRASP), first introduced by Feo and Bard ([100]). In our scheme, the GRASP is made reactive ([115, 202]), allowing to self-tune some parameters during the iterations of the algorithm.

The Reactive GRASP (RGRASP) basically consists of a loop embedding a construction phase, then hybridized with a local search phase that improves the trial solution provided by the first phase. The general structure of the RGRASP, is reported in Algorithm 3. Let g be a GRASP iteration counter, G an upper bound on the number of iterations of the algorithm and $\Gamma = \{\alpha_1, \dots, \alpha_M\}$ a discrete set of values for the value of the parameter α , which is used inside the construction phase (named $GRASP(\alpha)$) to control the randomness of the GRASP. During the iterations of the algorithm, the probability of picking a given α_m , $m = 1, \dots, M$ value is self-tuned (lines 8–15), making the scheme reactive. In particular, every G_{max} iterations, the probabilities $prob(\alpha_m)$, $m = 1, \dots, M$ associated to the values of $\alpha \in \Gamma$ are updated, to increase the probability associated to the most successful α values.

Algorithm 3: General structure of the RGRASP

```
1 Input:  $G_{max}$ ,  $G$ ,  $\Gamma = \{\alpha_1, \dots, \alpha_M\}$ 
2 Initialization:  $Z_{max} \leftarrow -\infty$ ;  $S_{max} \leftarrow null$ ;  $prob(\alpha_m) \leftarrow \frac{1}{M}$ ,  $\bar{z}_m \leftarrow$ 
    $0$ ,  $z_m \leftarrow 0$ ,  $\gamma_m \leftarrow 0 \ \forall m = 1, \dots, M$ .
3 for ( $g = 1$ ,  $g \leq G$ ,  $g++$ ) do
4   Pick randomly  $\alpha_{m^*} \in \{\alpha_1, \dots, \alpha_M\}$  according to probabilities
    $prob(\alpha_m)$ ,  $m = 1 \dots, M$ 
5    $s(\alpha_{m^*}) \leftarrow \mathbf{GRASP}(\alpha_{m^*})$ ;
6   Apply a local search obtaining a new solution  $s^*$  with value
    $z(s^*)$ 
7    $z_{m^*} \leftarrow (z_{m^*} + z(s^*))$ ,  $\gamma_{m^*} \leftarrow (\gamma_{m^*} + 1)$ 
8   if ( $g \bmod G_{max} == 0$ ) then
9     for ( $m = 1$ ,  $m \leq M$ ,  $m++$ ) do
10      if  $\gamma_m > 0$  then
11         $\bar{z}_m \leftarrow \frac{z_m}{\gamma_m}$ 
12         $prob(\alpha_m) \leftarrow \frac{1/\bar{z}_m}{\sum_{m=1}^M \frac{1}{\bar{z}_m}}$ 
13      end
14    end
15  end
16  if  $z(s^*) > z_{max}$  then
17     $S_{max} \leftarrow s^*$ ,  $z_{max} \leftarrow z(s^*)$ 
18  end
19 end
20 Return:  $z_{max}$ ,  $S_{max}$ 
```

In the construction phase, a feasible solution is built, composed by a set $R = \{\rho_1, \dots, \rho_K\}$ of routes. The pseudo code of the construction phase is shown in Algorithm 4.

At the beginning, all the nodes \bar{V} are candidate to be inserted, and are therefore collected into the set \mathfrak{U} of unvisited nodes and sorted in ascending order with respect to the following criterion $\frac{\lambda\mu_{0i} + (1-\lambda)\sigma_{0i}^2}{\lambda p_i}$ which accounts for both the distance from the depot and the collected profit. Note that the traveling time to reach a node and the associated profit contribute in opposite fashion to the objective function (2.20). One node is then inserted into each vehicle route, following the ordering criterion. Then, for each unvisited node $j \in \mathfrak{U}$, the possibility of being inserted in any position i over any route ρ is evaluated on the basis of its contribution into the objective function, denoted with $\delta(j, i, \rho)$. After the evaluation process, the aggregated contribution $\Delta_j = \sum_{\rho, i} \delta(j, i, \rho)$ as well as the best position i_j^* and the best route ρ_j^* are stored and all the eligible nodes are put into the list of candidate nodes, denoted by CL . The elements in CL are sorted following a decreasing order since, obviously, the nodes with higher aggregated contribution are more promising than the others. A *restricted candidate list (RCL)* is then built by considering $\Delta_{min} = \min_{j \in CL} \Delta_j$, $\Delta_{max} = \max_{j \in CL} \Delta_j$, which are, respectively, the minimum and the maximum value attained by the candidates. Hence, the RCL contains all elements whose values are above some threshold $\hat{\Delta}$ defined as follows: $\hat{\Delta} = \Delta_{max} - \alpha(\Delta_{max} - \Delta_{min})$, where $\alpha \in [0, 1]$. Nodes with the aggregated values greater than $\hat{\Delta}$ are selected and added into the RCL (lines 22–26) of Algorithm 4. Next, a random element is chosen from the RCL and the corresponding node is inserted in the best position in the best route (lines 27–29) of Algorithm 4. Once the insertion is executed, the candidate nodes are re-evaluated and a new RCL is built again. This stage ends when all the nodes have been inserted. Obviously, the parameter α controls the size of RCL , where the value of $\alpha = 0$ corresponds to the deterministic greedy algorithm, in which always the node associated with the best value is chosen. By increasing the value of α , more candidates are added to the RCL , improving the chances of escaping from local optimality. The adaptive component of the heuristic arises from the fact that the greedy algorithm is applied at each iteration, to reflect the changes implied by the selection of

the previous element. Algorithm 4 returns a solution $s(\alpha)$ representing the set of routes built in this phase.

The adaptive local search

The local search phase to be effective must be carefully designed in order to deeply investigate the portion of the solution space under investigation, exploiting the specific problem structure. The problem presents, indeed, a bi-level configuration, where the first level corresponds to a node selection problem, whereas the second one to a cumulative routing problem over the selected nodes. This suggests to adopt a two-level heuristic method that systematically applies two different exploration mechanisms, one for modifying the set of served nodes—since it is not mandatory to serve all of them— and another one to assign and order the nodes amongst different routes.

In the first level, since even a minor modification on the nodes to be selected might deeply affect final results, we have implemented a systematic exploration of 1-node neighborhoods, as a destroy mechanism. More in detail, the number of nodes to be served is iteratively reduced discarding the node with the worst contribution into the objective function, evaluated as the difference between the objective function with and without the node. Whenever a node is deleted, the route is repaired by connecting directly the two ending nodes. The process terminates whenever the number of nodes in the current solution s (denoted with $|s|$) falls below a given number.

An adaptive local search is then applied in the second level, where different moves are used to systematically change the neighborhood of the search. A scheme of the adaptive neighborhood search heuristic, performing T iterations, is shown in Algorithm 5.

Since the objective function of the problem is non-separable and non-additive due to the presence of the standard deviation, we propose efficient methods to evaluate each move.

Intra-route exchange operation

This operator exchanges the position of a pair of nodes within a vehicle route. Every pair is evaluated and the best move is then applied. In order to evaluate the contribution of the move, let i and j be two positions along the route ρ and let $[i]$, $[j]$ indicate the node that occupies position i , j , respectively. Given that within the local search phase no new node is added,

Algorithm 4: GRASP(α): The construction phase

```

1 Input: A value of  $\alpha \in \Gamma$ 
2 Initialization:  $s(\alpha) \leftarrow null$  ;  $\mathfrak{U} \leftarrow \bar{V}$ ;  $CL \leftarrow \emptyset$ ;  $RCL \leftarrow \emptyset$ ;
    $\Delta_j \leftarrow -\infty, \forall j \in \bar{V}, \rho_k \leftarrow \{0\}, k = 1, \dots, K$ 
3 Sort  $\mathfrak{U}$  in ascending order with respect to  $\frac{\lambda\mu_{0i} + (1-\lambda)\sigma_{0i}^2}{\lambda p_i}$ 
4 for ( $\rho \in R$ ) do
5    $c \leftarrow \arg \min_{i \in \mathfrak{U}} (\frac{\lambda\mu_{0i} + (1-\lambda)\sigma_{0i}^2}{\lambda p_i})$ 
6   Insert  $c$  in route  $\rho$  in the best position
7    $\mathfrak{U} \leftarrow \mathfrak{U} \setminus \{c\}$ 
8 end
9 while ( $\mathfrak{U} \neq \emptyset$ ) do
10  for ( $j \in \mathfrak{U}$ ) do
11     $\Delta_j \leftarrow 0$ 
12     $\bar{\Delta}_j \leftarrow -\infty$ 
13    for ( $\rho \in R$ ) do
14      for ( $i \in \{1, 2, \dots, length(\rho)\}$ ) do
15        Evaluate  $\delta(j, i, \rho)$ 
16        if ( $\delta(j, i, \rho) \geq \bar{\Delta}_j$ ) then
17           $\bar{\Delta}_j \leftarrow \delta(j, i, \rho)$ 
18           $i_j^* \leftarrow i, \rho_j^* \leftarrow \rho$ 
19        end
20         $\Delta_j \leftarrow (\Delta_j + \delta(j, i, \rho))$ 
21      end
22    end
23     $CL[j] \leftarrow \Delta_j$ 
24  end
25  Sort  $CL$  in decreasing order based on  $\Delta_j$ 
26   $\Delta_{min} \leftarrow \min_{j \in CL} \Delta_j, \Delta_{max} \leftarrow \max_{j \in CL} \Delta_j$ 
27   $\hat{\Delta} \leftarrow \Delta_{max} - \alpha(\Delta_{max} - \Delta_{min})$ 
28  for ( $r = 1, r \leq CL.size, r++$ ) do
29    if ( $CL[r] \geq \hat{\Delta}$ ) then
30       $RCL \leftarrow RCL \cup CL[r]$ 
31    end
32  end
33  Choose a random element  $\Delta_{\hat{j}}$  in  $RCL$ 
34  Insert node  $\hat{j}$  in position  $i_j^*$  in route  $\rho_j^*$ 
35   $\mathfrak{U} \leftarrow \mathfrak{U} \setminus \{\hat{j}\}$ 
36 end
37  $s(\alpha) \leftarrow R$ 
38 Return  $s(\alpha)$ 

```

Algorithm 5: Scheme of the local search

```
1 Input: The current solution  $s^*$ ,  $T$ 
2 Initialization:  $prob_\nu \leftarrow \frac{1}{\mathfrak{N}}$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ 
3 repeat
4   for  $t = 1$  to  $T$  do
5     if ( $t > \text{warm-up period}$ ) then
6       | Update the selection probabilities  $prob_\nu$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ 
7     end
8     Select a Neighborhood  $\hat{\nu}$  according to the probabilities
       $prob_\nu$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ 
9      $s \leftarrow \text{Local Search}(s^*, \hat{\nu})$ 
10    if ( $z(s) > z(s^*)$ ) then
11      |  $s^* \leftarrow s$ 
12    end
13    if No improvements after a certain number of iterations then
14      | Return  $s^*$ 
15    end
16  end
17  Apply the destroy mechanism and let  $i$  be the discarded node
18  Delete node  $i$  from  $s^*$  and repair the solution  $s^*$ 
19 until  $|s| \geq \lceil 0.75n \rceil$ ;
20 Return  $s^*$ 
```

and the profits collected are the same, we can simplify the difference in the objective function Δz as

$$\Delta z = -\lambda \Delta z_\mu - (1 - \lambda) (\sqrt{z_{\sigma^2}^{old} + \Delta z_{\sigma^2}} - \sqrt{z_{\sigma^2}^{old}}) \quad (2.32)$$

where $z_{\sigma^2}^{old}$ indicates the value of the objective function variance before the move, and Δz_μ and Δz_{σ^2} are, respectively, the differences in terms of expected values and variances which are defined as follows.

$$\Delta z_\mu = \begin{cases} a + b + c + d, & j \neq \text{length}(\rho) \\ a + b + c, & \text{otherwise} \end{cases}$$

and

$$\Delta z_{\sigma^2} = \begin{cases} \acute{a} + \acute{b} + \acute{c} + \acute{d}, & j \neq \text{length}(\rho) \\ \acute{a} + \acute{b} + \acute{c}, & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} a &= (\text{length}(\rho) - i + 1) (\mu_{[i-1][j]} - \mu_{[i-1][i]}), & b &= (\text{length}(\rho) - i) (\mu_{[j][i+1]} - \mu_{[i][i+1]}), \\ c &= (\text{length}(\rho) - j + 1) (\mu_{[j-1][i]} - \mu_{[j-1][j]}), & d &= (\text{length}(\rho) - j) (\mu_{[i][j+1]} - \mu_{[j][j+1]}), \\ \acute{a} &= (\text{length}(\rho) - i + 1)^2 (\sigma_{[i-1][j]}^2 - \sigma_{(i-1)i}^2), & \acute{b} &= (\text{length}(\rho) - i)^2 (\sigma_{[j][i+1]}^2 - \sigma_{i[i+1]}^2), \\ \acute{c} &= (\text{length}(\rho) - j + 1)^2 (\sigma_{[j-1][i]}^2 - \sigma_{[j-1][j]}^2), & \acute{d} &= (\text{length}(\rho) - j)^2 (\sigma_{[i][j+1]}^2 - \sigma_{[j][j+1]}^2). \end{aligned}$$

2-opt neighborhood This moves deletes two non-adjacent edges along the path and adds two other edges. Let consider two non-adjacent nodes $[i]$ and $[j]$; a 2-opt move is performed by deleting edges $([i], [i + 1])$ and $([j], [j + 1])$ and adding two new edges $([i], [j + 1])$ and $([i - 1], [j])$. While the difference in the objective function is evaluated using (2.32), the terms Δz_μ and Δz_σ are expressed as

$$\Delta z_\mu = \begin{cases} a_1 + b_1 + c_1, & j \neq \text{length}(\rho) \\ a_1 + b_1, & \text{otherwise} \end{cases}$$

and

$$\Delta z_\sigma = \begin{cases} \acute{a}_1 + \acute{b}_1 + \acute{c}_1, & j \neq \text{length}(\rho) \\ \acute{a}_1 + \acute{b}_1, & \text{otherwise} \end{cases}$$

where

$$a_1 = (\text{length}(\rho) - i + 1) (\mu_{[i-1][j]} - \mu_{[i-1][i]}), \quad c_1 = (\text{length}(\rho) - j) (\mu_{[i][j+1]} - \mu_{[j][j+1]}),$$

$$\begin{aligned}
b_1 &= \sum_{k=0}^{j-i-1} (\text{length}(\rho) - (i+k)) (\mu_{[j-k][j-k+1]} - \mu_{[i+k][i+k+1]}), \\
\acute{a}_1 &= (\text{length}(\rho) - i + 1)^2 (\sigma_{[i-1][j]}^2 - \sigma_{[i-1][i]}^2), \quad \acute{c}_1 = (\text{length}(\rho) - j)^2 (\sigma_{[i][j+1]}^2 - \sigma_{[j][j+1]}^2), \\
\acute{b}_1 &= \sum_{k=0}^{j-i-1} (\text{length}(\rho) - [i+k])^2 (\sigma_{[j-k][j-k+1]}^2 - \sigma_{[i+k][i+k+1]}^2),
\end{aligned}$$

Or-opt neighborhood

This move deletes any triple of non-adjacent edges within the route and reconnects them by adding three new edges such that the order of middle unchanged edges over the path is preserved. Let $([i], [i+1])$, $([j], [j+1])$, and $([k], [k+1])$ be a triple of edges such that $j-i > 1$, $k-j > 1$. An *or-opt* move is performed by deleting the aforementioned edges and by inserting three new edges $([i], [j+1])$, $([j], [k+1])$, and $([k], [i+1])$. Note that the order of the edges in this case is preserved. The difference terms of Δz_μ and Δz_σ can be calculated as follows.

$$\Delta z_\mu = \begin{cases} a_1 + b_1 + c_1 - d_1 + e_1 + f_1, & j \neq \text{length}(\rho) \\ a_1 + b_1 + c_1 + e_1 + f_1, & \text{otherwise} \end{cases}$$

and

$$\Delta z_\sigma = \begin{cases} \acute{a}_1 + \acute{b}_1 + \acute{c}_1 - \acute{d}_1 + \acute{e}_1 + \acute{f}_1, & j \neq \text{length}(\rho) \\ \acute{a}_1 + \acute{b}_1 + \acute{c}_1 + \acute{e}_1 + \acute{f}_1, & \text{otherwise} \end{cases}$$

where

$$\begin{aligned}
a_1 &= (\text{length}(\rho) - i) (\mu_{[i][j+1]} - \mu_{[i][i+1]}), & b_1 &= (\text{length}(\rho) - (i+k-j)) \mu_{[k][i+1]}, \\
c_1 &= (\text{length}(\rho) - j) \mu_{[j][j+1]}, & d_1 &= (\text{length}(\rho) - k) (\mu_{[j][k+1]} - \mu_{[k][k+1]}), \\
e_1 &= \sum_{s=1}^{k-j-1} (j-i) \mu_{[j+s][j+s+1]}, & f_1 &= \sum_{s=1}^{j-i-1} (k-j) \mu_{[i+s][i+s+1]}, \\
\acute{a}_1 &= (\text{length}(\rho) - i)^2 (\sigma_{[i][j+1]}^2 - \sigma_{[i][i+1]}^2), & \acute{b}_1 &= (\text{length}(\rho) - (i+k-j))^2 \sigma_{[k][i+1]}^2, \\
\acute{c}_1 &= (\text{length}(\rho) - j)^2 \sigma_{[j][j+1]}^2, & \acute{d}_1 &= (\text{length}(\rho) - k)^2 (\sigma_{[j][k+1]}^2 - \sigma_{[k][k+1]}^2), \\
\acute{e}_1 &= \sum_{s=1}^{k-j-1} (i-j) (i+j+2(s-\text{length}(\rho))) \sigma_{[j+s][j+s+1]}^2, & & \\
\acute{f}_1 &= \sum_{s=1}^{j-i-1} (k-j) (j-k+2(\text{length}(\rho) - (i+s))) \sigma_{[i+s][i+s+1]}^2, & &
\end{aligned}$$

Inter-route exchange neighborhood

This move exchanges a pair of nodes belonging to two different paths. Consider two routes of $\rho_1, \rho_2 \in R$ respectively and two nodes $[i]$ in ρ_1 and $[j]$ in ρ_2 . The terms of Δz_μ and Δz_σ are computed as follows.

$$\Delta z_\mu = \begin{cases} a_1 + b_1 + a_2 + b_2, & i \neq \text{length}(\rho_1), j \neq \text{length}(\rho_2) \\ a_1 + a_2 + b_2, & i = \text{length}(\rho_1), j \neq \text{length}(\rho_2) \\ a_1 + b_1 + a_2, & i \neq \text{length}(\rho_1), j = \text{length}(\rho_2) \\ a_1 + a_2, & \text{otherwise} \end{cases}$$

and

$$\Delta z_\sigma = \begin{cases} \acute{a}_1 + \acute{b}_1 + \acute{a}_2 + \acute{b}_2, & i \neq \text{length}(\rho_1), j \neq \text{length}(\rho_2) \\ \acute{a}_1 + \acute{a}_2 + \acute{b}_2, & i = \text{length}(\rho_1), j \neq \text{length}(\rho_2) \\ \acute{a}_1 + \acute{b}_1 + \acute{a}_2, & i \neq \text{length}(\rho_1), j = \text{length}(\rho_2) \\ \acute{a}_1 + \acute{a}_2, & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} a_1 &= (\text{length}(\rho_1) - i + 1) (\mu_{[i-1][j]} - \mu_{[i-1][i]}), & b_1 &= (\text{length}(\rho_1) - i) (\mu_{[j][i+1]} - \mu_{[i][i+1]}), \\ a_2 &= (\text{length}(\rho_2) - j + 1) (\mu_{[j-1][i]} - \mu_{[j-1][j]}), & b_2 &= (\text{length}(\rho_2) - j) (\mu_{[i][j+1]} - \mu_{[j][j+1]}), \\ \acute{a}_1 &= (\text{length}(\rho_1) - i + 1)^2 (\sigma_{[i-1][j]}^2 - \sigma_{[i-1][i]}^2), & \acute{b}_1 &= (\text{length}(\rho_1) - i)^2 (\sigma_{[j][i+1]}^2 - \sigma_{[i][i+1]}^2), \\ \acute{a}_2 &= (\text{length}(\rho_2) - j + 1)^2 (\sigma_{[j-1][i]}^2 - \sigma_{[j-1][j]}^2), & \acute{b}_2 &= (\text{length}(\rho_2) - j)^2 (\sigma_{[i][j+1]}^2 - \sigma_{[j][j+1]}^2), \end{aligned}$$

Delete-Insert neighborhood

This procedure deletes a node from a path and adds it to the other one.

$$\Delta z_\mu = \begin{cases} -a_1 - b_1 + c_1 + a_2 + b_2 + c_2, & i \neq \text{length}(\rho_1) \\ -a_1 - b_1 + a_2 + b_2 + c_2, & \text{otherwise} \end{cases}$$

and

$$\Delta z_\sigma = \begin{cases} \acute{a}_1 + \acute{b}_1 + \acute{c}_1 + \acute{a}_2 + \acute{b}_2 + \acute{c}_2, & i \neq \text{length}(\rho_1) \\ \acute{a}_1 + \acute{b}_1 + \acute{a}_2 + \acute{b}_2 + \acute{c}_2 & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} a_1 &= \sum_{k=0}^{i-2} \mu_{[k][k+1]}, & b_1 &= (\text{length}(\rho_1) - i + 1) \mu_{[i-1][i]}, \\ c_1 &= (\text{length}(\rho_1) - i) (\mu_{[i-1][i+1]} - \mu_{[i][i+1]}), & \acute{b}_1 &= (\text{length}(\rho_1) - i + 1)^2 \sigma_{[i-1][i]}^2, \\ \acute{a}_1 &= \sum_{k=0}^{i-2} (1 - 2(\text{length}(\rho_1) - k)) \sigma_{[k][k+1]}^2, & b_2 &= (\text{length}(\rho_2) - j + 1) (\mu_{[i][j]} - \mu_{[j-1][j]}), \\ \acute{c}_1 &= (\text{length}(\rho_1) - i)^2 (\sigma_{[i-1][i+1]}^2 - \sigma_{[i][i+1]}^2), & \acute{b}_2 &= (\text{length}(\rho_2) - j + 1)^2 (\sigma_{[i][j]}^2 - \sigma_{[j-1][j]}^2), \\ a_2 &= \sum_{k=0}^{j-2} \mu_{[k][k+1]}, & c_2 &= (\text{length}(\rho_2) - j + 2) \mu_{[j-1][i]}, \\ c_2 &= (\text{length}(\rho_2) - j + 2) \mu_{[j-1][i]}, & \acute{a}_2 &= \sum_{k=0}^{i-2} (1 + 2(\text{length}(\rho_2) - k)) \sigma_{[k][k+1]}^2, \\ \acute{a}_2 &= \sum_{k=0}^{i-2} (1 + 2(\text{length}(\rho_2) - k)) \sigma_{[k][k+1]}^2, & \acute{c}_2 &= (\text{length}(\rho_2) - j + 2)^2 \sigma_{[j-1][i]}^2. \end{aligned}$$

We should note that the performance of the proposed heuristic has been compared with another heuristic named as the hybrid iterated local search (HILS, for short), which is, in fact, our earlier contribution presented in the conference paper [54]. A scheme of the HILS is presented in Algorithm 6.

Algorithm 6: Scheme of the HILS

```

1 Initialization:  $PBL \leftarrow \emptyset$ ;  $prob_\nu \leftarrow \frac{1}{\mathfrak{N}}$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ ;  $prob(\alpha_m) \leftarrow$ 
    $\frac{1}{M}$ ,  $\forall m = 1, \dots, M$ ;  $s^* \leftarrow null$ ;  $z_{max} \leftarrow -\infty$ ;  $s_{max} \leftarrow null$ 
2 while  $|PBL| \leq \lceil 0.25n \rceil$  do
3   Pick randomly  $\alpha_{m^*} \in \{\alpha_1, \dots, \alpha_M\}$  according to probabilities
    $prob(\alpha_m)$ ,  $m = 1 \dots, M$ ,  $s(\alpha_{m^*}) \leftarrow \mathbf{GRASP}(\alpha_{m^*}, PBL)$ 
4   Adaptive local search()
5   for  $t = 1$  to  $T$  do
6     if  $(t > \text{Warm-up Period})$  then
7       Update the selection probabilities  $prob_\nu$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ 
8       Select a Neighborhood  $\hat{\nu}$  according to the probabilities
        $prob_\nu$ ,  $\forall \nu = 1, \dots, \mathfrak{N}$ 
9     end
10     $s^* \leftarrow \text{Local Search}(s(\alpha_{m^*}), \hat{\nu})$ 
11    if  $(z(s^*) > z(s(\alpha_{m^*})))$  then
12       $s(\alpha_{m^*}) \leftarrow s^*$ 
13    end
14    if No improvements after a certain number of iterations then
15      Return  $s^*$ 
16    end
17  end
18  if  $z(s(\alpha_{m^*})) > z_{max}$  then
19     $s_{max} \leftarrow s(\alpha_{m^*})$ ;  $z_{max} \leftarrow z(s(\alpha_{m^*}))$ 
20  end
21  Apply the destroy mechanism and let  $i$  be the discarded node
    $PBL \leftarrow PBL \cup \{i\}$ 
22 end
23 Return  $z_{max}$ ,  $s_{max}$ 

```

2.4.2 Computational results

The proposed metaheuristic was coded in C++ and the mathematical model was solved using the open source SCIP library, release 3.2.0. All the experiments have been performed on an Intel® Core™ i7 2.90 GHz, with 8.0 GB of

RAM memory, running under Windows operating system. To evaluate the performance of our algorithm, we assessed the quality of the upper bounds and of the heuristic solutions obtained by the hybrid RGRASP, with respect to the best solution obtained by SCIP within a time limit of 3600 seconds. The number of iterations G and T have been set to 400 and 1200, respectively. The *warm-up period* has been set to 100 iterations and the local search in the second level is interrupted after 20 iterations without any improvement. Also, the trade-off parameter λ is taken from the set $\{0.1, 0.5, 0.9\}$. The α values are selected from the set $\Gamma = \{0, 0.1, \dots, 0.9\}$ and the probabilities are updated every $G_{max} = 10$ iterations. In order to study the effect of the diversification implemented within the reactive mechanism of the GRASP, we also run the algorithm with $T = 1$ for each value of $\alpha \in \Gamma$, thus obtaining ten different solutions. This experiments allow to disentangle the effect of the reactive mechanism of the GRASP. We have also compared the algorithm with the results of HILS.

We have tested the heuristic approach on three set of instances, including the P-instances ([22]), the E-instances ([77]), already tested and the largest CMT-instances ([76, 188]) with a number of nodes and vehicles ranging from 100 – 199 and (7 – 17), respectively.

A first set of experiments has been carried out with the aim of investigating the effect of different move strategies within the local search ([196]). In particular, we have tested the best admissible (BA) strategy, the first best admissible (FBA) and the least best admissible (LBA). In the BA strategy, the best admissible improving solution within the neighborhood, if any, is chosen to update the incumbent solution, otherwise the incumbent is retained. In the FBA, the first improving solution within the neighborhood is selected, if exists, and otherwise the incumbent is retained. The LBA strategy allows the move to the least improving neighbor solution, if there is any improving one, otherwise the search over the neighborhood is terminated without the update of the incumbent. The results of these experiments are shown, for the instance Pn55k7 (similar results have been obtained for all the other test cases), in Table 2.5. The table reports both the *CPU* time and the solution quality (the percentage gap evaluated with respect to the best known feasible solution or, when SCIP is not even able to provide a feasible solution, to the best upper bound value obtained). The results of

Table 2.5: LBA versus BA and FBA for Pn55k7

	RGRASP		
	λ	<i>Gap%</i>	<i>CPU</i>
LBA	0.1	3.48	100.66
	0.5	1.84	104.69
	0.9	3.45	96.55
Avg.		2.92	100.63
BA	0.1	9.5	98.26
	0.5	7.64	100.18
	0.9	7.28	99.27
Avg.		8.14	99.23
FBA	0.1	3.84	105.51
	0.5	4.35	100.85
	0.9	4.07	103.25
Avg.		4.08	103.2

BA are superior, in terms of the computational time, to those provided by the BA and LBA, but the LBA strategy, on the contrary, outperforms the BA and FBA in terms of solution quality. In general, the LBA provides more chance to escape from local optimality, by moving to those solutions with the smallest improvements. On the other hand, the BA and the FBA strategies provide better solutions in the first iterations, but are more likely to be trapped into local-optimal solutions in the long term. To obtain good quality solutions, we have applied the LBA strategy in the computational experiments.

2.4.3 Results for the P-instances

Tables 2.6, 2.7, and 2.8 summarize the results obtained for the P-instances, for different values of λ . Columns 1, 2, and 3 refers to the name of instances, the number of nodes and the number of vehicles, respectively. Columns 4-15, classified in four categories, correspond to the HILS proposed in [54], the RGRASP, and the GRASP. In each category, it is reported the *CPU* time (in seconds) and the percentage gap evaluated with respect to the best known feasible solution or, when SCIP is not even able to provide a feasible solution, to the best upper bound value obtained. The optimality gap of SCIP has been reported in the dedicated column with heading *Opt%*. For the GRASP heuristic, the average, the minimum and the maximum *CPU*

Table 2.6: Results for the P-instances, $\lambda = 0.1$

Instance	n	K	HILS		RGRASP		GRASP						SCIP	
			CPU	Gap%	CPU	Gap%	CPU			Gap%			CPU	Opt%
							Avg	Max	Min	Avg	Max	Min		
Pn16k8	15	5	0.23	10.28	0.75	1.49	0.32	0.37	0.24	1.49	1.49	1.49	3.59	0
Pn19k2	18	2	1.07	6.36	3.78	0.76	1.00	1.18	0.82	1.82	7.71	0.06	360.91	0
Pn20k2	19	2	1.18	4.56	4.64	1.12	1.10	1.19	1.01	0.72	1.66	0.05	471.42	0
Pn21k2	20	2	1.52	9.01	5.38	0.87	1.28	1.46	1.15	5.75	9.6	0.42	822.1	0
Pn22k2	21	2	2.13	4.27	6.55	0.36	1.59	1.92	1.37	1.92	6.57	0.00	41.9	0
Pn22k8	21	8	0.77	4.97	1.58	2.43	0.59	0.62	0.57	1.79	1.79	1.79	14.3	0
Pn23k8	22	8	0.8	2.78	1.93	0.48	0.91	0.97	0.85	0.93	1.16	0.38	8.13	0
Pn40k5	39	5	10.18	2.65	35.08	1.85	7.29	7.72	6.54	1.03	1.59	0.24	3600	7.87
Pn44k5	44	5	16.08	3.79	57.91	0.8	9.13	10.08	7.9	1.09	2.79	0.15	3600	5.35
Pn50k7	49	7	16.36	9.57	90.72	0.96	11.68	13.23	10.37	2.49	3.62	1.29	3600	10.19
Pn50k8	49	8	17.8	9.2	64.27	1.87	10.37	12.94	7.84	2.33	5.51	0.99	3600	9.49
Pn50k10	49	10	17.72	6.13	59.4	1.22	11.31	12.72	9.75	1.58	2.53	0.71	697.51	0
Pn51k10	50	10	16.27	2.35	60.77	0.71	12.06	15.35	10.46	1.03	1.79	0.41	3600	2.15
Pn55k7	54	7	32.49	11.81	100.66	3.48	15.77	18.07	13.67	3.64	5.34	2.56	3600	9.36
Pn55k8	54	8	25.64	9.81	95.28	2.73	16.01	17.8	12.78	2.07	3.32	0.60	3600	10.03
Pn55k10	54	10	24.99	12.9	80.03	2.32	13.36	16.55	9.85	3.02	5.61	1.38	156.49	0
Pn55k15	54	15	17.62	7.73	51.75	1.8	13.28	13.96	12.59	0.99	1.52	0.42	3600	7.82
Pn60k10	59	10	32.12	6.56	127.78	1.12	17.67	20.56	14.07	2.45	4.24	0.93	595.33	0
Pn65k10	64	10	43.74	4.3	168.01	1.09	20.81	23.12	17.82	1.30	2.6	-0.04	3600	8.02
Pn70k10	69	10	57.53	5.66	227.97	1.49	29.11	32.88	25.8	1.63	2.57	0.17	3600	7.58
Pn76k4	75	4	147.96	2.38	485.79	2.29	73.01	77.51	65.75	2.25	2.7	1.97	3600	-
Pn76k5	75	5	138.99	1.78	452.06	1.74	59.01	64.92	51.71	1.79	1.98	1.68	3600	-
Avg			28.33	6.31	99.19	1.50	14.85	16.60	12.86	1.96	3.53	0.80		

and percentage GAP values $Gap\%$, obtained by running the GRASP with all the α values in Γ , have been also reported.

First, we benchmark the performance of the RGRASP algorithm and SCIP on these instances. A first general observation is that the results are influenced by the parameter λ : by decreasing its value more weight is put on the non-linear part of the objective function, reflecting a risk-averse behavior of the decision-maker. This claim is supported by the increase in the solution gap and the CPU times of the heuristic as well as SCIP. The heuristic performs nearly as good as SCIP in very short times. The highest CPU time of the heuristic is around 500 seconds, for the instance Pn76k4 with $\lambda = 0.9$. In general, the heuristic is much faster with a percentage speedup of 91.5%, 82.5%, and 64%, respectively for $\lambda = 0.1, 0.5$, and 0.9 . We observe that the greatest improvement in the CPU time spent is achieved for the instances reflecting a risk-averse behavior of the decision maker (small λ values). The gaps are very small, regardless the value of λ and always below 4. When the number of the nodes increases to 75, the heuristic outperforms SCIP also in terms of the solution obtained (we report in this case a negative gap).

Analyzing the performance of the GRASP with and without the reactive mechanism, reported in columns 8-13, we observe that the best solution obtained by the GRASP (which corresponds to a given fixed value of α) might be better than the one obtained with the reactive mechanism. We should consider that the best solution has been obtained by running the algorithm with all the ten different values of α and therefore, the total time spent is the average time (for instance, 14.85 seconds for $\lambda = 0.1$) multiplied by ten. Hence, a slightly better solution could be obtained by using GRASP with many values of α (considering the best solution over all the runs), but at the cost of a 50% *CPU* time increase, on average. Moreover, the worst solution can deteriorate the quality of the solution considerably, especially for $\lambda = 0.1$ and the deviation can be remarkable. For example for the instance Pn21k2, the worst gap is around 10% and the deviation between the max and the min gap found by the GRASP over ten runs with different values of α is 9.18%. On average the deviation is around 2.73% for $\lambda = 0.1$, and around 1.5% for the other values of λ .

The HILS approach proves to be successful in terms of time efficiency, but the solution quality deteriorates especially for $\lambda = 0.1$ (the average gap in this case is 6.31%). We observe that as λ increases, the average gap of the local search exhibits a downward trend, going from 6.31% to -1.35%. This might suggest to limit the use of the HILS heuristic to risk-neutral problems. As far as the time performance is concerned, we observe that the HILS heuristic is faster than the RGRASP, with an average *CPU* time of less than 30 seconds.

Finally, we have investigated the quality of the upper bounds proposed in Subsection 2.4. In Table 2.9, we have reported the gaps between the best objective values obtained either by SCIP or by the heuristic (*Best objective value*), and the upper bounds UB^i , $i = 1, \dots, 3$, calculated as

$$\Delta^i = \frac{UB^i - \text{Best objective value}}{\text{Best objective value}} \times 100$$

The first upper bound outperforms the third one in 15 out of 66 instances, although on average the third upper bound provides slightly better gaps (2.88% of improvement). The second upper bounds seems to be dominated by the other two. The quality of the upper bound improves as far as the number of nodes increases. For example, considering as a threshold a number

Table 2.7: Results for the P-instances, $\lambda = 0.5$

Instance	n	K	HILS		RGRASP		GRASP						SCIP	
			CPU	Gap%	CPU	Gap%	CPU			Gap%			CPU	Opt%
							Avg	Max	Min	Avg	Max	Min		
Pn16k8	15	5	0.33	3.78	0.86	0.77	0.28	0.34	0.22	1.01	1.01	1.01	1.99	0
Pn19k2	18	2	0.96	2.48	3.92	0.65	0.94	1.32	0.8	0.54	0.86	0.18	16.13	0
Pn20k2	19	2	1.35	4.35	4.72	2.57	1.20	1.31	1.08	2.12	4.17	1.47	35.97	0
Pn21k2	20	2	1.62	6.71	6.1	3.99	1.32	1.44	1.16	0.60	1.63	0.00	32.49	0
Pn22k2	21	2	1.9	4.44	6.47	0.58	1.77	2	1.46	1.29	3.26	0.27	45.68	0
Pn22k8	21	8	0.54	2.32	1.4	0.48	0.65	0.67	0.63	0.09	0.09	0.09	8.7	0
Pn23k8	22	8	0.75	1.34	1.92	0.21	0.77	0.87	0.69	0.19	0.21	0.06	6.53	0
Pn40k5	39	5	11.1	2.7	35.71	0.58	6.86	8.12	5.9	1.06	1.84	0.63	921.45	0
Pn44k5	44	5	17.27	2.71	58.07	1.03	9.49	10.6	8.45	1.15	1.76	0.52	3600	0.56
Pn50k7	49	7	18.93	8.25	95.13	2.08	12.33	14.15	8.93	1.94	3.43	0.74	3600	0.75
Pn50k8	49	8	20.47	5.07	68.31	0.93	11.06	13	8.86	1.64	2.96	0.36	639.97	0
Pn50k10	49	10	17.49	4.76	59.12	1.8	11.73	12.42	11.02	1.34	2.11	0.33	3600	0.14
Pn51k10	50	10	15.75	1.35	59.74	0.31	11.85	16.55	8.7	0.47	0.63	0.37	104.93	0
Pn55k7	54	7	28.08	6.69	104.69	1.84	16.01	17.9	12.85	2.56	5.17	1.17	3600	0.66
Pn55k8	54	8	26.6	4.55	91.93	0.99	15.64	17.89	13.08	2.35	3.38	1.08	3600	0.53
Pn55k10	54	10	21.77	9.2	78.84	1.6	13.67	15.85	11.82	1.80	2.9	0.46	215.05	0
Pn55k15	54	15	17.28	4.37	54.41	0.53	13.49	14.56	11.88	1.02	1.42	0.76	85.63	0
Pn60k10	59	10	27.74	4.74	123.29	1.24	17.75	20.05	16.59	1.64	2.75	0.74	3600	0.42
Pn65k10	64	10	45.14	4.75	169.33	2.1	21.56	26.11	16.26	1.65	3.05	0.46	3600	0.45
Pn70k10	69	10	63.01	4.3	238.03	1.51	30.73	33.8	26.15	1.84	2.62	1.12	3600	0.46
Pn76k4	75	4	162.97	-56.45	497.04	-56.61	73.80	81.58	69.93	-56.35	-56.02	-56.73	3600	58.49
Pn76k5	75	5	136.95	1.22	444.04	1.21	59.06	66.6	43.5	1.42	1.69	1.20	3600	-
Avg			29.00	1.53	100.14	-1.35	15.09	17.14	12.73	-1.30	-0.41	-1.99		

of nodes equal to 30, we observe that when the number of nodes is below the threshold, the average gap of the best upper bound is around 3.87%, and that it decreases to 0.31% when the number of nodes exceeds the threshold. Over all the instances, the average gap of the best upper bound is around 5%. As far as the computational time is concerned, it is neglectable in all the three cases.

2.4.4 Results for the E-instances

The computational results for the E-instances are reported in Tables 2.10, 2.11, and 2.12. The RGRASP is capable to obtain good quality solutions with an average gap around 0.5% for all the three values of λ and always less than 1.44% (instance En51k5 with $\lambda = 0.5$). The average running time required by the heuristic method to solve all the tested instances is approximately equal to 9%, 11%, and 17% of the time required by the SCIP solver, for $\lambda = 0.1, 0.5$, and 0.9 , respectively. While the heuristic algorithm completes its runs in less than 160 seconds, on average, SCIP takes close to

Table 2.8: Results for the P-instances, $\lambda = 0.9$

Instance	n	K	HILS		RGRASP		GRASP						SCIP	
			CPU	Gap%	CPU	Gap%	CPU			Gap%			CPU	Opt%
							Avg	Max	Min	Avg	Max	Min		
Pn16k8	15	5	0.28	3.47	0.81	0.85	0.35	0.42	0.25	0.85	0.85	0.85	1.96	0
Pn19k2	18	2	1.23	2.47	3.93	0.11	1.08	1.24	0.85	0.2	0.58	0.11	16.77	0
Pn20k2	19	2	1.46	2.32	4.71	1.07	1.21	1.36	1.05	0.43	1.07	0.00	19.62	0
Pn21k2	20	2	1.64	2.23	6.34	0	1.30	1.44	1.18	0.71	2.9	0.00	25.91	0
Pn22k2	21	2	2.12	3.37	6.46	0.41	1.73	1.93	1.55	0.72	3.56	0.20	25.89	0
Pn22k8	21	8	0.65	2.39	1.55	0.33	0.75	0.79	0.73	0.51	0.51	0.51	8.92	0
Pn23k8	22	8	0.84	1.47	1.94	0.31	0.68	0.69	0.66	0.15	0.15	0.15	6.04	0
Pn40k5	39	5	8.83	2.75	35.22	0.71	6.89	7.57	6.24	1.31	2.48	0.56	160.83	0
Pn44k5	44	5	15.18	2.48	56.25	1.07	9.29	10.94	7.47	1.06	1.58	0.42	379.37	0
Pn50k7	49	7	19.95	5.61	96.26	0.82	11.86	13.73	8.83	1.57	2.98	0.59	310.64	0
Pn50k8	49	8	19.65	5.12	67.86	1.72	11.26	12.35	9.97	1.42	2.11	0.64	205.71	0
Pn50k10	49	10	17.51	3.61	59.24	1.36	11.58	12.25	10.79	0.87	1.56	0.52	63.52	0
Pn51k10	50	10	17.58	1.65	62.33	0.64	12.07	18.56	9.85	0.54	0.77	0.34	122.08	0
Pn55k7	54	7	24.07	7.46	96.55	3.45	15.09	18.44	10.69	2.53	3.98	0.85	425.51	0
Pn55k8	54	8	27.89	5.44	96.82	1.73	17.16	18.28	16.09	2.7	3.86	1.28	301.08	0
Pn55k10	54	10	20.7	7.5	77.82	0.98	13.45	15.8	10.99	1.62	4.1	0.71	98.08	0
Pn55k15	54	15	17.11	4.61	55.55	0.54	13.09	13.91	11.31	1.05	1.86	0.65	100.10	0
Pn60k10	59	10	27.65	4.23	127.78	1.05	17.16	20	14.56	1.28	2.4	0.65	170.62	0
Pn65k10	64	10	45.42	4.4	168.98	1.45	21.62	25.69	18.53	1.69	2.66	0.46	858.86	0
Pn70k10	69	10	57.49	4.65	230.62	1.6	29.84	34.02	25.11	1.78	2.69	1.01	728.51	0
Pn76k4	75	4	173.45	-108.27	505.08	-108.46	74.20	77.46	65.24	-108.241	-107.83	-108.57	3600	110.28
Pn76k5	75	5	126.66	1.27	438.49	1.26	61.75	68.57	51.78	1.34	1.5	1.12	3600	-
Avg			28.52	-1.35	100.03	-3.95	15.16	17.07	12.90	-3.81	-2.89	-4.41	579.78	5.25

one hour for the largest test instances (more than 30 nodes), especially for risk-averse problems.

It is apparent that the use of the reactive strategy achieves more reliable solutions compared to the one without. More specifically, the results are improved by 0.28%, on average, when the reactive mechanism is executed. In addition, the solutions obtained for different values of α can be really different, with a maximum variation of around 6% for the test problem En23k3 with $\lambda = 0.1$.

The HILS heuristic provides roughly satisfying solutions with the average gap limited above by 2.63%.

Analyzing the effect of the parameter λ , the decrease in the value of λ , makes the problem more involved, as we have already observed for the P-instances. This claim is supported by the increase in the solution gap and the *CPU* times of the results of the proposed heuristics as well as the SCIP solutions.

Table 2.9: Upper bound performance for the P-instances

Instance	n	K	λ	Δ^1	Δ^2	Δ^3	Best Δ
Pn16k8	15	5	0.1	12.78	40.02	7.6	7.6
			0.5	4.21	23.35	3.13	3.13
			0.9	3.39	21.59	3.02	3.02
Pn19k2	18	2	0.1	19.31	15.96	8.87	8.87
			0.5	10.73	8.23	4.71	4.71
			0.9	9.72	7.31	4.13	4.13
Pn20k2	19	2	0.1	24.88	16.93	8.54	8.54
			0.5	11.01	5.58	2.21	2.21
			0.9	11.56	6.30	3.28	3.28
Pn21k2	20	2	0.1	26.84	12.20	5.92	5.92
			0.5	17.81	7.28	3.82	3.82
			0.9	16.81	6.77	3.61	3.61
Pn22k2	21	2	0.1	14.33	9.03	5.21	5.21
			0.5	8.98	4.70	2.37	2.37
			0.9	8.51	4.41	2.26	2.26
Pn22k8	21	8	0.1	2.87	20.16	5.49	2.87
			0.5	2.04	13.33	1.77	1.77
			0.9	1.90	12.59	1.34	1.34
Pn23k8	22	8	0.1	3.17	15.47	2.11	2.11
			0.5	1.34	10.16	2.57	1.34
			0.9	1.20	9.61	2.61	1.2
Pn40k5	39	5	0.1	12.79	9.49	6.42	6.42
			0.5	7.88	5.15	3.72	3.72
			0.9	7.47	4.80	3.53	3.53
Pn44k5	44	5	0.1	10.52	6.61	5.03	5.03
			0.5	7.37	4.44	3.38	3.38
			0.9	7.05	4.21	3.21	3.21
Pn50k7	49	7	0.1	14.39	16.33	10.66	10.66
			0.5	9.12	10.17	6.97	6.97
			0.9	8.60	9.54	6.59	6.59
Pn50k8	49	8	0.1	11.63	16.06	9.93	9.93
			0.5	7.06	9.51	5.89	5.89
			0.9	6.60	8.84	5.47	5.47
Pn50k10	49	10	0.1	6.93	14.65	7.62	6.93
			0.5	4.44	9.33	4.93	4.44
			0.9	4.07	8.64	4.52	4.07
Pn51k10	50	10	0.1	2.98	4.38	2.51	2.51
			0.5	2.00	2.83	1.48	1.48
			0.9	1.89	2.72	1.36	1.36
Pn55k7	54	7	0.1	15.26	15.10	10.42	10.42
			0.5	9.75	9.27	6.04	6.04
			0.9	9.20	8.80	5.65	5.65
Pn55k8	54	8	0.1	12.61	15.09	9.66	9.66
			0.5	7.44	9.57	5.96	5.96
			0.9	6.95	9.05	5.62	5.62
Pn55k10	54	10	0.1	12.51	23.20	13.23	12.51
			0.5	8.07	15.19	8.42	8.07
			0.9	7.46	14.41	7.75	7.46
Pn55k15	54	15	0.1	7.03	26.29	13.16	7.03
			0.5	4.05	18.32	9.25	4.05
			0.9	3.68	17.42	8.76	3.68
Pn60k10	59	10	0.1	12.12	19.53	11.1	11.1
			0.5	7.70	13.45	7.97	7.7
			0.9	7.18	12.75	7.57	7.18
Pn65k10	64	10	0.1	9.12	13.11	7.66	7.66
			0.5	5.54	9.20	4.93	4.93
			0.9	5.29	8.91	4.78	4.78
Pn70k10	69	10	0.1	10.65	13.20	9.88	9.88
			0.5	6.38	8.94	6.32	6.32
			0.9	6.03	8.59	6.08	6.03
Pn76k4	75	4	0.1	6.17	2.13	2.01	2.01
			0.5	4.91	1.61	1.54	1.54
			0.9	4.63	1.43	1.36	1.36
Pn76k5	75	5	0.1	4.41	1.87	1.7	1.7
			0.5	3.37	1.30	1.21	1.21
			0.9	3.23	1.22	1.14	1.14
Avg				8.29	10.87	5.41	5.02

Table 2.10: Results for the E-instances, $\lambda = 0.1$

Instance	n	K	HILS		RGRASP		GRASP						SCIP			
			<i>CPU Gap%</i>		<i>CPU Gap%</i>		<i>CPU</i>			<i>Gap%</i>			<i>CPU Opt%</i>			
							Avg	Max	Min	Avg	Max	Min				
En22k4	21	4	0.99	5.96	3	.53	0.29	1	.02	1.32	0.78	0.64	2.64	0.29	26.95	0
En23k3	22	3	1.84	7.45	5	.76	1.58	1	.25	1.63	1.04	3.14	6.62	0.66	18.56	0
En30k3	29	3	5.61	3.6	1	6.25	-1.84	3	.38	3.9	3.09	-1.07	0.02	-2.40	3600	17.45
En33k4	32	4	6.3	1.19	1	8.8	0.06	3	.97	4.5	3.2	-0.32	1.18	-1.41	3600	14.04
En51k5	50	5	22.93	2.79	8	6.89	2.2	1	3.61	15.61	10.77	0.85	1.46	0.18	3600	6.61
En76k7	75	7	108.8	3.77	3	71.74	3.22	4	8.82	53.17	43.85	3.86	4.74	3.45	3600	-
En76k8	75	8	95.38	-1.73	3	66.75	-2.01	4	1.45	49.39	32.05	-1.54	-1.08	-1.88	3600	5.42
En76k10	75	10	85.95	0.33	3	25.4	0.09	4	3.30	45.46	37.27	0.34	0.64	0.10	3600	2.56
En76k14	75	14	62.26	0.27	2	43.2	0.1	3	2.93	36.96	29.94	0.40	0.6	0.22	3600	1.64
Avg			43.34	2.63	159.81	0.41		21.08	23.55	18.00	0.70	1.87	-0.09			

Table 2.11: Results for the E-instances, $\lambda = 0.5$

Instance	n	K	HILS		RGRASP		GRASP						SCIP			
			<i>CPU Gap%</i>		<i>CPU Gap%</i>		<i>CPU</i>			<i>Gap%</i>			<i>CPU Opt%</i>			
							Avg	Max	Min	Avg	Max	Min				
En22k4	21	4	1.13	3.18	3.76	0.98	1.13	1.22	1.06	0.861	2.87	0.35	16.44	0		
En23k3	22	3	1.79	2.18	5.79	0	1.33	1.68	1.11	1.63	3.98	0.84	15.73	0		
En30k3	29	3	5.44	2.24	17.01	0.35	3.50	4.06	2.85	0.89	2.36	0.35	1885.35	0		
En33k4	32	4	7.12	1.22	20.53	0.62	4.14	4.54	3.59	1.68	2.63	1.08	3600	0.88		
En51k5	50	5	29.03	2.97	93.77	1.44	13.87	17.23	11.53	1.12	1.45	0.61	3600	0.56		
En76k7	75	7	112.89	0.07	381	-0.27	46.20	55.22	39.05	0.31	0.51	0.09	3600	1.11		
En76k8	75	8	96.15	1.12	358.16	0.86	42.24	48.57	34.69	0.76	1.4	0.49	3600	0.47		
En76k10	75	10	86.96	0.58	323.98	0.45	43.82	46.08	39.38	0.53	0.67	0.43	3600	0.26		
En76k14	75	14	64.17	0.25	248.58	0.16	32.55	34.57	30.49	0.29	0.38	0.16	3600	0.18		
Avg			44.96	1.53	161.40	0.51	20.98	23.69	18.19	0.90	1.81	0.49				

Table 2.13: Upper bound performance for the E-instances

Instance	n	K	λ	Δ^1	Δ^2	Δ^3	Best Δ
En22k4	21	4	0.1	16.76	19.94	15.14	15.14
			0.5	10.88	12.52	10.5	10.5
			0.9	10.2	11.69	9.93	9.93
En23k3	22	3	0.1	21.23	26.57	17.17	17.17
			0.5	14.47	15.96	12.09	12.09
			0.9	13.87	14.93	11.58	11.58
En30k3	29	3	0.1	16.14	27.52	23.13	16.14
			0.5	9.48	17.24	14.57	9.48
			0.9	8.81	16.13	13.62	8.81
En33k4	32	4	0.1	8.82	29.23	9.63	8.82
			0.5	2.76	17.14	6.5	2.76
			0.9	2.09	15.88	6.12	2.09
En51k5	50	5	0.1	11.61	6.69	5.44	5.44
			0.5	7.56	3.41	2.75	2.75
			0.9	7.21	3.13	2.54	2.54
En76k7	75	7	0.1	5.44	3.86	3.33	3.33
			0.5	3.79	2.49	2.15	2.15
			0.9	3.52	2.26	1.93	1.93
En76k8	75	8	0.1	4.43	3.79	3.13	3.13
			0.5	2.79	2.29	1.86	1.86
			0.9	2.73	2.25	1.84	1.84
En76k10	75	10	0.1	2.81	3.48	2.63	2.63
			0.5	1.7	2.19	1.59	1.59
			0.9	1.6	2.06	1.49	1.49
En76k14	75	14	0.1	1.06	3.17	1.91	1.06
			0.5	0.73	2.37	1.41	0.73
			0.9	0.67	2.26	1.34	0.67
Avg			7.15	10.02	6.86	5.84	

Table 2.12: Results for the E-instances, $\lambda = 0.9$

Instance	n	K	HILS		RGRASP		GRASP						SCIP	
			CPU	Gap%	CPU	Gap%	CPU			Gap%			CPU	Opt%
							Avg	Max	Min	Avg	Max	Min		
En22k4	21	4	1.29	2.19	3.92	0.19	1.02	1.32	0.78	0.925	1.54	0.19	18.05	0
En23k3	22	3	2.12	4.95	6.18	2.5	1.25	1.63	1.04	2.12	3.09	0.42	21.43	0
En30k3	29	3	5.65	2.48	16.48	0.89	3.38	3.9	3.09	0.98	2.35	0.17	128.36	0
En33k4	32	4	6.72	1.34	19.94	0.74	3.97	4.5	3.2	1.31	2.38	0.54	123.18	0
En51k5	50	5	25.22	2.38	91.66	0.77	13.61	15.61	10.77	1.18	2.32	0.36	520.88	0
En76k7	75	7	107.73	0.77	375.76	0.43	48.82	53.17	43.85	0.71	0.99	0.50	3600	0.38
En76k8	75	8	90.49	1.17	362.8	0.92	41.45	49.39	32.05	0.69	0.96	0.49	3600	0.31
En76k10	75	10	92.19	0.72	327.18	0.45	43.30	45.46	37.27	0.47	0.6	0.33	1539.27	0
En76k14	75	14	70.77	0.28	254.77	0.19	32.93	36.96	29.94	0.25	0.35	0.16	1436.60	0
Avg			44.69	1.81	162.08	0.79	21.08	23.55	18.00	0.96	1.62	0.35		

In Table 2.13 the gaps between the best objective values obtained either by SCIP or by the heuristic (*Best objective value*), and the upper bounds proposed in Subsection 2.4 have been reported for the E-instances.

The first upper bound outperforms the third one in 9 out of 27 instances. The third upper bound provides, on average, slightly better gaps (less than 0.3% of improvement). The second upper bounds seems to be dominated by the other two, as for the P-instances. As before, the quality of the upper bound improves as far as the number of nodes increases. For example, considering as a threshold a number of nodes equal to 30, we observe that when the number of nodes is below the threshold, the average gap of the best upper bound is around 10.3%, and that it decreases to 2.2% when the number of nodes exceeds the threshold. Over all the instances the average gap of the best upper bound is around 5.84%.

2.4.5 Large size instances

Further computational tests have been carried out on larger size instances, with up to 199 customers and 17 vehicles. SCIP was not able to find a feasible solution within four hours and went out of memory, so we did not try longer time limit. For this reason, we have omitted the corresponding columns and we only report the results for RGRASP heuristic to serve as a benchmark for future comparisons. The gaps have been obtained considering the best upper bound, hence the actual optimality gap is possibly less than the one reported.

Table 2.14: Results for the CMT-instances

Instance	n	K	λ	RGRASP	
				CPU	$Gap\%$
CMT3	100	8	0.1	569.1	1.29
CMT3	100	8	0.5	570.09	1.13
CMT3	100	8	0.9	570.05	1.17
CMT4	150	12	0.1	820.6	0.62
CMT4	150	12	0.5	1539.78	0.46
CMT4	150	12	0.9	795.31	0.45
CMT5	199	17	0.1	1918.91	0.32
CMT5	199	17	0.5	1858.53	0.24
CMT5	199	17	0.9	1902.41	0.23
CMT11	120	7	0.1	639.84	0.64
CMT11	120	7	0.5	733.11	0.38
CMT11	120	7	0.9	604.13	0.34
CMT12	100	10	0.1	542.81	0.95
CMT12	100	10	0.5	560.73	0.56
CMT12	100	10	0.9	558.65	0.6
Avg				945.6	0.63

The results in Table 2.14 show that the RGRASP metaheuristic is relatively stable in terms of solution quality but its running time becomes important on the largest instances with 199 demand nodes.

2.4.6 Deterministic K -TRPP

The K -TRPP under risk is a generalization of the deterministic TRPP. In this section, we apply our metaheuristic to solve instances of the TRPP, that will therefore constitute a new benchmark for this class of problems. The results are shown in Tables 2.15 and 2.16 for the P- and the E-instances, respectively. We report in this case also the objective function value and the number of visited nodes ($\#$) obtained by both the heuristics and SCIP. In particular, for the GRASP, we report the best, the worst, and the average values of the objective function (OF_{best} , OF_{worst} and OF_{avg} respectively), obtained by running the GRASP with different α values.

For the P-instances, first of all we should note that SCIP is able to find the optimal solution in all but the last two instances (Pn76k4 and Pn76k5), for which feasible solutions are found with an optimality gap of 6.54 and 1.11, respectively. The average solution times are smaller than those obtained for

the stochastic counterparts. The gap is in general smaller for the deterministic instances and for a few instances (namely Pn20k2, Pn21k2, and Pn22k2) the GRASP heuristic finds the optimal solution, for a given value of α . As far as the E-instances are concerned, we notice that SCIP is able to solve all but two instances to optimality within the time limit. These results are to some extent expected, since we can observe the same behavior in Table 2.12 for a value of $\lambda = 0.9$, which in fact assigns a low weight ($\lambda = 0.1$) to the variance of the objective function. Also in this case, the time spent for solving the deterministic problem heuristically is lower than that required by the mean-risk problem. We finally notice that the performance of the heuristic is still satisfying in terms of gap, even though it is not tailored to the solution of the deterministic problem.

Table 2.15: Results for the deterministic P-instances

Instance	RGRASP				GRASP							SCIP			
	<i>OF</i>	#	<i>CPU</i>	<i>Gap%</i>	<i>OF_{best}</i>	<i>CPU_{best}</i>	<i>OF_{worst}</i>	<i>CPU_{worst}</i>	<i>OF_{avg}</i>	<i>CPU_{avg}</i>	<i>Gap%</i>	<i>OF</i>	#	<i>CPU</i>	<i>Gap%</i>
Pn16k8	691	13	0.9	0.86	691	0.26	691	0.35	691	0.31	0.86	697	13	2.64	0
Pn19k2	2706	15	3.51	0.11	2706	0.58	2706	1.1	2706	0.85	0.11	2709	15	20.99	0
Pn20k2	2692	17	3.75	1.07	2721	1.06	2692	1.36	2715.2	1.24	0	2721	17	18.16	0
Pn21k2	2026	18	4.54	4.21	2115	1.03	2045	1.44	2088.9	1.27	0	2115	17	19.72	0
Pn22k2	3891	18	6.32	0.15	3897	1.28	3754	1.78	3867	1.53	0	3897	18	14.86	0
Pn22k8	1850	18	1.41	0.32	1843	0.55	1843	0.81	1843	0.65	0.7	1856	18	4.27	0
Pn23k8	2594	20	1.98	0.15	2594	0.68	2593	0.93	2593.2	0.85	0.15	2598	20	7.75	0
Pn40k5	6463	34	31.57	2.09	6540	5.32	6472	7.22	6499	6.40	0.92	6601	34	141.59	0
Pn44k5	9627	38	51.37	1.21	9672	5.46	9608	9.46	9646.1	7.84	0.75	9745	37	322.71	0
Pn50k7	4025	42	65.55	1.25	4037	12.59	3942	21.79	3986.4	16.91	0.96	4076	40	315.6	0
Pn50k8	4119	43	60.68	0.79	4116	10.77	3987	62.23	4074.1	19.32	0.87	4152	41	128.02	0
Pn50k10	4228	43	53.29	0.59	4227	9.64	4187	11.64	4202.9	10.7	0.61	4253	43	62.66	0
Pn51k10	11962	43	59.32	0.61	11981	14.25	11933	16.86	11969.3	14.98	0.46	12036	44	97.6	0
Pn55k7	4529	46	89.74	1.01	4528	11.65	4436	15.60	4504.7	13.65	1.03	4575	46	327.31	0
Pn55k8	4601	46	86.42	1.5	4642	11.63	4506	16.65	4585.1	13.89	0.62	4671	47	165.21	0
Pn55k10	2770	45	78.06	1.6	2797	10.40	2718	15.04	2772.2	12.75	0.64	2815	47	93.83	0
Pn55k15	2897	46	46.87	0.69	2897	9.62	2874	12.41	2888.9	11.65	0.69	2917	48	102.6	0
Pn60k10	4121	53	115.55	0.91	4102	12.26	4002	18.35	4071.9	15.51	1.37	4159	53	221.53	0
Pn65k10	5996	56	153.17	1.46	6006	16.74	5923	20.71	5955.3	18.96	1.3	6085	55	621.98	0
Pn70k10	6716	58	214.36	1.6	6766	22.90	6630	28.77	6693.8	25.73	0.86	6825	56	529.39	0
Pn76k4	60937	73	427.46	-5.5	60962	58.00	60775	81.09	60875.9	64.93	-5.55	57759	75	3600	6.54
Pn76k5	61735	74	385.38	-0.37	61870	41.09	61536	83.30	61711.3	54.73	-0.59	61506	74	3600	1.11
Avg			88.24	0.74		11.72		19.50		14.46	0.31			473.56	0.35

Table 2.16: Results for the deterministic E-instances

Instance	RGRASP				GRASP								SCIP			
	<i>OF</i>	#	<i>CPU</i>	<i>Gap</i>	<i>OF_{best}</i>	<i>CPU_{best}</i>	<i>OF_{worst}</i>	<i>CPU_{worst}</i>	<i>OF_{avg}</i>	#	<i>CPU_{avg}</i>	<i>Gap</i>	<i>OF</i>	#	<i>CPU</i>	<i>Gap</i>
En22k4	2300	18	3.51	0.9	2317	0.86	2285	1.38	2301.4	19	1.08	0.17	2321	20	14.95	0
En23k3	3564	19	5.65	2.44	3638	1.10	3541	1.42	3593.2	20	1.25	0.41	3653	19	20.33	0
En30k3	5468	27	13.82	0.71	5463	2.66	5394	3.44	5449.1	25	3.06	0.8	5507	27	92.4	0
En33k4	13517	31	17.28	0.79	13551	3.20	13411	4.08	13480.1	32	3.63	0.54	13624	31	130.84	0
En51k5	11620	46	78.51	0.93	11655	10.15	11557	15.05	11602.7	45	12.3	0.63	11729	46	420.52	0
En76k7	28770	72	346.64	-0.14	28667	36.69	28500	45.82	28594.8	75	39.55	0.22	28731	72	3600	0.73
En76k8	29010	72	324.61	0.29	28939	33.06	28805	41.95	28863.8	75	37.31	0.53	29093	72	3600	0.12
En76k10	29258	72	347.16	0.43	29249	31.42	29132	40.25	29211.7	75	37.36	0.46	29385	72	1326.53	0
En76k14	29628	72	227.02	0.09	29604	24.86	29513	32.31	29567.6	75	27.3	0.17	29654	72	623.15	0
Avg			151.58	0.72		16.00		20.63			18.09	0.44			1092.08	0.09

2.5 Conclusions and future research directions

We addressed the K -TRPP under uncertain travel times, which is an important problem in many customer-centric real-world applications. We developed different metaheuristics, tailored to cope with the peculiarities of the problem at hand and a local search procedure based on different effective moves. To enhance the performance, a move evaluation procedure over different neighborhood structures has been also introduced which speeds up the heuristic performance. The computational results show that our heuristics are able to provide good solutions, even when SCIP is not able to reach any feasible solution in 4 hours. In these cases, the proposed upper bounding procedures facilitated the assessment of the quality of the metaheuristic solutions.

Chapter 3

Uncertain latency-based routing problem with service level constraint: an application to disaster relief

3.1 Introduction and motivation

Based on the world disaster report published in 2016, from 2006 to 2015, about 1,917,556,951 people have been affected by different types of disasters all over the world ([11]). The total amount of estimated disaster damage during this time period is estimated to be about 1,424,814,000,000 dollars. Only within the years 2006 to 2015, about two billion people, all over the world, have been somehow affected by disasters that caused about one and a half trillion dollars of economic losses and many damages. The increasing growth in the number of natural and man-made disasters and the dramatic consequences on vulnerable population have put an increasing pressure on the operational research community to investigate different aspects related to the disaster management with the aim of providing effective supporting tools.

The use of mathematical modeling to tackle disaster management problems dates back to the 1950s, when the focus was on determining the optimal location for fire-fighting resources [255]. Later on, Knott developed pioneering vehicle routing models for relief management ([144, 145]). From that time, hundreds of papers have been published to create a consolidated stream of research on the emergency or humanitarian logistics which is fostered ev-

ery year by new and interesting contributions. The readers are referred, for example, to the survey papers ([64, 198]) for a general introduction on the topic. The emergency logistics literature is typically classified according to whether the disaster operations are performed before or after a disaster. Strategic planning problems, as facility location and stock pre-positioning and/or disaster mitigation and evacuation problems, play an instrumental role prior to the disaster occurrence (i.e. in the pre-disaster phase). After disaster occurrence (i.e. in the post-disaster phase) operational decisions should be planned and coordinated. These include last mile distribution of relief goods (such as medicine, food, water), and the routing of the rescue teams to affected areas.

Particularly, we focus on relief routing operations, which are recognized as a key element in the post-disaster phase with a crucial impact on the success level of the relief operations ([61]). The relief routing problem in the post-disaster exhibits distinguishing characteristics from routing in the business sector, and poses new challenges, as highlighted, for example, in [157]. First of all, commercial routing problems are known to be server-centric and cope with cost-based objectives, whereas humanitarian relief logistic operations are inherently customer-centric in the sense that customers' satisfaction and their priorities should be addressed.

Secondly, in the disaster emergency response context, the responsiveness is known as a common performance measure which leads to enhanced safety and welfare of the victims. Indeed, the sooner the affected areas are served, the more lives are saved and less suffering is experienced. This idea is supported by the humanitarian society emphasis on the importance of quick response within the first 72 hours after disaster ([219]), when the short-term demand, in terms of delivery of perishable commodities (such as medical medicine and blood products) to the affected areas, should be satisfied. In the relief routing literature, the emergency responsiveness is typically measured in terms of *arrival time* to the affected areas (see, for example, [61]) and the minimization of the total latency represents a primary goal for reducing deaths and losses.

Thirdly, because of the limited fleet size, in some situations visiting all the affected areas might require an excessive waiting time. In this case, the operational response planning should be performed according to a selective

framework in that only a subset of the affected areas should be visited. The choice of the areas should be carried out on the basis of an urgency level accounting for the population living there and the severity of the disaster in each area.

It is worth pointing out that, as in any operational context, even more in the emergency one, routing operations are complicated by the high level of uncertainty which is inherent in disasters. A severe problem that usually occurs after an earthquake is, for example, the partial road disruption that may limit the accessibility of some links. It is evident that ignoring uncertainty in the route definition can lead in this case to very poor solutions with detrimental consequences.

In this Chapter, we are aimed at addressing these challenges, presenting another variant of latency-based routing model at the presence of uncertainty.

We contribute to the scientific literature in many respects:

- To the relief routing literature, by proposing a mathematical model that recognizes the customer–centric nature of the problem aimed at minimizing the total arrival times of a fleet of vehicles to the areas affected by a disaster. The model is also enhanced with a service level constraint to guarantee that a minimum system performance is achieved. Affected areas are prioritized according to an urgency level, computed on the basis of the disaster severity and the population size. We mention that the issue related to the selection of a subset of customers on the basis of a service level constraint has been recently addressed in [58] in a general context, whereas [23] considered a minimum coverage ratio that is included into the objective function.
- To the stochastic Vehicle Routing Problem (VRP) literature, by proposing a risk–averse approach of the minimum latency problem, and designing an efficient solution approach, able to efficiently solve the problem.
- To the practice, by addressing a real case study based on the Haiti earthquake in 2010 and provides different managerial insights. The case study has been comprehensively examined and different plausible scenarios have been considered to further validate the model

- To the disaster management literature, by proposing a comprehensive review on some problems arising in post-disaster operations. In particular, we review the most recent papers, published after 2016, on the applications of logistic models for relief response planning in post-disaster phase. Particularly, we focus on different aspects of the literature and classify the recent literature based on the model type, the characteristics of the relief network, the types of decisions involved, the methodologies, the presence or the absence of uncertainty and the paradigm used to deal with uncertainty. In addition, we plan to stress the main streams followed in the literature, highlight the most important issues along with the potential gaps and the challenges ahead.

3.2 Disaster relief response and logistics planning: literature review

This section is aimed at conducting a purposive review of the most recent literature on post-disaster management, especially those addressing logistics aspects questions arising in the response phase. There are some recent reviews in the disaster management literature (see, for instance, [24, 50, 125, 125, 113, 26, 66, 134]) , but to the best of our knowledge, they refer to the papers published between the years 2005 and 2016. We present an updated version, restricting our attention to those studies on post-disaster response planning published from 2016 to Feb 2018 which are at least indexed in *Scopus* and accessible through *Google scholar*.

Among the gathered resources, we selected those containing at least one of our main keywords, including *disaster management*, *disaster response*, *post-disaster*, *logistics*, *disaster relief*, and *relief routing*.

Figure 3.1 shows the frequency of the 62 selected papers published in 31 operational research journals.

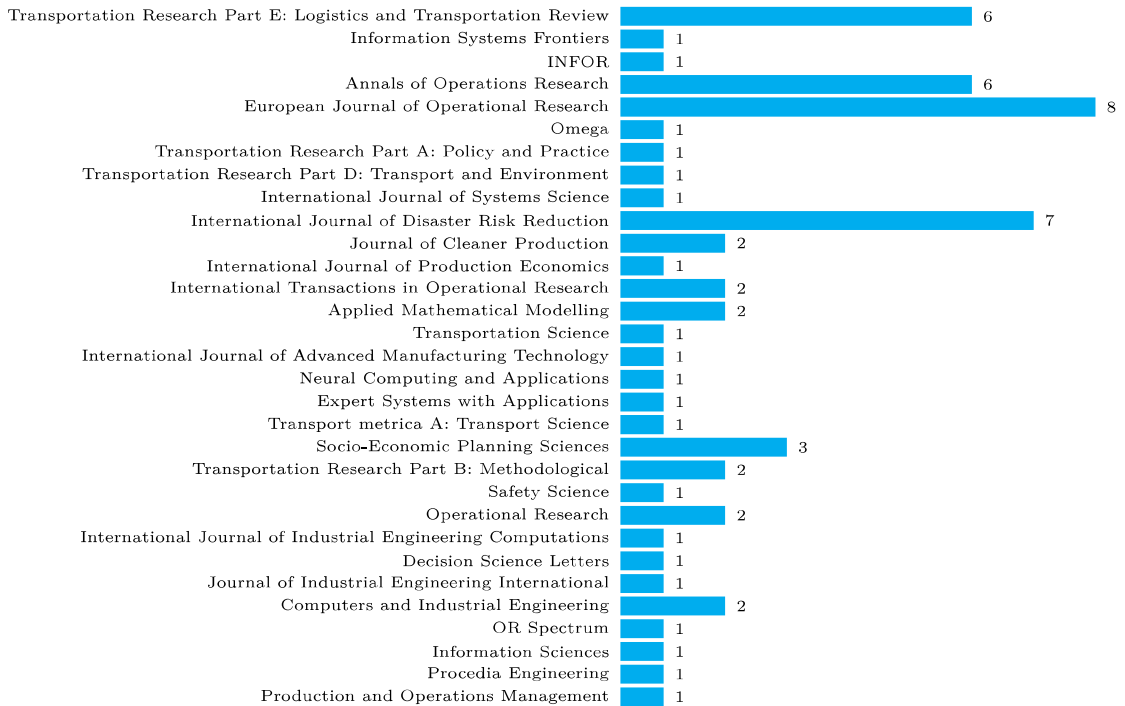


Figure 3.1: The frequency histogram of selected papers in journals

The rest of the Section is organized as follows. In Subsection 3.2.1, we review the literature on the deterministic mathematical models for humanitarian relief logistics. Subsection 3.2.2 provides a review on the non-deterministic contributions for relief logistics in which the uncertainty somehow has been addressed. Subsection 3.2.3 is devoted to a discussion on the contribution of technology applications such as Geographical Information System (GIS), simulation, and prediction models for humanitarian relief response. Finally, in Subsection 3.2.4, we present a discussion on findings, potential gaps, challenges ahead, and conclude the Section.

3.2.1 Deterministic models for humanitarian relief response

The OR community and the practitioners have investigated optimization-related issues in all the four components of the disaster management framework, from the pre-disaster planning (when the mitigation and preparedness operations are planned) to the post-disaster phase (when the relief response and recovery decisions are performed). The mitigation and preparedness

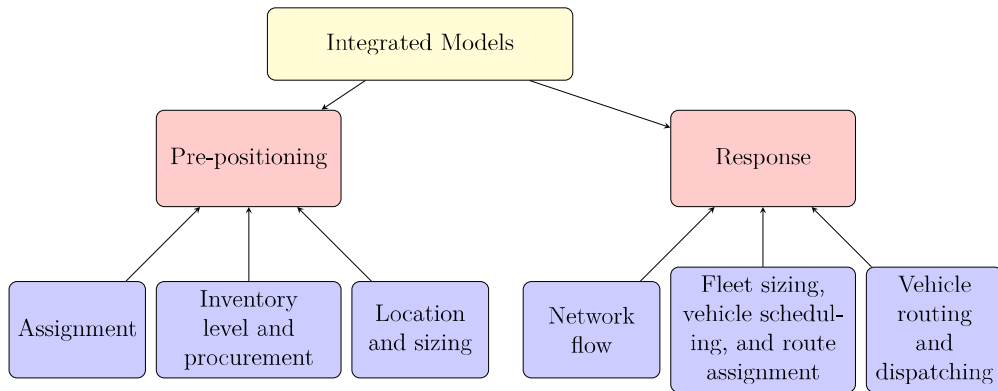


Figure 3.2: Decision layers in humanitarian relief network

phases encompass all actions performed before the occurrence of disaster with the aim of reducing, or possibly eliminating, the effects of hazards while the response and recovery phases begin after the disaster occurs and include all those operations performed to meet the needs of the affected people and to recover the damaged system and the infrastructure, respectively.

Figure 3.2 shows typical decisions in the pre-disaster and post-disaster phases. We have categorized the *decision types* addressed in the literature into one of the following types.

- Location and sizing
- Inventory level and procurement
- Assignment (allocation) and covering
- Network flow (transportation)
- Fleet sizing, vehicle scheduling, and route assignment (route design)
- Vehicle routing and dispatching

Location and sizing decisions are generally taken over the pre-disaster preparedness phase where the geographical configuration (location) of central depots, regional warehouses, distribution centers (DCs), and shelters as well as their capacities (sizing) for storing the relief items are determined.

After establishing the suppliers, the amount of relief commodities supplied in such places should be determined which is addressed as the *Inventory*

level (and procurement) decisions. In general, the inventory levels of suppliers are specified such that an estimated amount of needs over the very first days after the disaster, and before the arrival of other humanitarian donations, are satisfied. Usually the spoilage of some relief commodities necessitating the periodic replacement of such items as well as the damage of a percentage of commodities due to the disaster are taken into account.

The *Assignment (allocation)* decisions are common between both of the pre-disaster preparedness and the post-disaster response phase. Usually, the assignment decisions over the preparedness phase are about the allocation of major suppliers such as staging areas and central warehouses to the relief distribution centers and the regional warehouses while the allocation decisions over the response phase deal with the assignment of relief distribution centers to the affected population centers. The *covering* decisions are specific allocation types in which the allocation control the distance or travel time between the pair of supplier and demand.

The *Network flow* decisions deal with the amount of relief commodities transported between the suppliers, DCs, and affected people as well as the number of evacuees transported between the affected areas and the safe shelters considering the capacity of suppliers as well as the amount of demand in the system.

The *Fleet sizing* decisions are usually long-term decisions taken before the disaster, whereas *vehicle scheduling and route assignment* decisions are more tactical and planned as post-disaster decisions. In particular, *vehicle scheduling* decisions cope with the amount of vehicles and the number of trips to be dispatched between each pair of points such as supplier-relief distribution center or relief distribution center-affected area.

Route assignment (route design) is the process of determining a set of pre-specified routes for the vehicles ([228]). In the case of evacuation, there may be different roads connecting a population center, all different in terms of road conditions and capacity, traffic load, accessibility and safety. Hence the selection of routes connecting each pair of origins and destinations is a challenging problem to be addressed.

The operational and real-time decisions of arranging different routes and the order of visiting the nodes fall within the *vehicle routing* category, and *dispatching* deals with the choice of the most appropriate emergency vehicle

for a specific emergency response. We have classified the vehicle routing models into three categories: location-allocation models, network flow models, and fleet management models, based on the purposes the model was proposed for. The decisions on the *fleet sizing, vehicle scheduling, route assignment, vehicle routing and dispatching* all are issues related to the fleet management and, in our review, we refer to them as fleet management models.

Relief operations are focused on the tactical and operational decisions in the response phase, including (i) the establishment and the activation of regional warehouses, DC, and the points of distribution (PODs) which supply the affected people with relief commodities, (ii) the delivery and distribution of relief commodities (such as water, food, medical supplies, shelter, tent, clothing, etc) to the DCs, PODs, and the affected areas, (iii) the evacuation of affected people to safe shelters and the transport of casualties and injured people to the hospitals or temporary health centers. In addition, for all relief plans involving transportation actions, the operational decisions on emergency fleet routing and dispatching should be addressed as well. It is important to note that all of the aforementioned operational and tactical decisions are somehow connected to the strategic and long-term decisions adopted before the disaster occurs during the pre-disaster phase. For instance, the performance of the operational dispatching and routing policies are affected by the choice of the depot(s) (such as DCs or warehouses) to dispatch the emergency fleet from. Clearly, the decision on the location of such suppliers is a strategic decision to be taken in advance. This consideration has inspired the OR community to develop integrated models, as the state-of-the-art, in order to address both the operational and the tactical response decisions along with the long-term pre-disaster decisions simultaneously.

Therefore, even though the focus of this review is on the post-disaster phase, many studies present integrated models that address the strategic pre-disaster plans accounting for the *location and sizing* of warehouses, relief distribution centers, staging areas, shelters, etc, the *inventory level and procurement* of the relief commodities over the facilities (pre-positioning), along with the post-disaster tactical decisions on the relief distribution planning.

In Table 3.1, we have classified the papers based on different features, including the *decision(s)* involved, the *model type*, the definition of the *ob-*

jective(s), the *staticity or dynamicity*, the *disaster type* for which the model was proposed, and the *methodology* applied to solve the problem. We should mention that the deterministic and the non-deterministic studies are reviewed separately, and for those dealing with uncertainty, in addition to the aforementioned features, the *source of uncertainty* and the specific *paradigm* applied to deal with the uncertainty are considered as well.

Table 3.1: Classification of literature on deterministic models for humanitarian routing

Article	Model type	Decision type	Objective(s)	Static/Dynamic	Disaster type	Methodology
[23]	FM	AAs,VR	Coverage	Static	Earthquake	TS
[263]	FM	VR	Arrival time weighted by demands	Static	General	Hybrid ant Colony
[138]	L,NF ^{+,3-layer}	L,AL,RE, I,FS,CF (blood)	(MO)time, Cost	Dynamic (MP)	Earthquake	Lexicographic weighted Tchebycheff
[254]	L,FM ^{*,+}	L,CF, Shortage Road repair	(MO) Travel time, Cost(establishment,traveling,repair)	Dynamic (MP)	Earthquake	Metaheuristic (NSGAI + MOPSO)
[102]	L,FM	L,VR, Coordination	Average lead-time (arrival time)	Static	Flood	Simulation, Heuristic, TS
[118]	L,NF	L,CF, Shortage	Cost, Unmet demand, Wardrop equilibrium	Static	General	Exact
[89]	L,AL,NF	L,AL,CF, Population dynamics, Facility activation date	Population assisted	Dynamic (MP)	Flood	Decomposition, Heuristic, VNS
[13]	NF,FM	VR,CF, EF,Shortage	Shortage	Dynamic (MP)	General	Heuristic
[128]	NF ^{3-layer}	EF,CF,WF, Collected resource,I, Psychological penalty	Cost (monetary+ psychological penalty) (MO)	Dynamic (MP)	Earthquake	Exact
[211]	FM,NF	I,CF, Shortage, Number of trips, Budget	Shortage weighted by urgency level	Dynamic (MP)	Flood	Dynamic algorithm based on rolling horizon

*: multi-commodity, +: multi-modal transport, 2(3)-echelon: ²⁽³⁾, MP: Multi-period MO: Multi-objective, ϵ -C: ϵ -constraint, NF: Network flow, L:Location, LA: Location-Allocation, AL: Allocation, VR: Vehicle Routing, VS: Vehicle Scheduling, R: Relocation, I: Inventory, S: Sizing, FS: Fleet Sizing, RA: Route Assignment, FM:Fleet Management, TS: Tabu Search, CF: Commodity Flow, EF: Evacuation Flow, WF: Worker Flow

Most of the problems consider the distinction between different relief commodity types (multi-commodity is depicted by a * in Table 3.1). Some multi-commodity models, in Table 3.1, are bi-commodity and just consider two types of relief commodities, critical and non-critical [250, 254]. This reflects the fact that some relief commodities are perishable (such as blood and food) and should be treated differently in terms of delivery time. In addition, the demand of relief items over the affected region is not the same: for instance, areas close to the epicenter of a severe earthquake are more in need of medical supplies than those far from the epicenter.

Another common feature is the use of different vehicle types (multi-modal transport is depicted by + in Table 3.1): some isolated areas can be reached solely by aerial fleet such as helicopters. The relief network may consist different echelons, or layers with a hierarchical structure. For example, a typical 2-echelon network (depicted by ² in Table 3.1) might include the central warehouse, the relief distribution centers, and the affected areas and the relief commodities are flowed from the central warehouse to the relief distribution centers and, from there, to the affected areas.

The objective functions of the response relief literature can be classified into three groups addressing the *efficiency*, *effectiveness*, and *responsiveness* measures.

Finding the appropriate objective function in relief response is a challenging issue ([124]). As a popular trend, many studies incorporate multi-objective functions which is reasonable when different terms in different and possibly inconsistent measures, such as travel time and shortage or cost, are simultaneously considered. The most common objective is the minimization of total cost which may include the location and pre-positioning (inventory and procurement) costs, fleet usage and fleet sizing costs, the evacuation cost or the delivery cost of relief commodities, the holding cost of unused inventories stored in distribution centers, and the penalty cost for the shortage (unmet demands) and wastage (for perishable products such blood products [136]). In general, the cost-based objectives are classified as efficiency-based criteria. It should be noted that the penalty cost of unmet demands (shortage) is more related to the responsiveness criterion meeting the demands and cannot be interpreted as an efficiency index. In addition, even if we accept

that the penalized shortage costs can be added to the total cost, its cost coefficients should be expressed in terms of monetary loss to be consistent with other cost terms. To this end, [182] developed the idea of *deprivations costs*, interpreted as the monetary cost that the unmet demands, in the course of time, are willing to pay in order to meet their urgent needs. We should remind that the logistics models in response relief are, in nature, different from the logistics models in business context. Obviously, the definition of objectives in terms of cost, which is usual in the server-centric logistics model, is far from the scope of the humanitarian relief logistics that are supposed to be customer-centric.

Some studies dealt with the horizontal equity measure by minimizing the total unmet demands. Different studies has adopted different approaches to deal with equity in humanitarian relief. For example, Rezaei-Malek et al. ([209]) addressed the equity concept in the relief distribution by balancing and limiting the difference in the weighted shortage of commodities received by each pair of demand points. Similarly, In [213], the authors dealt with the equity concept over the preparedness phase, by minimizing the maximum unmet demands. Hu et al. ([127]) addressed the equity by incorporating a service level constraint to assure that the amount of relief commodities provided to each demand point, from the pre-disaster inventory levels and the post-disaster procurement, is above a given service level. In [89], Duhamel et al. evaluate the equity of relief distribution, as a post analysis step, in terms of the relative gap between the most and the least assisted affected areas over different periods. Noyan et al. ([192]) approached to equity from a different viewpoint by addressing the equity in accessibility, usually measured based on the response time, and the equity in supply allocation, expressed as the maximum proportion of unmet demands. The equity in accessibility is guaranteed by limiting the assignment of demand points to those points of distributions whose accessibility score is above a minimum threshold, where the accessibility scores for each link connecting the points of distribution to the affected areas are calculated based on the travel time, the risk of being close to a lake with the flooding danger, and a mobility score related to the demographical characteristics of the affected areas. In a similar work, Noyan et al. ([193]) presented a model for last-mile relief distribution, where the equity in the resource allocation is ensured by minimizing the deviation

from the desired level of the total amount of supply allocated to the points of distribution, striking the balance between the shortage over different points of distribution.

Compared to the *efficiency* and *responsiveness*, the *effectiveness*-based objectives have been less addressed [263, 102]. The *effectiveness* in relief response is related to the timely and effective delivery of relief aids to the affected people and it is closely related to the response time (latency or arrival time) criterion [263]. The reliability is another criterion addressed in the literature in which the possibility of damage in the infrastructure network is taken into account to deliver the relief items or to transfer the evacuees through safe routes [225, 226, 228, 227]. A study combining the efficiency, equity, and effectiveness-based objectives and metrics in relief routing can be found in [129].

The majority of the studies have recognized the dynamic nature of response operations to reflect variations in the demands pattern, supplies and inventory levels, travel times, and other input parameters over the operational horizon. The dynamic inputs are usually treated as time-dependent values over different time periods and a multi-period mathematical model is proposed to take the time-dependent decisions over each period. To the best of our knowledge, only Duque et al. ([90]) adopt a different approach to deal with the dynamicity of response relief. In fact, they presented an exact dynamic programming to arrange the scheduling and routing decisions for a repair crew responsible for road repair in order to facilitate the arrangement of relief response operations. Although the proposed problem is applied in the response phase to accelerate the relief operations, as the authors confirm, the model is more related to the recovery operations than to the relief distribution problem. We should highlight in this context, the importance of demand forecasting. Zhan and Liu ([270]) conducted a study to find the maximum response delay considering that more accurate demand estimations can be obtained with the arrival of new information over time.

A simple look at the disaster types of models in Table 3.1 shows that most of them focus on earthquake which is reasonable since earthquake is one of the most common disasters with devastating effects that might influence a larger population, compared with other disaster types such as floods or hurricanes. It is quite common to incorporate different scenarios, simulating

possible disaster conditions, based on the severity of disaster, the time it occurs, and the effects on the infrastructure transport network.

3.2.2 Stochastic models for humanitarian relief response

Most of the aforementioned studies in the deterministic context share a common idea about the importance of dealing with uncertainty. Uncertainty is known as the undeniable and inseparable characteristic of disaster relief operations originating from the unpredictability of disasters, in terms of severity, impacts, and the time and place it occurs. Uncertainty may affect many elements involved in relief operations, for instance, based on the magnitude of an earthquake, the demands for different types of relief items may vary from a few people to millions. The same reasoning holds for the possible variations in the delivery time and road reliability due to the disruption in the infrastructure network. Based on the disaster severity, even some warehouses, relief distribution points, shelters and emergency health centers may fail to operate (facility failure), and some portion of the pre-positioned relief stock stored in the pre-disaster phase may be damaged. This shows how the uncertainty experienced after the disaster may affect even the performance of pre-disaster policies implemented before. Many papers focus on the uncertainty of demands (see [65, 268, 136]), while uncertainty can be experienced in other elements, such as, supply and the usable portion of pre-positioned items, capacity, ([18, 127]), travel time, delivery cost, road reliability and availability, infrastructure ([205, 265, 227]), supply (influenced both by the behavioral pattern of society and by the humanitarian organizations ([80])). Uncertainty also affects some less addressed but important issues such as the the behavioral reactions of the affected population towards the relief network (the selection of the point of distribution to refer to ([118]), the acceptance of evacuation plans ([265]), and some safety related issues such as ransack and influx probability).

To cope with the uncertainty in relief operations, different approaches and assumptions are made. The most popular paradigm to model uncertainty in relief efforts is the stochastic framework, developed based on the *probability theory*, and among different stochastic paradigms, the two-stage stochastic programming approach has received most attention since it enables the modeler to link pre-disaster with the post-disaster decisions. A list

of studies incorporating the two-stage stochastic programming approach can be found in Table 3.2.

In the two-stage stochastic programming approach, the pre-disaster decisions are taken in the first stage, before the uncertainty is known. The stochasticity, represented by a set of discrete scenarios, affects the post-disaster decisions in the second stage which are *scenario-dependent*. In particular, the pre-disaster decisions on location of suppliers, sizing and inventory level are usually taken under uncertainty, whereas the second stage deals with the relief flow decisions after uncertainty disclosure.

As a shortcoming, we should mention that the two-stage models cannot represent a realistic scheme for the evolution of uncertainty over different periods as well as the relation between different response operations performed during different periods [205]. As an alternative, some studies developed models based on the multi-stage stochastic programming approach and considered different (dependent) decision stages over the response phase [265]. The popularity of multi-period modeling as an alternative to address the dynamicity of system, has motivated the development of two-stage multi-period stochastic models where the scenario-dependent second stage decisions are taken over a planning horizon ([101, 158, 51, 220, 18, 225, 226, 227]). The multi-stage stochastic programming is another faithful paradigm to cope with dynamicity where the response decisions are taken over a set of distinct stages showing the dependency of actions along the stages as well as the system evolution [265, 268]. Following another stream, Cook and Lodree ([80]) presented a stochastic dynamic programming approach in order to find the optimal dispatch for a single vehicle delivering relief commodities where the supply and demand are uncertain inputs. Of course the curse of dimensionality is one of the reasons why the stochastic dynamic programming has been less appealing, in comparison with the multi-period mathematical models, however, we believe that there is a potential to deal with stochastic models in dynamic programming context, even by adopting its approximate counterpart or the design of some heuristics to overcome the model complexity and computational intractability.

Following the stochastic framework, Noham and Tzur ([190]) adopted a scenario-based approach to characterize the stochasticity in demands and the number of local distribution centers.

Following the robust framework, some studies adopt the interval data robust approach proposed by Bertsimas and Sim ([44]) where the uncertain parameters takes their values from a continuous interval ([278, 225, 253]). Some studies adopt the scenario stochastic programming approach ([278]) to discretize the uncertainty set, in set of pre-specified scenarios ([253]).

Both in the robust framework and in the fuzzy counterpart, it is quite popular to propose hybrid models integrating different uncertainty approaches and uncertainty types. In [217], Salehi et al. proposed a robust two-stage stochastic model for the blood supply chain design, where the objective function is expressed as the linear combination of two terms, including the maximum deviation from the best scenario in terms of the total cost (related to the robustness) and the expected total cost where the cost is expressed as the summation of scenario-dependent and scenario-independent terms. Another contribution on the robust two-stage stochastic approach can be found in [209, 119], adopted from the original work of Mulvey et al. ([184]) and Yu and Li ([266]) to balance the *solution robustness* and *model robustness*. Following the possibilistic approach supported by the *probabilistic theory*, Shahparvari et al. ([226]) addressed vehicle scheduling, route design, and shelter selection decisions for the bushfire evacuation problem, where the evacuation time window, travel time, and shelter capacity are represented as triangular fuzzy numbers.

Addressing both types of possibilistic ambiguity and stochastic uncertainty has been the motivation for proposing the fuzzy scenario-based model proposed in [179], and the possibilistic two-stage and multi-stage stochastic programming models presented in [220] and [250, 268], respectively. The authors claim that the possibilistic and the stochastic information complement each other and modeling scenario-based stochastic parameters as imprecise fuzzy numbers is a reasonable alternative since, in general, the value of such parameters under each scenario cannot be precisely determined.

The uncertainty in route reliability has been addressed in [227] for an evacuation problem involving vehicle scheduling decisions and route design (for other variations and extensions of the model see [225, 226, 228]). The cumulative failure risk over each route, connecting a pair of population center and shelter, is calculated as a function of failure risk scores over each route segment. Most of the aforementioned studies are classified as risk-neutral

Table 3.2: Classification of literature on stochastic models for humanitarian routing

Article	Model type	Decision type	Objective(s)	Static/Dynamic	Uncertainty source	Uncertainty paradigm	Disaster type	Methodology
[65]	*L,NF	L,CF, Shortage	cost (transportation,shortage)	Static	Demand	Two-stage stochastic	Earthquake	Exact
[179]	L,AL	L,AL, Coverage	Demand-weighted travel distance, Coverage, Evacuees' failure (MO)	Static	Demand, Facility failure	Fuzzy stochastic programming	Flood	Lp-metrics technique
[182]	*+ L,NF	L,FS,NF, RA,I,Shortage	Cost (MO)	Dynamic (MP)	Incoming supply, Demand, Usable commodity, Route availability	Two-stage stochastic	Flood	Heuristic
[270]	*+NF,FM	CF,VS,FS	Loss	Dynamic (MP)	Demand	Bayesian updating	Typhoon	Exact
[268]	² NF	CF,I, minimum percentage of demands met	Prioritized cost	Dynamic (MP)	Demand	Multi-stage possibilistic stochastic programming	Earthquake	Exact
[278]	* ³ L,NF	L,CF, Shortage	Cost (transportation,L,Shortage)	Static	Demand, Supply, Cost	Robust (scenario stochastic programming)	Earthquake	Exact
[217]	* ³ L,NF	L,AL,RE, I(stocked),CF(blood), Shortage,Supply	Cost, Delivery time, Shortage	Dynamic (MP)	Demand, Capacity, Supply	Robust two-stage stochastic	Earthquake	Exact
[226]	FM	RA,VS, Shelter selection	Evacuees	Dynamic (MO)	Shelter capacity, Time window, Travel time, Network availability	Fuzzy possibilistic programming, Risk	Earthquake	Genetic algorithm

*: multi-commodity, +: multi-modal transport, 2(3)-level: $2^{(3)}$, MP: Multi-period MO: Multi-objective, SO:single objective, ϵ -C: ϵ -constraint, NF: Network flow, L:Location, AL: Allocation, VR: Vehicle Routing, VS: Vehicle Scheduling, R: Relocation, I: Inventory, S: Sizing, FS: Fleet Sizing, RA: Route Assignment, FM:Fleet Management, TS: Tabu Search, CF: Commodity Flow, EF: Evacuation Flow, WF: Workers Flow

models ([65]) while there are also some contributions on risk-averse models ([91]). Most studies in the risk-averse framework, adopt the Conditional Value-at-Risk (CVaR), as a coherent risk measure to cope with uncertainty; for example, Alem et al. ([18]) proposed a two-stage stochastic model for relief pre-positioning and distribution, where, in addition to the classical minmax-regret measure, the semi-deviation and CVaR measures were used. A similar contribution can be found in ([126]) and also in ([79]) in which the objective function is expressed as the summation of the first stage cost and the linear convex combination of the CVaR and the expected second-stage cost. By varying the relative weight of the CVaR and the expected term, different risk-aversion degrees of the decision maker can be modeled. The idea of the latter paper is also adopted in [91] presenting a pre-positioning and relief distribution model, where joint probabilistic chance constraints are used to assure the feasibility of second stage decisions.

As the only study on the *uncertainty theory*, Huang and Song ([130]) presented a customer-centric model for response operations where travel times are dealt by the uncertainty theory and the total arrival time is constrained to be as small as possible with a given confidence level.

3.2.3 Simulation, GIS, and prediction methods

Although the contribution of the OR community to address the optimization problems arising in disaster management is quite satisfying and valuable, we should note that the role of mathematical models should be regarded as complementary to other decision support tools such as simulation, GIS and qualitative models ([102, 276]).

Enabling technologies, such as GIS, are able to handle massive load of spatial-based data and information which can be managed for analysis purposes; for example, GIS can be helpful as a network analysis tool to obtain more accurate data about spatial elements and, hence, applied to determine the vulnerability risk over the infrastructure network and the severity of disaster spread on different areas. GIS is also widely used to find the location of potential suppliers and to design the set of pre-specified routes for the route assignment problem, taking into account real-time travel times. Last but not least, GIS can be implemented to generate different scenarios to be used for simulating and visualizing the planning scenarios.

Among those recent papers which use GIS for post-disaster management, we refer to the work of Zhao and Liu ([276]) who presented a decision support system, integrating multi-objective optimization with GISs in order to design a network of urban emergency rescue facility locations. The authors emphasize the necessity of addressing the problem in the geospatial multi-objective context and of dealing with big data on the spatial pattern and natural geographical conditions. They also proposed a heuristic based on a genetic algorithm to solve the three-objective optimization model, including service capacity, global efficiency, and equity. Recently, Rodríguez-Espíndola et al. ([213]) presented an integrated approach combining GIS and optimization for the disaster preparedness in order to find the optimal supplier locations and the pre-positioning policy. Specifically, under the case of a flood, the authors applied GIS to first, discard potential flood-prone facility locations (shelters and distribution centers), second, to assess the damage experienced in population centers, and third, to investigate the possible effects of disaster on infrastructure network disruption under different scenarios. Clearly, the solutions provided by the complex optimization models and tailored heuristic algorithms cannot be implemented in practice unless the

input parameters are generated and estimated in a reliable manner. Following this stream, Lu et al. ([155]) proposed a prediction module using interval data fusion (IDF) and interval Kalman filtering (IKF) methods to predict the uncertain and time-dependent input parameters, including the demands and the delivery times for the real-time relief distribution applications. The prediction relies on the multiple information sources provided with the stakeholders over the time. Then, a mathematical module is called to find the relief flow over the network for each time period using a rolling horizon-based approach. The proposed framework is applied on a case study on the 1999 Taiwan earthquake.

The reliability of the solutions provided by the mathematical models should be evaluated by simulation and under the supervision of disaster management experts. The simulation in humanitarian relief literature is used to evaluate the efficiency of models and/or heuristic algorithms ([206]) or in order to generate possible disaster scenarios (see [101] for a case of using simulation for disruption scenario generation) and some input data, such as demands and priorities (see [205, 155]) or to find the optimized values for the heuristic input parameters ([181]). [155] take the advantage of simulation to generate a set of initial data that were combined with other official resources in order to obtain the number of time-dependent fatalities over the affected areas. Salehi et al. ([217]) took the advantage of a Monte Carlo simulation for a blood supply chain network design problem, and evaluated the quality of solutions of the deterministic model in comparison with those provided from the robust stochastic counterpart.

In [102], Fikar et al. proposed a simulation-optimization approach to design a decision support system for the coordinated disaster relief distribution. They applied an agent-based simulation to evaluate the solution of the optimization approach within a loop.

3.2.4 Findings and potential gaps

Despite the significant contribution of the reviewed studies on humanitarian relief, still there are many open questions to be answered, many challenges to cope, and potential gaps to be addressed. Although most of the contributions, both in the deterministic and stochastic context, address integrated models dealing with at least two decision types, still there is a potential

to present more sophisticated integrated models in which all response relief decisions are somehow addressed with the aim of providing the stakeholders with a unified network. In addition, it would be interesting to see how some short-term post-response decisions— such as recovery— can be implemented in parallel with relief distribution actions in order to operate the response phase in a more successful manner. For example, the repair of some roads with minor disruption might help to perform the relief response over the recovered route segments, shortening the response time. To the best of our knowledge, there are only a few studies addressing both the road recovery and the relief distribution decisions [254, 90].

An underinvestigated area concerns the efforts to model the behavioral pattern of the affected people towards the stakeholders' decisions: socio-economical, cultural and demographical characteristics of the affected population should be taken into account in this respect (for example, see [23]). This not only enables the stakeholders to model the reaction of affected people in a better way, but also helps them to provide appropriate relief services to people with special needs, addressing vertical equity (horizontal fairness deals with providing relief services with all emergencies in a fair manner; vertical equity deal with providing different services based on different characteristics and needs). The work of Gutjahr and Dzubur ([118]) is an attempt to consider the behavioral pattern of the relief recipients. This is a realistic assumption since, in practice, the relief assignment pattern expected by the stakeholders may differ from what the relief recipients do in reality. In the proposed model, the upper level is controlled by the stakeholders with the aim of finding optimal distribution center locations while minimizing the location cost and the amount of unmet demands, whereas at the lower level the affected people should choose to which distribution centers refer. The model provides the affected people with a choice model to select their supplier based on both the travel cost and the amount of supply expected to receive to count for the relief scarcity.

As another potential gap, we can mention the development of bi-level models to cope with the fact that the decisions makers cannot always control the affected population behavior to implement the stakeholder's decisions as the evacuation orders or the relief distribution policies. The idea of developing bi-level models for humanitarian logistics is appealing, since it enables

the decision makers to incorporate the behavioral reaction of the affected people receiving relief aids. This is of great importance since the reaction of affected people versus the evacuation operations or the points of distributions providing them with relief commodities may be inconsistent with what the stakeholders expect. In the stochastic context, Yi et al. ([265]) presented a bi-level multi-stage stochastic model to arrange the evacuation orders under a hurricane and at the presence of uncertainty in hurricane evolution. The proposed bi-level model is another attempt to model the behavioral reaction of the people who does not follow the evacuation order and refuse to leave their properties. The model is formulated as a multi-stage stochastic problem taking the uncertainty of the travel time and time away from home as well as the travel risk and the risk of sheltering-at-home into account. We strongly believe that still there is a gap to be filled in bi-level models. In particular, an interesting future research stream can be devoted to the study of bi-level models in which the reactions and the behavioral patterns have a hierarchical structure, where, in addition to the stakeholders of public relief agencies, there are other private or non-governmental relief agencies (such as NGOs) that perform the humanitarian relief in parallel. The design of a three-level mathematical model, with two distinct levels for the public and private relief agencies and one level for the affected people, would be an appealing idea in order to systematize the cooperation between the relief teams at different decision-making levels over a centralized decision-making system and to address the behavioral reactions of the affected population. Obviously, this might prevent the probable inconsistency of relief operations conducted by different stakeholders in a decentralized and non-cooperated system.

As another potential gap, we should refer to the lack of customer-centric and efficacy-based models, especially those dealing with waiting times and delay in response. To the best of our knowledge, there are a few studies proposing customer-centric objectives expressed in terms of the response of arrival times ([130] in the stochastic framework and [263, 102] in the deterministic context). In addition, the literature on fleet management for relief efforts is more focused on fleet sizing, vehicle scheduling, and route assignment, leaving less contributions on the vehicle routing problem, especially, in the last mile relief logistics [192].

In the next Section, we present a latency-based vehicle routing problem for post-disaster relief routing under uncertainty.

3.3 A latency-based vehicle routing problem with service level constraint under uncertainty

In this section, we address the post-disaster relief operations planning problem, involving visiting the affected areas to promptly supply humanitarian aid to the victims of a disaster. This problem poses planning and operational challenges, caused by the uncertainty about the potential damages of the disaster and the number of victims, and the lack of information about the status of the transport infrastructure. Moreover, for the presence of scarce resources, not all the areas can be serviced and the relief plan should include a subset of sites to be visited immediately in the aftermath of a disaster.

Because of the richness of the scientific literature and of the partial analysis conducted in the literature review of the Chapter 3.2 (only the papers published from 2016 onwards have been reviewed), in the following, we shall focus on the papers relevant for us, highlighting our contribution with respect to the state of the art.

A vehicle routing problem arising in post-disaster humanitarian relief was proposed in 2015 by Sharif and Salari in [231]. Here, the demand of each affected area is satisfied either directly or through the assignment to another visited area which is within a given coverage distance. The problem, aimed at minimizing the total cost is solved by a GRASP algorithm. In [201], Pourrahmani et al. consider an earthquake evacuation routing problem from local shelters to regional ones for a long-term safe settlement using public vehicles. The main contribution relies on the definition of a dynamic evacuation routing approach that can update the routing plans by incorporating time-dependent travel times. The authors presented a simulated annealing algorithm for solving the mathematical formulation applied to a case study on routing evacuation in a district of Tehran, Iran. In another work [200], Pourrahmani et al. proposed a genetic algorithm to solve a similar model but in a static context, where the number of evacuees is a fuzzy number. Following the same stream, [264] presented a dynamic transportation routing

multi-period model capturing the fluctuations of travel times and demands for disaster relief operations. The model aimed at minimizing the total time traveled by vehicles, along the planning horizon, and the number of vehicles used. Also, a tabu search algorithm was proposed to solve the model. We also mention some contributions ([123, 154, 261]) that even though do not deal exactly with our problem, focus on the same dramatic event that hit Haiti in 2010. For example, [123] analyzes the performance of different post-disaster humanitarian logistic structures that arose in response to the Haiti earthquake. In [261], the authors focus on the distribution of emergency aid in disaster relief operations. In particular, the problem consists of designing routes for vehicles, choosing the types of vehicles and determining the flow of the aid. Different criteria, related to the specific conditions of the disaster, are considered in the route design. The contribution [154] extends the previous work and introduces a model that combines recovery operations of transportation infrastructure elements with aid distribution planning. The objective function considers multiple criteria, such as reliability and security.

The papers mentioned above consider, as the main criterion to be minimized, the total travel time (cost), thus neglecting the customer-centric nature that should be accounted for in the design of routing operations in relief effort. In this case, the minimization of the total arrival times at nodes should represent a primary goal to increase survivability. Notable exceptions are the work of Campbell et al. who proposed in [61] two latency-based formulations, the first one aimed at minimizing the maximum arrival time, and the second one focused on the average arrival time, the paper of Victoria et al. ([260]) who presented a cumulative capacitated vehicle routing problem for routing and relief distribution planning with time-dependent demands and the cumulative multi-depot routing model for relief distribution proposed in [263].

In our contribution, we introduce a customer-centric vehicle routing model to support the post-disaster relief activities, addressing an issue that has been only partially addresses by the extant literature.

In addition to the latency-based objective function, the other distinctive feature of our model is its selective nature. Selective vehicle routing problems have represented the subject of intensive research in the last ten years. We refer the interested readers to the very recent contribution [58] (and the

references therein), where the authors proposed a compact mathematical formulation aimed at minimizing the sum of transportation costs and lost profits. Both a Branch & Price algorithm and a hybrid genetic heuristic approach were tested on a large set of tests derived from benchmark instances. In the specific field of relief routing, we have already mentioned in Chapter 3.2 the recent paper ([23]), where the author presented a selective vehicle routing problem where only a subset of demand nodes is visited. The model was solved by a tabu search heuristic and tested on a real case study on earthquake in Turkey.

Besides integrating latency with selection, our contribution explicitly deals with uncertainty in the travel time. The analysis of the scientific literature of Chapter 2 reveals that, even in the general routing setting, latency based problems under uncertain travel times have not been fully investigated. Hence, we should expect that also in the disaster management context, the same claim is true. In fact, to the best of our knowledge, we could find only one paper ([130]) describing an emergency logistics distribution routing problem under uncertainty. In that contribution, travel times are dealt by the uncertainty theory and the total arrival time is constrained to be as small as possible with a given confidence level.

In the present contribution, we analyze the problem under a risk-averse perspective. Even this aspect is rather new in relief routing operations. An exception is represented by the very recent contribution of Elçi and Noyan ([91]) who proposed the use of the CVaR measure for the humanitarian relief network design.

The model is validated on a case study, which is based on real-world data of the Haiti earthquake in 2010. In addition, a heuristic approach is presented to solve larger cases.

The remainder of the Section is organized as follows. In Subsection 3.3.1, we briefly introduce the problem and present the mathematical formulation for optimizing the routing decisions under uncertain travel times. Subsection 3.3.2 is devoted to the description of the real case study of the Haiti earthquake. Subsection 3.3.3 present a thorough discussion about the model validity and behavior along with managerial insights. Subsection 3.3.4 introduces the heuristic approach designed to solve larger instances along with some preliminary computational results. Finally, Subsection 3.3.5 concludes

this study, summarizing the key findings and shedding light on future research.

3.3.1 The Mathematical Model

Let us denote by $G = (V, E)$ a complete undirected graph, where $V = \{0, 1, \dots, n\}$ and $E = \{(i, j) \subseteq V \times V\}$ represent the sets of vertices and edges, respectively. Node 0 corresponds to the depot where a fleet of K identical vehicles is based. Each node $i \in \bar{V} = V \setminus \{0\}$ represents an area affected by the disaster. From now on, terms “nodes”, “customers”, and “affected areas” are used interchangeably. With each node $i \in \bar{V}$ it is associated a coefficient θ_i computed as function of the severity level of the disaster affecting the area and the population living there. Obviously, those populated areas more severely affected by the disaster have a higher priority to receive service. Any edge $(i, j) \in E$ represents a possible link between two areas i and j with an associated travel time d_{ij} . Assuming that such values are known in advance is a very stringent hypotheses in any operative setting, and even more in the aftermath of a disaster. Thus, we assume that $\tilde{d}_{ij}, \forall i, j \in V$ are independent random variables with known first- and second-order moments, denoted by μ_{ij} and σ_{ij}^2 , respectively.

The goal is to find K disjoint routes, (from now on, route and path are interchangeably used since the return of vehicles to the depot is not explicitly required), starting at the depot and visiting a subset of areas selected to assure a given service level, in a such a way to minimize the random total latency. The explicit inclusion of the uncertainty changes the nature of the optimization problem from deterministic to stochastic. In what follows, we introduce the main decision variables and constraints and we discuss how to deal with a stochastic objective function.

In order to derive the mathematical formulation, we consider again the expanded layered network as proposed in [194] for the K -TRP and presented in Section 2.4. For the sake of completeness, we report hereafter the main notation. Starting from the number of nodes n and vehicles K , we derive an upper bound N on the number of visited nodes in a path as $N = n - K + 1$. The layered network contains $N + 1$ levels, where level $N + 1$ includes a copy of depot, whereas all the other levels include a copy of depot and all demand nodes, except for the level 1, which has only copies of the demand nodes and

corresponds to the last nodes visited over each path. In this network, each path ends by visiting a node in level 1 and starts in a copy of the depot in some level. It is therefore easy to identify each path by starting from the last visited node at level 1, tracking the node connected to it at level 2, and repeating this process for any other subsequent levels until a copy of the depot is visited. We consider two sets of decision variables. The first one, denoted by x_i^r , for each node $i \in \bar{V}$ and visiting level r , take the value 1 iff node i is visited at level r , and 0 otherwise. The second set contains the binary variables y_{ij}^r , for each edge (ij) , that take the value 1 if edge (ij) belongs to a path for some vehicle and the total number of nodes in this path, after node i , is exactly r . The model contains the following constraints:

$$\sum_{r=1}^N x_i^r \leq 1 \quad i \in \bar{V}, \quad (3.1)$$

$$\sum_{i \in \bar{V}} x_i^1 = K \quad (3.2)$$

$$\sum_{r=1}^N \sum_{j \in \bar{V}} y_{0j}^r = K \quad (3.3)$$

$$\sum_{\substack{j \in \bar{V} \\ j \neq i}} y_{ij}^r = x_i^{r+1} \quad i \in \bar{V}, \quad r = 1, \dots, N-1 \quad (3.4)$$

$$y_{0j}^r + \sum_{\substack{i \in \bar{V} \\ i \neq j}} y_{ij}^r = x_j^r \quad j \in \bar{V}, \quad r = 1, \dots, N-1 \quad (3.5)$$

$$y_{0j}^N = x_j^N \quad j \in \bar{V} \quad (3.6)$$

$$\sum_{r=1}^N \sum_{i \in \bar{V}} \theta_i x_i^r \geq \Gamma \quad (3.7)$$

$$x_i^r \in \{0, 1\} \quad i \in \bar{V}, \quad r = 1, \dots, N \quad (3.8)$$

$$y_{ij}^r \geq 0 \quad i \in V, \quad j \in \bar{V}, \quad r = 1, \dots, N-1 \quad (3.9)$$

Constraints (3.1) require that each demand node is visited at most once. Constraint (3.2) ensures that each vehicle is assigned to exactly one demand node at the end of its tour (level 1). Constraint (3.3) guarantees the dispatch of exactly k vehicles from the depot. The set of constraints in (3.4)–(3.6) are the connectivity constraints and show the relation between the binary variables x_i^r and y_{ij}^r . Constraints (3.4) require that any node (i) visited at the upper level ($r+1$) should be connected to exactly one upcoming visited node (let say j) by traversing edge (ij) at the lower level (r). The set of constraints (3.5) impose that any node j visited at level r should be linked to

exactly one recently visited node (let say i) by traversing edge (i, j) or linked directly to the depot by traversing edge $(0, j)$ at the same level. Constraints (3.6) require that each node visited at the highest level N should be the first visited node over the path which is connected to the depot by traversing edge $(0, j)$ at the same level. The service level constraint is represented by (3.7). Affected areas should to be selected to guarantee the satisfaction of a minimum service level threshold, represented by the parameter Γ properly defined by the decision maker. Finally, constraints (3.8)–(3.9) indicate the nature of variables. We note that the y -variables can be defined as continuous because of the total unimodularity property of the matrix constraints.

The definition of the routing plans is carried with the aim of minimizing the stochastic total latency. In order to deal with this more involved objective function, the straightforward approach simply consists in replacing the random quantities with their expected values. It is evident that in a highly uncertain setting, how the one experienced after a dreadful earthquake, the “average view” of the system can be misleading. More prudentially, the decision maker would be interested in controlling the risk associated to possible long delays that may compromise the decisions. To account for this wiser attitude, we address the problem under a risk-averse perspective. As the analysis of the scientific literature (mainly appeared in the financial field) shows, there are several ways to capture and model risk. Interested readers are referred to the survey paper [147] for a general introduction of the main risk measures. Here, we have chosen an intuitive and effective way, suggested by the classical Markowitz theory ([163]). In particular, we measure risk in terms of standard deviation. Differently from the variance, this measure is expressed in the same unit of the expected arrival time. The proposed objective function presents a mean–risk structure:

$$\begin{aligned} \min : \quad Z = & \lambda \left(\sum_{j \in \bar{V}} \sum_{r=1}^N r \mu_{0j} y_{0j}^r + \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \sum_{r=1}^{N-1} r \mu_{ij} y_{ij}^r \right) \\ & + (1 - \lambda) \sqrt{ \sum_{j \in \bar{V}} \sum_{r=1}^N r^2 \sigma_{0j}^2 y_{0j}^r + \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \sum_{r=1}^{N-1} r^2 \sigma_{ij}^2 y_{ij}^r } \end{aligned} \quad (3.10)$$

where the first term accounts for the expected total latency, whereas the second one for the standard deviation. Both the terms can be derived by

applying the standard formula of the expected value and variance of the sum of independent random variables. The factor $\lambda \in (0, 1]$ in (3.10) is used to weight the importance attributed to the two terms. Obviously, the lower is λ the greater is the importance attributed by the decision maker to the risk. We notice that the same mean standard–deviation structure of the objective function, except for the multiplicative coefficients, appears when considering the quantile-related (Value at Risk) risk measures under the assumption that the random variables are elliptically distributed (see, for example, [60]). Under this respect, the proposed risk-averse formulation is rather general.

3.3.2 The case study

On the 12th of January 2010, an earthquake of magnitude 7.0 of the Richter scale hit Haiti producing devastating consequences: 230,000 people were killed, 300,000 injured and about 2,000,000 became homeless. Almost one third of the Haitian population was affected by the disaster. Only in Port-au-Prince, the capital city, the 15% of the population died or was injured. The earthquake caused massive economical losses and the damage amounted to 117% of Haiti’s annual economic output [160]. Because of the severity of the earthquake and the massive losses, the earthquake was classified as one of the worst disasters experienced and ranked as the fourth worst earthquake, in terms of the fatalities, since 1900 [47]. What makes the earthquakes different from other disasters is the fact that they are highly unpredictable and often occur suddenly without warning, but what makes them even worse is the chain of the disasters they cause.

The earthquake caused a partial destruction of the transportation network and damages to many infrastructures. The Port-au-Prince airport canceled all the flights for three days after the earthquake, and the seaport was closed for ten days [123]. After the earthquake, all the hospitals in Port-au-Prince and many other healthcare facilities, transportation and communication systems, collapsed or were severely damaged. The governmental administrative building was destroyed, causing the death of 17% of the employees [32].



Figure 3.3: Haiti earthquake, Encyclopædia Britannica, Inc.

Figure 3.3 shows the map of Haiti depicting the intensity of shaking of the earthquake. In Haiti, there are four levels of government below the national level, including departments, arrondissements, communes, and communal sections. The geographical distribution of the affected areas (in blue) and of the depot sited near to Port-au-Prince, (in white), is shown in Figure 3.4.

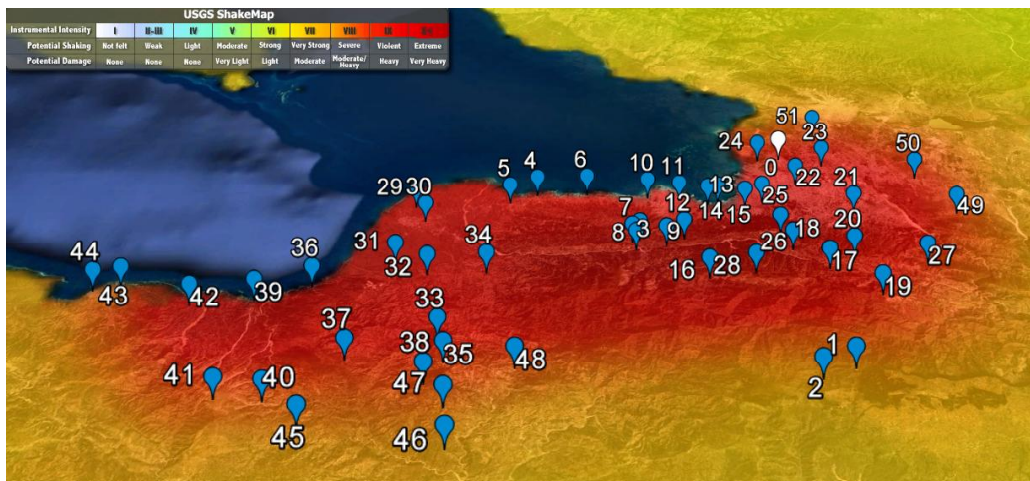


Figure 3.4: 51 communal sections with potential damage higher than moderate

The information about the severity of the earthquake experienced by each communal section has been taken from the shake intensity maps [1]. Among 571 communal sections, we have considered 51 sections within four classes

that we call severity levels (SLs, for short), associated with the instrumental intensity from *VII* to *X+* that caused a potential damage from *Moderate*, to *Very Heavy*.

The SL has been mapped into a numerical factor, ranging from 1 to 4 and the priority level θ_i of each node i has been determined by taking into account both the SL (multiplied 0.2) and the population living in the area (gathered from the 2015 national census reports [8]) normalized by the total population. The value of Γ used in the experiments has been set to 0.6.

Figure 3.5 depicts the geographical configuration of the affected areas in the Cartesian coordinate where the width of each point is proportional to its population (ranging between a minimum value of 504 to a maximum of 531,434), and the colour represents the SL of the area based on the colour bar scale.

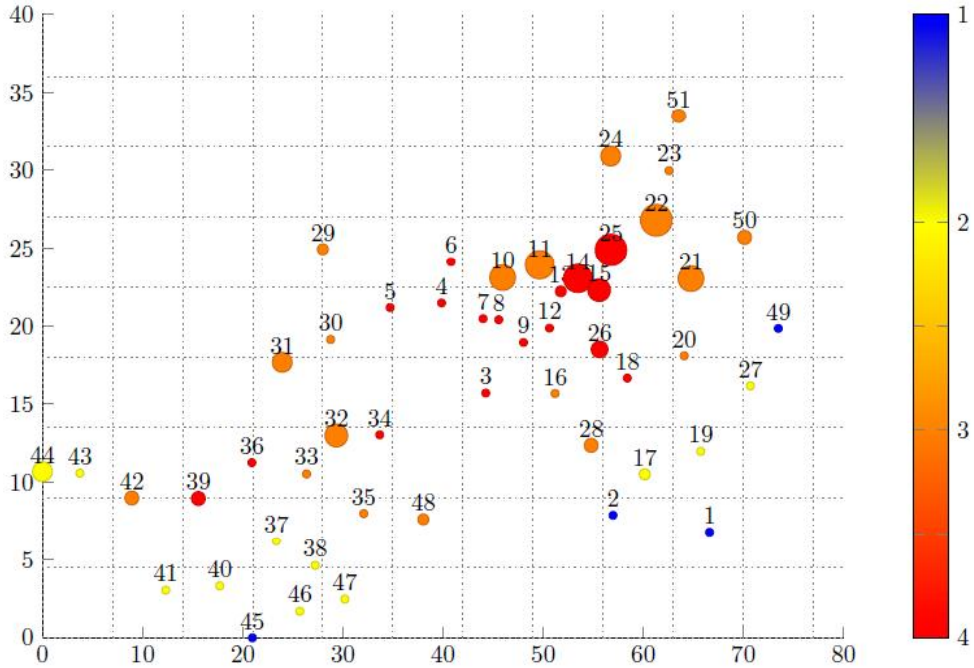


Figure 3.5: Classification of the affected areas based on population and the severity level

The travel distance along each pair of points has been calculated on the basis of the real-road network using the Google Map API [2]. Starting from the travel distances, we have computed the travel times by considering an average speed of $30km/h$. The variances σ_{ij}^2 have been computed as $[(\mu_{ij} b)^2]$,

where b is a random number uniformly distributed in the interval $[0.1, 0.32)$. In our experiments we have considered a fleet of 4 vehicles.

The next parts are devoted to the presentation of the numerical results collected on the considered real case study, with the aim of demonstrating how the proposed model can be used to support the routing planning decisions after a dramatic disaster. All the experiments have been performed on an Intel® Core™ i7 2.90 GHz, with 8.0 GB of RAM memory, running under Windows operating system. The proposed mathematical model and the heuristic were coded in C++ and the model was solved by the open source SCIP library, release 3.2.0.

3.3.3 Performance of the model

We present the results of a sensitivity analysis of the proposed formulation for the case of Haiti earthquake, as described in 3.3.2, with respect to some key parameters, namely, the service level and the fleet size. The final aim is to provide the decision maker with useful insights on resource planning, for example, how much the total latency will decrease if more vehicles are available and how the variation of the threshold Γ affects the selection of the areas to be serviced.

3.3.3.1 The impact of the risk

After an earthquake, there is usually a high level of uncertainty in the transportation network. In Haiti, apart from the damages produced by the earthquake to the road-network, the presence of rubble, even on the main roads, was a critical consequence, increasing the congestion level of the overall network [3]. In such a critical context, considering the travel time variations represents an important issue to address, since the explicit inclusion of the risk component can lead to different routing plans. More specifically, the lower the λ value, the more emphasis is put on the risk component. To mathematically evaluate the impact of the risk, we have performed a first set of experiments by varying the value of the parameter λ between 0.1 and 0.5, to weight the total expected arrival time (*EAT* for short), and its standard deviation (*STDAT*). Higher values of λ have not been reported since they do not reflect a risk-averse behaviour.

The following Figure 3.6 shows the total expected arrival time versus its standard deviation as function of λ . As expected, lower values of λ provide more stable routing plans at the expense of an increase on the expected total arrival time. The opposite happens for higher values attributed to the weighting parameter. The trade-off is favorable for the extreme values used in the experiments: we may observe a decrease of almost 27% in the standard deviation, when a greater importance is attributed to the risk component, with only an increase of around 3% of the expected arrival time. These results confirm the importance of explicitly accounting for risk in the model formulation.

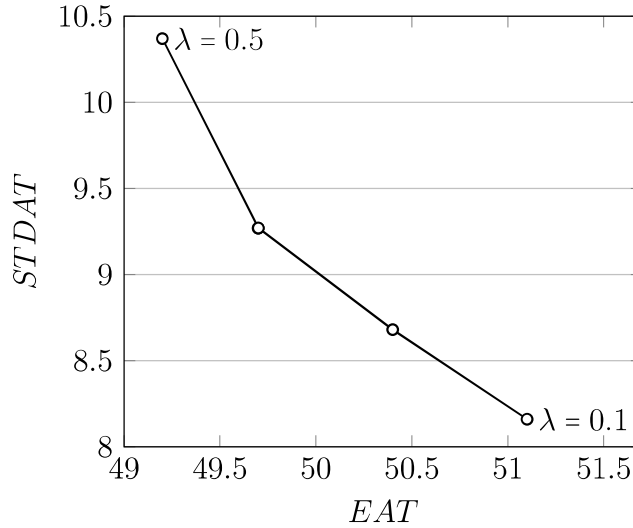


Figure 3.6: EAT versus $STDAT$

Analyzing the results in more details, we may appreciate how the risk aversion impacts on the selection of the visited areas. The following Table 3.3 shows the routes traveled by the vehicles and the total number of visited nodes (# Nodes) for different values of λ .

λ	0.1	0.2	0.3	0.4	0.5
# Nodes	20	20	21	21	21
Route 1	0-22-25-15-14-11-29-32	0-22-25-15-14-29	0-22-25-15-14-32	0-22-25-15-14-32	0-22-25-15-32
Route 2	0-24-13-10-31-39-44	0-24-13-11-10-31-32-39-44	0-24-13-11-10-4-31-39-42	0-24-13-11-10-4-31-39-42	0-24-14-13-11-10-4-31-39-42
Route 3	0-23-21-26-18-28	0-23-21-26-18-28	0-23-21-26-18-17-28	0-23-21-26-18-17-28	0-23-21-26-18-17-28
Route 4	0-51-50	0-51-50	0-51-50	0-51-50	0-51-50

Table 3.3: Sensitivity of the solution with respect to λ

We observe that, for all the λ values, the determined routes share a given set of nodes, namely $\{4, 10, 11, 13, 14,$

15, 17, 18, 21, 22, 23, 24, 25, 26, 28, 29, 31, 32, 39, 42, 44, 50, 51}. These are priority areas because of the number of inhabitants and the SL. Among them, the most populated area is represented by the node 25 (Turgeau in Port-au-Prince) with a population of more than half million. The least populated ones are the areas 4 and 18 with, respectively, 19,824 and 23,376 inhabitants. All the visited areas (with the exception of 17 and 44) have a SL at least 3 (areas 4, 13, 14, 15, 18, 25, 26 have a SL=4). It is interesting to note that other areas have SL=4, but they are not visited since their population is lower than 5000 people.

For the case with λ equal to 0.1, 98% of the total population affected by the earthquake with SL=4 is visited, which means that only 2% of population with the highest SL do not receive aid during the first response phase. The set of critical areas with the SL=3 receiving relief aids, (10, 11, 21, 22, 23, 24, 28, 29, 31, 32, 42, 50, 51), represent about 96% of the whole population in need of humanitarian aid.

We observe that, when increasing λ from 0.1 to 0.2, only the order of visiting the affected areas of 11 and 32 changes resulting into a slight improvement in the total expected arrival time (about 1%) with a deterioration of its standard deviation of about 6%. The number of visited areas for λ values greater than 0.3 increases slightly. In the case of λ equal to 0.3, despite visiting more affected areas, the total expected arrival time decreases, which is a valuable insight on the non-monotonic behavior of the total expected latency with respect to the number of visited nodes. Again, the results obtained with λ equal to 0.4 and 0.5 show that, although the same set of areas is visited, the area 14 is serviced by a different vehicle, leading to an improvement in the total expected arrival time and a deterioration of the standard deviation.

3.3.3.2 The fleet size

The results reported hereafter show how the fleet size k impacts the system performance expressed in terms of visited nodes and total latency. In particular, Table 3.4 reports the values collected by considering an Γ value equal to 0.6 and a λ value equal to 0.1, as function of different k values.

k	EAT	$STDAT$	$\#$ Nodes	Routes
4	51.1	8.16	20	0-22-25-15-14-11- 29 -32 0-24-13-10-31-39- 44 0-23-21-26-18-28 0-51-50
6	49	6.27	21	0-22-26-18- 17 -28 0-24-10-31-32 0-23-21 0-13- 4 - 42 0-25-15-14-11-39 0-51-50
7	48	6.00	21	0-22-26-18-28 0-24-10-31-39 0-13-11- 8 0-23-21- 17 0-14-32 0-25-15- 42 0-51-50
8	47.4	5.90	21	0-22-26-18- 17 -28 0-24-10-31-39 0-13-11- 8 0-23-50 0-21 0-14-32 0-25-15- 42 0-51

Table 3.4: Sensitivity of solution with respect to fleet size k

As confirmed by the results, the increase of the fleet size allows to reduce both the total expected arrival times and its standard deviation, since adding more vehicles provides more balanced routing plans where the visited nodes are split among new routes. This allows, in turn, to reach the affected areas sooner by traversing closer links from the depot or through other intermediate nodes, thus improving the total latency. For instance, in the case with 4 vehicles, the visit of the affected areas 25 and 13 is delayed to first visit areas 22 and 24, respectively, but with at least 6 vehicles, they can be visited directly from the depot. In terms of the number of visited areas, in case of 4 vehicles, the difference between the longest and the shortest routes is 5, whilst with 7 vehicles, this difference decreases to 2. We also note that the increase in the fleet size does not have a significant impact on the total number of visited nodes, which is rather related to the service level constraint, as will be shown in the next paragraph.

3.3.3.3 The service level

Another set of experiments has been carried out to investigate the selective nature of the proposed formulation. To this aim, by keeping fixed all the other parameters, we have performed additional experiments by varying the minimum service level Γ between 0.45 to 0.60.

Γ	% <i>EAT</i>	% <i>STAT</i>	<i>Visited areas</i>	<i>Routes</i>
0.55	49.31	42.57	{10, 11, 14, 15, 21, 22, 23, 24, 25}	0-22-25-15-14-32 0-24-13-11-10-31 0-23-21-26 0-51-50
0.5	69.27	57.12	{10, 11, 13, 14, 15, 21, 22, 23, 24, 25, 26, 31, 32, 50, 51}	0-25-15-14 0-24-13-11-10 0-22-21-26 0-51-50
0.45	81.10	72.0	{10, 11, 14, 15, 21, 22, 23, 24, 25}	0-22-25-15 0-24-10 0-23-21 0-14-11

Table 3.5: Sensitivity analysis as function of Γ

Table 3.5 reports the results collected for λ equal to 0.1 (similar results have been obtained also for the other values of λ). In particular, we report the percentage of reduction of the *EAT* and the *STDAT* for different Γ values evaluated with respect to the solution obtained for $\Gamma = 0.6$. An indication of the visited areas is also provided. As expected, the lower the imposed service level, the lower the *EAT* and the *STDAT*, since less nodes are required to be visited. Looking at the results, we may observe that with an increase of 15% in the minimum service level Γ , i.e. passing from 0.45 to 0.60, we can provide relief aid for 12 other affected areas, including a population of more than half million and covering the 98% of communes with SL equal to 4. This is justified by these communes are clustered around the depot. The situation for the areas with SL 3 is a bit different, since, as evident from Figures 3.4 and 3.5, they are scattered over the region within a larger radius from the depot. This implies that for $\Gamma = 0.45$, only 6 critical areas (10, 11, 21, 22, 23, 24) are visited, while other more populated areas, with the same SL, are not. In general, for small Γ values, it is quite probable that an attractive area with high population and high SL level is not visited due to its long distance from the depot. When the service level increases, more

communes with the same SL but less populated than those visited previously receive aid.

3.3.3.4 Model validation

In order to validate the model, we have first investigated the importance of using real data in the definition of the test case. In particular, we have compared the routing plans obtained by considering the real-road network with the ones obtained when the travel distance over each edge (i, j) is calculated as the Euclidean distance between affected areas i and j . We should remind that the real travel distance between each pair of nodes was calculated as the shortest distance provided by Google Map API [2], and thus, many factors such as the presence of physical obstacles (mountains, valleys, and rivers), cost, environmental impact of the road, are taken into account.

	λ	EAT	$STDAT$	<i>Routes</i>
Real	0.1	51.1	8.16	0-22-25-15-14-11-29-32
				0-24-13-10-31-39-44
				0-23-21-26-18-28
	0.5	49.2	10.37	0-51-50
				0-22-25-15-32
				0-24-14-13-11-10-4-31-39-42
Euclidean	0.1	36.9	7.53	0-23-21-26-18-17-28
				0-51-50
				0-22-26-18-28
	0.5	35.5	6.68	0-24-25-15-31
				0-23-51-50-21
				0-14-13-11-10-32-39-44
				0-22-26-18-28
				0-24-13-8-32-31-39
				0-25-15-14-11-10-29
				0-51-23-21-50

Table 3.6: Real-road network versus the Euclidean-based network

Table 3.6 shows the EAT and the $STDAT$ for Γ of 0.60 and λ equal to 0.1 and 0.5.

Looking at the results, we may appreciate how the Euclidean distance leads to an underestimation of the total latency (about 38% for the EAT and 55% for the $STDAT$). Moreover, as far as the number of visited nodes is concerned, the Euclidean case provides dominated solutions, with less nodes visited. This latter finding shows the importance of incorporating realistic

input data in order to obtain meaningful results which are also reliable in practice, and supports and validates the present case study, solved on a real-road network.

3.3.3.5 The analysis under a disruption scenario: the occurrence of a tsunami

To analyze the impact of disruption in the definition of the routing plans, we have defined two operational scenarios where some road segments, linking a subset of visited areas, are disrupted. The first scenario copes with the impact of a tsunami. In 2010, two of independent tsunamis hit the shores after the earthquake, one along the south coast near Jacmel and the other one along the Bay of Port-au-Prince (see Figure 3.7).

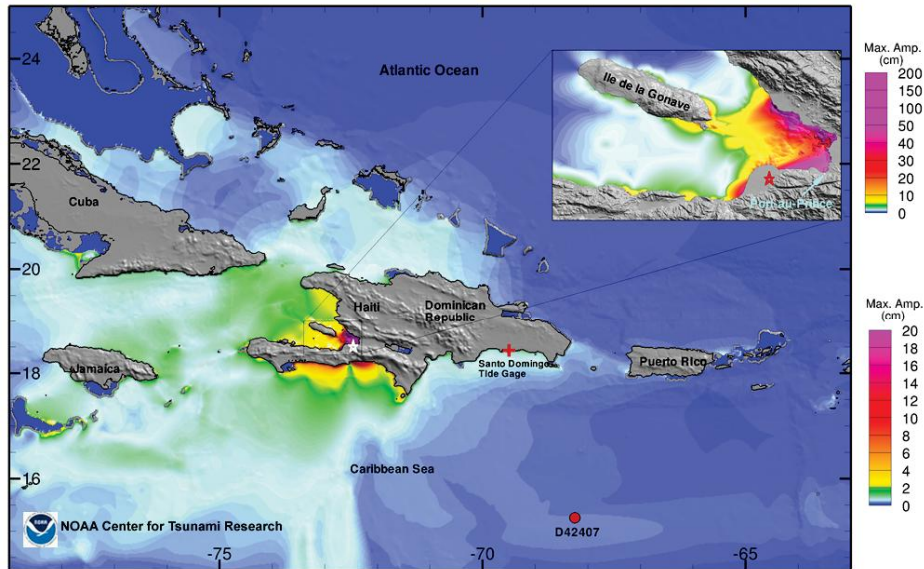


Figure 3.7: Tsunami event-NOAA Center for Tsunami Research

In this scenario it is assumed that some parts of road network near the shore are disrupted by the tsunami and some links are not any more traversable. Figure 3.8 shows the broken links in yellow. Looking at the routes for the ideal (non disrupted) case, for λ equal to 0.5 in Table 3.3, we observe that they involve two links between communes 10,11 and 13,14 which in this scenario cannot be traversed. Hence, without an adequate planning, not only areas 10,11,13 will not receive relief aid, but also all the

subsequent areas, including about 22% (688,028 people) of satisfied demands over all routes.

The model solved under this scenario re-routes the available vehicles to enable the relief operations. The optimal routes under the tsunami scenario are reported below:

- **Route 1** (0-22-25-15-14-11-32)
- **Route 2** (0-24-13-10-4-31-39-42)
- **Route 3** (0-23-21-26-18-17-28)
- **Route 4** (0-51-50)

Figure 3.9 shows the changes in the order of visiting nodes for both cases. The third and the fourth routes under the tsunami scenario are the same as those reported for the ideal case in Table 3.3. The first and the second routes are instead modified. In fact, under the disruption, the affected areas 10, 11, 13 cannot be reached from node 14. Hence, nodes 11, 14 are relocated and served after area 15 in the first route. In addition, areas 10, 13 are visited by traversing longer non disrupted links from the area 24.

In the modified solution, the expected latency of areas 11 and 14 increases by 36% and 45%, respectively. The increase in the individual expected latency values contributes to a 2% deterioration of the total expected latency, with respect to the ideal case.

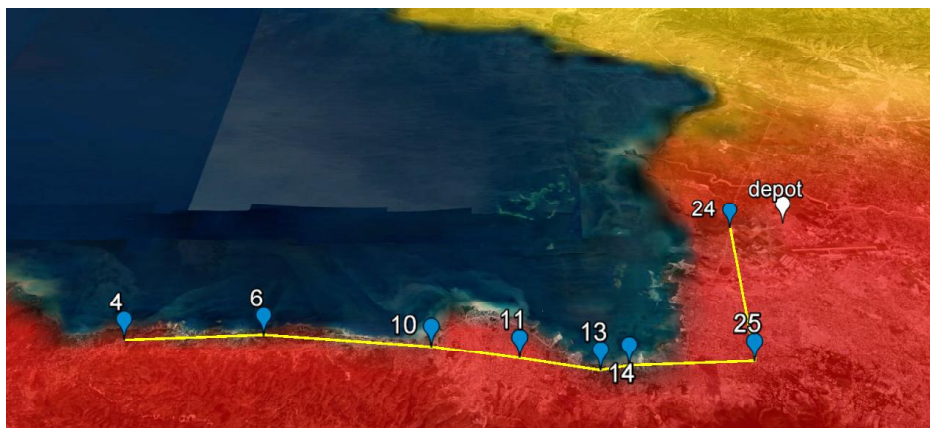
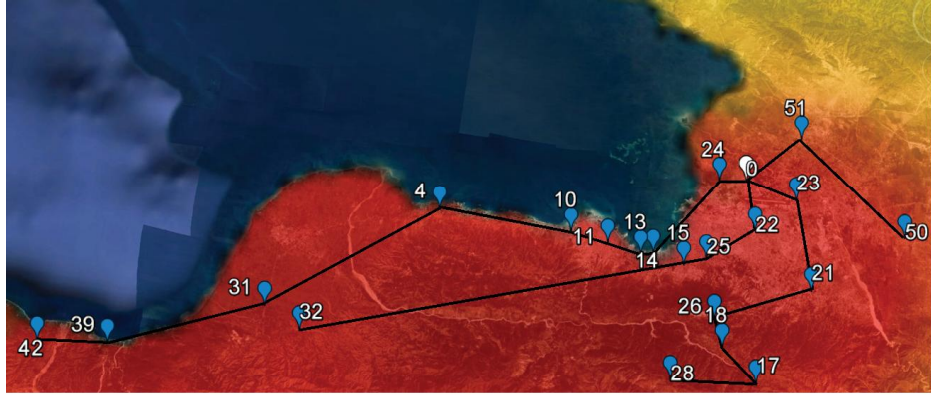
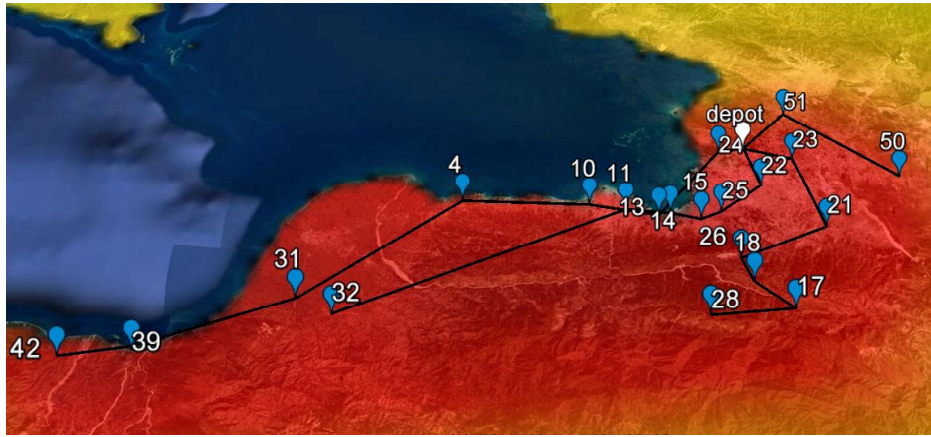


Figure 3.8: The broken links under the tsunami scenario



(a)



(b)

Figure 3.9: Optimal paths:
a) Ideal case b) Disrupted case under the tsunami scenario

3.3.3.6 The analysis under a disruption scenario: the failure of the depot

The analysis of the results from the previous scenario and the ideal case are based on the assumption that the transportation network surrounding the depot is fully reliable.

In order to depict another critical scenario, we have considered the disruption over the links connecting the depot to the critical areas within a radius of five kilometers (22, 23, and 24). In addition, we have also considered the disruption for road links emanating from the serviced critical areas which are within ten kilometers from the depot (21, 25, 51). In summary, we have considered the following disrupted links $\{(0, 22), (0, 23), (0, 24), (50, 51), (22, 25),$

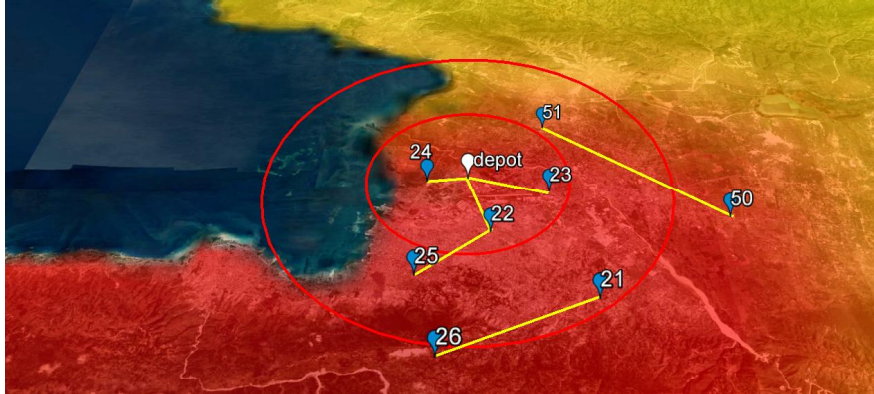


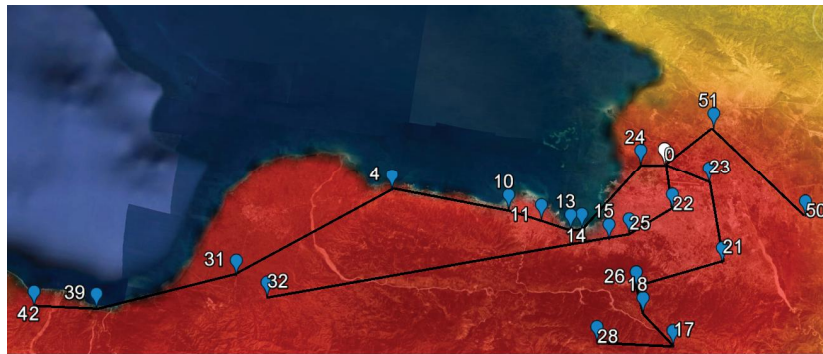
Figure 3.10: Disrupted links under the depot failure scenario

(21, 26)} (see Figure 3.10, where the broken links are highlighted in yellow).

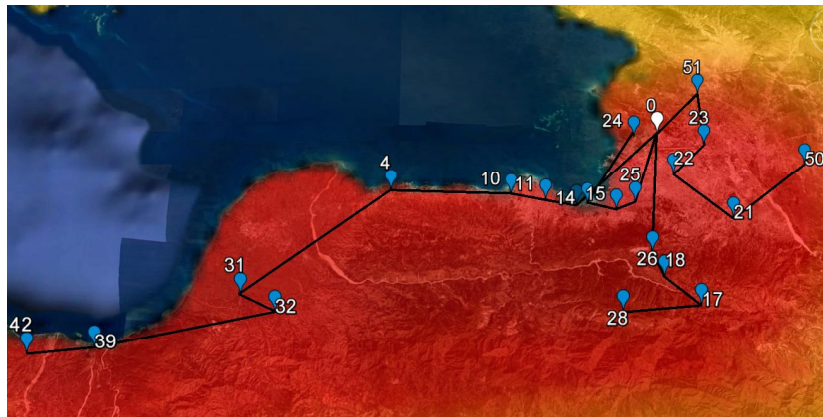
Looking at the results, collected by considering λ equal to 0.5, we may observe that the set of visited nodes is the same of those found in the normal case and that the model is able to rearrange the routing plans. Obviously, since all the disrupted links were part of the routes in the ideal case, an increase in the objective function value is registered. In particular, we have a deterioration of 12% and 4% in the expected latency and its standard deviation, respectively. The optimal routes in this case are as follows:

- **Route 1** (0-25-15-14-24)
- **Route 2** (0-13-11-10-4-31-32-39-42)
- **Route 3** (0-26-18-17-28)
- **Route 4** (0-51-23-22-21-50)

Figure 3.11 shows the optimal routes for both cases on the map. The disruption delays the visit of the affected areas 22, 23, 24 that cannot be visited anymore as the first nodes. More specifically, in order to visit the area 24, a longer path should be taken traveling through the areas 25 and 15, which is an alternative, but much longer way, of reaching the area 24. In a similar way, areas 22, 23 are visited via the longer path passing through the critical area 51 and also area 50 is visited by traversing its alternative path passing through areas 23, 22, 21, which is much longer. These alternative paths are, in fact, safe in the sense that they are not disrupted, but, obviously,



(a)



(b)

Figure 3.11: Optimal paths:
 a) Ideal case b) Disrupted case under the depot failure scenario

their expected latency and standard deviation are higher. This disruption scenario also affects the visit of other visited areas due to the rearrangement of visited nodes over the paths, for example, the latency of area 14 increases and the visit of areas 39 and 42 is delayed, since before them, area 32 should be served.

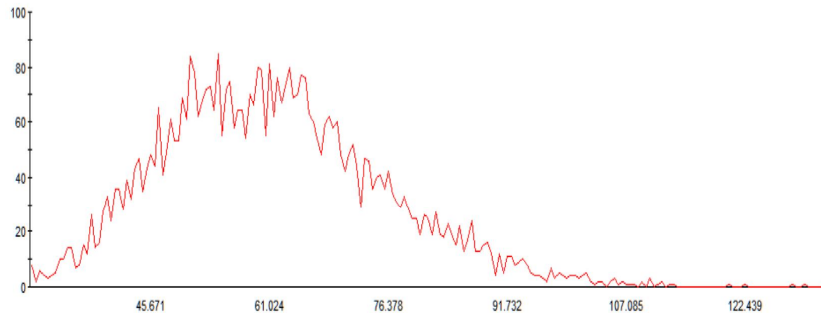
3.3.3.7 The Monte Carlo simulation

Additional experiments have been carried out with the aim of assessing the behaviour of the solutions provided by the stochastic model. To this end, we have evaluated how the stochastic solution behaves with respect to the deterministic one over a large set of possible scenarios. In particular, we have performed a Monte Carlo simulation generating, for each link, 5,000 different scenarios, taken from a normal distribution, representing different travel times values. For each scenario s , the total latency has been calculated according to the following formula:

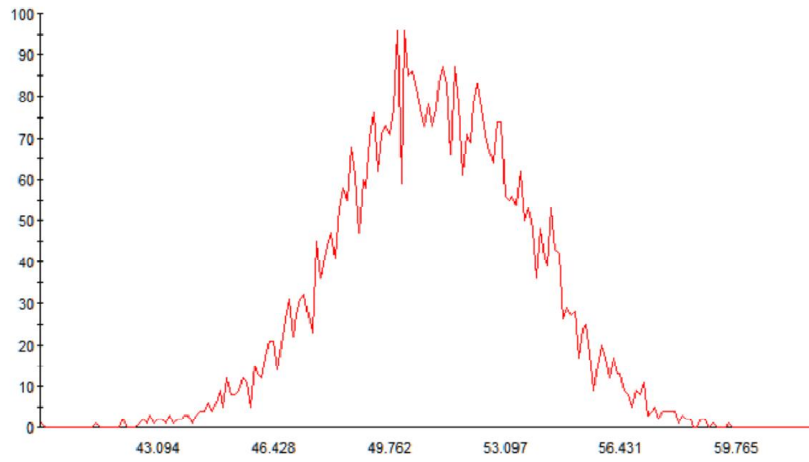
$$TL^s = \sum_{j \in \bar{V}} \sum_{r=1}^N r t_{0j}^s \bar{y}_{0j}^r + \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} \sum_{r=1}^{N-1} r t_{ij}^s \bar{y}_{ij}^r$$

where $\bar{\mathbf{y}}$ denotes the optimal solution and t_{ij}^s is the travel time over link the (i, j) under scenario s . We point out that the deterministic solution has been determined by solving the proposed model with λ equal to 1, whereas the stochastic solution refers to the case with λ equal to 0.1.

The following Figure 3.12 reports the frequency histograms for the simulated total latency of the proposed model and the deterministic one.



(a)



(b)

Figure 3.12: Frequency histograms:
a) Deterministic model b) Proposed model

The simulation results show that the solution of the deterministic case is highly unstable with high variations in the travel time (the long tail in the histogram of the deterministic solution underlines this unstable behavior).

We should note that the deterministic model ignores the variance completely and it is reasonable to expect high standard deviation values. In general, we should expect that, in terms of the total expected latency, the deterministic case provides better results than the proposed model. We have observed that the simulated values of the TL for the deterministic model are 21% larger than the corresponding values for the stochastic model.

	<i>Mean</i>	<i>Median</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
Proposed Model	51.074	51.083	59.580	39.860	8.067
Deterministic Model	61.870	60.776	133.186	27.104	42.846

Table 3.7: Simulation results

This consideration is also supported by the synthetic values reported in Table 3.7, namely, the arithmetic mean, the median, the max, and min values, and the standard deviation (SD), respectively. These values confirm that the results of the simulation are consistent with those provided by the proposed formulation.

To conclude, the proposed model provides realistic estimations for both the total expected latency and its variance, which are confirmed by the simulation results as well.

3.3.4 The heuristic approach

In terms of computational tractability, we have empirically found that SCIP was not able to optimally solve instances with more than 75 demand nodes in a reasonable amount of time. To be more precise, SCIP even failed to instantiate the model when more than 90 nodes are considered. These considerations have motivated the design of a heuristic approach exploiting the specific problem structure.

The proposed heuristic is a Variable Neighborhood Search (VNS) which systematically changes the size of the neighborhood in the attempt of escaping from a local minimum trap, as mentioned in [120]. It is composed of a construction phase providing an initial feasible solution, with respect to the set of constraints (3.1)–(3.9), which is then improved through a local search procedure. The pseudocode is sketched in Algorithm 7.

Algorithm 7: The proposed heuristic

```
1 Input:  $T_{Max}, \Delta_{Max}, \delta$ 
2 Initialization:  $t \leftarrow 1, s \leftarrow null$ 
3  $s \leftarrow Initial\ solution()$ 
4 while ( $t < T_{Max}$ ) do
5      $\nu \leftarrow 1$ 
6     while ( $\nu < \Delta_{Max}$ ) do
7          $s' \leftarrow Perturbation(s, \delta)$ 
8          $s'' \leftarrow Local\ Search(s', \nu)$ 
9         if ( $Z(s'') < Z(s)$ ) then
10             $s \leftarrow s''$ 
11             $\nu \leftarrow 1$ 
12        else
13             $\nu \leftarrow \nu + 1$ 
14    end
15     $t \leftarrow t + 1$ 
16 end
17 return  $s$ 
```

First, the *Initial solution()* function is called (line 3 in Algorithm 7) generating an initial feasible solution s which is specified by the set of visited nodes Π and distinct paths $\rho, \rho = 1, \dots, K$. At the beginning (see Algorithm 8), the set of visited nodes Π and unvisited nodes Ψ , ($\Psi = \bar{V}/\Pi$) are initialized, then each path $\rho, \rho = 1, \dots, k$ is filled by inserting a single node $c \in \Psi$ with the lowest value of $\lambda\mu_{0i} + (1 - \lambda)\sigma_{0i}, i \in \Psi$. The sets Π and Ψ are updated and the new path ρ is added to the solution. Next, if necessary, s is modified by adding more nodes in order to assure that the service level constraint (3.7) is satisfied. The modification is done by inserting exactly one node $c \in \Psi$, at each time, to each path ρ , based on the same criterion mentioned before, until the feasibility is achieved. By inserting just one node, at each time, to each path, we ensure that the generated paths are balanced. The initial solution is fed in Algorithm 7 that basically consists in an outer loop executed for a certain number of iterations T_{Max} (or equivalently for a pre-specified *CPU time*) and an inner loop, the core of heuristic, which includes the *Perturbation* and the *Local search* procedures.

The *Perturbation* (s, δ) function, playing the role of *shaking* mechanism in the VNS heuristic, is explained in detail in Algorithm 9. This procedure deletes a pre-specified number of visited nodes (let us say δ) from the current

solution and replaces them with a subset of unvisited nodes in Ψ provided that the feasibility with respect to the service level constraint (3.7) is kept. The set of deleted nodes are then put into a *blacklist* which is updated at the beginning of the procedure. The output of the *Perturbation procedure* (s, δ) , s' , is then considered as a reference solution to perform the local search around. The local search is performed over Δ_{Max} different neighborhoods indexed by ν . The local search over each neighborhood ν is performed based on the best improvement criterion and in the hope of finding an improved solution s'' . If the obtained solution s'' is better than s , the current solution is updated, and the index of the neighborhood to be searched over the next iteration is set to its initial value ($\nu = 1$); otherwise, another neighborhood is explored (ν is incremented by one). The inner loop terminates whenever none of the neighborhoods could improve the reference solution, or equivalently $\nu = \Delta_{Max}$. The algorithm used explores five different neighborhoods, specified by their move operators and classified as intra-route and inter-route moves. The intra-route neighborhoods include the *Swap*, the *2-opt*, and the *Or-opt*; the inter-route neighborhoods used are the *Relocation* and the *Exchange*.

The *Swap* move operator exchanges the position of two visited nodes over a single path. In the *2-opt* move, a pair of non-adjacent edges are removed from the path and replaced with two new edges reconnecting the path. The *Or-opt* move operator is similar to *2-opt*, but a triple of non-adjacent edges are removed and reconstructed. The *Exchange* move exchanges a pair of nodes between two different paths while *Relocation* move deletes a node from a path and inserts it to another path.

We should note that the *blacklist* is updated each time over each iteration and the nodes in the *blacklist* cannot be visited until the next perturbation

mechanism is called.

Algorithm 8: Initial solution

```

1 Initialization:  $\Pi \leftarrow \emptyset, \Psi \leftarrow \bar{V}$ 
2 for ( $\rho \leftarrow 1$  to  $K$ ) do
3    $c \leftarrow \arg \min_{i \in \Psi} (\lambda \mu_{0i} + (1 - \lambda) \sigma_{0i})$ 
4   Insert  $c$  in route  $\rho$ 
5    $\Pi \leftarrow \Pi \cup \{c\}, \Psi \leftarrow \Psi - \{c\}$ 
6    $s \leftarrow (\rho, \Pi)$ 
7 end
8 if ( $s$  is not a feasible solution with respect to constraint (3.7)) then
9    $temp \leftarrow 0$ 
10  while ( $temp == 0$ ) do
11    for ( $\rho \leftarrow 1$  to  $K$ ) do
12       $c \leftarrow \arg \min_{i \in \Psi} (\lambda \mu_{0i} + (1 - \lambda) \sigma_{0i})$ 
13      Insert  $c$  in route  $\rho$ 
14       $\Pi \leftarrow \Pi \cup \{c\}, \Psi \leftarrow \Psi - \{c\}$ 
15       $s \leftarrow (\rho, \Pi)$ 
16      if
17        ( $s$  is a feasible solution with respect to constraint (3.7))
18        then
19           $temp \leftarrow 1$ 
20          break
21      end
22    end
23  end
24  return  $s$ 

```

Algorithm 9: Perturbation (s, δ)

```

1 Initialization:  $t \leftarrow 1, blacklist \leftarrow \emptyset$ 
2 while ( $|blacklist| \leq \delta$ ) do
3   choose randomly two nodes  $i \in \Pi$  and  $j \in \Psi, j \notin blacklist$ 
4   such that if  $i$  is replaced by  $j$ , the feasibility is hold
5    $blacklist \leftarrow blacklist \cup \{i\}$ 
6   replace  $i$  by  $j$ 
7    $\Pi \leftarrow \Pi \cup \{j\}, \Pi \leftarrow \Pi - \{i\}$ 
8    $\Psi \leftarrow \Psi \cup \{i\}, \Psi \leftarrow \Psi - \{j\}$ 
9 end
10  $s \leftarrow (\rho, \Pi)$ 
11 return  $s$ 

```

The proposed heuristic approach was used to solve a set of large instances

all derived from the real data set, where more affected areas have been considered in the set \bar{V} . In particular, we have generated 4 more instances, with a number of nodes ranging from 75 to 90. The other parameters, namely the fleet size and the threshold, are the same of the other experiments. A time limit of 1200 seconds has been set both for SCIP and the heuristic method ($T_{Max} = 1200$). Moreover, Δ_{Max} has been set to 5 and $\delta = 10\%|\Pi|$.

λ	Nodes	$Gap_{LB}(\%)$
0.1	75	80
	80	64
	85	65
	90	52
0.3	75	28
	80	18
	85	19
	90	53
0.5	75	17
	80	13
	85	10
	90	10

Table 3.8: Results for larger cases: proposed heuristic versus SCIP

Table 3.8 presents the results obtained. SCIP was not able to find a feasible solution in all but two instances. In Column $Gap_{LB}(\%)$ we have reported the gap of the solution obtained by the heuristic with respect to the lower bound obtained by SCIP when the nonlinear function is replaced by a polyhedral outer approximation. When SCIP was able to provide a solution, (i.e., for 80 and 90 nodes with $\lambda = 0.3$) its solution was worse than the solution provided by the heuristic, respectively, by 92% and 45%. The gap is quite high for $\lambda = 0.1$. This might be due to the fact that the outer approximation is not tight when more weight is put on the non-linear part of the model. For the other values of λ the gaps are quite good, considering also the inability for SCIP to find a good solution, if any.

3.3.5 Conclusions and future research directions

In this Chapter, we have proposed a tool for supporting the decision process in the post-disaster planning phase. In particular, we have proposed a customer-centric vehicle routing problem that minimizes the total arrival time, acknowledging in this way the criterion of quick responsiveness of the relief phase. The model has a selective structure to be more consistent with the characteristics of short-time relief routing efforts, especially over the first 72-hours after the disaster, when only a subset of affected areas, based on their urgency levels, can receive humanitarian aid. The selective criterion is based on the population of the affected areas as well as their urgency levels. As a main contribution, the uncertainty of travel times was injected into the problem and a risk-averse approach was considered.

The computational experiments provided interesting managerial insights about the importance of adopting risk-averse policies, when the uncertainty of travel times is incorporated. We also presented also a heuristic able to solve large scale instances in reasonable time.

Future research could involve the incorporation of correlation among the travel times in the network, especially when network disruptions are more probable. The extension of the proposed model to a location-routing model integrating both pre-disaster and post-disaster decisions in relief logistic could be an interesting future research stream.

Chapter 4

Asymmetric latency-based vehicle routing problems with uncertain travel and service times

In this Chapter, we will present a distributionally robust methodology for an extended version of the problem earlier studied in Chapter 3. In particular, we study the asymmetric latency-based VRP with service level constraint under uncertainty, considering also the presence of the service time. We present a metaheuristic approach based on a large neighborhood search and variable neighborhood descent heuristic. We will use the distributionally robust optimization approach as an alternative to incorporate the uncertainty of random parameters (in our case, travel and service times). A large family of distribution functions is considered since the random parameters may belong to any ambiguous distribution with known first and second moments and with a non-negative support of the distributional set. In Section 4.1, we review some preliminaries related to the distributionally robust approach and also we present the new version of the latency-based routing problem for which we derive an equivalent deterministic model. Section 4.2 introduces a metaheuristic methodology to solve the problem.

Although we present the mathematical model in the routing context, this study also contributes to the machine scheduling literature, presenting a multi-machine scheduling model with a selective structure in which the uncertainty of both the setup and processing times are taken into account.

Beside the model is interesting in itself, it represents also a new approach for dealing with the deterministic selective machine scheduling problem.

In Section 4.3, we present the computational results applying the robust model in the machine scheduling context, and we provide an extensive set of experiments performed on instances adapted from the benchmark testbed. Moreover, we also show the efficiency of our formulation, considering deterministic parameters, when compared with the deterministic model present in the literature.

4.1 The generalized robust model

In this section, we first present some preliminaries on robust optimization (RO). The RO models do not require specification of the exact distribution of the exogenous uncertainties of the model. This is the general distinction between the approaches of robust optimization and stochastic programming toward modeling problems with uncertainties. In the framework of robust optimization, uncertainties are usually modeled as random variables with true distributions that are unknown to the modeler, but are constrained to lie within a known support. The uncertainty set can be selected as a continuous interval or a finite set of different values. In this latter setting, the problem of interest is in general the optimization of the performance in the worst case scenario. Three criteria have been introduced in the literature: absolute robustness, robust deviation and relative robust deviation. Absolute robustness considers minimizing the objective value of the worst case directly. Robust deviation (or absolute regret) minimizes the largest possible difference between the observed objective value and the optimal one, while relative robust deviation (or relative regret) deals with the ratio of the largest possible observed value to the optimal value. In the continuous case, the uncertainty sets are selected as continuous intervals. Under this assumption, the expected solution performance is typically optimized. However, this criterion assumes that the decision maker is risk-neutral and leads to solutions that may be questionable. In this case, the decision maker attitude towards a risk should be taken into account. A criterion called Conditional Value-at-Risk (CVaR), early applied to a stochastic portfolio selection problem, can be used. Using this criterion, the decision maker provides a parameter

α which reflects his attitude towards a risk. When $\alpha = 0$, then CVaR becomes the expectation but for greater values, more attention is paid to the worst outcomes, which fits into the robust optimization framework. In this section, we consider the worst-case CVaR in situation where the information on the underlying probability distribution is not exactly known. In fact, typically, the first- and second-order moments of the uncertain parameters may be known, but it is unlikely to have complete information about their distributions.

Let assume that the uncertain travel \tilde{t}_{ij} and service time $\tilde{\tau}_j$ are defined by random vectors \mathbf{t} and $\boldsymbol{\tau}$, respectively, We investigate a specific case where the ambiguity set is determined by the mean and covariance and the distributional set is a semi-infinite support set. Let $\mathbb{P}_{\mathbf{t}}$ and $\mathbb{P}_{\boldsymbol{\tau}}$ be the ambiguous distribution of random vectors \mathbf{t} and $\boldsymbol{\tau}$, respectively, which are described by their first and second moments as follows:

$$\mathbb{P}_{\mathbf{t}} = \{ \mathbb{P}^{\mathbf{t}} | \text{Sup}(\tilde{t}_{ij}) = [0, \infty), \forall (i, j) \in V \times \bar{V}, E(\tilde{t}_{ij}) = \mu_{\tilde{t}_{ij}}, \text{Var}(\tilde{t}_{ij}) = \sigma_{\tilde{t}_{ij}}^2 \} \quad (4.1)$$

$$\mathbb{P}_{\boldsymbol{\tau}} = \{ \mathbb{P}^{\boldsymbol{\tau}} | \text{Sup}(\tilde{\tau}_i) = [0, \infty), \forall i \in \bar{V}, E(\tilde{\tau}_i) = \mu_{\tilde{\tau}_i}, \text{Var}(\tilde{\tau}_i) = \sigma_{\tilde{\tau}_i}^2 \} \quad (4.2)$$

Following the risk-averse approach, we apply the *CVaR* risk measure at a given confidence level $\alpha \in (0, 1)$, denoted by $CVaR_{\alpha}$. This risk measure quantifies the expected loss of the random variable \mathbf{T} in the worst $\alpha\%$ of cases described as follows:

$$CVaR_{\alpha} = E[\mathbf{T} | \mathbf{T} \geq \inf\{t | P(\mathbf{T} > t) \leq 1 - \alpha\}] \quad (4.3)$$

where \mathbf{T} is a vector of random variables like \mathbf{t} or $\boldsymbol{\tau}$ in (4.1) or (4.2).

Therefore, considering the above definitions, we define the robust risk measure $CVaR_{\alpha}(\mathbf{T})$, denoted by, $RCVaR_{\alpha}(\mathbf{T})$, as follows:

$$RCVaR_{\alpha}(\mathbf{T}) = \sup_{\mathbb{P}^{\mathbf{T}} \in \mathbb{P}^{\mathbf{T}}} CVaR_{\alpha}(\mathbf{T}) \quad (4.4)$$

It is easy to see that the worst-case $CVaR_{\alpha}(\mathbf{T})$ is nothing but the robust $RCVaR_{\alpha}(\mathbf{T})$ which can be equivalently expressed as

$$\min_{(\mathbf{x}, \mathbf{y}) \in \mathbf{X}} RCVaR_{\alpha}(\mathbf{T}) = \min_{(\mathbf{x}, \mathbf{y}) \in \mathbf{X}} \sup_{\mathbb{P}^{\mathbf{T}} \in \mathbb{P}^{\mathbf{T}}} CVaR_{\alpha}(\mathbf{T}) \quad (4.5)$$

where \mathbf{X} is the solution space describing the set of constraints in (3.1)-(3.9).

We now adopt the same approach proposed in [67], and generalize it for an extension of the latency-based VRP with service level presented in Chapter 3.

Theorem 4.1.1. *For any random variable $\mathbf{T} \in \mathbb{R}^+$, with a distribution function $\mathbb{P}^{\mathbf{T}}$ belonging to the distributional set $\mathbb{P}_{\mathbf{T}} = \{\mathbb{P}^{\mathbf{T}} | \text{Sup}(\mathbf{T}) = [0, \infty), E(\mathbf{T}) = \mu_{\mathbf{T}}, \text{Var}(\mathbf{T}) = \sigma_{\mathbf{T}}^2\}$, the $\text{RCVaR}_{\alpha}(\mathbf{T})$ is calculated as follows:*

$$\text{RCVaR}_{\alpha}(\mathbf{T}) = \begin{cases} \frac{\mu_{\mathbf{T}}}{1-\alpha}, & \text{if } 0 \leq \alpha \leq \frac{\sigma_{\mathbf{T}}^2}{\sigma_{\mathbf{T}}^2 + \mu_{\mathbf{T}}^2} \\ \mu_{\mathbf{T}} + \sqrt{\frac{\alpha}{1-\alpha}} \sqrt{\sigma_{\mathbf{T}}^2}, & \text{if } \frac{\sigma_{\mathbf{T}}^2}{\sigma_{\mathbf{T}}^2 + \mu_{\mathbf{T}}^2} \leq \alpha \leq 1 \end{cases} \quad \text{Proof: See [67]. } \square$$

Theorem 4.1.1 provides a baseline to present an equivalent non-deterministic mixed integer mathematical model for the robust model that we are going to present.

Let $\acute{G} = (V, \acute{E})$ be a directed (asymmetric) graph where $V = \{0, 1, \dots, i, \dots, j, \dots, n\}$ represents the set of potential customers to be visited by a fleet of K homogenous vehicles, and $\acute{E} = \{(ij) \in V \times \bar{V} | i \neq j\}$ ($\bar{V} = V \setminus \{0\}$) define the set of arcs. To each arc (ij) , we assign a random travel time \tilde{t}_{ij} for which the first and the second moments, respectively, $\mu_{\tilde{t}_{ij}}$ and $\sigma_{\tilde{t}_{ij}}^2$ are known. Each potential customer i is categorized by its profit ψ_i and random service time $\tilde{\tau}_i$. The first and the second moments of $\tilde{\tau}_i$ are denoted by $\mu_{\tilde{\tau}_i}$ and $\sigma_{\tilde{\tau}_i}^2$, respectively. The aim is to select a set of customers to be visited such that a minimum revenue level is achieved and the total latency of visited customers is minimized. The binary variable y_{ij}^r is assigned to each arc $(ij) \in \acute{E}$ such that $i \in V, j \in \bar{V}$ and takes the value 1 if the arc $(ij) \in \acute{E}$ is used to link node i in level $r+1$ to node j in level r , otherwise its value is set to 0.

The objective function of the problem is as follows:

$$(I) \quad \min : z = \sum_{r=1}^N \sum_{j \in \bar{V}} r (t_{0j} + \tilde{\tau}_j) y_{0j}^r + \sum_{r=1}^N \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} r (t_{ij} + \tilde{\tau}_j) y_{ij}^r \quad (4.6)$$

The set of constraints are the same as those presented in (3.1)-(3.9) where the constraint in (3.7), just for the matter of being consistent with the application in Section 4.3, is expressed in an alternative way

$$\sum_{i \in \bar{V}} \sum_{r=1}^N \psi_i x_i^r \geq \Gamma \sum_{i \in \bar{V}} \psi_i \quad (4.7)$$

where $\Gamma \in (0, 1)$ is an input parameter specified by the decision-maker.

This idea is supported by Theorem 4.1.2.

Theorem 4.1.2. For any feasible solution $(\mathbf{x}, \mathbf{y}) \in \mathbf{X}$ described by the set of constraints in (3.1)-(3.9) with the distributional loss function $\mathbf{Z}_t, \mathbf{Z}_\tau$ subject to a distribution in

$\mathbb{P}^z = \{\mathbb{P}^z | \text{Sup}(\mathbf{Z}_{(\mathbf{x}, \mathbf{y})}) = [0, \infty), E(\mathbf{Z}_{(\mathbf{x}, \mathbf{y})}) = \mu_z(\mathbf{x}, \mathbf{y}), \text{Var}(\mathbf{Z}_{(\mathbf{x}, \mathbf{y})}) = \sigma_z^2(\mathbf{x}, \mathbf{y})\}$, the following result hold:

$\min_{(\mathbf{x}, \mathbf{y})} \text{RCVaR}_\alpha^z(\mathbf{Z}_{(\mathbf{x}, \mathbf{y})}) = \min_{(z_1^*, z_2^*)} (z_1^*, z_2^*)$, where

$$\min : z_1 = \frac{1}{1 - \alpha} \left(\sum_{r=1}^N \sum_{j \in \bar{V}} r [\mu(t_{0j}) + \mu(\tilde{\tau}_j)] y_{0j}^r + \sum_{r=1}^N \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} r [\mu(t_{ij}) + \mu(\tilde{\tau}_j)] y_{ij}^r \right) \quad (4.8)$$

$$\begin{aligned} \min : z_2 = & \sum_{r=1}^N \sum_{j \in \bar{V}} r [\mu(t_{0j}) + \mu(\tilde{\tau}_j)] y_{0j}^r + \sum_{r=1}^N \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} r [\mu(t_{ij}) + \mu(\tilde{\tau}_j)] y_{ij}^r + \\ & + \sqrt{\frac{\alpha}{1 - \alpha}} \times \sqrt{\sum_{r=1}^N \sum_{j \in \bar{V}} r^2 [\sigma^2(t_{0j}) + \sigma^2(\tilde{\tau}_j)] y_{0j}^r + \sum_{r=1}^N \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} r^2 [\sigma^2(t_{ij}) + \sigma^2(\tilde{\tau}_j)] y_{ij}^r} \end{aligned} \quad (4.9)$$

Proof: The proof is similar to what discussed in [67] in which two independent random vectors, instead of one, are considered. \square

As Theorem 4.1.2 shows, the optimal solution of the distributionally robust model in (4.6) is obtained by solving one linear and one non-linear mixed integer mathematical model with the objective functions in (4.8) and (4.9) and with the same set of constraints (3.1)-(3.9). Then, we should take the minimum between the optimal values of z_1^* and z_2^* . In what follows, we present a metaheuristic approach to solve the latter problems.

4.2 The metaheuristic approach

The computational intractability of the problem has motivated us to design a heuristic approach exploiting the specific problem structure. Particularly, we propose an Adaptive Large Neighborhood Search (ALNS) heuristic first introduced by Ropke and Pisinger [215] which has been widely used for solving combinatorial problems in different fields ([215, 210]). The ALNS is mainly described as an iterative approach involving a set of removal and insertion procedures. Each iteration starts by selecting a pair of removal and

repair methods: the removal policy partially destroys a part of the current solution where the insertion method tries to repair it. This, in general, provides us with a large set of neighborhood solutions and hopefully may lead into finding near-optimal solutions. In general, the selection criterion for both of the removal and repair methods is based on the weight values assigned to each method controlling how frequent each method is called. Very often, the performance of LNS is enhanced by adding an adaptive mechanism which tracks the performance of each method, in terms of the success or the failure in finding improving solutions over the past iterations, and use this information to dynamically adjust the weights. This gives those removal-insertion methods, which have been more successful, more chance to be selected in the next iterations. We develop our proposed algorithm based on the same idea of ALNS described in [215] but we have customized and modified the heuristic, based on the specific features of our problem. The most important distinctive features of the proposed heuristic are classified as follows.

We hybridize the ALNS with Variable Neighborhood Descent (VND) in order to intensify the search process. As a matter of fact, the removal and insertion operators essentially perform a blind search exploring large size neighborhoods. In simple words, the removal methods can destroy a large part of the solution and the neighborhood contains a large number of solutions. To overcome this disadvantage, some authors augmented the LNS with an intensification mechanism such as local search or hybridized it with other heuristics such as Tabu search and simulated annealing [215]. In our hybridized algorithm, after a certain number of iterations performing the ALNS heuristic, we run a VND-based procedure to deeply explore different neighborhoods of the current solution in the hope of moving to better solutions.

The general scheme of the proposed ALNS heuristic is shown in Algorithm 10. The algorithm is composed of a construction phase generating an initial feasible solution s , with respect to the set of constraints (see Section 3), which is then improved in the improvement phase over an iterative procedure.

The initial solution is generated in a greedy fashion started by selecting K different customers with the lowest service completion time CT , to be

scheduled as the first visited nodes. The service completion time of customer $i \in \bar{V}$ is calculated as $CT_i = \mu_{t_{0i}} + \mu_{\tau_i} + \sqrt{\sigma_{t_{0i}}^2 + \sigma_{\tau_i}^2}$. If this initial solution is a feasible one with respect to the service level constraint (4.7), the procedure is finished; otherwise, we continue adding the most profitable non-visited nodes, one by one, to each vehicle. This guarantees to have balanced paths in the sense that the difference in the number of visited nodes over each pair of paths is at most one. This process is repeated until the feasibility is gained and the obtained solution is fed in Algorithm 10.

The improvement phase consists a main loop running for It_{Max} iterations and each iteration starts with the selection of a pair of destroy (d) and repair (r) methods. After that, the destroy is applied to destroy part of the current solution s and then the repair method rebuilds a new solution s' . We have considered six different destroy and insertion methods, including both the original ALNS methods ([215]) and their modified versions customized to be consistent with the problem structure. The proposed heuristic allows a controlled exploration of intermediate infeasible solutions. To this end, we replace the objective function z by its penalized counterpart \bar{z} where $\bar{z} = z + \lambda [\Gamma \sum_{i \in \bar{V}} \psi_i - \sum_{i \in \bar{V}} \sum_{r=1}^N \psi_i x_i^r]^+$, $[x]^+ = \max(0, x)$, and λ is the penalty parameter defined by the decision-maker.

Hence, the solution s' is not necessarily a feasible one. If it is feasible, the current and the best solutions s and s^* can be updated. Otherwise, if the new solution s' neither outperforms the current solution, and consequently nor the best solution, it can be accepted with respect to the aspiration criterion, explained in Subsection 4.2.4. In any of the cases discussed above, the scores assigned to the selected operators d and r are also appropriately modified. These scores are used to update the weights of operators after each η iteration. The details on the update of scores and weights are given in subsection 4.2.3.

The iterative VND-based heuristic is called in after every δ_2 segments, exploring different neighborhood structures around the current solution s in order to find an improving solution, if any. This procedure is halted whenever the input solution s cannot be improved anymore. Finally, the update of current and best solutions is performed ending one iteration of the main loop.

Algorithm 10: Hybrid ALNS-VND heuristic

```
1 Input:  $It_{Max}, T_0, \theta, \delta_1, \delta_2$ 
2 Initialization:  $t \leftarrow 1, T \leftarrow T_0, s \leftarrow null, s_{best} \leftarrow null,$ 
3 Initialize the scores  $\xi_i$ , the values  $n_i$ , and weights  $\omega_i$  assigned to
  destroy and repair operators
4  $s \leftarrow Initial\ solution()$ 
5  $s_{best} \leftarrow s$ 
6 for ( $t \leftarrow 1$  to  $It_{max}$ ) do
7   Select a pair of destroy and repair operators  $(d, r)$ 
8    $s' \leftarrow r(d(s))$ 
9   if ( $\bar{z}(s') < \bar{z}(s)$ ) then
10     $s \leftarrow s'$ 
11    if ( $\bar{z}(s) < \bar{z}(s_{best})$  &  $s$  is feasible) then
12       $s_{best} \leftarrow s$ 
13      update the scores of  $r$  and  $d$  by  $\phi_1$ 
14    else
15      update the scores of  $r$  and  $d$  by  $\phi_2$ 
16    end
17  else
18    if ( $s'$  is accepted) then
19       $s \leftarrow s'$ 
20      update the scores of  $r$  and  $d$  by  $\phi_3$ 
21    end
22  end
23  if ( $it \% \delta_1 == 0$ ) then
24    Update the weights  $\omega_i$  and reset  $\xi_i$  and  $n_i$ 
25  end
26  if ( $it \% \delta_2 == 0$ ) then
27     $s' \leftarrow VND(s)$ 
28    if ( $\bar{z}(s') < \bar{z}(s)$ ) then
29       $s \leftarrow s'$ 
30      if ( $\bar{z}(s) < \bar{z}(s_{best})$ ) then
31         $s_{best} \leftarrow s$ 
32      end
33    end
34     $T \leftarrow T \times \theta$ 
35  end
36 end
37 return  $s_{best}$ 
```

4.2.1 Removal heuristics

The removal heuristics try to destroy a part of current solution by deleting a specified number of selected nodes, let say γ , one by one, in the current solution where $\gamma = \lceil \delta |\bar{V}| \rceil$ and $\delta \in (0, 1)$ is an input parameter. The higher is the value of δ , the higher is γ , the bigger neighborhood is searched. Here, we present six different removal methods applied in the proposed heuristic.

Worst removal heuristic This destroy operator starts with an empty set for deleted nodes $U = \emptyset$ which is going to be filled by γ nodes from the set of visited nodes, let say L . To this end, for each visited node i , we calculate the change in the objective function $\Delta \bar{z}^i$, if i is being removed from the current solution. The elements in L are sorted in non-decreasing order with respect to the $\Delta \bar{z}^i$ values and the i^* -th element in L such that $i^* = \lfloor \pi^\rho |L| \rfloor$ is selected to be deleted from the current solution where π is a random number in $[0, 1]$ and ρ is an input parameter greater than 1 controlling the randomized degree of heuristic. After each node deletion, the $\Delta \bar{z}^i$ values are evaluated with respect to the last destroyed solution.

Worst removal heuristic: greedy version

This heuristic is similar to the previous one but it does not use the randomized mechanism. The node with the largest $\Delta \bar{z}^i$ value is deleted from the current solution.

Shaw removal with respect to the expected completion time and profit

The Shaw removal heuristic proposed by Shaw [232] and Ropke and Pisinger [215] accounts for the similarity between nodes. Nodes that are more similar to each other, with respect to a specified criterion, are chosen to be deleted from the current solution. The Shaw removal starts with a node (the reference node) let say r , randomly selected from the set of non-visited nodes, $|U|$. Then, for each visited node i in the current solution, we calculate the expected latency of node i (ECT_i) (composed of the expected travel times and the expected service time up to node i) as well as the expected arrival time of node r (ECT_r), if visited instead of node i . The similarity score $SR(i, r)$ between nodes i and r is defined as $SR(i, r) = \left| \frac{ECT_i - ECT_r}{ECT_i + ECT_r} \right| + \left| \frac{\psi_i - \psi_r}{\psi_i + \psi_r} \right|$. Then, all

selected nodes are sorted in non-decreasing order with respect to the similarity scores $SR(i, r)$ and the i^* -th element in $\bar{V} - U$ such that $i^* = \lfloor \pi^\chi |\bar{V} - U| \rfloor$ is select to be deleted from the current solution. Here π is a random number in $[0, 1]$ and $\chi \geq 1$ is an input parameter. The ending arcs of the path are re-connected after deleting a node. This process is continued until ρ nodes are deleted.

Shaw removal based on profit

This heuristic works like the the above *Shaw removal* where the only difference is in the way the similarity score is defined: $SR(i, r) = |\psi_i - \psi_r|$.

Shaw removal based on set up time The similarity score used in this heuristic is defined as $SR(i, r) = |\frac{\tau_i - \tau_r}{2}|$. The other parts of this heuristic are the same as explained in the *Shaw removal*.

Random removal This destroy operator randomly selects a visited node in the current solution to be deleted. The same stopping criterion explained before holds here as well.

4.2.2 Insertion heuristics

We use six different insertion heuristics to repair the partially destroyed current solution by adding some nodes into the solution. Due to the selective structure of the problem, we should look for an appropriate subset of nodes to be inserted into the solution; to this end we let the heuristic choose the appropriate nodes from the set of all non-visited nodes, including those present in the current solution but destroyed over the removal heuristics as well as those nodes that were not selected in the current solution. The node insertion is continued until either γ nodes are inserted or a feasible solution is obtained. We should note that the repair methods consider the possibility of inserting both the selected nodes deleted over the destroy phase and the non-selected nodes. Let U be the set of non-visited nodes.

Basic greedy insertion heuristic

This insertion heuristic inserts nodes in U one by one, starting from the first element, in the best possible position (order) and path resulting in the least change of objective function \bar{z} . After inserting each node, the current solution is modified and the changes in the objective function for other candidate nodes are evaluated with respect to the repaired solution.

Deep greedy insertion heuristic

This heuristic is similar to the *Basic greedy insertion* but instead of inserting each non-visited node in the best possible order, looks for the best non-visited nodes to be inserted in the best order and path such that the total increase in \bar{z} is minimized. Similar to the basic greedy heuristic, the current solution is repaired gradually by inserting node.

Regret insertion

In this heuristic, for each non-visited node i in U , we calculate its regret value $Regret_i = \Delta\bar{z}^{i,2} - \Delta\bar{z}^{i,1}$ expressed as the changes in the penalized objective function \bar{z} , if node i is inserted in its first best and second best positions. Then, the node maximizing this value $i^* = \arg \min_{i \in U} (Regret_i)$ is chosen to be inserted in its best position in the corresponding path.

Random insertion

This insertion heuristic selects randomly a non-visited node $i \in U$ and inserts it in a random position in a randomly chosen path.

Most profitable insertion greedy

This heuristics ranks all elements in U based on their profit values in non-ascending order, then starting from the first element, assigned with the highest profit, the non-visited nodes are inserted one by one in a randomly chosen position over a random path. Like other insertion heuristics, this process is continued until the stopping criterion is met.

Most profitable insertion randomized

Similar to its greedy version, this heuristic first orders all the elements in U based on their profits. Then, the i -th element in U_p such that $i = \left\lceil \frac{\pi^\kappa |U_p|}{2} \right\rceil$ is chosen to be inserted in a randomly selected position in a randomly selected

path, where π is a random number in $[0, 1]$, $\kappa \geq 1$ is a parameter, and U_p is the set of elements in U sorted based on profit value ψ_i in non-ascending order.

4.2.3 Adaptive mechanism

At each iteration of the ALNS, a pair of removal and insertion methods are selected to be performed on the current solution generating a set of its neighborhoods. The selection is done in an adaptive way since the performance of each removal or insertion is recorded giving more chance to the more successful neighborhoods to be selected in the next iterations. To this end, the search iterations N_{max} is divided into N different segments, each segment including δ_1 iterations ($N_{max} = N \delta_1$). During the first segment, we assign equal weights $\omega_i = 1$ to each removal or insertion heuristic i and a *roulette wheel* mechanism is used to select the pair of heuristics based on weights. The weights ω_i over each segment are fixed but the performance of each heuristic is carefully recorded and the score values of selected heuristics (ξ_i) are updated. Here, we consider three different score values $\phi_1 = 50$, $\phi_2 = 20$ and $\phi_3 = 0$ that initially are set to zero and if the solution found using the pair of removal-insertion heuristics is better than the best solution, their score values are independently increased by ϕ_1 , if it does not dominate the best solution but is better than the current one, their scores are increased by ϕ_2 , and finally if it does not hold any of the latter cases but results in a deteriorated but accepted solution, the scores are increased by ϕ_3 . At the end of each segments the weights ω_i are updated as follows:

$$\hat{\omega}_i = \begin{cases} \omega_i & \text{if } n_i = 0 \\ (1 - \eta) \omega_i + \eta \frac{\xi_i}{n_i} & \text{if } n_i > 0 \end{cases}$$

where n_i is the number of times heuristic i has been called over the latter segment and $\eta \in [0, 1]$ is the reaction factor controlling the weight adjustment with respect to the trade-off between the current weight and the heuristic performance. Given the above discussion, the probability of selecting each insertion (removal) heuristic \hat{i} , is calculated as $\frac{\omega_{\hat{i}}}{\sum_{i=1}^{n(\cdot)} \omega_i}$ where $n(\cdot)$ is the total number of insertion (removal) heuristics.

4.2.4 The acceptance criterion

The acceptance criterion is in the attempt to diversify the search process and escape from being stuck in a local minimum. In this framework, we also accept non improving solutions satisfying the acceptance criterion. Here, we use the simulated annealing as a common acceptance criterion used in ALNS. If the new solution s' is better than the current solution s , it is always accepted; otherwise, it is accepted with probability $\exp(\frac{\bar{z}(s)-\bar{z}(s')}{T})$ where $T > 0$ is the current temperature parameter which is gradually decreased at the end of each iteration by a factor equal to the cooling rate $\theta \in (0, 1)$. Obviously, the initial temperature, let say T_0 , should be set appropriately such that over the initial iterations, more non improving solutions are accepted giving more chance to explore promising areas in search region. With the increase in the number of iterations, the moves towards the improving solutions are prioritized.

4.2.5 The VND-based heuristic

We propose a VND-based heuristic to intensify the search process every δ_2 iterations. The classical VND follows the simple idea of switching to different neighborhood structures starting from the simplest neighborhood type. This process is stopped whenever all neighborhoods are explored and the current solution cannot be improved anymore.

Our VND-based heuristic is different from the traditional VND in two aspects. First, our proposed heuristic is composed of two main loops instead on one and more importantly, whenever we found an improving solution s' with respect to the neighborhood $\nu = \nu_0$, in the next iteration, we keep exploring the same neighborhood ν_0 while VND switches to the first neighborhood $\nu = 1$. Similarly to the traditional VND, the proposed approach switches to the next neighborhood $\nu + 1$, if no improving solutions over the neighborhood ν are found. The outer loop runs the VND-based heuristic in an iterative way, until the initial solution is improved. The general structure of the proposed heuristic is sketched in Algorithm 11. More details about the local search are given in Subsection 4.2.6.

Algorithm 11: The VND-based heuristic

```
1 Input:  $s, N_{max}$ 
2 Initialization:  $s^* \leftarrow \infty, s' \leftarrow null$ 
3 repeat
4    $s^* \leftarrow s$ 
5    $\nu \leftarrow 1$ 
6   repeat
7      $s' \leftarrow Local\ Search(s, \nu)$ 
8     if ( $z(s') < z(s)$ ) then
9        $s \leftarrow s'$ 
10    else
11       $\nu \leftarrow \nu + 1$ 
12    end
13  until ( $\nu > N_{max}$ );
14 until ( $z(s) < z(s^*)$ );
15 return  $s$ 
```

4.2.6 Local search procedure

A local search is applied around the solution s with respect to neighborhood ν , seeking for the best improving solution, if any. The applied local search involves eight different neighborhood structures classified as intra-route and inter-route move operators. The repair methods consider the opportunity of adding both the non-selected nodes and the discarded nodes. In the local search, once again, we allow the possibility of modifying the set of selected nodes. To be more precise, we consider three different move operators, including moves between the class of selected nodes in the solution, moves between the set of selected and non-visited nodes that can be added into the solution (see *Replace* in below), and moves among the group of non-selected nodes competing for being inserted (see *Insert unvisited* in below).

1. Intra-route neighborhoods, including *Swap*, *2-opt*, and *Or-opt* moves.
 - *Swap* move: exchanges the position of two visited nodes over a single path.
 - *2-opt* move: removes a pair of non-adjacent arcs from the path and replaces it with two new arcs reconnecting the path.
 - *Or-opt* move: deletes a triple of non-adjacent arcs and rebuilds them such that the order of intermediate arcs are preserved.

- *Delete visited* move: deletes a visited node from its path.
 - *Insert unvisited* move: inserts a non-visited node into a path.
 - *Replace* move: replace a visited node by a non-visited one.
2. Inter-route composed of *Exchange* and *Relocation* operators.
- *Exchange* move: exchanges a pair of nodes between two different paths.
 - *Relocation* move: deletes a node from a path and inserts it to another path.

4.3 Application in the multi-machine scheduling context

4.3.1 Application description

Here, we specifically, discuss how the model in (3.1)-(3.9) can be used to describe a selective multi-machine scheduling problem.

There is a close relation between problems addressed in the vehicle routing and in the machine scheduling fields, confirmed by the large number of similarities in terms of the mathematical models and methodologies ([46]).

Let us consider the same asymmetric graph $\vec{G} = (V, \vec{E})$ presented in Section 4.1, where V includes a set of potential jobs (instead of customers) to be processed (instead of being visited) on K identical machines (instead of vehicles) and the set of arcs \vec{E} represents the precedence relationships among the jobs. The travel time \tilde{t}_{ij} assigned to $(i, j) \in \vec{E}$ is not interpreted as a travel time, but, in fact, is the set up time, which is equivalent to the minimum required time to set up the machine after processing job i and before starting job j . Each potential job i requires a processing time τ_i and has an associated profit value ψ_i . For the dummy job $0 \in V$, representing the start of the action over each machine, $\tau_0 = \psi_0 = 0$. The aim is to select a subset of jobs in \vec{V} minimizing total completion time, such that the total benefit gained by the chosen jobs, in percentage, is above a minimum required level Γ determined by the manufacturer. Hence, in the model presented in Section 4.1, the objective function (4.6) represents the total completion times for all the processed jobs, including the set up and the processing times.

4.3.2 Computational experiments

To test the heuristic approach, we have chosen 750 instances from the benchmark test used for the order acceptance problem in the single machine context ([195]). To the best of our knowledge, the order acceptance model in [195] is the first contribution with a selective structure in scheduling context. The instances are divided into classes (of 250 instances each) of different size (10, 15, and 25 nodes). The setup times, the processing times, and the profits are the same as those reported in the benchmark and the setup time variances $\sigma_{t_{ij}}^2$ as well as the processing time variances $\sigma_{\tau_i}^2$ were set to $\sigma_{t_{ij}}^2 = \lceil \zeta_t^2 \rceil$ and $\sigma_{\tau_i}^2 = \lceil \zeta_\tau^2 \rceil$ where ζ_t and ζ_τ are random numbers uniformly distributed in intervals $[1, \frac{1}{2}(\min_{i \in V, j \in \bar{V}} t_{ij} + \max_{i \in V, j \in \bar{V}} t_{ij})]$ and $[1, \frac{1}{2}(\min_{i \in \bar{V}} \tau_i + \max_{i \in \bar{V}} \tau_i)]$, respectively. The value of Γ is set to 0.6 and the number of machines (K) varies from one to six depending on the size of instance.

The proposed heuristic was implemented in C++ and the models were solved using the open source SCIP library, released 3.2.0. The experiments were executed on an Intel® Core™ i7 2.90 GHz, with 8.0 GB of RAM memory. The heuristic solutions were assessed with respect to the best solution obtained by SCIP within a time limit of 3600 seconds. For each instance, SCIP is called to solve two different problems, as described in Section 4.1. The heuristic parameter settings are reported in Table 4.1.

Parameter	definition	Value
It_{max}	Number of iterations in ALNS	2000
δ_1	Segment of iterations to update the weights	40
η	Reaction factor controlling the weight update	0.5
δ_2	Segment of iterations to run VND-based heuristic	50
N_{max}	Number of move operators (neighborhoods)	8
T_0	Initial temperature in the SA acceptance criterion	0.4
θ	Cooling factor	0.99
δ	Destroy fraction	{0.1, 0.2}
κ	Randomization parameter in <i>Most profitable insertion randomized</i> method	2
χ	Randomization parameter in <i>Shaw removal</i> -based methods	2
ϕ_1	Increment rate of score ξ_i assigned to method i , if the best solution is updated	50
ϕ_2	Increment rate of score ξ_i assigned to method i , if the current solution is updated	20
ϕ_3	Increment rate of score ξ_i assigned to method i , if the degenerated solution is accepted	0

Table 4.1: Heuristic settings

Table 4.2 summarizes the results for 250 instances with 10 jobs and a single machine organized in 25 data sets each containing 10 instances with the same size. For each data set, we have reported the average gap Gap_{Avg}

calculated with respect to the best solution obtained by SCIP and expressed as $Gap = \frac{z_{heu} - z_{SCIP}}{z_{SCIP}} \times 100$ where z_{heu} and z_{SCIP} are the heuristic and SCIP best solutions, respectively. In a similar way, the speed up of the solution time for each instance is calculated as $\Delta T = \frac{CPU_{heu}}{CPU_{SCIP}} \times 100$ where CPU_{heu} and CPU_{SCIP} represent the computational times of heuristic and SCIP, respectively. We should mention that the lower bound assigned to each instance (z_{SCIP}) is obtained by solving two models and taking the minimum of their optimal values. Hence, the SCIP solution time CPU_{SCIP} is the total computational time spent by SCIP on both models. In addition, since for the set of instances with 10 and 15 nodes in Tables 4.2 and 4.3, SCIP was able to find the optimal solutions, we did not report the gap zero for the sake of brevity. The column with heading *#opt* shows the number of instances (out of 10) in each data set for which the heuristic gap (Gap) is zero.

In terms of the solution quality, the heuristic provides near-optimal solutions with an average gap around 4% which increases slightly with an increase in the risk level α . The heuristic finds optimal solutions for 2 to 9 instances in each data set and, in total, for 135, 140, and 147 (out of 250) instances with α of 0.1, 0.5, and 0.9, respectively. In 19 out of 25 existing data sets, the heuristic solves at least half of the instances to optimality. In terms of computational time, with respect to all the risk levels, the heuristic is on average about 3 and up to 8 times faster than SCIP. In addition, the slight higher average gap for $\alpha = 0.9$ is compensated by the lower speed up rate as reported in Table 4.2. In fact, the heuristic average speed up decreases with the increase in risk level. The average speed up for $\alpha = 0.9$, is above 8% and 9% lower than cases with $\alpha = 0.5$ and 0.1, respectively. This could be an insight showing that the proposed heuristic works well for higher risk levels.

Table 4.2: Results for the instances with $n = 10$ nodes and $k = 1$

Instance	$\alpha = 0.1$			$\alpha = 0.5$			$\alpha = 0.9$			
	<i>Gap</i> _{Avg} %	ΔT _{Avg} %	# <i>Opt</i>	<i>Gap</i> _{Avg} %	ΔT _{Avg} %	# <i>opt</i>	<i>Gap</i> _{Avg} %	ΔT _{Avg} %	# <i>opt</i>	
1	10orders-Tao1R1	4.98	36.03	6	4.85	35.27	5	5.44	33.97	3
2	10orders-Tao1R3	5.33	38.06	6	4.69	39.8	6	7.06	35.06	4
3	10orders-Tao1R5	3.46	46.95	5	2.6	45.85	5	2.4	40.31	5
4	10orders-Tao1R7	3.62	40.3	7	3.39	42.04	6	0.52	39.95	9
5	10orders-Tao1R9	4.04	38.89	7	6.16	39.47	5	5.55	34.82	4
6	10orders-Tao3R1	2.79	40.91	7	3.89	42.48	5	4	42.09	6
7	10orders-Tao3R3	1.14	35.11	6	0.26	34.28	8	2.9	29.01	5
8	10orders-Tao3R5	3.02	42.94	6	3.69	41.78	5	3.77	35.48	6
9	10orders-Tao3R7	6.39	44.46	3	6.06	42.69	3	5.15	38.78	4
10	10orders-Tao3R9	1.6	47.6	8	0.32	42.15	8	0.84	42.21	6
11	10orders-Tao5R1	1.54	46.51	8	4.39	46.24	7	2.27	41.89	8
12	10orders-Tao5R3	4.22	40.32	5	4.22	39.2	6	6.37	38.58	5
13	10orders-Tao5R5	8.18	45.17	2	9	41.51	3	12.04	48.41	2
14	10orders-Tao5R7	3.02	36.59	3	1.42	38.82	4	2.18	35.47	6
15	10orders-Tao5R9	4.81	47.22	4	6.64	43.95	3	6.72	44.68	3
16	10orders-Tao7R1	2.32	45.1	8	2.11	45.49	7	5.31	45.8	6
17	10orders-Tao7R3	4.55	37.53	6	7.38	39.5	6	5.39	32.93	5
18	10orders-Tao7R5	3.76	43.69	5	3.56	40.88	4	3.25	35.7	7
19	10orders-Tao7R7	1.68	41.31	8	1.38	37.12	6	1.59	30.85	6
20	10orders-Tao7R9	1.5	27.95	8	2.37	27.58	6	2.37	25.34	5
21	10orders-Tao9R1	2.42	31.36	7	0.96	33.6	8	3.92	28.26	5
22	10orders-Tao9R3	4.22	44.21	7	3.39	45.54	7	2.15	39.96	9
23	10orders-Tao9R5	2.83	34.28	5	1.74	33.75	8	1.13	30.3	6
24	10orders-Tao9R7	6.48	34	3	6.6	32.64	3	3.5	29.59	5
25	10orders-Tao9R9	4.97	47.75	7	6.83	46.24	6	5.67	40.18	5
Avg.		3.71	40.57		3.92	39.91		4.06	36.78	

The results for data sets with 15 jobs are reported in Table 4.3. Compared with the case with 10 nodes, the average gap decreases with the increase in risk level α . The gap is 6.04 for $\alpha = 0.1$ and decreases to 2.27 for $\alpha = 0.5$, and then slightly increases to 2.39 for $\alpha = 0.9$. The heuristic finds optimal solutions for 104, 129, and 120 out of 250 instances for α of 0.1, 0.5, and 0.9, respectively. In terms of solution times, the average speed up decreases slightly with respect to the increase in risk level. In particular, for $\alpha = 0.1$ and 0.5, the heuristic is at least 2 up to 7 times faster than SCIP. Even for $\alpha = 0.9$ the heuristic provides near-optimal solutions up to 10 times faster than SCIP. Clearly, the heuristic and SCIP solution times for larger instances with 15 nodes are considerably higher than instances with 10 nodes, but the heuristic speed up rates are much lower, showing that heuristic computational time performance increases with the increase in the size of problem.

Table 4.3: Results for the instances with $n = 15$ nodes and $k = 1$

Instance		$\alpha = 0.1$			$\alpha = 0.5$			$\alpha = 0.9$		
		$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#Opt$	$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#opt$	$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#opt$
1	15orders-Tao1R1	1.71	27.32	5	2.03	27.72	3	1.27	25.26	5
2	15orders-Tao1R3	2.94	24.96	5	3.94	24.14	4	2.1	23.3	5
3	15orders-Tao1R5	1.66	26.67	8	0.74	27.74	7	2.03	25.52	5
4	15orders-Tao1R7	1.07	28.97	7	0.62	26.93	8	1.18	24.59	5
5	15orders-Tao1R9	2.26	22.14	5	2.42	24.25	5	0.88	22.49	4
6	15orders-Tao3R1	1.85	27.71	4	2.47	29.06	5	2.67	27.73	3
7	15orders-Tao3R3	2.41	24.24	6	2.36	23.64	6	2	23.6	5
8	15orders-Tao3R5	1.5	24.45	6	0.95	26.08	7	4.16	24.17	6
9	15orders-Tao3R7	1.03	29.11	6	1.95	27.47	4	2.08	25.16	5
10	15orders-Tao3R9	2.77	25.98	4	1.98	25.78	6	4.28	25.35	4
11	15orders-Tao5R1	5.36	26.3	2	3.63	25.02	2	4.05	22.75	4
12	15orders-Tao5R3	10.09	27.74	3	1.38	25.59	7	2.61	24.31	6
13	15orders-Tao5R5	5.21	29	4	2.31	28.74	4	1.96	25.26	6
14	15orders-Tao5R7	16.11	25.19	3	1.9	27.33	6	3.92	26.17	4
15	15orders-Tao5R9	14.73	26.69	3	0.85	25.43	8	1.77	24.42	5
16	15orders-Tao7R1	13.86	29.39	3	1.72	25.7	6	1.71	26.05	5
17	15orders-Tao7R3	3.88	27.66	5	1.75	28.2	5	0.64	26.6	6
18	15orders-Tao7R5	7.38	26.16	2	3.59	23.47	3	3.43	22.73	3
19	15orders-Tao7R7	15.65	28.65	3	3.75	25.2	5	1.37	26.5	7
20	15orders-Tao7R9	15.49	24.55	4	3.54	23.57	3	3.89	23.31	2
21	15orders-Tao9R1	5.37	25.17	3	3.29	23.15	5	2.62	23.91	5
22	15orders-Tao9R3	11.59	27.54	1	1.9	26.33	5	1.02	25.29	5
23	15orders-Tao9R5	1.86	28.91	3	2.74	28.23	4	2.35	28.81	4
24	15orders-Tao9R7	4.3	28.58	2	3.85	26.94	4	3.69	27.11	4
25	15orders-Tao9R9	0.9	28.75	7	1.01	26.33	7	2	25.29	7
Avg.		6.04	26.87		2.27	26.08		2.39	25.03	

Table 4.4 shows the results for the set on instances with 25 jobs and 2 machines where the column with heading Gap_{SCIP} represents the optimality gap, in percentage, reported by SCIP. Obviously, those cases for which the Gap_{SCIP} values are zero have been solved to optimality by SCIP.

Table 4.4: Results for the instances with $n = 25$ nodes and $k = 2$

Instance		$\alpha = 0.1$			$\alpha = 0.5$			$\alpha = 0.9$			$Gap_{SCIP}\%$
		$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#Opt$	$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#opt$	$Gap_{Avg}\%$	$\Delta T_{Avg}\%$	$\#opt$	
1	5orders-Tao1R1	5.19	91.11	1	5.18	67.48	1	3.5	49.29	2	1.11
2	25orders-Tao1R3	3.67	89.3	1	4.38	87.3	1	5.96	64.42	0	0
3	25orders-Tao1R5	2.74	75.25	3	3.41	68.83	3	2.29	39.49	1	1.16
4	25orders-Tao1R7	3.65	101.86	1	2.64	96.16	1	2.04	62.66	4	1.07
5	25orders-Tao1R9	6.5	82.7	1	5.53	60.75	2	4.29	51.49	0	3.89
6	25orders-Tao3R1	4.12	79.85	2	2.47	91.28	3	1.73	78.55	3	0
7	25orders-Tao3R3	2.75	66.89	3	1.45	65.97	5	1.59	43.51	1	2.34
8	25orders-Tao3R5	1.97	79.13	3	2.62	66.06	2	1.54	46.24	4	1
9	25orders-Tao3R7	2.56	75.71	3	2.18	58.82	4	3.33	53.11	2	2.42
10	25orders-Tao3R9	4.24	93.5	0	3.31	67.57	0	4.83	45.01	2	2.91
11	25orders-Tao5R1	3.6	91.77	3	3.24	69.88	2	3.92	46.41	3	2.41
12	25orders-Tao5R3	3.07	87.06	3	3.03	64.28	1	2.85	53.66	2	1.59
13	25orders-Tao5R5	3.72	123.03	1	2.93	101.74	2	3.34	65.69	2	0
14	25orders-Tao5R7	1.27	66.64	3	2.01	75.69	3	2.99	53.28	2	0
15	25orders-Tao5R9	3.16	63.33	1	2.37	48.99	2	3.03	50.49	2	0
16	25orders-Tao7R1	4.54	74.31	1	5.16	90.37	2	3.97	45.45	1	4.84
17	25orders-Tao7R3	2.37	105.03	4	3.97	81.75	2	2.44	54.22	2	0
18	25orders-Tao7R5	2.71	97.36	2	2.42	90.95	2	3.01	70.92	1	0
19	25orders-Tao7R7	5.06	104.83	1	4.12	71.72	1	5.43	51.66	1	0.91
20	25orders-Tao7R9	4.95	95.63	1	3.28	101.07	2	2.36	62.17	1	1.66
21	25orders-Tao9R1	5.02	95.67	1	5.17	104.36	1	2.62	62.18	1	0
22	25orders-Tao9R3	4.78	86.16	2	3.27	75.35	2	2.39	41.65	4	2.54
23	25orders-Tao9R5	6.02	75.59	0	3.33	82.92	2	3.62	41.19	2	2.28
24	25orders-Tao9R7	10.07	107.57	1	3.9	78.02	1	3.83	43.88	1	0
25	25orders-Tao9R9	10.04	90.39	2	3.72	110.76	2	2.38	66.63	4	1.34
Avg.		4.31	87.99		3.4	79.12		3.17	53.73		1.34

4.3.3 Proposed mathematical model versus the traditional order acceptance model

The mathematical model presented in this section, is a robust counterpart of a model defined on a layered graph and used in the routing context. Therefore, a natural question arises, arguing if this formulation, in this deterministic version, is a good or a bad formulation for the selective parallel machine scheduling problem under total completion time minimization. We were not able to find an exact deterministic equivalent for our model, hence we adapted the order acceptance scheduling model presented in [195], which shares the same idea of job acceptance or rejection with our model. To make the comparison in a fair way, we replace the constraints related to the tardiness and deadlines in the order acceptance model with the service level constraint (4.7) and set the objective function equal to the total completion time. Since the order acceptance model in [195] is, indeed, designed for the single machine case, we set $K = 1$ in our proposed model. Both models were implemented in AIMMS 4.1 and solved by the CPLEX solver considering

a time limit of 1000 seconds. We perform a set of experiments on 20 instances selected from the benchmark. Table 4.5 summarizes the obtained results. Columns 1 and 2 represent the instance name and the number of jobs, respectively; Columns 3 and 4 report the best objective value and the computational time (in seconds) for our model, followed by its relative linear gap (in percentage). In a similar way, Columns 6-8 present the same information of the Columns 3-5, but for the model taken from the literature. Column 9 shows the optimality gap for the order acceptance model and, finally, Column 10 indicates the speed up in solution time calculated as $\Delta = \frac{CPU_{\text{Proposed Model}}}{CPU_{\text{Order acceptance Model}}} \times 100$. In terms of solution quality, our model was able to find the optimal solution, verified by the zero gap values in Column 5, in a time limited to 132 seconds.

For the order acceptance model, CPLEX found the optimal solutions only in 9 instances, including 4 cases for which the optimality did not proved (verified by the non-zero gaps in Column 8). In 7 cases, CPLEX provided only near-optimal solutions with different optimality gaps varying from 0.31% to 12.73% and for the 4 last instances with the largest size, CPLEX did not find any feasible solution (a dash is reported in the table). Also, with respect to the solution time, the time limit for all cases but those with 10 jobs was reached. Apart from that, the considerable difference between Gap_{Opt} and Gap_{LB} in Columns 8 and 9 is an informative insight showing that the linear relaxation of the order acceptance model is very weak even for moderate instances with 25 jobs. In summary, our model outperforms the known model in terms of both the solution quality and the computational time.

Table 4.5: Comparing the results of two models

Instance	#Jobs	Proposed Model			Traditional Model				
		Obj Val.	CPU(s)	$Gap_{LB}(\%)$	Obj Val.	CPU(s)	$Gap_{LB}(\%)$	$Gap_{Opt}(\%)$	$\Delta(\%)$
10Tao-R1-1	10	131	0.15	0	131	4.1	0	0	3.66
10Tao-R3-1	10	150	0.05	0	150	9.36	0	0	0.53
10Tao-R5-1	10	106	0.05	0	106	1.46	0	0	3.42
10Tao-R7-1	10	155	0.09	0	155	5.94	0	0	1.52
10Tao-R9-1	10	192	0.09	0	192	11.85	0	0	0.76
15Tao-R1-1	15	175	0.1	0	175	1000	29.7	0	0.01
15Tao-R3-1	15	287	0.4	0	287	1000	44.9	0	0.04
15Tao-R5-1	15	311	0.24	0	311	1000	46.3	0	0.02
15Tao-R7-1	15	244	0.23	0	282	1000	74	13.48	0.01
15Tao-R9-1	15	269	0.13	0	269	1000	40	0	0.01
25Tao-R1-1	25	646	2.24	0	648	1000	79.3	0.31	0.21
25Tao-R3-1	25	566	1.56	0	569	1000	77.9	0.53	0.16
25Tao-R5-1	25	772	2.24	0	843	1000	82.2	8.42	0.22
25Tao-R7-1	25	555	1.43	0	557	1000	78.9	0.36	0.14
25Tao-R9-1	25	620	2.43	0	649	1000	77.2	4.47	0.24
50Tao-R1-1	50	1501	43.7	0	1720	1000	91.2	12.73	4.34
50Tao-R3-1	50	1977	24.24	0	-	1000	∞	∞	2.33
50Tao-R5-1	50	2300	131.24	0	-	1000	∞	∞	13.12
50Tao-R7-1	50	1854	23.01	0	-	1000	∞	∞	2.3
50Tao-R9-1	50	2216	49.39	0	-	1000	∞	∞	4.94
Avg			14.15						1.9

4.4 Conclusions and future research directions

We proposed a distributionally robust approach for a latency-based VRP in which the uncertainty affects both the service and travel times. We formulated the problem as a risk-averse model. From the solution viewpoint, we proposed a hybrid ALNS-VND based heuristic which uses different destroy and repair methods exploring the order-sensitive structure of the problem. We also presented an application of the problem in the machine scheduling context and tested the efficiency of the proposed approach on a large set of benchmark instances. The proposed methodology provided good quality solutions for the set of small and moderate instances.

Chapter 5

Incorporating equity into emergency medical service strategic planning

Latency and equity are two closely related performance measures both affecting the customer's satisfaction. In this Chapter, we have applied the equity concept into a strategic location-allocation model. As an elective applicative field, we have chosen a problem arising in emergency medical services (EMSs) which are one of the most important health care services, since they play a vital role in saving people's lives and reducing the rate of mortality and morbidity. The importance and sensitivity of decision making in the EMS field have been recognized by operations research scientists, EMS planners, and health care practitioners who studied many problems arising in the management of the EMS systems since the 1960s. In this Chapter, we first present a part of a broad literature review on EMS that appeared in the authored published paper [21]. In the second part of the Chapter, we present a novel location-allocation model balancing efficiency and equity in strategic EMS design.

5.1 Locating Emergency Medical Services: literature review and new challenges

The EMS systems have attracted considerable attention from researchers since the 70s. The primary concern is to efficiently design a system which is able to respond to emergency calls in a timely fashion.

Traditionally, EMS problems deal with the selection, the location and dimensioning of sites and the allocation of demand points to the facility sites. Hence, numerous reviews appeared over the last years on models and methodologies for locating EMS vehicles. One of the first reviews papers that addressed EMS location models is the paper of ReVelle et al. [207]. In [52], Laporte et al. presented a paper on ambulance location and relocation problems, and classified the existing models in three main groups: static and deterministic models, probabilistic models, and dynamic models. In [111], Goldberg focused on modeling aspects of the deployment of EMS vehicles. In a recent review paper, Li et al. ([153]) studied papers addressing different types of covering models for EMS planning, hypercube queuing models, and dynamic allocation and relocation models. Finally, Başar et al. ([28]) presented a taxonomic framework for EMS location problem.

Amongst the papers on EMS systems, there are only a few review papers that have a broader view on the management of an EMS system. In [114], Green and Kolesar surveyed the role of operations research and management science in improving emergency responsiveness over time, which lead to new policies and practices. In [121], Henderson presented a discussion on challenges in EMS, highlighting the role of system-status management (re-deployment strategies) in improving the EMS systems. Finally, Ingolfsson ([132]) surveyed research on planning and management for the EMS, emphasizing four topics: (i) demand forecasting, response times, and workload; (ii) measuring performance; (iii) choosing station locations; and (iv) allocating ambulances to stations based on predictable and unpredictable changes in demand and travel times.

The structure of the review paper [21] is based on the concept of Emergency Care Pathway (ECP), which is an emerging trend that shifts the central role from health care providers to patients. A typical ECP starts when the EMS receives an emergency request. After determining the urgency of the incident, an ambulance is dispatched. The ambulance should reach the emergency scene as soon as possible to provide first-aid and to transport the patient to the ED of a hospital. Once the patient is discharged from the hospital, the ECP finishes. The review covers many decision problems in the ECP as shown in Figure 5.1.

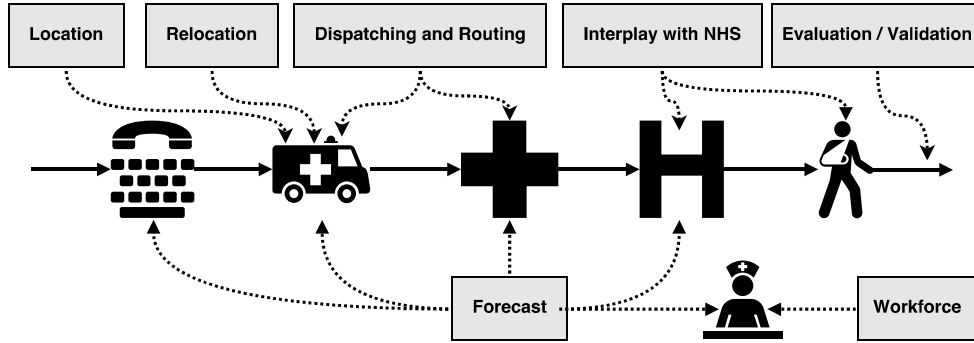


Figure 5.1: Emergency Care Pathway and related problems

In this Chapter, we will include just the first block, related to the locational decision making in the EMS. This critical topic plays an outstanding role in designing efficient EMS systems in which early response to emergency calls is provided.

Differently from other reviews, we classify the existing contributions based on two key concepts: *equity* and *uncertainty*. In what follows, challenges, gaps and new trends are highlighted in boxes.

5.1.0.1 Incorporating equity

Equity is one of challenging concerns in the healthcare sector and especially in the EMS systems, since it evaluates the fairness of how resources (notably EMS vehicles) are allocated to patients.

Since the main aim of the EMS planners is to provide early response, equity is usually expressed as a function of distance or Response Time (RT) traveled by the EMS vehicles. Although there are some standard equity indicators in location literature (mostly distance-based deviation measures [173]), there exist only few studies that incorporate equity in the EMS location problems (see also [153]) and no general consensus exists on appropriate equity measures.

To evaluate the performance of different Response Time Thresholds (RTTs), McLay and Mayorga [170] designed a location model which maximizes the overall marginal increase of the fraction of high priority calls that meet the performance standards. They also incorporate the concept of patient survival in their model by simplifying the survival function of [149]. To address the equity concept in their model, they include a coefficient in the objective

function which corresponds to the proportion of rural demand at each demand node. The model is tested on data of the Hanover County EMS system in Virginia. The results show that “longer RTTs (9 and 10 min in the case of Hanover county) result in more equitable patterns of patient survival. In these solutions, patient lives were saved in the rural areas at the expense of losing patient lives in the urban areas, which reduced the disparity between patient survival rates in urban and rural areas. Since a 9 min response time threshold is the most common performance measure used by EMS systems in the United States, this suggests that many EMS systems implicitly consider patient equity or fairness” [page 135] [170].

Interestingly, in contrast to the contribution of Mclay and Mayorga in [170], there is some evidence showing that the EMS planners are not much concerned about equity in services. For instance, Erkut et al. ([94]) mention that the existence of different actual performance as well as the RT standards (especially in rural and urban areas) support the claim that the EMS system managers are against the provision of equal access. They also state that “although a policy of equal access seems difficult to criticize, such a policy implies that lives are valued differently in different areas, because the cost of saving a life can be much higher in sparsely populated rural areas than in urban centers” [page 43] [94]. The development of a consensus on the definition of equity in healthcare provision should rely on the composition of two different equity concepts referred to as horizontal equity and vertical equity [236]. Horizontal equity implies that all demand nodes are considered in an equal manner. It means that suburban nodes are treated without discrimination. Vertical equity is concerned with the distribution of the service between different groups. By this definition, EMS systems are equitable if they favor disadvantaged groups, compensating for overall inequities. Horizontal equity has been researched extensively in health economics, while vertical inequity is receiving growing attention in the last years [140].

To decrease the level of disparity for EMS systems in rural areas, Chantart et al. [69] presented three different bi-objective location models solved by the ϵ -constraint method [176]. First, the demand nodes are classified in rural and urban nodes based on their geographical positions. Then, the authors investigate the trade-off between efficiency and equity criteria by maximizing the expected number of covered calls on one hand, and by choosing one of the

following three criteria on the other hand: the minimization of the number of uncovered rural demand zones, minimization of the maximum distance between uncovered demand zones and their closest open stations, and the minimization of the number of uncovered zones. It is worthwhile to note that maximizing the expected covered demand favors the location of ambulances in densely populated areas, which results in longer response times for patients in more rural areas. To compensate this drawback, the first two additional objective functions try to implement vertical equity, whereas the last one focuses on horizontal equity. As the authors acknowledge, comparing the three model solutions is very difficult given the lack of a unitary measure of vertical and horizontal equity.

In [68], Chanta et al. uses the *Gini coefficient* [88] to measure horizontal equity. The Gini coefficient is a measure of statistical dispersion, which was firstly developed for evaluating inequity in economic and social welfare literature, and is the most commonly used measure of inequality. Extending the minimum-envy model proposed by Espejo et al. ([96]) in which the priority levels of serving stations are incorporated, Chanta et al. [68] develop a minimum p -envy location problem to provide equitable services in emergency response location problems. The contribution of this model is modeling customer dissatisfaction as a new distance-based metric denoted by the envy. The envy is a function of distance, which indicates the dissatisfaction level of a demand node with its serving station in comparison to another demand nodes for the same level of priority. The model then minimizes the overall envy value. There are several reasons supporting the use of satisfaction as a significant factor in health care. First and above all, satisfaction is an important health outcome in its own right, enhancing the compliance and the cooperation among customers. Second, evaluating satisfaction is important for continuous quality monitoring and improvement in healthcare delivery. To show the validity of the model, the authors compared their results with those obtained through the p -center model ([84]) and a model where the Gini coefficient is minimized [88], which are both accepted equity models in location literature. Following their previous work, Chanta et al. ([70]) presented a modified version of the p -envy model in which the distance-based objective function is replaced with a survival function, dependent on the RT [170]. The model imposes lower bounds on the individual survival rates provided

by first priority servers as well as on the system-wide survival rate. The model handles the unavailability of EMS vehicles by using the hypercube model (see the references in 5.1.0.2).

The Gini coefficient, together with the *squared coefficient of variation* [53], has also been used in [251] to measure equity among servers. Here, it evaluates the relative differences among the workload of different stations obtained through the approximation procedure of [57]. Toro-Díaz et al. [251] adopt an interesting fairness perspective that equalizes the performance of the system by reducing disparities in the mean response time of different demand zones, as well as in the workloads of the servers (the EMS vehicles). To account for efficiency, the model also considers *the mean response time* and *the expected coverage*.

Besides the balance between efficiency and equity, an important issue which deserves thorough investigation is the trade-off between horizontal and vertical equity. Following this stream, Khodaparasti et al. [140] present an integrated framework for siting facilities in an equitable and efficient manner. The results of this research show that, especially for the applied equity measures, horizontal and vertical indices are consistent with each other.

Another way to address equity is by applying cooperative location games to the EMS planning. Most of the previous research addresses non-cooperative location games while the incorporation of server cooperation, especially in heavy loaded systems, is inevitable. This issue is of interest in EMS planning due to the close relation of server cooperation with the busy fraction as we will show in Subsection 5.1.0.2.

In [104], Fragnelli and Gagliardo report a game theoretic approach for the EMS location problem with interaction between candidate locations for housing EMS vehicles. Their method orders the potential locations based on two location games, namely the *coverage game* and the *multi-coverage game*. The focus in the coverage game is on maximizing the coverage of the area while the multi-coverage game aims at reducing the overlap of covered zones. The model is able to deal with the equity criterion by maximizing the area covered by all ambulances without considering the number of calls received. This approach can be considered as a starting point for further investigation of applying game theory to the EMS location planning. Since one area can be

covered by more than one EMS vehicle, the cooperative game should take the marginality issue into account, i.e., the contribution of each EMS vehicle to the total coverage. In cooperative game theory, the Shapley value (see, e.g., [229]) accounts for the total surplus generated by the coalition of all players and provides an index of the importance of each player concerning the overall cooperation. Fragnelli and Gagliardo [104] define the Shapley value based on two fairness criteria: the coverage indifference and the demand indifference. Two algorithms are provided to compute the Shapley value of the games in polynomial time.

Our literature review indicates that, even though some effort has been made in the last years to incorporate equity principles into EMS systems, there still exist important under-investigated areas on equity concepts that deserve attention. The existing literature on the incorporation of equity in EMS planning does not follow a straight direction and neglects important realistic and practical aspects of equity. One of the reasons for this might be the absence of a general consensus on equity measures for the EMS.

Future research should focus on different aspects of equity in EMS planning and on the definition of a set of widely accepted evaluation metrics for developing equity based models in EMS planning. To this end, studying the equity concept from both a horizontal and vertical point of view might be a good starting point.

The focus of equity in the EMS literature has mostly been on the geographical position of demand, which initially imposes the division of demand nodes into urban and suburban/rural zones. From this perspective, the efficiency of EMS systems has been evaluated by the quality of service (the provision of timely service) in urban areas and the fairness of service is, in general, evaluated by the coverage provided in suburban/rural zones. Even though spatial distribution is the most commonly used characteristic to differentiate between demand zones, also other characteristics can be used. Even demand zones located in the same urban zone can be different from each other in some characteristic, which calls for different treatment of these demand zones. Finally, since EMS systems are subject to dynamic pressures and often the actual configuration does not comply with the original design of the system, it is important to embed the temporal aspect into the equity concepts and to carefully investigate this issue.

5.1.0.2 Incorporating uncertainty

Uncertainty plays a big role in the EMS analysis. There is uncertainty about the amount and location of demand, travel times, severity of incidents, availability of EMS vehicles, length of stay at the ED, etc. However, despite these uncertainties, costly decisions with long term implication need to be taken, which can be a challenging task. In addition, parameter estimates might be inaccurate due to poor measurements. Therefore, making strategic decisions under uncertainty requires tailored approach to exploit strategically relevant uncertain information. To address this issue of uncertainty the in EMS literature, the existing literature adopts one of the following approaches: the *probabilistic paradigm*, the *stochastic programming approach*, the *robust counterpart*, and the *fuzzy framework*.

Probabilistic paradigm. EMS literature has focused on the uncertainty in three types of factors that can play a role in planning decisions: demand, availability of EMS vehicles, and response times.

The probabilistic paradigm mainly focuses on the last two sources of uncertainty. This results in a probability that a node is covered, which depends on the availability of EMS vehicles and uncertainty in RT. Two different streams in this framework can be distinguished. One of these streams is based on the results of the descriptive Hypercube Queuing Model (HQM). This model relies on queuing theory to estimate the busy fractions of EMS vehicles and other system performance measures. The other stream incorporates uncertainty directly into system parameters such as RT, service time, delay in response, busy fraction, etc., and is based on mixed-integer linear programming.

The descriptive HQM was developed by Larson ([150]) to evaluate steady-state busy fractions, loss probabilities, average RTs, and expected coverage, for any fixed configuration of facilities. Even though the hypercube model addresses cooperation between EMS vehicles as well as variation in workload for EMS vehicles, it relies on several assumptions for dispatching rules and it requires that the (exponentially distributed) service time is the same for all EMS vehicles, regardless the type of emergency call and the location of the vehicle. Moreover, HQM ignores many practical policies that most organizations employ regularly as, for instance, EMS vehicles being dispatched on

the road, diversions of EMS vehicles from low to high priority calls and many other features of EMS systems such as the presence of multiple depots. In spite of these disadvantages, HQM works well in practice and is considered to be a viable alternative to simulation. In fact, most researchers use it as a post-solution analysis tool for the evaluation of performance measures (see [95]). In [151], Larson presents an approximate HQM to avoid the computational difficulty of HQM. Other contributions to HQM are motivated by the need to overcome the restrictive assumptions of HQM.

Burwell et al. [59] present a modified version of Larson’s approximate HQM in which the ties in preferences for EMS vehicles is captured. In [135], Jarvis presents a generalized HQM in which service time distributions can be dependent both on base locations for EMS vehicles and on the type of emergency calls. In [29], Batta et al. develop correction factors for the busy fractions in an embedded HQM. They include the fact that EMS vehicles do not operate independently, and thus, may have different busy fractions which depend on each other and on the location of the EMS vehicle. In [247], Takeda et al. apply the hypercube approach to investigate the effect of the decentralization of EMS vehicles over the service area by applying the proposed model to the urban EMS of Campinas in Brazil. The results of this study show that by increasing the number of EMS vehicles that are partially decentralized, the performance measures and especially the RTs improve while variation in the workload of EMS vehicles is negligible. Further, the results of the model show that the total decentralization policy may not produce satisfactory results for the decision makers. In [131], Iannoni et al. propose two hypercube queuing-based models to evaluate the EMS performance on highways. The model can deal with three priority levels for calls, different EMS vehicle types, partial backup, and a multiple dispatch policy.

In most previous works, the busy fraction is considered as an exogenous input parameter which is estimated through site-specific or area-specific formulations, overlooking the fact that a (site-specific) busy fraction is one of the model outputs and should be estimated after knowing the exact configuration of facilities (see, e.g., [52, 75]). In order to address this shortcoming, in [75], Cho et al. propose an EMS location model for siting trauma centers and helicopters in which the busy fraction of each helicopter is endogenized

as a variable depending on the number of patients transported by the helicopter per time unit. In view of the fact that vehicle specific busy fractions depend on the deployment of EMS vehicles, In [230], Shariat-Mohaymany et al. present a strategic and tactical model for minimizing the costs of locating EMS vehicles and their base stations while imposing an upper bound on the busy fraction of each deployed EMS vehicle. This upper bound depends on the minimum number of EMS vehicles required to guarantee a desired reliability level as suggested by ReVelle and Hogan in [208].

In the following, we consider probabilistic models based on mixed-integer linear programming. The apparent difficulty in this is that the stochastic elements of EMS systems result in non-linear expressions, which need to be approximated. In practice, this can compromise the accuracy of the model predictions and may lead to sub-optimal solutions for the original non-linear models.

The seminal contribution is undoubtedly the Maximal Expected Covering Location Problem (MEXCLP) proposed by Daskin [83]. The MEXCLP uses a system-wide busy fraction obtained by dividing the total workload of the system by the total workforce of EMS vehicles available. Assuming that EMS vehicles operate independently, it is possible to estimate the reliability of service at each node using probability rules. The assumption of EMS vehicles being uniformly busy throughout the area, regardless spatial variations of demand, is relaxed in the Maximum Availability Location Problem (MALP) formulation proposed by ReVelle and Hogan ([208]), where local busy fractions depend on both the number of EMS vehicles available and the aggregated level of demand within each local area. In [239], Sorensen and Church presented a probabilistic location model for EMS planning which integrates the area-specific busy fraction of MALP in MEXCLP. The simulation results show the superiority of the proposed model over MALP and MEXCLP in terms of the overall percentage of calls that receive coverage within the RTT.

The server independence assumption is relaxed in the Queuing based Probabilistic Location Set Covering Problem (Q-PLSCP) and the Queuing-based Maximum Availability Location Problem (Q-MALP) formulations of Marianov and ReVelle [161]. To evaluate the reliability of service at each demand node, Q-MALP makes use of an M/G/s-loss queuing model. Although

Q-MALP relaxes the server independence assumption, it still relies on local service areas and on location-independent service times. In [133], Ingolfsson et al. recognize that, beside ambulance availability, randomness has an impact on delays and travel times. This study emphasizes that by ignoring the randomness in delays, the coverage performance might be overestimated.

These findings are confirmed by Erkut et al. ([95]), who evaluated five probabilistic covering models, namely the Maximal Covering Location Problem (MCLP) proposed by Church and Reville ([78]) and a variation of it including probabilistic response times (MCLP+PR), the MEXCLP, and two non-linear formulations that include probabilistic response times and base location specific busy probabilities (MEXCLP+PR and MEXCLP+PR+SSBP). In [95], Erkut et al. compare the five models on three performance measures, namely expected coverage, loss probability and average RT, by applying them to the case of Edmonton, Canada. They conclude that the two non-linear models, MEXCLP+PR and MEXCLP+PR+SSBP, perform better in terms of coverage compared to MCLP+PR and MEXCLP, which confirms that the accuracy in modeling complex systems can lead to substantial improvements in performance. In particular, based on the obtained results as well as on their previous research in [94], Erkut et al. conclude that models that include uncertainty “*not only result in better coverage estimates, but also cause coverage to be a better proxy for lives saved*” [page 64] [170].

We should warn the reader that simplifying assumptions might come at a cost. This is, for example, the case for the MALP model where using a local busy fraction requires the assumption of a local service area. Which simplified assumption poses the main limitation on finding an optimal solution for the true problem is difficult to determine. Another limitation of mixed integer probabilistic problems is the needed computation time. To efficiently solve these optimization problems, tailored metaheuristic approaches have to be developed. However, it is not guaranteed that these solutions methods lead to an optimal solution. Therefore, simulation is needed to determine how well the found solutions perform in practice. Simulation might also shed some light on which simplified assumptions poses the main limitation on finding the optimal solution.

Stochastic programming paradigm. In the literature, different stochastic programming problems have been presented to explicitly consider the dependence of the system reliability on the randomness in demand. In the stochastic programming paradigm, the model parameters are assumed to be random variables with a known probability distribution function. This assumption can be justified by the availability of historical data (or outputs of a simulation model) from which the empirical probability distribution might be estimated. Under this assumption, two different strategies for making decisions can be considered. In the chance constrained paradigm, “here and now” decisions are taken under uncertainty and implemented whatever the future is. Since a probabilistic guarantee is imposed on the feasibility of the solution, there is a chance that the implemented solution is not feasible in practice. A different approach is represented by the stochastic programming paradigm with recourse. In this case, some of the decisions must be made under uncertainty, whereas other decisions, the recourse decisions, can be postponed until the realizations of the random variables become known. Hence, the decisions taken in the first stage can be corrected in other stages by the recourse decisions, which explicitly depend on the specific value (realization) assumed by the random variables. Both the probabilistic paradigm and the recourse paradigm have their advantages and disadvantages. The probabilistic paradigm is most appropriate when the reliability of the system (for example, in terms of coverage) is most important. However, the two-stage nature of facility location problems – where locations are chosen given uncertain future demand, but customers are assigned to facilities once the uncertainty has been resolved – advocates the use of the recourse paradigm. In [25], Ball and Lin developed a model with separate chance constraints, where the probability of failure (the inability to be serviced within the RTT) of each demand point should be less than a given threshold value. They consider the worst case busy fraction, which occurs when each server is attending all calls from its neighborhoods. In [36], Beraldi et al. developed a location model based on joint probabilistic constraints to capture the inherent uncertainty in demand. The probabilistic constraints are jointly imposed on all demand points, which ensures that the reliability of the entire geographical area is kept above the prescribed level. In [275], Zhang and Li developed another strategic and tactical location model in which the probabilistic chance

constraints on the capacity of base locations are approximated by their equivalent second-order cone constraints, assuming that random demands can be modeled by continuous random variables. Even though the distribution of emergency calls is naturally discrete and follows a Poisson distribution, the considered approximation can be justified by the computational tractability of the resulting model.

To capture the inherent two-stage nature of location decisions, Beraldi and Bruni ([36]) present a two-stage stochastic model with embedded joint probabilistic constraints. The paper relaxes the independence assumption among servers at the expense of discretizing the uncertainty into a finite set of scenarios. Following the same structure, Noyan ([191]) proposed a two-stage stochastic location model for EMS planning in which the uncertainty in demand is taken into consideration by integrated chance constraints. These integrated chance constraints can be considered to be a relaxed version of separate chance constraints and allows a convex approximation of the non-convex feasible set defined by the probabilistic constraints.

Since full discretization of uncertainty in scenarios results in solving huge mixed integer problems, often a limited number of scenarios is generated by using proper scenario generation techniques. These techniques aim at finding a good trade-off between 1) including enough scenarios to guarantee an accurate representation of the underlying stochastic process and 2) including not too many scenarios such that the computational tractability of the resulting scenario-based mixed integer problem is preserved.

Regardless the tractability of the model, we should first investigate whether it is worthwhile to include randomness into EMS models. Even though uncertainty plays a big role in EMS models, it is not evident that this should indeed be included as is done in the stochastic programming paradigm. Although well known measures can be used to evaluate the advantages of using a stochastic model over a deterministic model (namely the *Value of the Stochastic Solution* and the *Expected Value of Perfect Information*), these quantities are not useful to decide whether a stochastic model should be used in practice. The validation of potential advantages of stochastic models over deterministic ones should be critically assessed (For more information see also [21]). In addition, stochastic models usually require information on distributions and other values that might not be known. As a conclusive remark, the art of modeling amounts to

describing the crucial aspects of a problem and approximating the less relevant ones. Including randomness might turn out to be not the most important issue to address.

Robust optimization and fuzzy paradigms. A more recent approach for optimization under uncertainty is “robust optimization”. While there are many high level similarities with the stochastic paradigm, robust optimization is a distinct field that seeks a solution that is feasible for any realization in a given uncertainty set instead of immunizing the solution against stochastic uncertainty. This paradigm is successful in various application areas because of its computational tractability. Moreover, this approach is the only reasonable alternative when information about the probability distribution is not readily available.

Adopting a robust counterpart approach, Zhang and Jiang ([274]) proposed a bi-objective location model for strategic and tactical EMS planning in which uncertain demand is addressed. The uncertainty in the number of emergency calls is captured by an ellipsoidal uncertainty set and the weighting method is applied to make a trade-off between costs and responsiveness.

Another computational tractable way of addressing uncertainty is the fuzzy paradigm. This framework is mostly applied when the probabilistic paradigm or stochastic framework cannot be used. In the absence of historical data, qualitative information extracted from expert’s observations is a crucial aspect in the decision making process. As a matter of fact, some service level concepts in EMS decision making such as patient satisfaction and staff performance are inherently known as qualitative information. The fuzzy paradigm facilitates the use of qualitative data as well as expert-based knowledge by characterizing them as linguistic terms. There are only a few papers in this stream. Some papers adopt the fuzzy set theory concept to deal with uncertainty, others apply the fuzzy programming approach to solve multi-objective location models.

Araz et al. [19] presented a three-objective location model for strategic EMS planning in which different Fuzzy Goal Programming (FGP) solution methods are applied. The first and the second criteria in the objective function maximize the total number of demand nodes covered at least once and

twice, respectively. The third criterion minimizes the total traveled distance for the uncovered demand nodes. The authors use the fuzzy goal programming approach to express the aspiration levels as imprecise values (fuzzy numbers). The efficiency of the proposed approach, in terms of first and backup coverage, is shown by comparing the results of different proposed FGP models with the lexicographic multi-objective technique. Although, the proposed FGP models provide better results with respect to the backup criterion, the performance of such techniques highly depends on the choice of the weights and on the priority levels assigned to different criteria. To allow the decision maker to find near-optimal solutions for different problem inputs in a short period of time, tailored heuristics could be developed. Following this idea, Uno et al. ([252]) presented an interactive FGP approach combined with the particle swarm optimization method to solve an emergency location problem.

In [146], Koc and Bostancioglu presented a fuzzy location model for EMS planning based on the Double Standard Model (DSM) of [109]. The authors used linguistic variables to capture the uncertainty in both travel times and RTTs. A triangular fuzzy number is associated with each linguistic term and a probability-possibility transformation is applied to find the probability distribution of each fuzzy variable. A Monte Carlo simulation was applied to show the efficiency of the proposed approach for reproducing unknown data.

A clear advantage of the fuzzy paradigm is that it does not require much information on the probability distribution of uncertain parameters. However, relying only on expert knowledge or qualitative data may lead to severe loss of information about the behavior of the system, which in turn may lead to inaccurate and unrealistic results. Therefore, it is best to only use the fuzzy approach in cases where the incorporation of qualitative data as well as expert-based opinions is unavoidable.

Remarks. In summary, the type of uncertainty as well as the type of available information should be considered when choosing an appropriate solution approach. Applying a combination of approaches that includes different types of uncertainty, could help EMS managers to come up with a

more realistic model in which the complexity of the system is fully captured and a better overview of the whole system is provided.

The EMS models reviewed so far, focus either on minimizing the expected costs/response time or on maximizing the expected coverage. This focus is appropriate for risk-neutral decision makers as these models are insensitive to costs/coverage variations. However, arriving at an incident location only a few seconds earlier can already save a human's life. Thus, instead of minimizing expected values, risk averse decision makers might accept higher costs in return for higher protection against losing lives. Unfortunately, traditional location models fail to meet the needs of risk-averse planners as only a few papers suggest mechanisms to reduce the chance of unfavorable large RTs or to increase the probability of arriving at the scene on time.

It is important to incorporate the notion of risk aversion in EMS location models, for example, by providing risk-based measures especially designed for EMS systems. This is one of the most interesting directions for future research within the stochastic framework. Furthermore, there is an increasing need for developing models that simultaneously incorporate realistic information and several uncertainty sources. Most existing studies in the literature focus on only one or two stochastic aspects and there is a lack of location models handling all sources of uncertainty simultaneously. One reason for this deficiency is the complexity of the resulting models, which necessitates the design of tailored solution methods with high computational efficiency. In addition, future research within the robust stream could enrich the literature in the EMS field and could also provide a comparative framework in which the distinguishing features of stochastic and robust models can be investigated.

5.2 Balancing Efficiency and Equity in location-allocation models with an application to strategic EMS design

In this section, we present an integrated bi-objective location-allocation model balancing efficiency and equity criteria. The new formulation combines two domains: facility location and data envelopment analysis. To support the decision maker with more realistic solutions based on the op-

timal location-allocation decisions, we endogenize the outputs of the model as a function dependent on the allocation variables. The resulting model is very difficult to deal with since it turns out to be a bi-objective mixed integer nonlinear problem. An exact solution approach is then proposed to address its solution. To illustrate the viability of the proposed approach, we investigated the potential application of the model to the design of an emergency medical service system.

5.2.1 Introduction

Major economic changes have placed increasing pressure on facility location decisions. They require large investments that can not be recovered in the short term and influence the competitive capacity of the company, affecting not only costs but also the income. For instance, considering a service business, market proximity is a critical success key to attract customers, whereas for a manufacturing business, facility location might affect procurement costs, product delivery time, and customers' service levels. In the public sector, the implications of poor location decisions are not confined to cost considerations. The effect of wrong facility location decisions may extend to disastrous consequences as increased mortality and morbidity. Thus, facility location modeling has even greater importance when applied to the siting of public, notably health care, facilities. Moreover, the public ownership of the organization and the nature of the facilities is a distinctive feature that might invalidate the use of classical facility location models. For instance, if the siting of private facilities is usually driven by the minimization of costs, the goals of public agencies, especially in health care operations, are more difficult to capture and interpret, calling for tailored approaches.

Many researchers and practitioners, since the 80s, have tried to address this issue, incorporating into the models important considerations on the equity or fairness of the service. Mayhew and Leonardi ([167]) presented a model that trades off equity and efficiency, with application to health care resource allocation in London. In [159], Mandell incorporated both equity and effectiveness measures into the distribution of library books in public libraries throughout a region. As already observed in the previous Section, the issue of equity in the EMS literature has been only recently incorporated, even if this concern has a long history in the location literature [164, 168].

The majority of the models in the literature deals with the issue of efficiency by focusing on performance measures. Our model captures the multi-criteria nature of the service performance, by evaluating the efficiency through the data envelopment analysis (DEA, for short). DEA is a non-parametric approach widely applied to assess the relative performance of a set of decisions making units (DMUs) that use several inputs to produce several outputs. It is worth noticing that the performance of the DMUs (the facilities in our case) is, in general, related to the allocation pattern and it can not be determined regardless of the allocation choices. To overcome this drawback, we endogenize the produced outputs, that therefore, become a function of the allocation decisions.

To the best of our knowledge, this is the first study in the location-allocation literature, and especially in EMS area, addressing both issues of *efficiency* and *equity* through a DEA model with endogenized outputs.

The resulting framework is quite novel (we mention a few papers accounting for equity and efficiency simultaneously, like [74]- where system equity is measured by the opportunity to receive medical services, and efficiency is represented by consumer and producer welfare- and [236]- where a range of discrete hierarchical location models with bi-criteria efficiency/equity objectives is presented), and of general applicability, and it provides an integrated framework for siting facilities in an equitable and efficient manner.

To illustrate the applicability of the proposed model, we consider a real case study related to the strategic design of an EMS system.

The rest of the Section is organized as follows. In Subsection 5.2.2, some literature review on the use of DEA in facility location models is presented. 5.2.3, the mathematical formulation of the proposed model is presented. In Subsection 5.2.4, we describe the solution approach for the proposed model. In Subsection 5.2.5, we apply the proposed model on a real case study. Finally, conclusions and discussions are drawn in Subsection 5.2.6.

5.2.2 The use of DEA in facility location models

Despite the DEA has been used as a pure performance assessment methodology, only recently some attempts have been made with the aim of incorporating the evaluation of the locations into allocation decisions [177]. Thomas et al. ([248]) studied an obnoxious facility location problem considering both

proximity and DEA scores by running the location and the DEA models iteratively. Narasimhan et al. ([186]) embedded branch offices efficiencies into a model, which considers capacities, demand requirements and budget constraints. Klimberg and Ratick ([143]) proposed two bi-objective location-allocation models by developing a simultaneous DEA model and merging it to the location model. In the proposed models, any potential allocation link is considered as a decision making unit. One of the objectives expresses the total cost of allocation while the other objective indicates the total DEA score assigned to the open links. In another research on facility location, Sahin et al. ([216]) presented a hierarchical model for siting blood service facilities. The results of their study show that some inefficient locations might be included in the optimal solution. The authors then suggested to improve the efficiency of optimal facilities associated with low DEA scores, reconsidering the solution of the location model on the light of the efficiency of the locations. In a recent paper, Mitropoulos et al. [177] presented a bi-objective DEA-location model in health care sector. The model includes three objectives: the minimization of total distance traveled by patients from all communities to their nearest located hospital, the minimization of underachievement in soft constraint and the maximization of the mean efficiency scores. The DEA score of facilities are computed a priori and then they are entered into the location model as parameters.

The distinctive feature of our model, with respect to the aforementioned literature, and especially with respect to the Klimberg's and Ratick's model [143], is the endogenization of the location outputs as a function of the allocation variables. In this respect, it is worthwhile noticing that the efficiency of the candidate locations is related to the allocation pattern and it can not be determined before allocating the demands. By endogenizing the outputs, we provide the decision maker with a more realistic solution which is dependent on the optimal allocation of the demand, rather than on rough estimation of future performance.

5.2.3 Designing an equitable and efficient system

In this part, we present a mathematical formulation for the equity and efficiency based location-allocation problem. Formally, the problem is defined on a graph with a set I of demand points. Each node $i \in I$ can be covered by

a candidate location $j \in J$ in case the facility is located within a service radius τ . Let $J(i)$ be a subset of J containing the list of candidate locations that can cover node i . The aim is the optimal location of p facilities among a set J of possible candidate locations. The proposed model integrates two parts: one related to the location-allocation problem, the other concerned with the performance evaluation. For the location-allocation part of the model, we define two sets of binary variables. Variables x_j are equal to one only if the location $j \in J$ is opened, whereas variables y_{ij} refer to the allocation and are equal to one if the demand node $i \in I$ is assigned to the location $j \in J$.

In order to measure the efficiency of each candidate location j , referred in the foregoing as DMU, we use the DEA methodology. We assume that each DMU $j \in J$ has a set A of different inputs denoted by $I_{aj}, a \in A$ and a set B of outputs denoted by $O_{bj}, b \in B$. The efficiency, is defined as the ratio of the weighted sum of outputs to the weighted sum of inputs, where the weights are variables to be optimized. We denote, for each $j \in J$, by v_{aj} the variables associated with the weight of the a -th input I_{aj} and by w_{bj} the one associated with the b -th output O_{bj} . In location-allocation models, very often the performance of the locations depends on the assignments made in the design phase. Indicating by \mathbf{y} the vector of the y_{ij} variables, we can define $O_{bj}(\mathbf{y})$ as the b -th endogenized output for candidate location $j \in J$. The aim is to determine the best location-allocation pattern maximizing at the same time the efficiency and the equity of the system. We observe that equity is also a function of some measure related to the assignment pattern. Although the location literature shows the use of different equity measures (like for example coverage and system reliability), there is no concerted agreement on how equity should be measured ([30]). Thus, we generally denote it as $H(\mathbf{y})$.

By considering the notation introduced above, the mathematical formu-

lation of the proposed bi-objective model can be expressed as follows:

$$\max \quad f_1 = \sum_{j \in J} (1 - \theta_j) \quad (5.1)$$

$$\max \quad f_2 = H(\mathbf{y}) \quad (5.2)$$

s.t.

$$\sum_{j \in J(i)} y_{ij} \leq 1 \quad \forall i \in I \quad (5.3)$$

$$y_{ij} \leq x_j \quad \forall i \in I, \forall j \in J \quad (5.4)$$

$$\sum_{j \in J} x_j \leq p \quad (5.5)$$

$$\sum_{a \in A} v_{aj} I_{aj} = x_j \quad \forall j \in J \quad (5.6)$$

$$\sum_{b \in B} w_{bj} O_{bj}(\mathbf{y}) + \theta_j = 1 \quad \forall j \in J \quad (5.7)$$

$$\sum_{b \in B} w_{bj} O_{bj'}(\mathbf{y}) - \sum_{a \in A} v_{aj} I_{aj'} \leq 0, \quad \forall j, j' \in J, \quad j \neq j' \quad (5.8)$$

$$v_{aj} \geq \delta x_j \quad \forall a \in A, j \in J \quad (5.9)$$

$$w_{bj} \geq \delta x_j \quad \forall b \in B, j \in J \quad (5.10)$$

$$y_{ij} \in \{0, 1\} \quad \forall i \in I, \quad \forall j \in J \quad (5.11)$$

$$x_j \in \{0, 1\} \quad \forall j \in J \quad (5.12)$$

$$\theta_j \geq 0 \quad \forall j \in J \quad (5.13)$$

The objective function (5.1) maximizes the total efficiency scores assigned to facilities, whereas the objective function (5.2) maximizes the system-wide equity performance $H(\mathbf{y})$. Constraint (5.3) ensures that each demand point should be assigned to at most one facility. Constraints (5.4) allow assignments only to open facilities. Restriction (5.5) imposes a limit on the number of sited facilities. Constraints (5.6)-(5.10) are related to the DEA part of the model. Constraint (5.6) states that the sum of weighted inputs for any sited facility should be equal to 1. Restrictions (5.7) define the inefficiency variable θ_j as a function of the weighted outputs. The set of constraints in (5.6) in combination with restrictions (5.7) and (5.8) ensure that the efficiency assigned with closed facilities is zero. Note that the outputs are dependent on the assignment vector \mathbf{y} . Constraints (5.8) ensure, for open facilities, that the weights are assigned in a way that the efficiency of any other facility is not greater than 1, if it uses the same weights. Moreover, constraints (5.9)-(5.10), where δ is a non-Archimedean number, guarantee

that only non-dominated efficient solutions in the DEA model are investigated. Finally, restrictions (5.11)-(5.13) define the nature of decision variables. Our model resembles the Klimberg's and Ratick's [143] model, even though some distinguishing features can be highlighted. Firstly, our model considers each facility location as a potential DMU, whereas their model consider each assignment variable in a transportation network as a potential DMU. Secondly, the outputs in their model are not allowed to change and they are considered to be fixed whereas in our model the outputs are endogenized. Thirdly, the objective function has its own originality and unlike the Klimberg's and Ratick's model, does account for equity. Moreover, in order to deal with the endogenization of the outputs, we modified the constraints accordingly. These distinctive features characterize the model, completely different in many aspects from previous works, and that can be considered as a meaningful research extension.

5.2.3.1 The EMS case

In this section, we detail the general model for determining the optimal location of the EMS facilities so as to assure a given quality of service. We consider each potential ambulance station as a DMU which uses two inputs and produces two outputs. The inputs are the capacity, measured in terms of maximum number of ambulances which can be sited at the station, and the operational cost. The outputs are related to the RT and the coverage that each station can guarantee.

Since the problem is not concerned with the more tactical problem of the EMS station dimensioning, we consider the demand arising from each station as a deterministic parameter. On the other hand, the allocation is influenced by the RT, that, as well recognized in the scientific literature, is subject to potentially dangerous fluctuations.

We model the RT between point i and station j as a random variable to faithfully represent the real behavior of the system. In particular, we assume that the RT is lognormally distributed, as commonly reported in the literature [95].

Then, station j is considered to be able to cover a demand point i only if the coverage probability, that is the probability that the RT t_{ij} is within the response time threshold (RTT, for short) τ , is greater than a reliability

level $1 - \alpha$, in which α is the risk level defined by the decision maker [161]. Hence, a demand point i can be covered only by stations $j \in J(i)$, where

$$J(i) = \{j | P(t_{ij} \leq \tau) \geq 1 - \alpha\}$$

We should mention that, although this probabilistic performance target (corresponding to the $(1 - \alpha)$ -quantile of the RT distribution) allows to exclude the more risky situations (those for which the probability of arriving on time is less than $1 - \alpha$), it is not able to exclude, even for the covered nodes the possibility that the service is provided too late. It is generally useful, in order to predict performance measures that are more closely related to patient outcomes [94, 95, 170], to consider the tail of the RT distribution to capture the delay of the RT with respect to the target.

Instead of considering a risk-neutral perspective leading to the use of the expected RT, we explicitly introduce the risk of a late answer to the emergency call. For a given random variable t_{ij} , the conditional expectation of the RT with respect to the threshold τ can be written as $\hat{t}_{ij} = E[t_{ij} | t_{ij} > \tau] = \int_{\tau}^{\infty} t_{ij} f_{ij}(t) dt$, where f_{ij} is the probability density function of t_{ij} and E represents the expected value. This measure give us useful information by focusing on the mean of worst case realizations, given that the threshold is violated.

For each DMU $j \in J$, we denote by O_{1j} the first output of our model representing the total expected response time, conditioned on the fact that it is greater than τ :

$$O_{1j}(\mathbf{y}) = \sum_{i \in I} \hat{t}_{ij} y_{ij}, \quad (o1)$$

It is notable that this output is an undesirable one. To deal with this case, we have regarded undesirable outputs as inputs [233]. Other methods can be used to incorporate undesirable outputs into the DEA model [99].

The second output is related to the expected demands covered by each station j which is denoted by:

$$O_{2j}(\mathbf{y}) = \sum_{i \in I} p_{ij} d_i y_{ij}, \quad (o2)$$

where p_{ij} is the probability that the RT is within the threshold τ and d_i is the demand associated with the demand point i .

To address the issue of equity in our model, we use the equity criterion of minimizing the total number of uncovered demand zones. Chanta et al. [69] used this criterion as a measure of fairness for the ambulance location problem in a tactical level.

The EMS location-allocation model balancing equity and efficiency, which hereinafter is called E^3MS , for the sake of brevity, is then defined by problem (5.1)-(5.13) amended with constraints (o1) and (o2), where $H(\mathbf{y}) = \sum_{i \in I} [\sum_{j \in J(i)} y_{ij}]$. We should remark that the focus and the contribution of our model are the combination of equity objectives with efficient location decisions for the design of service systems in the public sector and not the choice of the particular equity function to be used.

5.2.4 The exact solution approach

The proposed model is very difficult to deal with. It is a bi-objective mixed integer non-linear problem. In order to address the non-linearity, we adopt a standard linearization technique to transform the model into a bi-objective mixed integer problem. In particular, we can apply linear McCormick [169] envelopes to linearize the non-linear terms in constraints (5.7)-(5.8) resulting from the product of continuous and binary variables.

Defining the auxiliary variable z_{bij} , as $z_{bij} = w_{bj}y_{ij}$, we can come up with the following set of linear constraints

$$z_{bij} \geq w_b^{max}y_{ij} + w_{bj} - w_b^{max}, \quad \forall b \in B, \quad \forall i \in I, \quad \forall j \in J \quad (5.14)$$

$$z_{bij} \geq 0, \quad \forall b \in B, \quad \forall i \in I, \quad \forall j \in J \quad (5.15)$$

$$z_{bij} \leq w_b^{max}y_{ij}, \quad \forall b \in B, \quad \forall i \in I, \quad \forall j \in J \quad (5.16)$$

$$z_{bij} \leq w_{bj}, \quad \forall b \in B, \quad \forall i \in I, \quad \forall j \in J \quad (5.17)$$

where w_b^{max} is an upper bound of the output weight w_{bj} .

The remaining part is devoted to the presentation of an exact method for our bi-objective model.

Different solution approaches have been proposed in the literature for computing the Pareto front for the class of mixed integer bi-objective problems, among which the weighting method is the most popular one. The method scalarizes the objective function vector, reducing the problem to a single objective one. The method exhibits several drawbacks, for example,

the dependence of the performance on the weighting coefficients and the lack of guarantee that the optimal Pareto front is obtained. Another well-known and widely used method is the so called ϵ -constraint method which consists in the repeated solution of a sequence of single objective problems, amended with additional constraints. In particular, one objective is chosen to be optimized and the other objective is incorporated as an inequality constraint where the right hand side (the parameter ϵ) assumes different values in a certain range. Since the method can also find weak efficient solutions, a modification of ϵ -constraint, specifically tailored for multi-objective problems, has been presented in [166] by Mavrotas.

Our model is a bi-objective mixed integer optimization problem, where one objective consists in the maximization of the total number of covered demand points and the other in the maximization of the total system efficiency. By choosing the efficiency criterion as the main objective function and considering the equity criterion as a constraint, the ϵ -constraint problem $P(\epsilon)$ is defined as

$$\begin{aligned} \max \quad & f_1(X) \\ \text{s.t.} \quad & \\ & f_2(X) \geq \epsilon \quad P(\epsilon) \\ & X \in \mathcal{X} \end{aligned}$$

where X denotes the set of variables defined in the mathematical model of Section 5.2.3 plus the auxiliary linearization variables introduced above. In addition, \mathcal{X} is the set of constraints in (5.3)-(5.13), (o1), and (o2).

The idea is to construct a sequence of ϵ -constraint problems based on a progressive reduction of ϵ . The set of values that ϵ may take is finite since the first objective represents the number of covered demands which is finite and integer.

In order to define the range of variations for ϵ , we first solve the E^3MS problem considering only the objective function related to the equity, to find the ideal point $f_2^I = \max_{X \in \mathcal{X}} f_2(X)$. Then, we optimize the efficiency related objective function alone to obtain the ideal point $f_1^I = \max_{X \in \mathcal{X}} f_1(X)$. Next, we apply the lexicographic optimization to obtain a Pareto optimal

(non-dominated) solution. In practice, the lexicographic optimization is performed by solving the following problems

$$f_1^N = \max_{X \in \mathcal{X}} f_1(X)$$

$$s.t. f_2(X) = f_2^I$$

and

$$f_2^N = \max_{X \in \mathcal{X}} f_2(X)$$

$$s.t. f_1(X) = f_1^I$$

The points f_1^N and f_2^N are called the nadir values.

After this calculation, we obtain, for the second objective function, a discrete set of integer points in $[f_2^N, f_2^I]$, whose values are iteratively assigned to ϵ . Let the cardinality of this range be C and ϵ^c be the element in the c -th position in the range. The scheme of the algorithm is reported in Algorithm 12.

Algorithm 12: The exact ϵ -constraint method

```

1 Let  $EFF$  be the set of generated efficient solutions.
2  $EFF \leftarrow \emptyset$ .
3 for  $c = C, \dots, 1$  do
4   Solve  $P(\epsilon^c)$  and let  $f_1^*(\epsilon^c)$  and  $S^*(\epsilon^c)$  be the optimal objective
   function value and the optimal solution, respectively.
5   if  $c \neq C$  then
6     if  $f_1^*(\epsilon^c) < f_1^*(\epsilon^{c-1})$  then
7        $EFF \leftarrow EFF \cup S^*(\epsilon^c)$ .
8     end
9   end
10 end
11 return  $EFF$ .
```

Lemma 5.2.1. The point (f_1^N, f_2^I) is an efficient solution.

Proof: Suppose that the point (f_1^N, f_2^I) is not efficient. Therefore, it should exist another point (z_1, z_2) such that either:

1. $z_1 = f_1^N$ and $z_2 > f_2^I$
2. $z_1 > f_1^N$ and $z_2 > f_2^I$

3. $z_1 > f_1^N$ and $z_2 = f_2^I$

Cases (1) and (2) are in contradiction with the fact that f_2^I is an ideal point. The case (3) is in contrast with the fact that f_1^N is a nadir point. Therefore, the point (f_1^N, f_2^I) is an efficient point. Similarly, we can prove that the point (f_1^I, f_2^N) is also an efficient solution.

Theorem 5.2.1. *The set of solutions produced by the Algorithm 12 generates the exact Pareto front of our model.*

Proof: In order to prove the claim, we should first prove that each solution belonging to the set EFF is Pareto optimal and second, we should prove that the set of solutions produced by Algorithm 12 gives the exact Pareto front.

In order to prove the first part, let us consider a solution $S \in EFF$. S is the optimal solution of a problem $P(\epsilon^c)$ for some $c = C, \dots, 1$. We might have two different cases: either $f_2(S) = \epsilon^c$ or $f_2(S) > \epsilon^c$. If $f_2(S) = \epsilon^c$, S is a Pareto optimal solution, as shown for the general case (see [176]).

On the other hand, if $f_2(S) > \epsilon^c$, since f_2 involves only binary variables and ϵ^c is integer, $f_2(S) = \epsilon^f$, with $f > c$ and therefore, S is the optimal solution of a problem solved before problem $P(\epsilon^c)$.

To prove the second part, we should show that any other solution $S \notin EFF$ is not efficient. Assume that there is a solution S for the E^3MS problem which does not belong to the efficient set EFF . Since S is feasible, $f_2(S)$ assumes an integer value. Let $f_2(S) = \epsilon^c$ for some $c = C, \dots, 1$ and therefore, it is an efficient solution, which contradicts the hypothesis.

5.2.5 Computational results

In this section, we report on the results collected by applying the proposed model on a real case study for strategic EMS design in Edmonton city, Canada.

The information about the case study and data set is based on the previous researches carried out by Erkut et al. [94, 95] and Erdogan [93]. The data set includes the average and the standard deviation of lognormally distributed response times, the average number of calls, and the capacity of stations for a set of 180 demand points and 16 potential ambulance

stations. The interested reader can find the data on the following link: "http://professor.business.ualberta.ca/armanningolfsson/Research/Data". Since the aforementioned data set does not include the cost of siting stations, we generated them from the interval [10, 32].

To evaluate the performance of solutions, concerning the fairness criterion, we apply the Gini coefficient of the Lorenz curve, the within-group Gini indices associated with covered ($Gini_1^w$) and uncovered ($Gini_0^w$) zones, and the gross between-group index ($Gini^{gb}$) in Dagum's Gini decomposition.

It is notable that the within-group Gini indices $Gini_1^w$ and $Gini_0^w$ quantifies the inequality among each individual group of covered and uncovered zones, respectively while the gross within-group Gini index captures the inequalities between the set of covered and uncovered demand zones.

Based on Dagum's Gini decomposition, the Gini index is decomposed as follows [185]:

$$Gini = Gini_0^w + Gini_1^w + Gini^{gb}$$

$$Gini_0^w = \frac{\sum_{i \in N_0} \sum_{k \in N_0} |\sum_{j \in J} t_{ij} y_{ij} - \sum_{j \in J} t_{kj} y_{kj}|}{2|I| \sum_{i \in I} \sum_{j \in J} t_{ij} y_{ij}}$$

$$Gini_1^w = \frac{\sum_{i \in N_1} \sum_{k \in N_1} |\sum_{j \in J} t_{ij} y_{ij} - \sum_{j \in J} t_{kj} y_{kj}|}{2|I| \sum_{i \in I} \sum_{j \in J} t_{ij} y_{ij}}$$

$$Gini^{gb} = \frac{\sum_{i \in N_1} \sum_{k \in N_0} |\sum_{j \in J} t_{ij} y_{ij} - \sum_{j \in J} t_{kj} y_{kj}|}{|I| \sum_{i \in I} \sum_{j \in J} t_{ij} y_{ij}}$$

where $N_0 = \{i \in I | \sum_{j \in J} y_{ij} = 0\}$ and $N_1 = \{i \in I | \sum_{j \in J} y_{ij} = 1\}$.

To capture the randomness in our model, we define the Gini index based on the quantile of lognormally distributed RT by replacing the distance values in the traditional Gini index with the quantile of RTs [88]. In addition, we classify demand points into two separate groups of covered and uncovered zones. This enables us to investigate the inequality among both covered and uncovered zones. Although, due to the spatial configuration of potential stations, some demand zones are not covered within the RTT, in general and in the absence of auxiliary servers, the closest open station would respond to them. As a result, we assign the uncovered zones to their closest open station enabling us to capture the inequality among uncovered zones as well as between covered and uncovered zones.

The algorithm and problem modeling have been implemented in AIMMS 4.1 [48] and solved by ILOG CPLEX 12.6 [10]. The experiments were executed on a laptop Intel core i7 with a 2.7 GHz processor and 4 GB RAM. The average solution time is around 900 seconds.

The first set of computational experiments is devoted to the analysis of the behavior of the model as a function of different risk levels α and RTTs.

Figure 5.2 reports the trade-off between the risk level α and the equity components of the Gini coefficient discussed above for the single objective problem optimizing f_2 . The instance solved has $p = 12$ and the RTT equal to 10 minutes. The results show that lower values for equity measures of $Gini_0^w$, $Gini^{gb}$ are obtained for higher risk levels. This can be explained by considering that, the higher is the risk level that the decision maker is willing to bear, the larger is the number of covered demand points. For instance, by varying the risk level from 0.05 to 0.15, the total number of covered zones increases from 98 to 141. This, in turn, results in a more balanced system configuration (lower values for $Gini^{gb}$), where equality in the covered demands is sacrificed to achieve better service for the uncovered zones. The decrease in the within-group Gini coefficient $Gini_0^w$ and, on the contrary, the increase in the $Gini_1^w$ support this claim. We would like to remark that the total Gini index will increase slightly, indicating that the decrease in the reliability level $1 - \alpha$ (the increase in the risk level α) will not necessarily produce more equitable solutions.

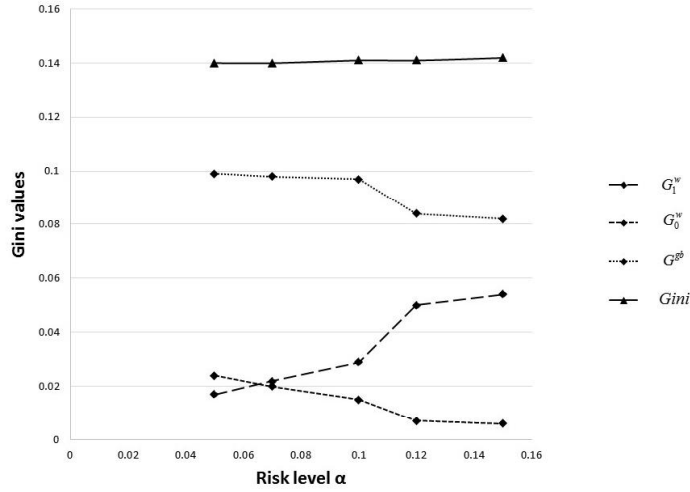


Figure 5.2: Risk level α vs. within- and between-group Gini indices

A note of caution should be in order. The results show a high sensitivity of the number of covered zones to the variations of the risk level. As an example, a 5%-increase in the risk level (from 0.05 to 0.1) will increase the number of covered demands by 18% while, by increasing α from 0.1 to 0.15, about 22% of uncovered zones can be covered. On the other hand, there is a dependence between the risk level and the RTT. By increasing the RTT from 10 to 15 minutes, the number of covered demands will increase to a different extent, considering different risk levels. For α equal to 0.05, 0.10, and 0.15, we obtain an increase of about 70%, 49%, and 24%, respectively, in the number of covered zones. This highlights the importance of choosing the appropriate values for the risk level and the RTT.

The second set of experiments is devoted to the validation of the proposed model, with respect to the pure equity- and pure efficiency- based models.

With this aim, we solve the E^3MS model, considering separately the two single objectives f_1 and f_2 considering $\alpha = 0.05$, $RTT = 10$ minutes and $p = 12$. The results have been reported in the Table 5.1, where the headers have the following meaning. The first column represents the criterion to be optimized, the second and third columns are the objective functions evaluated in the optimal solutions obtained. For the sake of clarity, we recall that the first two rows of the Table are related to the ideal points discussed in Section 3, whereas the last two rows are equivalent to the lexicographic

optimization, and the corresponding solutions are the nadir points. As we can see, the nadir point f_1^N outperforms the ideal point f_2^I with respect to the efficiency criterion. This reveals the fact that applying the combined E^3MS model enables the decision maker to improve the efficiency criterion while the highest equity level is guaranteed. The resulting configuration of the open stations as well as the assignment pattern exhibits the best performance for the equity criterion while producing more outputs and therefore increasing the overall efficiency. In a similar way, the nadir point f_2^N outperforms the ideal point f_1^I in the equity criterion. In this case, we are able to provide higher values for the equity measure even when the overall efficiency is maximum.

Table 5.1: E^3MS in comparison with pure efficiency and equity models

Optimization criterion	f_1	f_2
f_1	12	80
f_2	10.67	98
$f_1 f_2 = f_2^I$	11.299	98
$f_2 f_1 = f_1^I$	12	91

Table 5.2 reports the efficiency and equity performance measures of the Pareto optimal solutions for the case in which $\tau = 10$ minutes and $\alpha = 0.05$.

Table 5.2: The results for the Pareto front with $\alpha = 0.05$ and $RTT = 10$ minutes

solution	f_1	f_2	$Gini_1^w$	$Gini_0^w$	$Gini^{gb}$	$Gini$	SEC (%)	SCRT
1	11.327	97	0.017	0.025	0.098	0.140	91.90	307.21
2	11.357	96	0.017	0.026	0.097	0.140	92.80	307.29
3	11.392	95	0.016	0.027	0.097	0.140	93.69	308.42
4	11.434	94	0.016	0.028	0.096	0.140	94.58	308.51
5	11.478	93	0.015	0.029	0.095	0.139	95.54	309.67
6	11.540	92	0.015	0.030	0.094	0.139	96.35	310.86

The columns 2 and 3 show the values of efficiency and equity criteria for the Pareto optimal solutions. The other columns, but the last two, represent the equity components in Dagum's decomposition as well as the Gini's coefficient. In the last two columns, we reported the percentage of the system-

wide expected coverage SEC and the average RT above the threshold, SCRT in seconds.

The results in Table 5.2 indicate that the decrease in the number of covered demands will result in higher values for within-group Gini index assigned with uncovered zones $Gini_0^w$. In this situation, the disparity in uncovered zones is increased at the price of improving the efficiency. This agrees with our concept of equity, since the emphasis is on the uncovered demands. Therefore, it is reasonable that the equality in the uncovered group is higher, when the equity function is maximized. Higher values for the within-group Gini index for covered demands $Gini_1^w$, when our equity measure is maximized (solution 1), is an evidence of the fact that our equality measure is focused on the uncovered zones. In fact, the access to the service can be extended to cover more points, at the expense of a slight decrease in equality in covered demands. This shows also the trade-off between horizontal and vertical equity. Horizontal equity implies that all the demand points are considered in an equal manner. It means that suburban nodes are treated without discrimination. Vertical equity is concerned with the distribution of the service between different groups. By this definition, EMSs are equitable if they favor disadvantaged groups, compensating for overall inequities. In our case, horizontal equity is not in contrast with vertical equity, since, when suburban points are disadvantaged in terms of equity in access to the service (solution 6), the disparity between covered and uncovered zones reduces. This implies a decrease in the disparity between covered and uncovered zones. As a last remark, we should remark that the total Gini index exhibits only a slight variation over the Pareto front. This verifies that, although our main equity criterion is not explicitly known as a RT-based equity measure, it is not in contrast with the results of Gini performance measure.

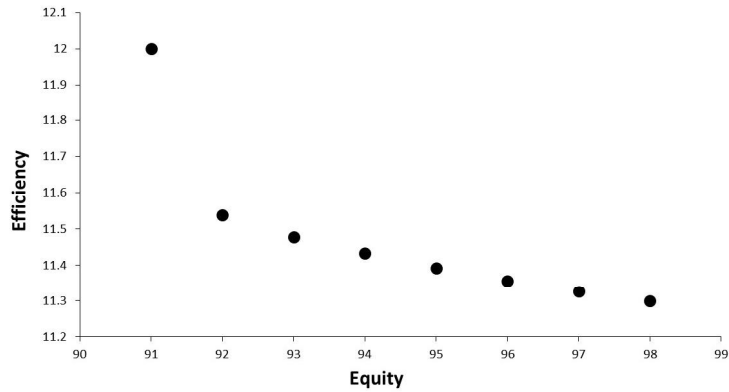


Figure 5.3: The trade-off between efficiency and equity criteria in Pareto front

The results in the last two columns show the trade-off between two outputs in the DEA model. It is evident that sacrificing the system equity will lead to better results for the system-wide expected coverage performance. In particular, the system-wide outputs of SEC and SCRT improve when the efficiency in the system is maximized. We should notice, as a shortcoming of our model, that the choice of the weights in the DEA model, and consequently the outputs, is out of the decision maker's control, and all the outputs are considered equal.

Figure 5.3 represents the exact Pareto front for the case in which $\alpha = 0.05$ and $\tau = 10$ minutes.

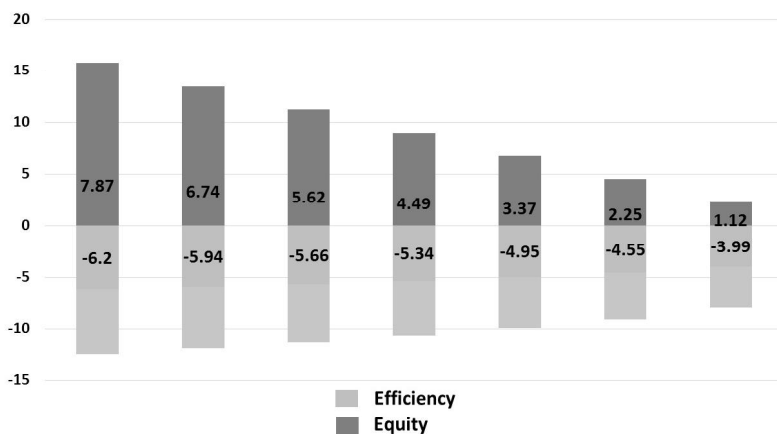


Figure 5.4: The percentage of trade-off: efficiency versus equity criteria in Pareto front

The trade-off between equity and efficiency for the Pareto efficient solutions is shown in Figure 5.4.

The last Pareto efficient solution in the Pareto front, which has the best value for the efficiency criterion, is considered as the baseline of this comparison. The positive values show the percentage of improvement in the equity criterion while the negative values indicate the decrease in the efficiency criterion. This provides the decision maker a framework to find that how much loss should undertake in one criterion for the benefit of a certain improvement in another criterion.

5.2.6 Conclusions and future research directions

We presented an integrated location-allocation model for balancing equity and efficiency with an embedded a DEA model. By endogenizing the outputs of the DEA, a better trade-off between equity and efficiency criteria is made. This issue is highly important since in location-allocation problems, the output of each DMU is dependent on the assignment variables. It is obvious that before siting facilities, we can not estimate the output parameters.

As already noticed, the choice of weights in the DEA model, and consequently the outputs, is out of the decision maker's control. In some cases, in which the decision maker has some priority levels on the outputs, this may probably result in undesirable results. To overcome this drawback, one suggestion can be the incorporation of weights restrictions into DEA models.

As another limitation of the proposed model, we can refer to the need for developing a tailored solution method which considers the peculiarity of the class of location-allocation models. This can be a motivation for future researches to explore the structure of the problem in more detail.

An application of the model to the strategic design of EMS was presented. The model has its own contribution in the EMS literature, since captures the uncertainty in RT in a novel way through the incorporation of a risk averse approach. In this applicative context, an interesting issue would be the exploration of the trade-off between equity and efficiency also in tactical-operational models, as for instance in location-relocation models (see [14, 222] and the references therein), where a stable performance should be guaranteed.

Although a general form for the equity function has been considered, there is an essential need for the investigation of equity measures designed especially for EMS systems. Future research will also focus on the application of the proposed model to other interesting applicative contexts in the health sector and in other fields.

Chapter 6

Location-allocation models in the primary health care sector

In this Chapter, we will present two location-allocation models supporting the decision-making in two interesting problem in the primary health care. At a first glance, it may seem that these two contributions are completely different from each other since they address separate problems with different characteristics and are related to different sectors in the health care system. For example, the first contribution presented in Section 6.1 is a multi-objective, static, and deterministic model which addresses a strategic-tactical decision problem while the other contribution, presented in Section 6.2, is, in fact, a single objective, multi-period and stochastic model, designed for a strategic decision-making problem. Despite all these distinctions, they share some similarities; for example, both models deal with location-allocation decisions, address the efficiency of health care systems as well as the accessibility of the service. The models belong to the same class of covering models, and are inspired by real case studies arising in the same geographical region. In the following, we will briefly introduce each contribution which will be later discussed in more detail in the corresponding Section.

Community Based Organizations (CBOs) are important health system stakeholders with the mission of addressing the social and economic needs of individuals and groups in a defined geographic area, usually no larger than a county. The access and success efforts of CBOs vary, depending on the integration between health care providers and CBOs but also in relation to the community participation level. To achieve widespread results, it is important to carefully design an efficient network which can serve as a bridge

between the community and the health care system. In Section 6.1, we address this challenge through a location-allocation model that deals with the hierarchical nature of the system explicitly. To reflect social welfare concerns of equity, local accessibility, and efficiency, we develop the model in a multi-objective framework, capturing the ambiguity in the decision makers' aspiration levels through a fuzzy goal programming approach. We also report the findings for the real case of Shiraz city, Fars province, Iran, obtained by a thorough analysis of the results.

In Section 6.2, we propose a multi-period location-allocation problem arising in nursing home network planning. In fact, we present a strategic model in which the improvement of service accessibility through the planning horizon is appropriately addressed. Unlike previous research, the proposed model modifies the allocation pattern to prevent unacceptable deterioration of the accessibility criterion. In addition, the problem is formulated as a covering model in which the capacity of facilities as well as the demand elasticity are considered. The uncertainty in demands within each time period is captured by adopting a distributionally robust approach. The model is then applied to a real case study for nursing home planning network in Shiraz city, Iran.

6.1 Enhancing Community based Health Programs in Iran: a multi-objective location-allocation model

6.1.1 Introduction

Over the past decades, major sociological and economic changes, such as the growing population and the increased urbanization, have placed increasing pressure on the Iranian health system [85, 148, 171]. In this context, the role of the health care system has been challenged ([171]), fostering considerable transformations. One of the most visible breakthroughs has been the establishment of a Primary Health Care network that has led to remarkable achievements in various areas, ranging from health education to endemic disease control [148, 171]. To reduce the gap between health outcomes in urban and rural areas, given the shortage of human and capital resources, the Primary Health Care system has relied on the community participation.

Locally-based groups, referred to as CBOs, which have operated in Iran for centuries, are playing an increasingly important role in social and economic developments, strengthening the foundations of an emergent civil society. In the case of Iran, CBOs are active in meeting and coping with the critical needs of underserved and vulnerable population. They promote community health and education, provide counseling services, drug addiction prevention, and improve service delivery, thereby explicitly reflecting social welfare concerns of equity and local accessibility.

Whether effective and equitable development can be ultimately achieved by CBOs depends on the community participation they foster, its main determinant being the geographical proximity [27].

The aim of the paper is to explore the potential for a structured quantitative approach for the strategic location of CBOs, considering the geographical distribution of specialized health care services.

The addressed challenges are manifold. From a modeling viewpoint, it investigates the optimal location of CBOs, proposing a novel multi-objective hierarchical location-allocation model. To the best of our knowledge, this is the first attempt to develop a mathematical model for this problem. Moreover, the model contributes to the location-allocation literature, since it combines the multi-objective and the hierarchical paradigms, thereby providing a faithful representation of the system and facing the concerns of the different stakeholders involved.

From a practical viewpoint, the insights derived from the research provide a systematic analysis of the trade-offs in the above mentioned applicative domain and shed light on determinant factors affecting the system efficiency and equity.

The first phase of this project consisted in a field-based research to better understand the system from an operational perspective. In this respect, for carrying out data collection, we collaborated with CBOs of Shiraz city, Fars Province, Iran. This collaboration with the practitioners was an essential component of this research rooted in a real-life application.

The rest of the section is organized as follows. In Subsection 6.1.2, the relevant literature is briefly reviewed. In Subsection 6.1.3, the characteristics of the system, the nature of the services provided, and the stakeholders' goals are discussed. Subsection 6.1.4 presents a mathematical programming

model for the specific case study. In Subsection 6.1.5, a solution methodology for solving the proposed model is described. Subsection 6.1.6 is devoted to the implementation of the proposed approach on a real case study enhanced by our findings and the policy implications suggested by the results in Subsection 6.1.7. Finally, in Subsection 6.1.8, some conclusions are drawn.

6.1.2 Literature review

We present a new multi-objective hierarchical location-allocation model for a health care network. To place the contribution of the paper in the right perspective, we restrict our focus to the literature on location-allocation and multi-objective models applied in health care.

6.1.2.1 Location-allocation models

There is a vast literature on the application of location-allocation models in the health care sector [204].

In [244], Syam and Corte presented a location-allocation model for specialized health care services, providing treatment and rehabilitation for strokes or traumatic brain injuries. The model minimizes the total cost, taking into account the effects of factors such as service centralization, facility overload costs, target utilization levels. In [33], Benneyan et al. introduced a location-allocation model for long-term decision makings in Veterans Health Administration sector, considering the fluctuation in demands. The objective function is a weighted sum of conflicting criteria including travel time, unoccupied capacity, and uncovered demands. In another paper, Zhang et al. ([273]) investigated the impact of client choice behavior in the location of preventive care facilities. The main aim is the maximization of the participation level which is proportional to the geographical proximity. In [116], a location-allocation model for the design of a primary health care network is presented. Three criteria, including the maximization of the coverage, the participation, and the total traveled distance are separately considered as accessibility measures. In [238], Song et al. introduced a new location-allocation model for the design of long-term health care services where the preferences of the patients are incorporated through closest assignment constraints. Kim et al. ([142]) proposed a location-allocation model for locating new public health care services in a network of existing private and public

facilities. The model deals with the preferences of low-income and high-income patients as well as the competition level among private and public facilities.

Following the hierarchical stream in location-allocation models, Galvao et al. studied the problem of locating perinatal facilities in Rio de Janeiro, where a nested hierarchy structure between different types of facilities exists [106, 107]. In [174], Mestre et al. presented a hierarchical model in which the improvement in the geographic equity of access is followed by the minimization of weighted distance traveled by the users within a hierarchical multi-service health care system. The study incorporates the efficiency of service as well as the operational costs. Sahin et al. proposed a 2-level hierarchical location model for locating facilities which provides blood services in Turkish Red Crescent [216]. In [237], Smith et al. presented a hierarchical location model for locating community health facilities in developing countries.

6.1.2.2 Multi-objective models in health care sector

The importance of the simultaneous consideration of multiple objective functions has been acknowledged in some of the location-allocation models reviewed in Subsection 6.1.2.1 (see also [33, 116]). Recognizing the compelling necessity of considering different criteria, Mohammadi et al. ([180]) proposed a bi-objective location-allocation model for the design of a reliable health care network. The first objective minimizes the total cost of treatment, transportation, and the expected cost of failure, while the second objective minimizes the sum of maximal accumulated travel time. In [242], Sun et al. presented a bi-objective allocation model for the optimal assignment of patients to hospitals during pandemic influenza outbreak. The two considered criteria are the minimization of the total distance traveled by patients to hospitals as well as the minimization of the maximum distance traveled by a patient to the assigned hospital. In [31], a multi-objective mathematical model for the allocation of beds to hospitals with uncertain demands has been presented. The model investigates the trade-off between three criteria: the cost of creating new beds, the number of nurses and physicians. In a recent research, Steiner et al. presented a multi-objective model to aggregate the health services offered in different municipalities into some microregions.

The aggregation is done to facilitate the management of resources [240]. The model provides a trade-off between three conflicting criteria, including maximizing the variety of services provided in each microregion, minimizing the inter-microregion travel distances, and maximizing the homogeneity of population in the microregions.

In [117], Guo et al. presented a bi-objective location-allocation model for the evaluation of community based health services. The model investigates a trade-off between cost and service where service is expressed as the total number of demand nodes that receive service within a given distance threshold. In another research, Mitropoulos et al. ([177]) developed a three-objective location model to find the most effective locations for locating health centers in Greece. The model investigates the trade-off between the total distance traveled by customers to their closest hospital, the underachievement in the minimum workload requirement of hospitals, and the average Data Envelopment Analysis (DEA) scores assigned to open hospitals. In [86], Davari et al. presented a bi-objective model for health care design in which the equity criterion has been considered. They also implemented fuzzy goal programming approach to solve the model. The model addresses the trade-off between two criteria, including the aggregate participation level in the network and the equity. To this end, they incorporate the attractiveness concept as a negative exponential function of travel time or distance. The participation criterion is evaluated by maximizing the total attractiveness captured by served demands, and equity criterion is expressed by the maximization of the minimum attractiveness in the network. In another work, Graber-Naidich et al. ([112]) presented a three-objective location-allocation model to account for cost, accessibility, and appropriateness of provided care for a primary care network design problem. In [63], Cardoso et al. presented a location-allocation model addressing long-term care network design in which three types of equity are considered, including equity of access, geographical equity, and socioeconomic equity. In [243], Syam and Cote developed a location-allocation model for establishing not-for-profit health care organizations. As the authors mention, the proposed model is an implicitly multi-objective model in which the objective function takes into account the cost of service and the penalty cost for the total unmet demands. The model also requires that a minimum service level be provided.

A hierarchical multi-objective location model for the design of hospital networks has been proposed by Mestre et al. [175]. The model considers the uncertainty in demands. In [236], Smith et al. introduced a bi-objective hierarchical location-allocation model with equity and efficiency criteria. The model is especially focused on public health services. Recently, another approach balancing equity and efficiency has been provided with an application to emergency service design [140].

Despite the quite rich literature reviewed, there are potential gaps and open issues yet to be investigated. Despite the large number of applications for location-allocation models in the health care field, there are only a few papers addressing the hierarchical nature of health care services [106, 107, 174, 216, 237], two of which, [175, 237], incorporate the multi-objective framework. Moreover, our multi-service hierarchical location model incorporates the preferences of referred recipients to upper level facilities. To this end, referral is only limited to a subset of upper level facilities which are within a specified threshold. This will facilitate the referral and encourage the recipients to continue their treatment which, in turn, will increase the level of participation. We recognized that except equity and accessibility, which have already been addressed ([175, 237]), the efficiency of existing upper level facilities is another important factor affecting the system performance. This issue is of paramount importance especially in hierarchical systems, where the quality of professional services provided at upper levels has a direct influence on the final outcome. The incorporation of efficiency helps the managers to recognize the most efficient facilities that deserve more financial support. On the other hand, considering the competition level component is another contribution of the present paper. The model tries to decrease the competition level among first level facilities by increasing the distance between any pair of those facilities. This gives servers located at optimal locations more chance to develop their service and to attract more recipients in a stable condition.

Another important issue usually neglected or only partially addressed in the health care literature is the incorporation of ambiguity in different stakeholders' preferences at different levels. To overcome this shortcoming, we

have considered the fuzzy set theory that enables the Decision Maker (DM) to express the preferences imprecisely and even as linguistic terms.

6.1.3 Problem description and formulation

The State Welfare Organization in Iran coordinates three different service providers: CBOs, Consulting Centers (CCs), and Addiction Treatment Clinics (ATCs). CBOs are small organizations with different headquarters in the city. The services provided by CBOs can be classified into two main components: basic consulting services and addiction prevention programs. Other services provided by CBOs are somehow related to these two types of services. They also hold workshops teaching life and work skills. The establishment of CBOs is essential for linking the district office and the local community. As a matter of fact, in recent years, CBOs have been successful in decreasing crime and addiction rates in poor neighborhoods [12, 203]. As a result, the local district supported the establishment of CBOs, technically and financially, providing human resources such as social workers and consultants. It should be mentioned that the CBOs neither interfere in addiction treatment nor provide the patients with methadone. In addition, they do not give the recipients professional consulting services. Instead, CBOs refer a portion of their recipients to other higher level facilities where they can receive more professional and cut-rate services, based on the needs and the severity of recipients' problems. In particular, professional CCs (both public and private) offer consulting services in different fields such as personal, family, child care and development as well as educational problems. They work under the supervision of district authorities that require accountability. ATCs work under either the supervision of district authorities or medical universities, and they offer inpatient and outpatient addiction treatment services.

The nature of this organization imposes a hierarchical structure into the system where CBOs are at the lowest level and the CCs and the ATCs provide service at the highest level. The service providers operate under a successively exclusive service hierarchy. This means that any upper level facility provides just its own level of service. In fact, although ATCs and CCs can also offer lower level services, they do not do so in most cases. It

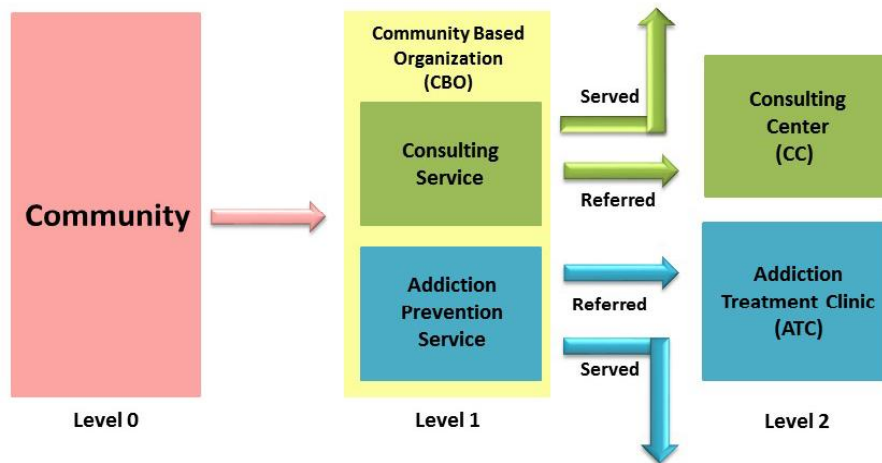


Figure 6.1: The system structure

should be noted that both CCs and ATCs have their own recipients, and the recipients referred by CBOs consist of only a portion of their customers.

The structure of the system is explained in Figure. 6.1. Note that level 0 denotes customers.

Many factors should be considered when designing a health care system of this kind, where the participation and completion of treatment rates are of paramount importance. The activity level of a facility is determined by the number of people who choose to seek its services. Therefore, service areas of the facilities cannot be neglected in the design process. One of the major determinants of participation to the programs offered by CBOs is the ease of access to that facility. The factors that impact the accessibility include the number, type, and location of the facilities. As well recognized [177, 259], the probability of participation and the motivation for continuing treatment decrease with the distance traveled. As a matter of fact, many patients stop their addiction treatment or do not attend their consulting sessions because of long distance.

In order to facilitate the accessibility of patients to the second-level facilities, the total distance traveled by recipients referred to the second-level facilities should be minimized. In addition, the referral is limited only to those second-level facilities which are within a critical distance. To improve the performance of the service network, some other criteria should be taken into account. For example, to improve the service level of any first-level facility and to increase system safety, first-level facilities should be dispersed throughout the city, avoiding strenuous competition levels. This implements an equity concern since it ensures equal access to CBOs to people living in different areas. In addition, the maximum distance between first- and second-level facilities should be minimized to assure that the second-level facilities are well dispersed over the area.

Besides this spatial efficiency, another concern of the stakeholders is the facility efficiency, expressed as a fraction of outputs produced for a given level of inputs. In this way, economical concerns, important especially in this context where the providers operate with limited resources, are also considered.

To address the efficiency and productivity of sited facilities, we apply the DEA framework, as a well-known tool for quantitative efficiency assessment which is widely applied in the operations research field. For more information about the DEA, the interested reader is referred to [81, 177].

6.1.4 The mathematical model

The mathematical model for the design of the system is presented in this section. We start by introducing the notation used throughout the article, and then present the objective functions and the constraints.

Sets/Indices:

I : is the set of demand points indexed by i

F : is the set of potential locations to site first-level facilities indexed by f

$F_i(T) \subseteq F$: denotes set of first-level facilities such that $d_{if} \leq T$, where T is the maximum threshold distance that recipients should travel to reach first-level facilities

K : is the set of service types (addiction treatment or specialized consulting)

provided by second-level facilities indexed by k

S_k : is the set of potential locations to site second-level facilities providing service type k

$S = \cup_{k \in K} S_k$: is the set of locations to site second-level facilities indexed by s
 $S_{kf}(\hat{T}) \subset S_k$: set of the second-level facilities providing service k such that $d_{fs} \leq \hat{T}$, where \hat{T} is the maximum threshold distance that recipients referred from a first-level facility should travel to reach a second-level facility

Input parameters:

h_i : denotes the amount of demand at demand point i , $i \in I$

$d_{a,b}$: denotes the distance between two arbitrary points or locations a and b

β_i^k : denotes the percentage of recipients from demand point i requiring service type k at the second-level

c : denotes the capacity (the maximum number of demand nodes) of the first-level facility f

C^k : denotes the capacity of the second-level facility s providing service k

η_s : denotes the amount of inefficiency associated with the second-level facility s

p : denotes the maximum number of first-level facilities

q^k : denotes the maximum number of second-level facilities providing service type k

M : denotes the maximum distance between first-level facilities ($M = \max_{f, f' \in F} d_{ff'}$)

m : denotes the minimum distance between first- and second-level facilities ($m = \min_{f \in F, s \in S} d_{fs}$)

Decision variables:

$$y_{if}^1 = \begin{cases} 1, & \text{if demand node } i \text{ is allocated to the first-level facility } f \\ 0, & \text{otherwise} \end{cases}$$

$$x_f^1 = \begin{cases} 1, & \text{if a first-level facility is established at potential site } f \\ 0, & \text{otherwise} \end{cases}$$

$$x_s^2 = \begin{cases} 1, & \text{if a second-level facility providing service } k \text{ is established at potential site } s \\ 0, & \text{otherwise} \end{cases}$$

w_{fs}^{12} = the amount of demand referred from the first-level facility f to the second-level facility s

D_1 : the minimum distance between any pair of first-level facilities

D_{12}^k : the maximum distance between any pair of first-level and second-level facilities offering service k .

The model reads as follows.

$$\min O_1(\mathbf{w}) = \sum_{f \in F} \sum_{k \in K} \sum_{s \in S_k} d_{fs} w_{fs}^{12}, \quad (6.1)$$

$$\min O_2(\mathbf{x}) = \sum_{k \in K} \sum_{s \in S_k} \eta_s x_s^2, \quad (6.2)$$

$$\min O_3(\mathbf{D}) = \sum_{k \in K} D_{12}^k - D_1, \quad (6.3)$$

$$\sum_{f \in F_i(T)} y_{if}^1 = 1, \quad i \in I \quad (6.4)$$

$$\sum_{i \in I} y_{if}^1 \leq c x_f^1, \quad f \in F \quad (6.5)$$

$$\sum_{f \in F} w_{fs}^{12} \leq C^k x_s^2, \quad k \in K, s \in S_k \quad (6.6)$$

$$\sum_{s \in S_{kf}(\hat{T})} w_{fs}^{12} - \sum_{i \in I} \beta_i^k h_i y_{if}^1 = 0, \quad f \in F, k \in K \quad (6.7)$$

$$D_{12}^k + (x_f^1 + x_s^2)(m - d_{fs}) \geq 2m - d_{fs}, \quad f \in F, k \in K, s \in S_k \quad (6.8)$$

$$D_1 + (x_f^1 + x_{f'}^1)(M - d_{ff'}) \leq 2M - d_{ff'}, \quad f, f' \in F (f \neq f') \quad (6.9)$$

$$\sum_{f \in F} x_f^1 \leq p, \quad (6.10)$$

$$\sum_{s \in S_k} x_s^2 \leq q^k, \quad k \in K \quad (6.11)$$

$$x_f^1, x_s^2 \in \{0, 1\}, \quad f \in F, s \in S \quad (6.12)$$

$$y_{if}^1 \in \{0, 1\}, \quad i \in I, f \in F \quad (6.13)$$

$$w_{fs}^{12} \geq 0, \quad f \in F, k \in K, s \in S_k \quad (6.14)$$

$$D_{12}^k \geq 0, \quad k \in K \quad (6.15)$$

$$D_1 \geq 0 \quad (6.16)$$

Departing from the rationale explained at the beginning of this section, the three objectives of accessibility, efficiency, and equity are operationalized into three different objective functions. The accessibility is represented by O_1 in (6.1) where the total travel time for individuals accessing second-level facilities, once referred from CBOs, is minimized. This assures spatial efficiency and maximizes at the same time the success probability of the program or the participation level.

The efficiency criterion is represented by O_2 in (6.2), which minimizes the total inefficiency score of second-level facilities in the system. The equity is represented by O_3 in (6.3). It includes two components: the maximum distance between any pair of open facilities offering similar services but at different levels, D_{12}^k , and the minimum distance between any pair of open CBOs, D_1 .

Constraints (6.4) state that recipients should be referred only to CBOs within a critical distance. Constraints (6.5) impose a limit on the number of quarters that any CBO can serve. Constraints (6.6) impose a limit on the number of referral from any open CBO to the upper level facilities. The next restrictions in (6.7) are associated with service referrals. They enforce recipients to be referred to second-level facilities which are within a critical distance. Constraints (6.8) define distance variable D_{12}^k , $\forall k \in K$, as the maximum distance between any pair of open facilities offering similar services at different levels. These restrictions in combination with (6.3) provide the minimum dispersion between any pair of lower and upper open facilities. Constraints (6.9) define dispersion variable D_1 as the minimum distance between any pair of first-level facilities. In a similar way, constraints (6.9) in combination with (6.3) assure that open CBOs are as much as possible dispersed throughout the area. Constraints (6.10) impose a limit on the number of first-level facilities to be opened. The set of constraints in (6.11) put an upper bound on the total number of upper level facilities to be established. Finally, constraints (6.12)-(6.16) define the type of variables.

In the next section, we will describe an approach for dealing with the multi-objective nature of the proposed model.

6.1.5 The fuzzy goal programming approach

For solving a multi-objective problem, a wide variety of multi-objective programming techniques can be applied. These methods differ from each other with respect to their different ways of converting the multi-objective model into a single objective one [241].

The goal programming approach is one of the most popular and practical approaches used [71]. In the goal programming framework, the DM imposes an aspiration (target) level on each criterion. Then, the deviations from these target levels are minimized. The most challenging issue in goal programming is the determination of aspiration levels as precise values. In fact, the main question to answer here is: how can the DM set precise values for targets, especially at the presence of some degrees of uncertainty which is inevitable in most real-world problems? To capture the ambiguity in the aspiration levels, the fuzzy set theory can be used [277]. The resulting method, known as fuzzy goal programming approach, enables the DM to express the target levels imprecisely and even as linguistic terms such as "approximately greater (less) than" or "approximately equal to".

In this case study, we apply the weighted additive model introduced by Tiwari et al. ([249]) and also the priority preemptive approach developed by Chen and Tsai ([73]). Both these models guarantee that the obtained solution is a Pareto efficient one.

Based on the main idea behind fuzzy goal programming approach, the mathematical model in (6.1) -(6.16) can be expressed as the problem of finding a feasible solution for the set of constraints in (6.17)-(6.18), as follows:

$$O_h(.) \leq O_h^I \quad h = 1, 2, 3 \quad (6.17)$$

$$(6.1) - (6.16) \quad (6.18)$$

where O_h^I represents the aspiration levels assigned with fuzzy goal in (6.17) $h = 1, 2, 3$. Note that \leq denotes the fuzziness in the fuzzy goal and can be interpreted as "approximately less than or equal to". Fuzzy goals in (6.17) represent fuzzy sets and can be identified by their membership functions indicating the degree of utility in achieving the target values.

The linear membership functions assigned to fuzzy goals in (6.17) are defined

as follows:

$$\mu_{O_h}(\cdot) = \begin{cases} 1 & O_h(\cdot) \leq O_h^I \\ 1 - \frac{O_h(\cdot) - O_h^I}{O_h^N - O_h^I} & O_h^I \leq O_h(\cdot) \leq O_h^N \\ 0 & O_h(\cdot) \geq O_h^N \end{cases} \quad (6.19)$$

where O_h^I and O_h^N are, respectively, the aspiration level (ideal solution) and the upper bound (nadir solution) assigned to fuzzy goal $\mu_{O_h}(\cdot)$, $h = 1, 2, 3$.

To obtain the ideal solution for the h^{th} criterion, we solve a single objective problem including criterion $O_h(\cdot)$ and the set of constraints in (6.1)-(6.16). The nadir solution, for the h^{th} criterion, is the maximum value that $O_h(\cdot)$ takes with respect to the ideal solutions of the other two criteria.

Applying the weighted additive model (Tiwari et al. [249]), an equivalent formulation for the initial fuzzy model in (6.17)-(6.18) is obtained as follows:

$$\max \sum_{h=1}^3 \omega_h \mu_{O_h}(\cdot) \quad (6.20)$$

$$(6.1) - (6.16) \text{ and } (6.19) \quad (6.21)$$

$$\mu_{O_h}(\cdot) \leq 1 \quad h = 1, 2, 3 \quad (6.22)$$

$$\mu_{O_h}(\cdot) \geq 0 \quad h = 1, 2, 3 \quad (6.23)$$

where ω_h ($\omega_h > 0$, $\sum_{h=1}^3 \omega_h = 1$) denotes the weight associated with the h^{th} fuzzy goal.

In multi-objective context, very often, the DM considers a priority structure in which some fuzzy goals have a higher priority for the achievement over the others and the fuzzy goals are ranked into different priority levels. Following this stream, Chen and Tsai ([73]) proposed the preemptive priority fuzzy goal programming model capturing the priorities imposed over goals. For our special case, the managers have considered two different priority levels for the fuzzy goals. The fuzzy goal associated with the criterion O_1 is ranked as the 1-priority, whereas the fuzzy goals associated with criteria O_2 and O_3 are ranked as the 2-priority. The preemptive priority model can be formulated

as follows:

$$\max \sum_{h=1}^3 \mu_{O_h}(\cdot) \quad (6.24)$$

$$\mu_{O_2}(\cdot) \leq \mu_{O_1}(\cdot) \quad (6.25)$$

$$\mu_{O_3}(\cdot) \leq \mu_{O_1}(\cdot) \quad (6.26)$$

$$(6.1) - (6.16) \text{ and } (6.19) \quad (6.27)$$

$$\mu_{O_h}(\cdot) \leq 1 \quad h = 1, 2, 3 \quad (6.28)$$

$$\mu_{O_h}(\cdot) \geq 0 \quad h = 1, 2, 3 \quad (6.29)$$

The algorithmic scheme of the presented fuzzy goal programming approach is reported below.

Algorithm 13: The exact fuzzy goal programming method

- 1 **for** $h = 1, \dots, 3$ **do**
 - 2 Solve the optimization problem with criterion $O_h(\cdot)$ and constraints
 (6.1)-(6.16).
 - 3 Let opt_h^* be the optimal solution.
 - 4 **end**
 - 5 **for** $h = 1, \dots, 3$ **do**
 - 6 Let $O_h^I = O_h(\text{opt}_h^*)$ and $O_h^N = \max_{i \neq h} (O_h(\text{opt}_i^*))$
 - 7 Build the membership functions $\mu_{O_h}(\cdot)$.
 - 8 **end**
 - 9 Solve the appropriate fuzzy goal programming model (either (6.20)-(6.23)
 or (6.24)-(6.29)).
 - 10 Return the obtained Pareto efficient solution.
-

Theorem 6.1.1. *The solution produced by the Algorithm 13 is a Pareto efficient solution for the proposed model.*

Proof: Let Z^* be the optimal solution of (6.20)-(6.23), where Z^* denotes the set of optimal variables $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{w}^*, \mathbf{D}^*)$. If Z^* is not an efficient solution for the model in (6.1)-(6.16), there exists another feasible solution Z such that $O_h(Z) \leq O_h(Z^*) \quad h \in \{1, 2, 3\}$ and for some $j \in \{1, 2, 3\}$ we have $O_j(Z) < O_j(Z^*)$. This results in $\mu_{O_h}(Z) \geq \mu_{O_h}(Z^*) \quad h \in \{1, 2, 3\}$ and $\mu_{O_j}(Z) > \mu_{O_j}(Z^*)$ for some $j \in \{1, 2, 3\}$. Since ω_h are strictly positive values, we conclude that $\sum_{h=1}^3 \omega_h \mu_{O_h}(Z) > \sum_{h=1}^3 \omega_h \mu_{O_h}(Z^*)$ which is a contradiction with the optimality of Z^* . The same results hold for the preemptive priority model.

6.1.6 Case study

As the sixth most populous city in Iran and the capital of Fars province, Shiraz city is well-known for being an important historical, cultural, political, economical, and commercial center. The specific geographical and economical characteristics of Shiraz city and its potential have encouraged many people to choose this city for immigration. Based on the national census of population and housing reports, prepared by statistical center of Iran in 2011, the population of the municipality of Shiraz is about 1700687 people and in the last 5 years, 175081 people have immigrated to this city.

The social structure of the municipality of Shiraz is also very complex due to the immigrants of different socio-economic backgrounds from other parts of the country, especially from rural areas and small cities in Fars province. This fact resulted in higher population density and increased exposure to deteriorated living conditions.

Most immigrants, coming from small cities and rural parts, are settled in suburban, poor, and crime-prone neighborhoods. The unemployment dilemma, addiction and drug-related crimes are serious concerns threatening the immigrant community. As a matter of fact, Fars province is among the first 12 provinces having the highest addiction prevalence rate [4]. In this situation, the main aim of the district authorities is to provide the residents with appropriate social care services, including addiction prevention and treatment programs as well as consulting services for improving life quality. The municipality of Shiraz includes nine municipal zones which are depicted in Figure. 6.2. To choose both the set of demand points and candidate facility locations, we concentrate on the municipal zones associated with higher crime rates. In 2010, Taghvaii et al. conducted a research to evaluate the crime rate in eight existing municipal zones of Shiraz [246]. Based on their findings, zone 2 reports the highest rate for drug-related crimes with 42.6%, followed by zones 5, 7, and 8 with 14.6%. The rate of drug-related crimes in other zones varies between 1.1% to 4.4%. Zone 9 is not surveyed in the research because it has been recently included in municipal zone divisions. Nevertheless, we consider it as a crime-prone area with a rather high rate of immigration located at the outskirts of the city. We discretized the municipal area in 84 population centers, representing the demand nodes. Currently, 24 CBOs, most of which located in the crime-prone areas with a

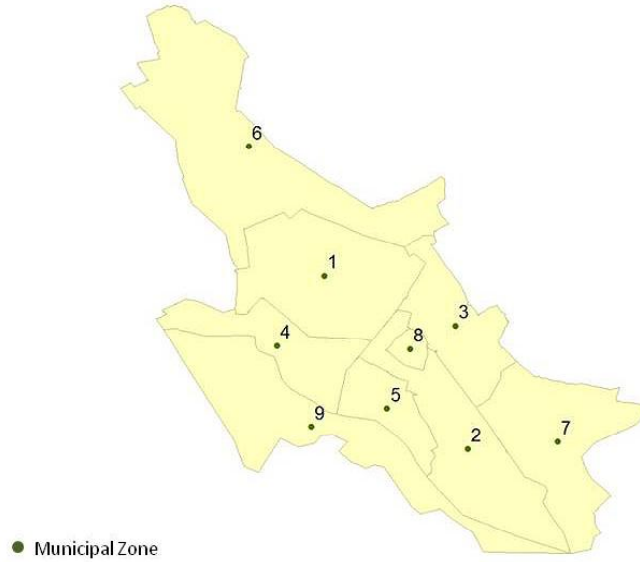


Figure 6.2: 9 Municipal zones of Shiraz city

high rate of immigration, provide lower level services concerning consulting programs and addiction prevention plans. Besides the 24 existing locations, we have considered 20 new locations in order to investigate the potential system improvement. Although some existing CBOs in municipal zones 1 and 6 are included in our study, we did not consider any new candidate CBO or demand node in these areas. The reason is that some CBOs in these areas did not have any considerable experience of referring their recipients to upper level facilities.

At present, 30 active CCs, including 29 private centers and a public one, are run under the supervision of the local authority. Currently, 10 ATCs, including 9 private clinics and a public one, provide addiction treatment services. Besides these 10 locations, we have considered 8 extra potential sites for establishing new ATCs.

Figure. 6.3 illustrates the spatial representation of demand nodes as well as both existing and potential sites for locating first-level and second-level facilities performed by ArcMap. We assume that at most 35 CBOs, 14 CCs, and 13 ATCs can be sited, i.e., $p = 35$, $p_1 = 14$ and $p_2 = 13$. To evaluate

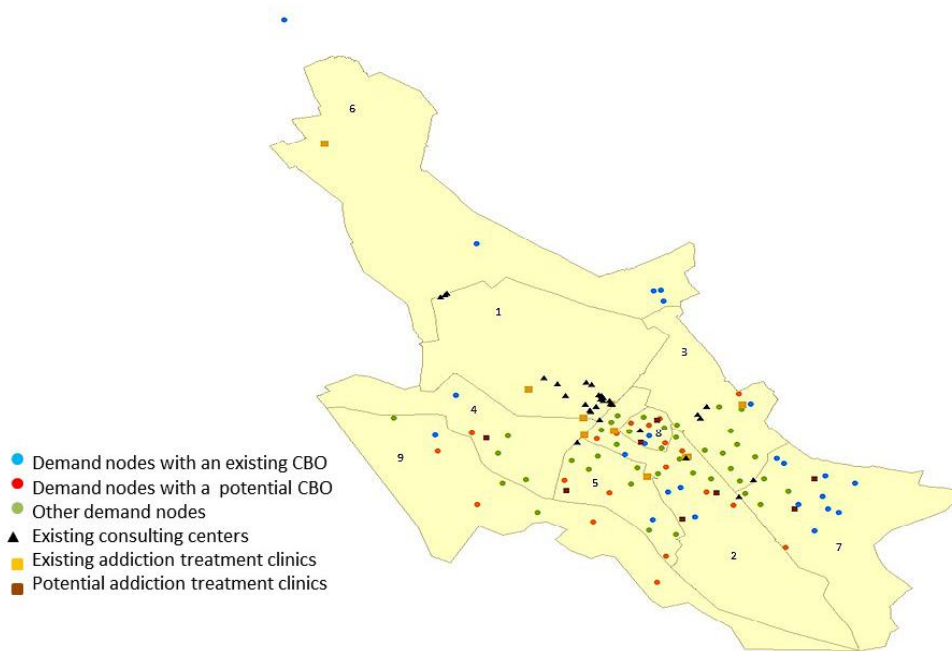


Figure 6.3: Spatial distribution of population centers and existing and potential locations of facilities

the efficiency score of each CC, we have considered the input and output parameters of Table 6.1. Then, the Klimberg’s simultaneous DEA model (see [143]) has been used to compute the DEA scores. At first sight by looking at

Table 6.1: Input and output parameters for CCs

Inputs	Outputs
Number of doctors	Number of non-referred treated recipients
Number of psychology consultants	Number of treated recipients referred by CBOs
Maximum capacity	

Figure. 6.3, it seems that CCs have not been sited in proper locations. For instance, we noticed that 17 out of 30 existing CCs are located in municipal zone 1, where demand for CBO services is considerably low. In other words, 57% of CCs are not well spread out over the city. We also found that the rate of referral from existing CBOs to some CCs is very low. These findings are confirmed by the efficiency scores of the CCs reported in Table 6.2.

Table 6.2: The DEA score of CCs

CC	Location coordinates(Km) (x, y)	DEA score	CC	Location coordinates (Km) (x, y)	DEA score
1	(-0.1438, 0.0215)	0.78	16	(-1.9436, 1.4414)	1
2	(-1.3876, 1.7628)	0.47	17	(1.7256, -1.554)	1
3	(-1.273, 1.5428)	0.74	18	(-1.6591, 1.928)	0.38
4	(2.3214, 0.758)	0.90	19	(2.2119, 0.954)	0.47
5	(3.9066, -3.7832)	0.52	20	(-8.0871, 7.8846)	0.63
6	(-1.8279, 2.1117)	0.60	21	(2.5888, 1.4285)	0.64
7	(-2.3633, 1.5692)	1	22	(-8.0174, 7.979)	0.57
8	(-1.7971, 0.6594)	0.64	23	(-2.3497, 2.7964)	0.60
9	(-2.1426, 1.1446)	1	24	(-1.7125, 2.0336)	0.38
10	(4.4771, -2.817)	0.41	25	(-3.1988, 2.0379)	0.74
11	(-2.1804, 1.2142)	0.43	26	(-3.5129, 2.7087)	0.74
12	(-2.7024, -6.695)	1	27	(-1.7049, 2.0256)	1
13	(-8.2694, 7.7555)	0.34	28	(-1.3378, 1.5832)	0.87
14	(-4.0818, 3.053.9)	0.35	29	(-2.1285, 2.6982)	0.59
15	(-1.7451, 1.8545)	0.35	30	(-1.6301, 1.7652)	1

The efficiency of ATCs has not been considered in our study given the limited number of existing ATCs. In order to instantiate the model, we have considered a coverage distance for CBOs equal to 2.5 kilometers to facilitate the residences' access and encourage them to interact with CBOs constantly. The appropriate coverage distance threshold for the second-level facilities has been set to 4.5 and 6 kilometers for CCs and ATCs, respectively.

To gather the needed information about the rate of referral from existing CBOs to upper level facilities and the number of demands at each demand point, we asked CBO managers to fill in our questionnaire. Other data about CBOs and ATCs were collected by contacting the managers of the local authorities. Collecting data as well as modeling the case study began in May 2013 and ended in March 2014.

6.1.6.1 Presentation of the results

This section is devoted to the presentation of the results obtained by solving the optimization problem for the case of Shiraz. The optimization problem was modeled using AIMMS 4.1 and solved by CPLEX 12.6, running on a 2.7 MHz personal laptop with 4 G RAM. The CPU time for all cases did not deviate from 3 seconds.

We present three research questions, relevant for the DMs of Shiraz, hereafter called scenarios *A*, *B*, and *C*. To provide the managers with alternative

solutions and to compare the current system with the solutions obtained by the model, we amended the problem (6.1)-(6.16) with the following constraint:

$$|J_e| - \sum_{f \in J_e} x_f^1 \leq q_e \quad (6.30)$$

where J_e denotes the set of existing CBOs' locations and q_e is the maximum number of existing CBOs to be closed. Hence, $|J_e| - \sum_{f \in J_e} x_f^1$ represents the number of existing CBOs which are closed and this number is limited above by q_e .

6.1.6.2 The current system

In this part, we evaluate the current network based on the three main criteria, including accessibility, facility efficiency, and equity.

Looking at Figure. 6.3, we observe that 9 out of 24 existing CBOs (the 38% of first-level facilities) have been located at municipal zone 7, while there are only 4 active CBOs (the 17% of first-level facilities) at municipal zone 2, which has the highest rate for drug-related crimes. In addition, municipal zones 5 and 8 have the same rate for drug-related crimes as municipal zone 7, but they are hosting only 1 and 2 active CBOs, respectively. This shows that the current spatial configuration of the first-level facilities is not consistent with the demands at different areas.

Moreover, the current location of CBOs is not balanced. As a matter of fact, the minimum distance among all pairs of the 24 existing CBOs is only 0.29 km, indicating a high competition level among CBOs.

Considering the current configuration of CBOs and the limitation on their capacities, the managers are able to cover only 52% of demand nodes within the threshold of 2.5 km. The percentage of uncovered demand nodes is around 44% of the total demands in the network. The minimum distance traveled by these recipients to their closest CBO is 4.72 kilometers, whereas the average is 3.69 kilometers. This is an evidence for the fact that, using existing CBOs, it is impossible to provide service for all quarters.

The location of uncovered demand nodes and their neighborhoods can guide the managers to find appropriate potential sites for establishing new facilities. There is no complete information about the referral pattern which is used in the system. However, by gathering the data, especially in the case of

CCs, we noticed that some recipients are referred to CCs, which are very far from their closest CBO. Even supposing the ideal case in which recipients are referred to their closest second-level facilities, the distance traveled to reach their closest CC and ATC is 6.14 and 8.48 kilometers, respectively. In the rest of this section, we investigate the optimal configuration of facilities under three different scenarios, referred to Case *A*, *B*, and *C*, that provide meaningful answers to a specific questions posed by the managers. For each case, a comparison with both the current system and other cases are provided.

6.1.6.3 Case A

This case answers the question: how should the current network be reorganized?

In other words, the managers adopt a risk-prone approach accepting the consequences of closing some of the existing CBOs in order to get the best possible spatial configuration for the whole system. In this case, the corresponding model is exactly the same as the model in (6.1)-(6.16). First of all, the solution obtained by the fuzzy goal programming method is compared with the solutions obtained by considering separately the three objective functions. The results are shown in Table 6.3. The first three rows in Table 6.3 represent the optimal objective function values resulting from the solution of single objective problems with criteria O_1 , O_2 , and O_3 , respectively while the last row shows the results for the fuzzy goal programming model, denoted by FGP.

Columns 5 – 7, in Table 6.3, show the utility degree in achieving the target value for each criterion.

Table 6.3: The optimal results for case A

	O_1	O_2	O_3	μ_1	μ_2	μ_3
O_1^*	5890.540	7.670	68.394	1	0	0
O_2^*	11518.775	2.721	67.559	0.11	1	0.25
O_3^*	12237.342	4.730	65.090	0	0.59	1
FGP	6998.799	2.827	65.350	0.83	0.98	0.92

The results show that, when O_1 is optimized, the highest utility degree for the accessibility criterion ($\mu_1 = 1$) is achieved at the price of sacrificing the

other two criteria completely ($\mu_2 = \mu_3 = 0$). Similarly, the utility degree associated with the efficiency criterion is at the highest level ($\mu_2 = 1$), while for the other two criteria this value is low ($\mu_1 = 0.11, \mu_3 = 0.25$) when O_2 is optimized.

Similar conclusions can be drawn for the third row. Unlike the first three cases, the optimal solution of the fuzzy goal programming model provides significantly higher utility degrees in target achievement for all the criteria. The optimal spatial configuration for first-level and second-level facilities obtained by the fuzzy goal programming method is shown in Figure 6.4. In addition, Table A.1 in Appendix A reports the open facilities for both the lower and the upper levels.

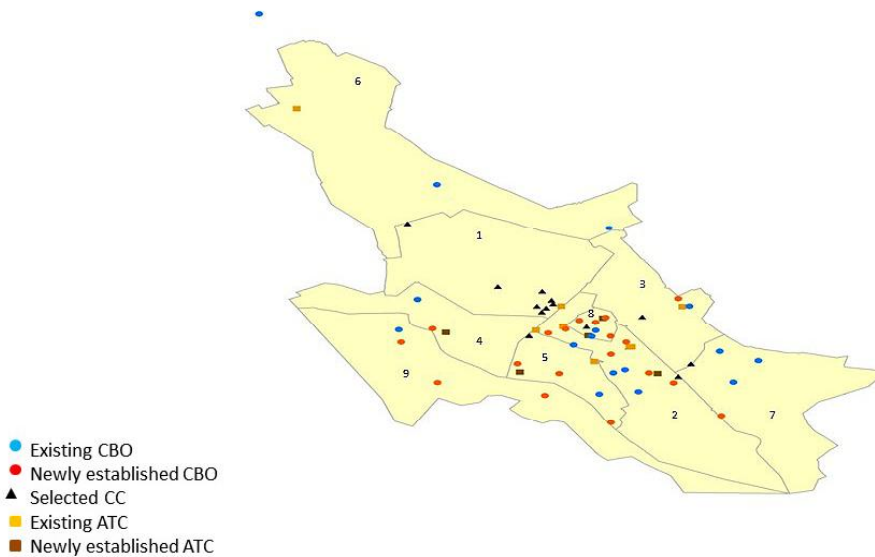


Figure 6.4: Spatial distribution of optimal locations for case A

Comparing the configuration with the current one, we observe that in the former case the minimum distance between any pair of open CBOs is 0.48, whereas in the latter is 0.29 kilometers. It is interesting to note that although 11 extra CBOs are established, to provide extra coverage, the minimum distance between CBOs increases by 65%. This finding shows that, for this

special case study, a higher number of CBOs does not exacerbate necessarily the competition level between them.

Hence, within an optimal configuration of CBOs, we are able to provide coverage for all demand nodes and at the same time to decrease the competition level between CBOs. By relocating $\frac{1}{3}$ of existing CBOs, the system reaches higher utility degrees for both accessibility and equity criteria.

Moreover, the results show that existing ATCs 4 and 9 are not included in the optimal solution and only 14 out of 30 CCs are selected. We should emphasize that, since 50% of active CCs in the new configuration are efficient, the recipients have more chance to be appropriately served than before.

In addition, the maximum distance between any pair of CBOs and CCs and ATCs, linked with referrals, are 4.340 and 5.80 kilometers, respectively. The average of the aforementioned distance for CCs (ATCs) is 2.953 (2.419). Comparing with the current system, this shows about 0.29% (20%) decrease in the maximum distance traveled by the referred demands to their assigned CCs (ATCs).

6.1.6.4 Case B

This case is relevant when the managers adopt a risk averse policy for the first-level facilities. In particular, this scenario let the DMs know how the current network could be upgraded without closing existing CBOs and establishing 11 new CBOs to provide coverage for the under-serviced areas.

The model associated with case *B* is (6.1)-(6.16) amended with the set of constraints $x_f^1 = 1, f \in J_e$ to preserve the existing CBOs. Table 6.4 shows the objective function values for single objective problems as well as the results for the fuzzy goal programming model.

Table 6.4: The optimal results for case B

	O_1	O_2	O_3	μ_1	μ_2	μ_3
O_1^*	6929.891	7.670	68.394	1	0	0
O_2^*	11518.775	2.721	67.559	0.04	1	0
O_3^*	12237.342	4.730	65.090	0	0.56	1
FGP	8324.161	2.827	65.541	0.73	0.98	1

Conclusions similar to the previous case can be drawn from Table 6.3.

Using the multi-objective approach, the total utility degree assigned to all

criteria is significantly higher than the total utility achieved for each single objective problem.

Comparing the solutions with the case A , we can notice a deterioration of both accessibility and equity criteria in the case B . In particular, the deterioration in the accessibility criterion is about 19% while the equity criterion is slightly (0.3%) worse.

The deterioration is related to the fact that the case A has more flexibility in finding the best configuration of first-level facilities. The value of the facility efficiency criterion is the same in both cases. This is an expected behavior, since the policy taken for the establishment of first-level facilities does not affect the efficiency of the second-level facilities (CCs).

To investigate the possibility of improving the accessibility criterion, we apply the preemptive priority model (6.24)-(6.29) considering the case that the accessibility criterion is prioritized over both facility efficiency and equity criteria.

The results of the preemptive priority model are summarized in Table 6.5, where PPFGP denotes the preemptive priority model.

Table 6.5: The optimal results for the case B with priority structure

	O_1	O_2	O_3	μ_1	μ_2	μ_3
PPFGP	7789.475	10.859	66.128	0.83	0.92	0.79

The results show that an improvement of 6.4% in the accessibility criterion can be obtained at the price of deteriorating facility efficiency (2.8%) and equity criteria (0.9%).

The spatial distribution of the optimal locations is shown in Figure. 6.5, whereas a Table A.1 is provided in the Appendix A.

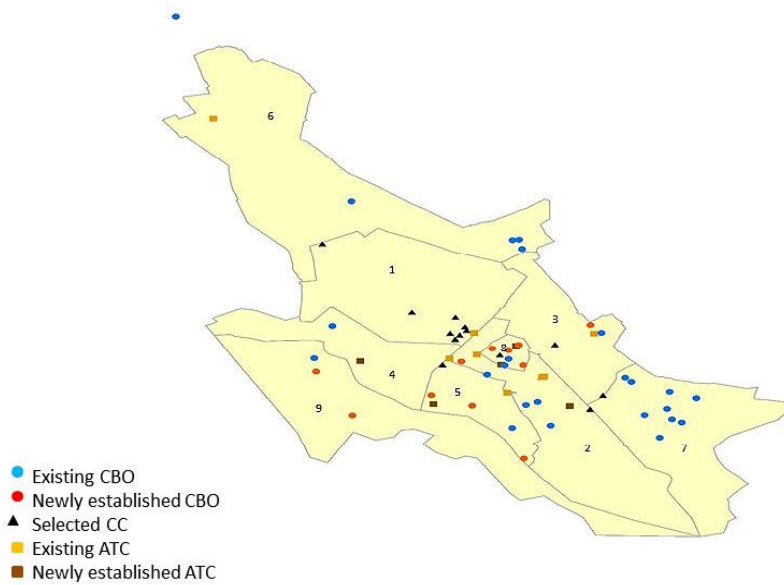


Figure 6.5: Spatial distribution of optimal locations for the case B

Comparing the optimized configuration with the current configuration we can notice that 8 out of 10 of the existing ATCs are located at optimal locations. This can guide the managers to relocate ATCs positioned in non-optimal locations.

6.1.6.5 Case C

Another relevant question is whether the current network could be upgraded while limiting the number of existing CBOs to be closed. The results have been obtained by solving the model (6.1)-(6.16) amended with the constraint (6.30), allowing at most 4 existing CBOs to be closed. The spatial distribution of the optimal locations is shown in Figure. 6.6. See also Table A.3 in Appendix A.

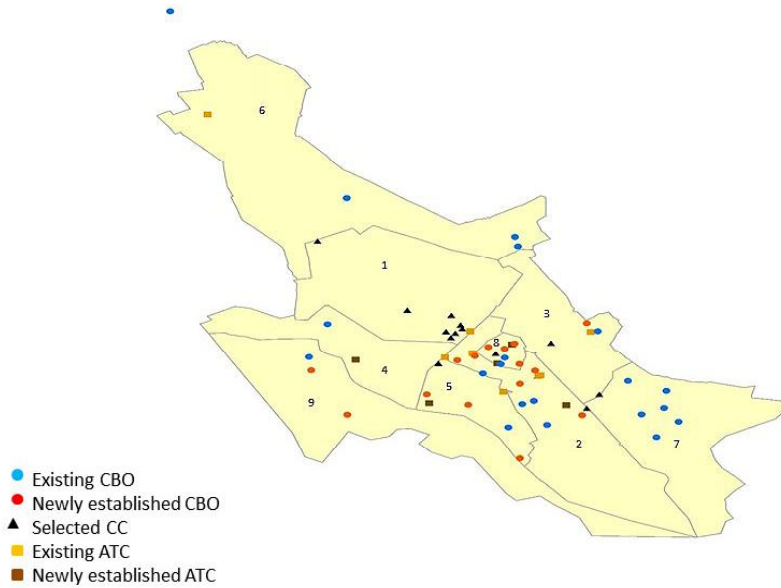


Figure 6.6: Spatial distribution of optimal locations for the case C

Table 6.6 indicates the optimal objective function values for the fuzzy goal programming problem. Comparing the results of the last row in Table 6.6, with similar results in Table 6.4, we can draw the following conclusions:

Table 6.6: The optimal results for case C

	O_1	O_2	O_3	μ_1	μ_2	μ_3
O_1^*	6025.212	8.141	68.394	1	0	0
O_2^*	11618.733	2.721	68.394	0.02	1	0
O_3^*	11749.450	5.02	65.264	0	0.58	1
FGP	7190.792	2.827	65.350	0.80	0.98	0.97

In terms of accessibility, this case provides better results than the case B , with an increase of 14%. The equity criterion is also improved, since the minimum distance between any pair open CBOs is 65% higher. As far as the efficiency criterion is concerned, both cases provide the same results. On the other hand, the case A provides higher accessibility levels (2.7% higher) than case C , whilst both cases provide the same results for the other two criteria.

6.1.7 Discussion

The analysis of these different planning contexts suggested important managerial insights. First of all, they showed that the existing CBOs are far from being able to meet the community needs, leaving many areas without access.

Regarding the accessibility criterion, scenario *A* provides the best results. This is expected as, in this case, the managers accept the risk of closing some existing CBOs. The model uniformly distributes the resources across the city, thereby ensuring the provision of services to the highest number of individuals.

Regarding the equity criterion, the maximum distance between the first-level and second-level facilities for all cases are the same and the only difference is in the minimum distance between CBOs. In this respect, cases *A* and *C* provide better values than the case *B*. This, on one hand, underlines a high competition level between CBOs in the current system and, on the other hand, it highlights the importance of adopting quantitative approaches in the configuration of these services.

In addition, the scenario *A* has the lowest value for the average distance traveled by recipients to reach ATCs (32% and 12% improvement over cases *B* and *C*, respectively), whereas in the current system, both the total weighted distance traveled to the second-level facilities and the average distance traveled to reach CCs/ATCs are high.

All the cases provide the same results for the efficiency criterion. This result is reasonable as this criterion is independent from the policy adopted for the selection of CBOs. It is remarkable that, for all cases, all efficient CCs are included in the optimal solutions.

6.1.8 Conclusions and future research directions

We proposed new combined multi-objective hierarchical location model arising in the context of community based services. The model accounts for three different policy objectives: equity, accessibility, and efficiency. The notion of ambiguity in the decision makers desiderata' has been taken into account, underlying the importance of the human component in the optimization phase. This is a salient model feature for the DMs, who were

able to express the objectives in linguistic terms. In addition, the investigation of the trade-off between these criteria enabled the managers to explore the structure of the system in detail and take decisions in which both the managers' perspectives and public viewpoints are considered. The proposed model was validated using real data for Shiraz city in Iran. By implementing the solution of the mathematical model, the managers could practically evaluate the improvement in the completion of therapy sessions and addiction treatments, especially for this case study.

The model built is quite general and can easily be adapted to different contexts with similar characteristics, although it was inspired by the situation of Shiraz city. As a future research deserving detailed examination, we mention the incorporation of qualitative measures in a comprehensive model. It would be interesting to see the trade-off between qualitative and quantitative measures for performance assessment.

Another interesting avenue for future research is the application of the VIKOR method to find the compromise solution that better suits the DM's preferences [234, 269].

6.2 A Multi-period Location–Allocation Model for Nursing Home Network Planning Under Uncertainty

6.2.1 Introduction

An aging revolution is taking place world–wide. Increased longevity is one of the most important success of our era but it also raises a challenge for health systems that are put under the pressure to reform their care organization to be sustainable for an aging society. These increased expectations should be reconciled with the limited resources available. In 2012, people aged 60 years and more were 0.8 billion, 11% of the world population. By 2030 they will number 1.4 billion and will represent up 17% of the world population. By 2050, this number will rise up to 22% [9]. In addition, increased urbanization and migration will result in older people living alone. Each country will need a comprehensive approach to make the necessary transformation and to meet the challenge. Improving services for older people entails to

consider each component of care system as, for instance, specialized clinics, home care services, and nursing homes.

The impact of the aging society will fall predominantly on the long-term care sector requiring appropriate re-design of the long-term care services, including supported self-care, home-based and, especially, home nurses. This contribution is motivated by the real problem of creating an efficient nursing home network to satisfy the elderly care demands in emerging countries. In particular, we shall focus on the Middle East considering Iran. In the last years, the Iranian society has experienced an exceptional increase of life expectancy: in 1960 this index was around 44 years against 73 years in 2012. It is supposed that even greater thresholds could be achieved with a widespread provision of public health care services in which all age groups are served equitably, including the elderly population (people aged 60 and over). The need of designing an efficient long-term care network is a quite new requirement in Iran, that for tradition, is a family-centered society. If in the past the idea of moving elderly family members to nursing home has always been disapproved, in recent years this trend has been inverted due to social changes, increased socio-economic difficulties and the opportunity to receive more professional care.

The problem of designing an efficient nursing home network has a strategic nature and involves decisions that have an impact over an extended planning horizon. The limited financial resources are typically spread over a time horizon making the adoption of a myopic policy, that ignores the inherent dynamic nature of the problem, highly inefficient. On the contrary, a long-sighted view, can be vital for the system financial survival.

To address this issue, we propose a multi-period location model that incorporates the dynamic evolution of the system throughout the planning horizon; with the provision of new financial resources, new facilities can be located and additional demand can be gradually satisfied. To this aim, in consecutive periods, newly established facilities are located and the assignment pattern is improved assigning all previously covered demands to closer facilities, if possible. In general, by applying this incremental approach, the distance between covered demands and facilities sequentially reduces over time and the accessibility (perceived as an important service level) is improved over the entire planning horizon. Indeed, in presence of finite financial resources the

system can fall, relatively easily and quickly, into very poor service levels, especially when the demand variability is high.

To the best of our knowledge, there is not any previous research on location literature addressing the strategic design of the nursing home network. In addition, although there is a vast literature on multi-period location models, only Albareda-Sambola et al. ([17, 15]) address the important issue of satisfying the demands in an incremental fashion, whereas the variability of the demand of service, along the horizon, has been typically neglected [62, 175]. This poses a challenge, since the determination of the optimal location of facilities at the beginning of the horizon should be made before the actual amount of demand is available. We deal with this issue by adopting a distributionally robust approach. We consider a general case in which only the mean and the deviation of the stochastic demands over each time period are known. Moreover, we opt for a risk-averse view point, considering the service levels as probabilistic constraints to be satisfied with a given reliability level [37, 55, 35]. We point out that there is a strong background on applying the chance constrained approach for strategic planning in the health care sector, supported by its risk-averse characteristic [36, 105]. As a risk-averse approach, the chance constrained paradigm allows the decision maker to capture the demand uncertainty and to exclude the more risky situations depending on the aversion level. Compared with risk-neutral approach ([62, 175]), in which only the expected values are considered, the risk-averse framework takes also the deviations of the uncertain parameters into account.

Another distinctive feature of the proposed model is the incorporation of the elasticity of the demand with respect to the distance traveled by the users (distance-elasticity of demand). There is a general consensus on the elasticity of demand in the health care sector [258, 259]. The explicit consideration of the distance-elasticity of the demand enables the managers to gain valuable information about the participation level for care services, supporting in a more realistic fashion the decisions of upgrading or extending the facilities in the long-term.

To address the elasticity of demands, we define a user-specific distance threshold reflecting the preferences of users to access the service. This threshold is different from the manager-specific threshold which is related to the

covering nature of the problem and represents the manager’s preferences.

The main contributions of the proposed model are as follows:

- (1) The model proposes a multi–period framework in which demand nodes are incrementally served.
- (2) Unlike previous related research ([17, 15]), the proposed model modifies the allocation pattern to prevent unacceptable deterioration of the accessibility criterion. In addition, the problem is formulated as a covering model in which the capacity of facilities is also considered.
- (3) The uncertainty in demands within each time period is captured by adopting a distributionally robust chance constrained approach. In addition, the model also incorporates the distance–elasticity of demands.

The rest of the section is organized as follows: Subsection 6.2.2 presents a brief review on existing relevant literature. Subsection 6.2.3 describes the problem and presents a stochastic formulation along with its deterministic equivalent counterpart. Subsection 6.2.4 presents the real case study and shows the improvements achievable when implementing the recommendations provided by the model in terms of location–allocation configuration in the nursing home network in Shiraz city. Finally, conclusions and findings are reported in Subsection 6.2.5.

6.2.2 Literature Review

There is a vast literature on the development of single period location–allocation models in the health care context ([34, 36, 140, 141]). Many existing researches have shown the advantage deriving from the adoption of a multi–period programming framework, when compared with a single period one, to deal with strategic location–allocation decisions. In particular, there is a wide literature on the application of multi–period location models for the public service sector ([197, 221]) and, especially, in the health care field [21].

Nevertheless, there are only a few papers in the literature addressing multi–period location–allocation models at the presence of stochasticity [162]. Among them, we refer to a recent work of Markovic et al. [162]. Recognizing the sparseness of literature, the authors presented a multi–period model to locate a set of flow–capturing facilities aimed at intercepting the stochastic

traffic flows with evasive behavior. The proposed model allows the adjustment of facility locations over different time periods. A Lagrangian relaxation heuristic is proposed and tested on two road networks.

Adopting a risk-neutral approach, Albareda-Sambola et al. proposed a multi-period location-allocation model under cost uncertainty [16]. They considered two alternative strategies, including the scenario-dependent case in which the decision locations are made gradually with the evolution of randomness over the planning horizon and the scenario-independent case, where the locations decisions are made in an a priori fashion at the beginning of the horizon. For the scenario-dependent strategy, they presented a multi-stage stochastic location model while for the a priori case, a two-stage model is defined.

In another paper, Nickel et al. presented a multi-period model for the facility location problem in supply chain, where demands and interest rates are affected by uncertainty and represented by a set of scenarios [189]. The problem is formulated as a multi-stage stochastic model in which the objective is expressed as the maximization of the total benefit and the achieved service level.

Hernandez et al. ([122]) studied a multi-period mathematical model for the prison selection problem in which the demands are represented as stochastic parameters. The model determines the location and the size of new facilities for each time period and the capacity upgrade for both existing and new prisons over the planning horizon. The objective function minimizes the opening and expansion costs, the costs of transferring the convicted inmates from the court to their assigned prison, and the cost of overpopulation in prisons which is, indeed, a penalty term. In order to address the accessibility issue, the objective function accounts for the closeness of the inmate's prison to the court which facilitates the frequent visit of inmates and their families. To incorporate the uncertainty in demands, the authors applied a scenario tree generation approach and then solved the resulted model using a branch-and-cluster coordination method. As a case study, the Chilean prison system has been considered.

To get the reader familiar with the main issues arising in the health care sector, in the following we review those researches addressing the location-allocation planning of health care facilities.

In [238], a mathematical model for designing a network of long-term care facilities is presented. To reflect the patient's preference, the model imposes the closest assignment property in which patients are assigned to the nearest open facility. They also recognized the changes in the demand pattern and suggested developing a multi-period model in which the variation of demand through different periods is captured.

In [267], Zahiri et al. proposed a multi-period location model for an organ transplant problem in which the uncertainty in input data (cost) is handled by using a robust probabilistic programming approach. They also extended the model to a bi-objective one in which the minimization of total traveled time is considered, underlining the importance of dealing with distance-based measures even when the allocation is not directly done between pair of facilities and users, but between pair of facilities.

Benneyan et al. ([33]) proposed a location-allocation model, as well as its extended multi-period counterpart, to address the fluctuation of demands over time, for Veterans Health Administration facilities. They considered the objective function as a weighted sum of conflicting criteria, including travel time, unoccupied capacity, and uncovered demands.

In [187], Ndiaye and Alfares presented a multi-period location-allocation model for the establishment of seasonal health care facilities serving transient populations. The objective function minimizes the sum of opening and operating costs as well as the total traveled distance. The adoption of the multi-period framework enables the managers to handle the seasonal variability in operating costs and demands. Although the coverage issue has not been considered in the model, improvement of the accessibility is obtained by the incorporation of a distance threshold.

Rodriguez-Verjan et al. proposed in [214] a multi-period location-allocation model for home care services to minimize the total cost in a multiple resource system. What distinguishes the paper from other works is the modelization of some peculiarities of health care systems, like the authorization, different resources, pathologies and their evolution in time.

In [110], Ghaderi and Jabalameli presented a multi-period location model considering budget constraint on investment during each period. The objective function minimizes the total travel and operating costs. Both fixed and operating costs for located facilities and constructed links over each period

time are considered. As a case study, they also presented an application of the model in the health sector.

Two different two-stage stochastic programming models for multi-period hospital network planning are presented in [62, 175]. The uncertainty in demand and supply is captured using different scenarios embedded in a two-stage stochastic framework. In the first model, the allocation decisions are postponed in the second stage, when the uncertainty realizes, whereas in the second model both location and allocation decisions can be taken in the first stage.

In [223], a multi-period location-allocation model for emergency blood supply scheduling problem was presented. A set of temporary blood facilities are located and assigned to the blood donors such that the total cost is minimized. The cost function is expressed as the total cost of transporting blood from blood facilities to the center as well as the cost of relocating blood facilities within consecutive periods. In addition, the total amount of unmet demands is penalized in the objective function. A coverage distance is imposed limiting the allocation of blood donors to blood facilities within the coverage radius. The number of available blood facilities over the planning horizon is fixed and a demand coverage constraint is imposed to guarantee the satisfaction of a specified percentage of demands. Enhanced with a Lagrangian relaxation solution approach, the model was implemented on a case study.

In a recent paper, Correia and Melo ([82]) proposed a multi-period location model in which the sensitivity of customers to delivery lead times has been incorporated. The novelty of the model comes from differentiating the customers who make the most contribution to the company's profit—and that should be responded on time—, from the others, for whom a maximum allowed delay is considered. Meanwhile, a subset of time periods over the planning horizon is specified in which strategic decisions such as the opening of new facilities, the closure of the existing ones, and the capacity acquisition decisions for new facilities are made. The tactical decisions about the distribution of services to customers can be made in any time period. Some additional assumptions, like considering different capacity levels for the facilities sited at potential locations or limiting the number of times that

the customers are responded with delay over the planning horizon, are also considered.

The most significant contribution of the reviewed literature relies on modeling features of health care problems which had not been addressed before. Despite their undeniable novelty, there are still some potential gaps to be filled. For instance, none of the aforementioned models addresses the importance and the possibility of improving accessibility and service level through the planning horizon. Neglecting the modification of the allocation pattern, when there is such a possibility, may result in an overestimate of the system performance. To partially address this issue, Albareda–Sambola et al. ([17]) proposed a multi–period incremental location model to serve demands incrementally over a discrete planning horizon. Later on, Albareda–Sambola et al. introduced, in [15], three different multi–period incremental location models, differing in the definition of variables and presented some computational comparisons.

To the best of our knowledge, the aforementioned papers are the only existing ones addressing incremental demand serving and budget limitation in a multi–period location problem. Although both contributions recognize that the allocation pattern might change through different time epochs, they do not address its negative results nor provide a solution for that. In addition, service levels were not considered.

Moreover, all the aforementioned studies, more or less, recognize the stochastic nature of problems in health care [33, 110, 214, 238], but only a few of them deal with uncertainty [62, 175, 267].

In addition, they share the same idea of minimizing the total cost or/and total traveled distance without considering the importance of the coverage concept in public health sector. It is notable that covering a particular demand node within the manager–specific distance threshold does not necessarily mean that all citizens of that zone (or at least a significant portion of them) will refer to the assigned facility, unless the preferences of the users in some way are incorporated. Hence, it might be impossible to come up with a single distance threshold in which both user and manager preferences are taken into consideration. In addition to the manager–specific distance threshold, which is related to the covering nature of the model, a user–specific

distance threshold can be defined reflecting the user preferences. This also facilitates the injection of distance–elasticity of demands.

We try to address these important issues, only partially investigated in the scientific literature.

6.2.3 Problem description and mathematical formulation

The nursing homes problem can be modeled as a covering location–allocation model in which a finite number of demand nodes (population centers) should be served by a number of facilities (nursing homes). Allocation of demand nodes to facilities is carried out by taking into account a (manager–specific) distance threshold, D , which prevents the assignment of demands to distant facilities, representing the covering nature of the model. The location of facilities is chosen from a set of prespecified potential sites. The assumptions of the proposed model are as follows:

- A demand zone is satisfied provided that a nursing home is located within the manager–specific distance threshold.
- Whenever a demand zone is satisfied, its demand should be also fulfilled in the subsequent periods.
- Each demand node must be served by at most one facility during any time period (single assignment property).
- Due to budget restrictions, at any time period, a limited number of nursing homes can be established.
- Once a nursing home is opened in a time period, it should be kept open for all subsequent periods.
- Each nursing home can host only a limited number of people which is fixed over all periods.

Using the multi–period framework, with the establishment of new facilities, some demand nodes, which were not covered during previous periods, might be served. To investigate the possibility of enhancing the service levels, the reassignment to farther facilities is prevented for all subsequent

periods, whereas previously covered demands can be only reassigned to closer facilities. Hence, throughout the planning horizon, the distance between demands and facilities is sequentially reduced. In general, this strategy helps to improve the accessibility criterion over consecutive periods.

In order to address the distance–elasticity of demands, we introduce another distance threshold, denoted by \hat{D} , which represents the preferences of users. Then, we define the "correction function", which is a function dependent on the user–specific threshold that estimates the expected portion of the demands from a population center that actually refers to the assigned facility. Obviously, all the people living in a covered population center will not necessarily refer to the facility assigned to. Hence, it is reasonable to have an estimation about the real value of referred demands.

In deterministic strategic planning, uncertainty is usually ignored and uncertain quantities are typically replaced by a single value forecast. While this approach could be accepted for single period problems, it is not realistic in multi–period problems where the horizon may span fifteen years. In these cases, a wrong decision may have serious consequences for many years, causing a deterioration of the system performance. Given the long–term nature of the problem, even forecasting the demand is difficult for the presence of unforeseen fluctuations in the population as well as inaccurate predictions of death rates. In our model, therefore, we explicitly account for uncertainty in the demand.

There are different approaches to deal with uncertainty: robust and worst–case methods often provide very conservative solutions. Chance constrained programming explicitly limits the probability of constraints violations. Since the main goal of the model is to ensure the provision of the service, we formulate the nursing homes problem as a probabilistic model with chance constraints.

In particular, our model includes, for each time period, a probabilistic constraint assuring that the stochastic demand can be covered by the opened nursing homes with a given probability value. This, in turn, enables the constraint to be violated with an acceptable violation probability, which is the risk the decision maker is willing to bear.

Since very often the assumption of full knowledge of the distribution of the random parameters fails, the uncertain demands are represented as ran-

dom variables with unknown probability distribution function, but known expected value and variance. Under this assumptions, we formulate a distributionally robust problem, in which the nursing home network is designed to minimize the total number of uncovered demand points, while the chance constraints on the capacity of each nursing home are formulated considering any distribution with the given mean and variance. This approach is especially beneficial for cases in which scant information about nursing homes demand is available.

6.2.3.1 The multi-period probabilistic location-allocation model

The following notation is used in the model formulation:

Sets and indices:

I : set of demand nodes indexed by i

J : set of potential facility sites indexed by j

$H = \{0, 1, \dots, T\}$: set of time periods indexed by t (time period 0 represents a dummy period)

Input Data and Parameters:

d_{ij} : shortest distance from demand node i to facility j

D : maximum acceptable service distance from the decision maker's point of view,

\hat{D} : maximum acceptable service distance from the user's point of view, ($\hat{D} \leq D$)

a_{ij} : element of the covering matrix equal to 1 if $d_{ij} \leq D$ and to 0 otherwise

$h_{it}(\omega)$: random demand generated at node i during period t

λ_{ij} : correction function, $\lambda_{ij} = (1 - \frac{d_{ij}}{D}) \rho_0$, where ρ_0 is the participation probability when travel distance is negligible and $\hat{d}_{ij} = \min(d_{ij}, \hat{D})$

Q_j : maximum amount of capacity for facility j

α : risk level

p_t : maximum number of facilities to be opened at period t

Decision Variables:

$$x_{ijt} = \begin{cases} 1 & \text{if demand node } i \text{ is allocated to facility } j \text{ during period } t \\ 0 & \text{otherwise} \end{cases}$$

$$y_{jt} = \begin{cases} 1 & \text{if a facility is located at site } j \text{ during period } t \\ 0 & \text{otherwise} \end{cases}$$

Considering the above notation, the mathematical formulation of the proposed multi-period probabilistic location-allocation model (*MPLM*) can be expressed as follows:

$$\min : \sum_{t=1}^T \sum_{i \in I} (1 - \sum_{j \in J} x_{ijt}) \quad (6.31)$$

s.t.

$$x_{ijt} \leq a_{ij} y_{jt} \quad \forall i \in I, \forall j \in J, t = 1, \dots, T \quad (6.32)$$

$$\sum_{j \in J} x_{ijt} \leq \sum_{j \in J} x_{ij(t+1)} \quad \forall i \in I, t = 1, \dots, T-1 \quad (6.33)$$

$$\sum_{j \in J} d_{ij} x_{ij(t+1)} \leq D(1 - \sum_{j \in J} x_{ijt}) + \sum_{j \in J} d_{ij} x_{ijt} \quad \forall i \in I, t = 1, \dots, T-1 \quad (6.34)$$

$$\sum_{j \in J} x_{ijt} \leq 1 \quad \forall i \in I, t = 1, \dots, T \quad (6.35)$$

$$P\left(\sum_{i \in I} \lambda_{ij} h_{it}(\omega) x_{ijt} - Q_j y_{jt} \leq 0\right) \geq 1 - \alpha \quad \forall j \in J, t = 1, \dots, T \quad (6.36)$$

$$y_{jt} \leq y_{j(t+1)} \quad \forall j \in J, t = 1, \dots, T-1 \quad (6.37)$$

$$\sum_{j \in J} (y_{j(t+1)} - y_{jt}) \leq p_{t+1} \quad t = 0, \dots, T-1 \quad (6.38)$$

$$x_{ijt} \in \{0, 1\} \quad \forall i \in I, \forall j \in J, t = 1, \dots, T \quad (6.39)$$

$$y_{j0} = 0 \quad \forall j \in J \quad (6.40)$$

$$y_{jt} \in \{0, 1\} \quad \forall j \in J, t = 1, \dots, T \quad (6.41)$$

The objective function (6.31) minimizes the number of times that a demand node is not covered during the time horizon. Constraints (6.32) state that each demand node can be assigned only to open facilities which are within the distance threshold D . The elderly are not the only users of the system

and the maximum acceptable service distance D was incorporated into the model not only for the sake of the residents of the nursing homes but their visitors and families too. In practice, the stay in nursing homes is more about years to days or months and families frequently travel to the nursing homes. Looking for possible ways to improve the accessibility of the systems, we came with the idea of imposing a maximum acceptable service distance, as a common idea in location–allocation literature. With the improvement of accessibility, the families are encouraged to go to the nursing homes more frequently which, in turn, has positive effects on the elderly’s life as well. To be accessible, the daily nursing homes should not require traveling more than 5 or 6 km per day. This is a common idea which has been addressed in the public facility location context; for instance, in [122], the closeness of the prisons assigned to the prisoners with their families is taken into account.

Constraints (6.33) state that whenever a demand is covered, it should be covered for all upcoming periods. The next set of restrictions in (6.34) imply that, for each period, any previously covered demand is reallocated to a closer facility if possible; otherwise, the demand is covered by the same previous facility. Restrictions (6.35) state that each demand node, at any period, is served by at most one facility. The probabilistic capacity constraints in (6.36) ensure that the probability of not exceeding the capacity of each candidate facility, during each period, should be greater than or equal to a prespecified reliability level $1 - \alpha$. It is worthwhile remarking that period-dependent capacities could be easily incorporated into the model by simply replacing Q_j with Q_{jt} . In our case study, based on the instructions imposed by the law to the governmental organization, the capacity of the nursing homes is fixed, and cannot be upgraded over the planning horizon.

Note that the term $\lambda_{ij}h_{it}(\omega)$ indicates the number of residents at demand node i who would, if assigned, effectively use the facility j during period t considering the distance–elasticity measure (see also [177, 259]).

Constraints in (6.37) state that once a facility is located, it should remain open for all the subsequent periods.

Restrictions (6.38) impose a limit on the maximum number of newly established facilities at any period. Note that $\sum_{j=1}^n (y_{j(t+1)} - y_{jt})$ indicates the number of newly established facilities in period $t + 1$ and y_{j0} represents the variable corresponding to the dummy period 0 which its value is set to zero.

Finally, restrictions (6.39)-(6.41) define the nature of decision variables. We should mention that, based on the application at hand some modifications might be made on the model. For instance, the single assignment property can be mitigated allowing to split the demands of a node over different facilities reflecting the users' preferences. This can be easily incorporated by relaxing the binary variables x_{ijt} , leading to a more tractable problem. It is worthwhile mentioning that the users' preferences are incorporated into the model through the elasticity of the demands, and that, very often, the single assignment property is required by managers willing to allocate the cumulative demands of each demand zone to a single facility.

We should remark that different attitudes (egalitarian and utilitarian) may be considered in any health care model. The egalitarian approach considers an equal weight for different target demand points, whereas the utilitarian one focuses on high populated demand zones. The current assumptions of the model reflects an egalitarian approach, covering as much zones as possible, along the horizon, regardless of the variability of the demands in each zone. The utilitarian approach could be implemented incorporating into the objective function the expected demand, as usual in the maximal covering literature. Since the proposed model was motivated by a real case study, the objective function reflects the decision makers' preferences, and focuses on covering as many areas as possible. This was also motivated by the fact that there is not a significant difference among demand levels over different areas.

6.2.3.2 The deterministic equivalent formulation

We show that the problem *MPLM* actually admits an explicit conic reformulation, which can then be conveniently solved using an outer approximation technique.

Let assume that the random demand value $h_{it}(\omega)$ follows an arbitrary distribution function but its mean (μ_{it}) and its variance (σ_{it}^2) are known. Specifically, we show that for any α value within $(0, 1)$, the distributionally robust chance constraint

$$\inf_{h_{it}(\omega) \sim (\mu_{it}, \sigma_{it}^2)} \mathbb{P} \left(\sum_{i \in I} \lambda_{ij} h_{it}(\omega) x_{ijt} - Q_j y_{jt} \leq 0 \right) \geq 1 - \alpha \quad (6.42)$$

is equivalent to the convex second-order cone constraint [40, 60]

$$\sqrt{\beta_\alpha \sum_{i \in I} x_{ijt}^2 \hat{\sigma}_{ijt}^2} - Q_j y_{jt} + \sum_{i \in I} \hat{\mu}_{ijt} x_{ijt} \leq 0, \quad \forall j \in J, t = 1, \dots, T \quad (6.43)$$

where $\hat{\mu}_{ijt} = \lambda_{ij} \mu_{it}$, $\hat{\sigma}_{ijt} = \lambda_{ij} \sigma_{it}$, and $\beta_\alpha = \frac{1-\alpha}{\alpha}$.

In fact, we can rewrite (6.42) as

$$\inf_{\mathbf{h}_t \sim (\hat{\mathbf{h}}, \Gamma)} \mathbb{P}(\mathbf{h}^T \tilde{\mathbf{x}} \leq 0) \geq 1 - \alpha \quad (6.44)$$

where \mathbf{h}_t is the vector of $(h_{1t}, h_{2t}, \dots, h_{|I|t}, -Q_j y_{jt})^T$ and $\tilde{\mathbf{x}}_{jt}$ denotes the vector $(\lambda_{1j} x_{1jt}, \lambda_{2j} x_{2jt}, \dots, \lambda_{|I|j} x_{|I|jt}, 1)$. In a similar way, $\hat{\mathbf{h}}$ and Γ represents the vector of expected values and the covariance matrix associated with \mathbf{h}_t . From now on, for the sake of simplicity, we denote \mathbf{h}_t and $\tilde{\mathbf{x}}_{jt}$ by \mathbf{h} and $\tilde{\mathbf{x}}$, respectively.

Let assume that $\mathbf{h} = \hat{\mathbf{h}} + \Gamma_f \mathbf{z}$, where $\mu_{\mathbf{z}} = \mathbf{0}$, $\sigma_{\mathbf{z}}^2 = \mathbf{I}$, and Γ_f is the full-rank factorization matrix such that $\sigma^2(\mathbf{h}) = \Gamma_f \Gamma_f^T$. The following two cases are possible.

1. $\Gamma_f^T \tilde{\mathbf{x}} \neq \mathbf{0}$. In this case, we have

$$\sup_{\mathbf{h}_t \sim (\hat{\mathbf{h}}, \Gamma)} \mathbb{P}(\mathbf{h}^T \tilde{\mathbf{x}} \leq 0) = \sup_{\mathbf{z} \sim (\mathbf{0}, \mathbf{I})} \mathbb{P}(\mathbf{z}^T \Gamma_f^T \tilde{\mathbf{x}} \geq -\hat{\mathbf{h}}^T \tilde{\mathbf{x}}) = \frac{1}{1 + q^2} \quad (6.45)$$

where $q^2 = \inf_{\mathbf{z}^T \Gamma_f^T \tilde{\mathbf{x}} > -\hat{\mathbf{h}}^T \tilde{\mathbf{x}}} \|\mathbf{z}\|^2$ (The last equality holds based on Theorem 9 in [165]).

- (a) If $\hat{\mathbf{h}}^T \tilde{\mathbf{x}} > 0$, then by taking $\mathbf{z} = \mathbf{0}$, we can obtain the infimum $q^2 = 0$.
- (b) If $\hat{\mathbf{h}}^T \tilde{\mathbf{x}} \leq 0$, then the problem is expressed as determining the squared distance from the origin of the hyperplane $\{\mathbf{z} | \mathbf{z}^T \Gamma_f^T \tilde{\mathbf{x}} = -\hat{\mathbf{h}}^T \tilde{\mathbf{x}}\}$ which is solved by taking $q^2 = \frac{(\hat{\mathbf{h}}^T \tilde{\mathbf{x}})^2}{\tilde{\mathbf{x}}^T \Gamma_f^T \tilde{\mathbf{x}}}$.

So we have

$$q^2 = \begin{cases} 0 & \tilde{\mathbf{x}}^T \hat{\mathbf{h}} > 0 \\ \frac{\tilde{\mathbf{x}}^T \hat{\mathbf{h}}^2}{\sigma(\tilde{\mathbf{x}}^T \hat{\mathbf{h}})} & \tilde{\mathbf{x}}^T \hat{\mathbf{h}} \leq 0 \end{cases}$$

which represents a closed-form expression for q^2 . Hence, the constraint in (6.44) holds iff $\frac{1}{1+q^2} \leq \alpha$. The last equation holds iff $\tilde{\mathbf{x}}^T \hat{\mathbf{h}} \leq 0$ and $(\tilde{\mathbf{x}}^T \hat{\mathbf{h}})^2 \geq \sigma(\tilde{\mathbf{x}}^T \hat{\mathbf{h}}) \frac{1-\alpha}{\alpha}$ which is equivalent to (6.43).

2. $\Gamma_{\mathbf{f}}^T \tilde{\mathbf{x}} = 0$. In this case, we simply conclude that $\tilde{\mathbf{x}}^t \Gamma \tilde{\mathbf{x}} = 0$ which results in

$$\inf_{\mathbf{h} \sim (\hat{\mathbf{h}}, \Gamma)} \mathbb{P}(\mathbf{h}^T \tilde{\mathbf{x}} \leq 0) = 1, \text{ if } \tilde{\mathbf{x}}^T \hat{\mathbf{h}} \leq 0 \quad (6.46)$$

Considering $\tilde{\mathbf{x}}^t \Gamma \tilde{\mathbf{x}} = 0$, the equivalency of (6.42) and (6.43) is obtained.

By replacing restrictions (6.36) with (6.43) in *MPLM*, we come up with an integer non-linear deterministic equivalent formulation named as (*NDMPLM*). To solve the model, a linearization of the model could be applied (as shown in Subsection A.0.2 in Appendix A) and then off-the shelf softwares such as CPLEX could be used.

Although, in theory, a linearized problem is computationally more attractive, in practice, the solution of the linearized MIP model, even for medium size problems, can be significantly time-consuming. In fact, the linearized model involves $3 |I| |J| (|H| - 1)(|I| + 1)$ more constraints and $|I| |J| (|H| - 1)(|I| + 1)$ more binary variables than the nonlinear model. Considering the special structure of the model, we can show that its continuous relaxation is convex.

Lemma 6.2.1. *The function $F_{jt}(\mathbf{x}, \mathbf{y}) = \sqrt{\beta_{\alpha} \sum_{i \in I} x_{ijt}^2 \hat{\sigma}_{ijt}^2} - Q_j y_{jt} + \sum_{i \in I} \hat{\mu}_{ijt} x_{ijt}$ is convex.*

Proof: We can rewrite F_{jt} as $Z(\mathbf{x}) - Q_j y_{jt} + \sum_{i \in I} \hat{\mu}_{ijt} x_{ijt}$ in which $Z(\mathbf{x}) = \sqrt{\beta_{\alpha} \sum_{i \in I} x_{ijt}^2 \hat{\sigma}_{ijt}^2}$, or equivalently, $Z(\mathbf{x}) = \sqrt{\mathbf{x} \beta_{\alpha} \hat{\sigma}^2 \mathbf{x}^T}$. As $Q_j y_{jt}$ and

$\sum_{i \in I} \hat{\mu}_{ijt} x_{ijt}$ are linear terms, it suffices to show the convexity of $Z(\mathbf{x})$. Since $\beta_{\alpha} \hat{\sigma}^2$ defines a semidefinite positive matrix, there is a Cholesky decomposition for it as $\beta_{\alpha} \hat{\sigma}^2 = L L^T$. Hence, we can rewrite Z as $Z = \sqrt{\mathbf{x} L L^T \mathbf{x}^T} = \|\mathbf{x} L\|$, where $\|\cdot\|$ is the Euclidean norm.

Assume $\mathbf{x}_1, \mathbf{x}_2$ as two arbitrary vectors and $\lambda \in (0, 1)$. We have

$$\begin{aligned} \|(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) L\| &\leq \|\lambda \mathbf{x}_1 L\| + \|(1 - \lambda) \mathbf{x}_2 L\| \\ &= \lambda \|\mathbf{x}_1 L\| + (1 - \lambda) \|\mathbf{x}_2 L\|. \end{aligned}$$

and the proof is complete.

Therefore, we may apply the outer approximation algorithm (AOA) on *NDMPLM* to obtain the global optimal solution ([224]).

6.2.4 Case Study

6.2.4.1 Case study description and input data

In this section, we apply the proposed model on a real case study for the nursing home network design problem in Shiraz city, the sixth most populous city in Iran and the capital of Fars province. The model has been implemented in AIMMS 4.1 and solved by AOA [48]. The experiments were executed on a laptop Intel core i7 with a 2.7 GHz processor and 4 GB RAM. The average solution time, for all experiments, was less than 50 seconds.

For the current case study, an extended planning horizon including three periods from 2015 to 2025 has been considered.

Currently, seven nursing homes provide the residents with elderly care services such as rehabilitation, education, and welfare services [5]. Nursing homes are allowed to admit at most seventy recipients. The municipality of Shiraz, including nine municipal zones, is divided into 76 population centers, based on postal divisions [6]. Each population center represents a demand node in this study.

In addition to the location of the seven existing nursing homes, 17 more candidate locations have been considered. This enables the managers to site new facilities as well as relocating the existing ones, if necessary.

Tables A.4, A.5, and A.6 in Appendix A, show the location coordinates of all the population centers and the candidate facility locations. See also Figure. 6.7.

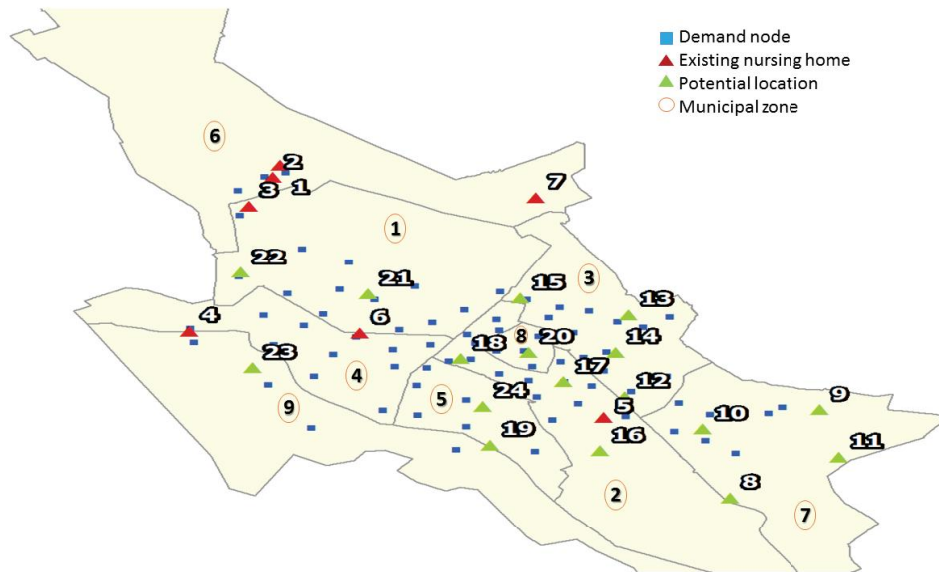


Figure 6.7: Spatial distribution of demand nodes and the locations of nursing homes in Shiraz city

To estimate the Shiraz population, we applied the population projection data extracted from the UN population projection reports [7]. To estimate the elderly population during different years, we set the fertility rate based on the medium fertility rate scenario, introduced by the Population Division of the Department of Economic and Social Affairs of the United Nations [183].

Analyzing the statistical reports published by the Shiraz municipality [6], from 2006 to 2009, we found approximately an identical trend for population distribution of each municipal zone over different periods. The estimated demand at each demand zone was considered as its expected value and the variance was set to 0.2 of the expected value. We also assumed that the demand of each municipal zone is uniformly distributed among its nodes. Table 6.7 shows the elderly population at any municipal zone for the next years.

Table 6.7: Estimated elderly population

Municipal zone	Elderly population during different years		
	2015	2020	2025
1	17096	21743	27840
2	17695	22505	28817
3	16188	20589	26363
4	17889	22751	29132
5	12778	16251	20809
6	13823	17581	22511
7	12772	16244	20799
8	5029	6396	8190
9	9843	12519	16029

The Euclidean distance was used to measure the travel distance between demand nodes and the nursing homes sites (see Appendix A). Other characteristics of the problem are summarized in Table 6.8.

Table 6.8: Case study inputs

p_t									
$ I $	$ J $	T	$t = 1$	$t = 2$	$t = 3$	α	D	\hat{D}	Q_j
76	24	3	2	3	5	0.05	5	5	70

6.2.4.2 Results and findings

Figure. 6.8 represents the optimal location of nursing homes, obtained by solving the model with $\alpha = 0.05$. The circles around the optimal locations 3, 13, 19, and 23 represent the considered coverage radius.

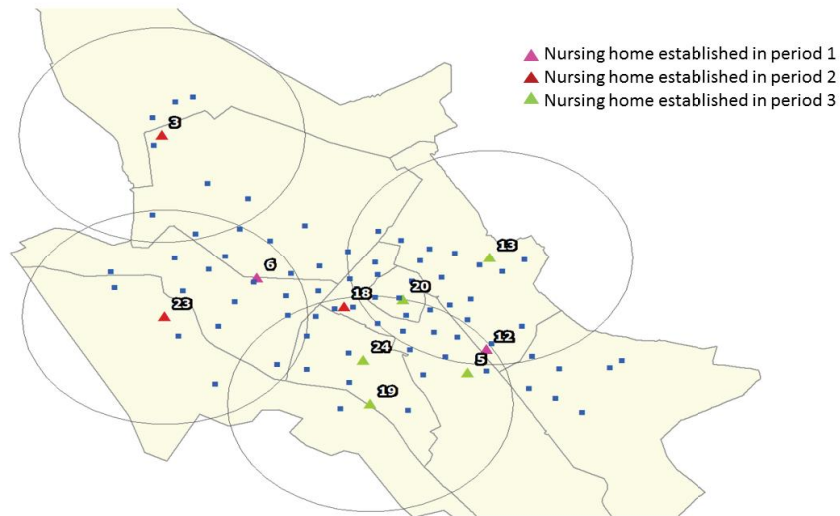


Figure 6.8: Spatial distribution of optimal nursing homes' locations during each period

In addition, Table 6.9 shows the optimal location of active nursing homes for different periods.

Table 6.9: The optimal nursing homes' sites

Nursing home number	Time period		
	$t = 1$	$t = 2$	$t = 3$
location3		*	*
location5			*
location6	*	*	*
location12	*	*	*
location13			*
location18		*	*
location19			*
location20			*
location23		*	*
location24			*

Based on the obtained results in Table 6.9, only three out of seven existing nursing homes are present in the optimal solutions. This shows that the current configuration of system can be improved by relocating nursing homes 1, 2, 4, and 7.

By calculating the distance between nodes and their assigned facility, we observed that the distance traveled by all covered demands is gradually decreasing through the planning horizon or at least is constant. The mean distance traveled by the residences is equal to 2.69, 2.61, and 2.02 kilometers

with a deviation of 1.51, 1.27, and 1.12 kilometers, along the periods 1, 2, and 3, respectively (see Table A.7 in Appendix A). All demand nodes during periods 2 and 3 are covered and the number of uncovered nodes within the first period is limited to 9, including nodes 1, 2, 6, 27, 28, 29, 31, 57, and 58. Table 6.10 classifies the nodes based on the relative improvement in the traveled distance within four categories. The relative improvement in the accessibility criterion, $s(i, t)$, for each covered node i over each period t , is calculated as follows.

$$s(i, t) = \frac{\sum_{j \in J} d_{ij} x_{ij(t-1)} - \sum_{j \in J} d_{ij} x_{ijt}}{\sum_{j \in J} d_{ij} x_{ij(t-1)}}, \quad t = 2, 3.$$

Table 6.10: Classifying covered nodes based on the relative improvement in accessibility criterion

Relative accessibility improvement	Time period	
	$t = 2$	$t = 3$
$s(i, t) \leq 25\%$	19, 30, 40, 41, 71, 73	13, 17, 25, 41, 43, 47, 48, 57, 59, 62, 66, 68
$25\% < s(i, t) \leq 50\%$	7, 9, 10, 32, 45, 67, 76	1, 7, 12, 14, 19, 44, 46, 56, 58, 70
$50\% < s(i, t) \leq 75\%$	8, 11, 23, 24, 25	3, 5, 6, 8, 18, 20, 26, 60, 61, 65
$s(i, t) > 75\%$	21, 22	45, 64

The first row in Table 6.10 represents the index of demand nodes with up to 25% improvement in the accessibility criterion over periods 2 and 3. The second row shows the nodes with at least 25% and at most 50% improvement in $s(i, t)$. In a similar way, other rows present similar results for higher values of $s(i, t)$.

The maximum relative improvement in the accessibility criterion was, respectively, about 91% and 92% for demand nodes 21, 22 and 45, 64 over periods 1 – 2 and 2 – 3 (last row in Table 6.10). Obviously, the traveled distance of the demand nodes not reported in Table 6.10 does not change over different periods.

Since we are addressing the multi-period service level based location problem in a coverage context, it can be possible that some demand nodes are not satisfied at the end of the planning horizon. Such demands can be assigned to the closest open facility provided that the facility capacity is increased or some external resources are available to serve unsatisfied demands.

Of course, the coverage radius strongly influences the system behavior (Figure. 6.9). By increasing the distance threshold values from 5 to 6 kilo-

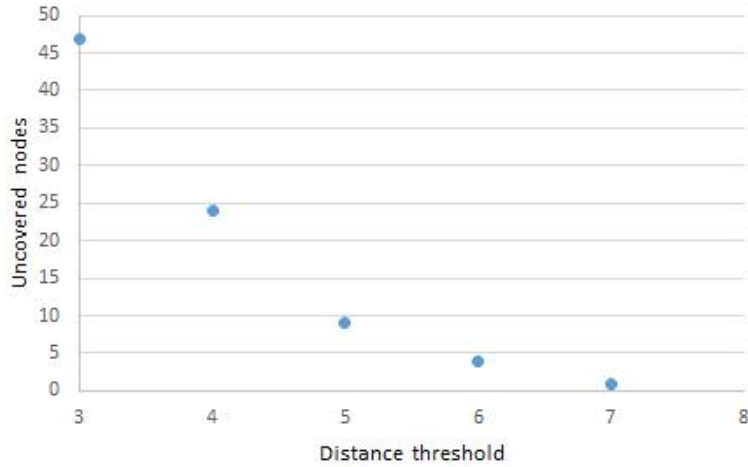


Figure 6.9: Coverage versus distance threshold

meters, it is possible to double the coverage, but nasty results are obtained for very narrow coverage radius values.

We also investigated how the changes in the risk level α affects the coverage performance. Clearly, when the risk level is low, or equivalently, a high reliability level is required, the probabilistic capacity constraints (6.36) are tight, since the demands of covered nodes should be within the capacity of facilities with a high probability. In order to satisfy the capacity constraints, less demand nodes might be covered. On the contrary, when the decision-maker accepts higher risk levels, and lower reliability level are required, it is more likely that the capacity constraints (6.36) are violated so the number of uncovered nodes decreases and we are able to cover more demand points. For the case with a distance threshold equal to 5 and a risk level of 0.01, the number of uncovered nodes over all periods is 13, but by increasing the risk level to 0.08, for the same distance threshold, the number of uncovered nodes decreases to 9 which shows about 31% improvement in the objective function.

Table 6.11: Sensitivity analysis with respect to α , ρ_0

α	ρ_0	D	Obj.	Expected satisfied demands
0.10	0.90	9	9	609
0.10	0.95	9	9	644
0.02	0.90	5	11	611
0.02	0.95	5	13	642

As another experiment, we have evaluated the impact of the participation probability ρ_0 . For example as a result of advertising programs, this probability can increase, determining a higher participation level in the health program. Based on our observations, by a 5% increase in the participation probability (from 0.90 to 0.95), the expected amount of satisfied demands increased up to 6% (5%) for risk level equal to 0.10 (0.02) and distance coverage of 9 (5) kilometers. See Table 6.11.

As a final consideration, the role of constraints in (6.36) in improving the accessibility performance has been assessed by comparing the model solution with and without these set of constraints.

Although the objective function value in both cases is equal, about 27% of covered demands during periods 2 and 39% of covered demands in period 3 experience up to 93% increase in the traveled distance.

6.2.4.3 Current system evaluation

We have also carried out a set of experiments to provide some managerial insights about the current system performance. Specifically, we have considered the operational scenario that the managers are not strongly motivated to upgrade the system by adding more facilities, probably due to financial crisis, and instead, are interested to run the system with only the existing nursing homes. This requires keeping all the seven existing nursing homes active and banning the establishment of new facilities over the whole planning horizon. This in the mathematical model (6.31)-(6.41) can be expressed by imposing the additional set of constraints

$$y_{j1} - 1 = 0, j = 1, \dots, 7$$

and solving the problem with $p_1 = 7, p_2 = p_3 = 0$.

Note that this new set of constraints in combination with (6.37) require the activation of existing nursing homes over all periods. The resulting model turned out to be infeasible due to the violation in the probabilistic capacity constraints (6.36). This supports the claim that the system should definitely be equipped with more facilities to address the increasing demand. Of course the capacity constraints are the most challenging constraints of the model and the managers might be willing to evaluate a capacity expansion, while

keeping the current configuration of the facilities. To investigate this possibility, we removed the capacity constraints (6.36) from the aforementioned augmented model and evaluated the coverage performance of system. Although in this case the problem is feasible, 24 demand nodes will not be covered over different periods showing that even in the absence of capacity constraints, with the current configuration of nursing homes, not all the demand areas can be covered. Interestingly, the latter results in terms of coverage performance are still 54% worse than the results reported in Table 6.11 for the case with capacity constraints (6.36).

Finally, we investigated the multi-period behavior of the model assuming that the system upgrade is allowed from the second time period with $p_2 = 1$, $p_3 = 2$ and that the first period is run with all the currently existing facilities ($p_1 = 7$). Again, the latter assumption requires adding a set of constraints to the model. The optimal objective value is 12 in which 9, 1, and 2 zones are not covered over periods 1, 2, and 3, respectively, while the optimal objective value obtained by our model was 9 in which only 9 zones are not covered in the first period and all the demands are covered over the next periods. This again support our initial claim that the current configuration of facilities is not optimal and even after upgrading the system, some demand nodes will never be covered. This will encourage the managers to modify the current system configuration and to relocate some facilities. Although the relocation of strategic facilities is costly and may involve unwanted consequences, it will improve the system performance.

6.2.4.4 Probabilistic versus deterministic and time-invariant model

In order to validate the probabilistic model, we have compared it with its deterministic counterpart, obtained by replacing the random variables with their expected value in constraints (6.36). Table 6.12 shows the resulting optimal sites. We observed that the assignment pattern associated with the deterministic model will result in the infeasibility of the problem at the presence of uncertainty. This shows the importance of incorporating the deviation of demands and adopting a probabilistic approach and provide evidence for the superiority of a risk-averse perspective over risk-neutral ones.

Table 6.12: The optimal nursing homes' sites for the deterministic model

Nursing home number	Time period		
	$t = 1$	$t = 2$	$t = 3$
location2		*	*
location5			*
location6	*	*	*
location9			*
location12		*	*
location14			*
location17	*	*	*
location21			*
location23		*	*
location24			*

Apart from that, we also evaluated the left-hand side of the reliability constraints (6) obtained by the solution $(\hat{x}_{ijt}, \hat{y}_{jt})$ of the deterministic model expressed as

$$P\left(\sum_{i \in I} \lambda_{ij} h_{it}(\omega) \hat{x}_{ijt} \leq Q_j \hat{y}_{jt}\right) = F_{\Phi}(Q_j \hat{y}_{jt}), \quad \forall j \in J, t = 1, \dots, T$$

where $\Phi = \sum_{i \in I} \lambda_{ij} h_{it}(\omega) \hat{x}_{ijt}$ is a normally distributed random variable with the cumulative distribution function $F_{\Phi}(\cdot)$ and \hat{x}_{ijt} and \hat{y}_{jt} are the optimal values of the deterministic model. As shown in Table 6.13, for some constraints, the probability of not exceeding the capacity is low and, considering a reliability level of 0.95 (corresponding to the risk level of α equal to 0.05), it is five times below the minimum required value. This cases are highlighted in bold in Table 6.13. This, again, supports our previous claim about the necessity of incorporating the stochasticity of uncertain parameters into the model.

Table 6.13: The reliability level of the deterministic model

Nursing home number	Time period		
	$t = 1$	$t = 2$	$t = 3$
location2	-	0.99	0.82
location5	-	-	1
location6	0.99	0.59	0.94
location9	-	-	1
location12	-	1	1
location14	-	-	1
location17	0.84	0.69	0.99
location21	-	-	1
location23	-	1	1
location24	-	-	1

We also investigated the importance of considering the temporal dependency of the stochastic demand.

A comparison between the coverage performance resulting from the proposed model and the same model, where the mean and the variance of the demand are considered constant over time (stochastic time-invariant demand model), is presented in Figure. 6.10.

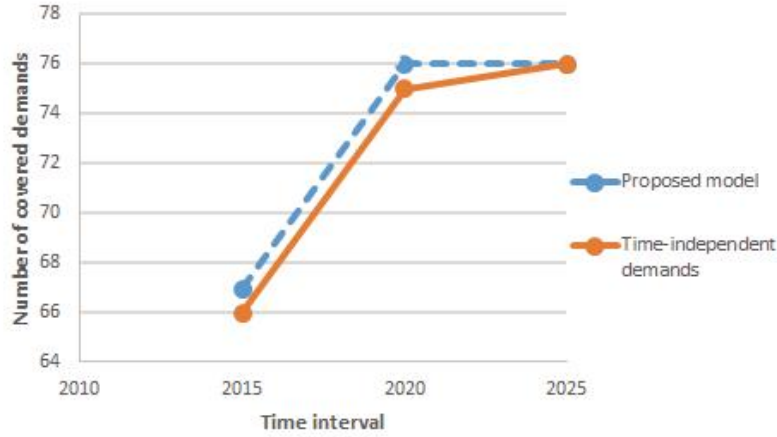


Figure 6.10: Stochastic time-invariant demand model versus the proposed model

The proposed model outperforms the stochastic time-invariant demand model, in terms of coverage performance. In particular, in the first period, the latter model overestimates the coverage performance by 14%.

6.2.4.5 Monte Carlo simulation

The last part of this section is devoted to a Monte Carlo simulation investigating the validity of the proposed model with respect to the probabilistic chance constraints (6.43). The simulation results are expected to provide informative insights about the effectiveness of the proposed risk-averse approach, allowing to test how frequently the demand can be expected to exceed the capacity. To run the simulation, for each pair of candidate location j and period time t in (6.43), and corresponding to each uncertain demand $h_{it}(\omega)$, $i \in I$, we generated 50000 different random values h_{its} drawn from the normal distribution $N(\mu_{it}, \sigma_{it})$ where each random value represents a scenario indexed by s . To validate the results of the stochastic model, the assignment and location variables x_{ijt} and y_{jt} were set to their optimal values, and the frequency of violation in constraints $\sum_{i \in I} \lambda_{ij} h_{its} - Q_j y_{jt} \leq 0$, $j \in J$, $t = 1, \dots, T$, $s = 1, \dots, 50000$ for different risk

level values was calculated. Performing this procedure, we observed that for risk level values $\alpha \in \{0.01, 0.02, \dots, 0.07\}$, all the constraints over all scenarios are satisfied, or equivalently, the frequency of violation is zero. These results are expected since lower risk levels are more conservative and it is more unlikely to experience any violation. The results of simulation for bigger values of α are reported in Table 6.14.

Table 6.14: The probability of violation in the Monte Carlo simulation

α	t	j																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0.08	1	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-
	2	-	-	0	-	-	0	-	-	-	-	0	-	-	-	-	-	0.009	-	-	-	-	-	0	-
	3	-	-	0	-	-	0	-	-	-	-	0	0	-	0	0	-	0	-	-	-	0	-	0	0
0.10	1	-	-	-	-	-	0	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-	-	-	0	-	-	-	-	0	-
	3	0	-	-	0	0	0	-	-	-	-	0	-	-	0	-	-	-	0.009	-	-	0	-	0	0
0.2	1	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-
	2	-	-	0	-	-	0	-	-	-	-	0	-	-	-	-	-	0.009	-	-	-	-	-	0	-
	3	-	-	0	-	-	0	-	-	-	-	0	-	0	-	0	0	0	0	-	0	0	-	0	-
0.3	1	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-
	2	-	-	0	-	-	0	-	-	-	-	0	-	-	-	-	-	0.009	-	-	-	-	-	0	-
	3	-	-	0	-	-	0	-	0	0	-	0	-	0	-	-	0	-	-	0	0	-	0	-	-

A few violations are experienced which are still less than the risk level. Since this violations are related to the candidate facilities 17 and 18, by adding other facilities to the network, possibly near the neighborhood of facility 17, the violation would be eliminated.

6.2.5 Conclusions and future research directions

In this section, we proposed a multi-period model for the nursing homes facility location problem. The multi-period perspective was adopted for handling the budget constraints as well as the fluctuation of demands over time.

The improvement of the accessibility performance was followed by the dynamic modification of the assignment pattern, if possible, while the deterioration of service level was strictly prohibited over the planning horizon.

We also discussed about the possibility of incorporating both the preferences of users and managers within a covering framework. This enabled

us to address the elasticity of demands, based on the distance parameter, as well.

The imprecise nature of demands was tackled by applying a probabilistically constrained approach on the capacity constraints to satisfy with a given probability. Additionally, the deterministic equivalent formulation of the model as well as its linearized counterpart were introduced. The model was implemented on a real case study for nursing home location planning problem in Shiraz city, Iran. The analysis of the results provided us with important managerial insights about the current configuration of nursing home facilities and the possibility of improving the current performance.

It is mentionable that although we developed the model for nursing home planning network, it can also be applied for other strategic location decisions arising in the public sector. Extending the proposed model to address the issue of fairness will be interesting as a future research topic. This can be investigated through the division of demand nodes into different subsets (categories) based on their characteristics and special needs for health services. Hence, the incorporation of different constraints or the definition of other objectives specified for each demand category could be possible. For instance, an option could be the division of demand zones based on their geographical locations into marginal and non-marginal zones and the problem could be modeled as a bi-objective problem by adding another objective which minimizes the amount of uncovered demands associated with marginal zones over all periods. In addition, different coverage radii or types of facilities for zones in which the access to health care services is limited could be introduced.

Appendix A

Appendix

A.0.1 Open facilities for Cases A, B, and C

Table A.1: Optimal locations for Scenario A

	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_{1s}^2	x_{1s}^2	x_{1s}^2	x_{2s}^2	x_{2s}^2	x_{2s}^2
1	9	18	23	29	34	40	1	10	20	1	7	15	
3	10	19	25	30	35	41	4	12	27	2	8	17	
4	11	20	26	31	36	42	5	14	29	3	10	18	
5	13	21	27	32	37	43	7	16	30	5	13	-	
7	15	22	28	33	38	44	9	17	-	6	14	-	

Table A.2: Optimal locations for Scenario B

	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_{1s}^2	x_{1s}^2	x_{1s}^2	x_{2s}^2	x_{2s}^2	x_{2s}^2
1	6	11	16	21	33	38	1	10	20	1	7	15	
2	7	12	17	22	34	40	4	12	27	2	8	17	
3	8	13	18	23	35	41	5	14	29	3	10	18	
4	9	14	19	24	36	42	7	16	30	5	13	-	
5	10	15	20	32	37	43	9	17	-	6	14	-	

Table A.3: Optimal locations for Scenario C

x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_f^1	x_{1s}^2	x_{1s}^2	x_{1s}^2	x_{2s}^2	x_{2s}^2	x_{2s}^2
1	7	13	20	27	33	38	1	10	20	1	7	15
3	8	15	21	28	34	40	4	12	27	2	8	17
4	9	17	22	30	35	41	5	14	29	3	10	18
5	10	18	23	31	36	42	7	16	30	5	13	-
6	11	19	24	32	37	43	9	17	-	6	14	-

A.0.2 Linearization

Since the term under the square root in Eq. (6.43) is non-negative, if $\sum_{i \in I} \hat{\mu}_{ijt} x_{ijt} \leq Q_j y_{jt}$, we can rewrite it as follows:

$$\left(\sqrt{\beta_\alpha \sum_{i \in I} x_{ijt}^2 \hat{\sigma}_{ijt}^2} \right)^2 \leq \left(Q_j y_{jt} - \sum_{i \in I} \hat{\mu}_{ijt} x_{ijt} \right)^2 \quad \forall j \in J, t = 1, \dots, T \quad (\text{A.1})$$

which can be simplified in Eq. (A.2):

$$\begin{aligned} & \sum_{i \in I} \beta_\alpha \hat{\sigma}_{ijt}^2 x_{ijt} - Q_j^2 y_{jt} + 2 Q_j \sum_{i \in I} \hat{\mu}_{ijt} x_{ijt} y_{jt} \\ & - \sum_{i \in I} \sum_{k \in I} \hat{\mu}_{ijt} \hat{\mu}_{kjt} x_{ijt} x_{kjt} \leq 0 \quad \forall j \in J, t = 1, \dots, T \end{aligned} \quad (\text{A.2})$$

We introduce the auxiliary variables z_{ijt} and w_{ikjt} denoting the bilinear terms $x_{ijt} y_{jt}$ and $x_{ijt} x_{kjt}$ in (A.2), respectively. The set of constraints (A.3) - (A.11) are also added to obtain a set of equivalent linear constraints for

their non-linear counterparts in (A.2):

$$\sum_{i \in I} \beta_{\alpha} \hat{\sigma}_{ijt}^2 x_{ijt} - Q_j^2 y_{jt} + 2Q_j \sum_{i \in I} \hat{\mu}_{ijt} z_{ijt} - \sum_{i \in I} \sum_{k \in I} \hat{\mu}_{ijt} \hat{\mu}_{kjt} w_{ikjt} \leq 0$$

$$\forall j \in J, t = 1, \dots, T \quad (\text{A.3})$$

$$\sum_{i \in I} \hat{\mu}_{ijt} x_{ijt} \leq Q_j y_{jt} \quad \forall i \in I, \forall j \in J \quad (\text{A.4})$$

$$z_{ijt} \geq x_{ijt} + y_{jt} - 1 \quad \forall i \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.5})$$

$$z_{ijt} \leq x_{ijt} \quad \forall i \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.6})$$

$$z_{ijt} \leq y_{jt} \quad \forall i \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.7})$$

$$w_{ikjt} \geq x_{ijt} + x_{kjt} - 1 \quad \forall i, k \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.8})$$

$$w_{ikjt} \leq x_{ijt} \quad \forall i, k \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.9})$$

$$w_{ikjt} \leq x_{kjt} \quad \forall i, k \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.10})$$

$$z_{ijt}, w_{ikjt} \in \{0, 1\} \quad \forall i, k \in I, \forall j \in J, t = 1, \dots, T \quad (\text{A.11})$$

The mathematical model *NDMPLM* amended with constraints (A.3) - (A.11), and the auxiliary variables, define the proposed model.

A.0.3 Coordinate transformation

The transformation in (A.12) was applied in order to convert the GPS coordinates of the demand nodes and facility locations, specified on the map, into the Cartesian coordinates, which is consistent with the Euclidean distance axiom.

$$x = R \times \cos(C \times 3.14/180) \times (s \times 3.14/180)$$

$$y = R \times t \times 3.14/180 \quad (\text{A.12})$$

where (x, y) and (s, t) represent the Cartesian and the GPS coordinates, respectively, R is the approximate earth radius, and C shows the latitude of a hypothetical center point over the region.

A.0.4 Location coordinates of population centers and facilities

Table A.4: Location coordinates of population centers at zones 1-3

Zone	Population center	Location coordinates		Zone	Population center	Location coordinates	
		x	y			x	y
1	(1)	5069.3	3294.5	2	(9)	5077.1	3289.4
1	(2)	5069.3	3297.8	2	(10)	5076.3	3290.2
1	(3)	5071.2	3296	2	(11)	5075.6	3290.1
1	(4)	5072.6	3295.3	2	(12)	5076.4	3291.1
1	(5)	5070.8	3293.6	2	(13)	5076.2	3291.5
1	(6)	5072.3	3293.9	2	(14)	5077.1	3291.7
1	(7)	5071.8	3292.6	2	(15)	5077.1	3292.3
1	(8)	5074.1	3291.8	3	(1)	5085.7	3287.7
1	(9)	5075.1	3292.1	3	(2)	5081.1	3288.5
1	(10)	5073.4	3293.3	3	(3)	5080.3	3289.6
1	(11)	5076.1	3292.8	3	(4)	5079.7	3290.3
1	(12)	5077.2	3293.8	3	(5)	5078	3293.3
2	(1)	5078.7	3287	3	(6)	5079	3292.9
2	(2)	5078.3	3288.2	3	(7)	5079.8	3292.7
2	(3)	5080.9	3287.2	3	(8)	5079.4	3291.6
2	(4)	5079.5	3287.9	3	(9)	5078.6	3292.4
2	(5)	5079.9	3288.8	3	(10)	5080.4	3290.6
2	(6)	5079.1	3289	3	(11)	5082.2	3289.3
2	(7)	5079	3290.1	3	(12)	5081.5	3291.9
2	(8)	5078	3289.1	3	(13)	5082.3	3292.4

Table A.5: Location coordinates of population centers at zones 4-9

Zone	Population center	Location coordinates		Zone	Population center	Location coordinates	
		x	y			x	y
4	(1)	5080.7	3292.2	6	(1)	5070.7	3300
4	(2)	5071.6	3289.3	6	(2)	5070.1	3299.8
4	(3)	5074	3289.8	6	(3)	5069.3	3299.1
4	(4)	5072.1	3290.5	7	(1)	5078.3	3291.4
4	(5)	5070.3	3291	7	(2)	5084.3	3285.3
4	(6)	5070.1	3292.5	7	(3)	5083.3	3286
4	(7)	5071.3	3292	7	(4)	5082.4	3286.4
4	(8)	5073.6	3287.5	7	(5)	5082.5	3287.9
4	(9)	5072.8	3291.4	7	(6)	5083.5	3287.3
4	(10)	5073.9	3290.7	7	(7)	5085.2	3287.4
4	(11)	5075.1	3290.9	8	(1)	5074.6	3294
5	(1)	5078.2	3285.4	8	(2)	5078.1	3289.8
5	(2)	5076.1	3286.7	8	(3)	5077.1	3290.6
5	(3)	5074.7	3287.3	8	(4)	5077.9	3290.6
5	(4)	5076.1	3288.1	9	(1)	5075.8	3285.5
5	(5)	5074.7	3288.8	9	(2)	5071.5	3286.6
5	(6)	5077.4	3288.5	9	(3)	5070.2	3288.9
5	(7)	5075	3289.8	9	(4)	5068	3291.1

Table A.6: The coordinates of potential facility sites

Facility	Location coordinates		Facility	Location coordinates	
j	x	y	j	x	y
1*	5071	3300	13	5081	3293
2*	5070	3300	14	5081	3291
3*	5070	3298	15	5078	3293
4*	5068	3292	16	5080	3285
5*	5080	3287	17	5079	3289
6*	5073	3292	18	5076	3290
7*	5078	3299	19	5077	3286
8	5084	3283	20	5078	3291
9	5087	3288	21	5073	3294
10	5083	3287	22	5069	3295
11	5087	3285	23	5070	3290
12	5081	3288	24	5077	3288

* Existing nursing homes

Table A.7: Distance traveled by covered demands

Demand node	period			Demand node	period		
i	$t = 1$	$t = 2$	$t = 3$	i	$t = 1$	$t = 2$	$t = 3$
1	-	4.76	2.405	39	1.815	1.815	1.815
2	-	3.586	3.586	40	3.396	2.582	2.582
3	3.866	3.866	1.014	41	4.753	3.726	3.318
4	4.992	4.992	4.992	42	2.969	2.969	2.969
5	3.941	3.941	1.368	43	3.202	3.202	2.567
6	-	3.546	1.226	44	4.55	4.55	2.934
7	4.62	3.21	2.017	45	3.867	1.963	0.154
8	4.796	2.169	0.605	46	4.094	4.094	2.635
9	3.229	1.901	1.901	47	4.448	4.448	4.392
10	3.521	2.253	2.253	48	3.324	3.324	3.289
11	2.719	1.101	1.101	49	2.35	2.35	2.35
12	2.501	2.501	1.553	50	1.64	1.64	1.64
13	2.663	2.663	2.284	51	2.706	2.706	2.706
14	1.019	1.019	0.665	52	4.404	4.404	4.404
15	1.471	1.471	1.471	53	4.758	4.758	4.758
16	1.148	1.148	1.148	54	0.327	0.327	0.327
17	1.985	1.985	1.874	55	1.512	1.512	1.512
18	2.67	2.67	1.075	56	2.42	2.42	1.659
19	3.023	2.35	1.48	57	-	3.703	3.182
20	3.977	3.977	1.435	58	-	4.022	2.126
21	3.62	0.311	0.311	59	4.631	4.631	4.606
22	3.066	0.357	0.357	60	3.688	3.688	1.689
23	3.522	0.939	0.939	61	4.735	4.735	1.908
24	3.239	1.292	1.292	62	2.392	2.392	2.353
25	4.21	1.892	1.453	63	1.653	1.653	1.653
26	4.182	4.182	1.991	64	3.68	3.68	0.778
27	-	2.092	2.092	65	4.392	4.392	1.231
28	-	1.626	1.626	66	3.936	3.936	3.12
29	-	0.867	0.867	67	2.649	1.917	1.917
30	4.691	3.742	3.742	68	2.054	2.054	2.015
31	-	0.56	0.56	69	1.366	1.366	1.366
32	4.755	2.768	2.768	70	2.66	2.66	1.362
33	3.748	3.748	3.748	71	3.025	2.785	2.785
34	2.991	2.991	2.991	72	1.724	1.724	1.724
35	2.38	2.38	2.38	73	4.109	3.571	3.571
36	1.49	1.49	1.49	74	0.229	0.229	0.229
37	1.209	1.209	1.209	75	1.314	1.314	1.314
38	2.255	2.255	2.255	76	2.231	1.148	1.148

Bibliography

- [1] URL: <https://earthquake.usgs.gov/earthquakes/eventpage/usp000h60h>. 78
- [2] URL: <https://www.google.com/maps>. 79, 85
- [3] URL: <http://www.countryreports.org/travel/Haiti/traffic.htm>. 80
- [4] (in Persian). URL: <http://www.farsnews.com/newstext.php?nn=13931020000880>. 172
- [5] (in Persian). URL: <http://www.behzisti-fars.ir/portal/show.aspx?page=17475>. 200
- [6] (in Persian). URL: <http://www.eshiraz.ir/infotech/fa/index>. 200, 201
- [7] URL: <http://esa.un.org/unpd/wpp/>. 201
- [8] Population totale, population de 18 ans et plus ménages et densités estimées en 2015. URL: <http://www.ihsi.ht/pdf/projection/Estimat-PopTotal-18ans-Menag2015.pdf>. 79
- [9] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Ageing 2015 (st/esa/ser.a/390). 184
- [10] CPLEX. ILOG, C.P.L.E.X.: 6.5: Users manual. CPLEX Optimization, Inc., Incline Village, NV. 1999. 149
- [11] World Disasters Report. The International Federation of Red Cross and Red Crescent Societies (IFRC). 2016. 50

- [12] M Ahmadzad-Asl, H Abtahi, M Moradpoor, V Mehrabani, N Mashaii, and R Bidaki. Community based organizations as an appropriate tool for harm reduction and prevention of drug abuse: The report of an experience. *Iranian Journal of Psychiatry and Clinical Psychology*, 16, 2010. (in Persian). [163](#)
- [13] Nader Al Theeb and Chase Murray. Vehicle routing and resource distribution in postdisaster humanitarian relief operations. *International Transactions in Operational Research*, 24(6):1253–1284, 2017. [59](#)
- [14] Ramon Alanis, Armann Ingolfsson, and Bora Kolfal. A markov chain model for an ems system with repositioning. *Production and operations management*, 22(1):216–231, 2013. [154](#)
- [15] Maria Albareda-Sambola, Antonio Alonso-Ayuso, Laureano F Escudero, Elena Fernández, Yolanda Hinojosa, and Celeste Pizarro-Romero. A computational comparison of several formulations for the multi-period incremental service facility location problem. *Top*, 18(1):62–80, 2010. [186](#), [187](#), [191](#)
- [16] Maria Albareda-Sambola, Antonio Alonso-Ayuso, Laureano F Escudero, Elena Fernández, and Celeste Pizarro. Fix-and-relax-coordination for a multi-period location–allocation problem under uncertainty. *Computers & operations research*, 40(12):2878–2892, 2013. [188](#)
- [17] Maria Albareda-Sambola, Elena Fernández, Yolanda Hinojosa, and Justo Puerto. The multi-period incremental service facility location problem. *Computers & Operations Research*, 36(5):1356–1375, 2009. [186](#), [187](#), [191](#)
- [18] Douglas Alem, Alistair Clark, and Alfredo Moreno. Stochastic network models for logistics planning in disaster relief. *European Journal of Operational Research*, 255(1):187–206, 2016. [63](#), [64](#), [66](#)
- [19] Ceyhun Araz, Hasan Selim, and Irem Ozkarahan. A fuzzy multi-objective covering-based vehicle location model for emergency services. *Computers & Operations Research*, 34(3):705–726, 2007. [134](#)

- [20] Aaron Archer and Anna Blasiak. Improved approximation algorithms for the minimum latency problem via prize-collecting strolls. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 429–447. SIAM, 2010. [6](#)
- [21] Roberto Aringhieri, Maria Elena Bruni, Sara Khodaparasti, and JT Van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368, 2017. [3](#), [121](#), [122](#), [133](#), [187](#)
- [22] Ph Augerat, Jose Manuel Belenguer, Enrique Benavent, A Corberán, D Naddef, and G Rinaldi. Computational results with a branch-and-cut code for the capacitated vehicle routing problem. 1998. [38](#)
- [23] Burcu Balcik. Site selection and vehicle routing for post-disaster rapid needs assessment. *Transportation research part E: logistics and transportation review*, 101:30–58, 2017. [52](#), [59](#), [69](#), [73](#)
- [24] Burcu Balcik, Cem Deniz Caglar Bozkir, and O Erhun Kundakcioglu. A literature review on inventory management in humanitarian supply chains. *Surveys in Operations Research and Management Science*, 21(2):101–116, 2016. [53](#)
- [25] Michael O Ball and Feng L Lin. A reliability model applied to emergency service vehicle location. *Operations research*, 41(1):18–36, 1993. [132](#)
- [26] Ruth Banomyong, Paitoon Varadejsatitwong, and Richard Olorun-toba. A systematic review of humanitarian operations, humanitarian logistics and humanitarian supply chain performance literature 2005 to 2016. *Annals of Operations Research*, pages 1–16, 2017. [53](#)
- [27] Abigail Barr, Marleen Dekker, and Marcel Fafchamps. The formation of community-based organizations: An analysis of a quasi-experiment in zimbabwe. *World Development*, 66:131–153, 2015. [158](#)
- [28] Ayfer Başar, Bülent Çatay, and Tonguç Ünlüyurt. A taxonomy for emergency service station location problem. *Optimization letters*, 6(6):1147–1160, 2012. [122](#)

- [29] Rajan Batta, June M Dolan, and Nirup N Krishnamurthy. The maximal expected covering location problem: Revisited. *Transportation Science*, 23(4):277–287, 1989. [129](#)
- [30] Rajan Batta, Miguel Lejeune, and Srinivas Prasad. Public facility location using dispersion, population, and equity criteria. *European Journal of Operational Research*, 234(3):819–829, 2014. [140](#)
- [31] F Ben Abdelaziz and Meryem Masmoudi. A multiobjective stochastic program for hospital bed planning. *Journal of the Operational Research Society*, 63(4):530–538, 2012. [160](#)
- [32] Ernest Benjamin, Adel M Bassily-Marcus, Elizabeth Babu, Lester Silver, and Michael L Martin. Principles and practice of disaster relief: lessons from haiti. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, 78(3):306–318, 2011. [77](#)
- [33] James C Benneyan, Hande Musdal, Mehmet Erkan Ceyhan, Brian Shiner, and Bradley V Watts. Specialty care single and multi-period location–allocation models within the veterans health administration. *Socio-economic planning sciences*, 46(2):136–148, 2012. [159](#), [160](#), [189](#), [191](#)
- [34] Patrizia Beraldi and Maria Elena Bruni. A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research*, 196(1):323–331, 2009. [187](#)
- [35] Patrizia Beraldi and Maria Elena Bruni. An exact approach for solving integer problems under probabilistic constraints with random technology matrix. *Annals of operations research*, 177(1):127–137, 2010. [186](#)
- [36] Patrizia Beraldi, Maria Elena Bruni, and Domenico Conforti. Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research*, 158(1):183–193, 2004. [132](#), [133](#), [186](#), [187](#)
- [37] Patrizia Beraldi, Maria Elena Bruni, and Francesca Guerriero. Network reliability design via joint probabilistic constraints. *IMA Journal of Management Mathematics*, 21(2):213–226, 2009. [186](#)

- [38] Patrizia Beraldi, Maria Elena Bruni, and Sara Khodaparasti. A fast heuristic for routing in post-disaster humanitarian relief logistics, the correlated case. *Technical report, De-Health Lab-Dimeg, Unical*, 2018. [12](#)
- [39] Patrizia Beraldi, Maria Elena Bruni, and Sara Khodaparasti. Post-disaster relief routing under uncertainty: The case of Haiti earthquake. *submitted*, 2018. [2](#)
- [40] Patrizia Beraldi, Maria Elena Bruni, Demetrio Laganà, and Roberto Musmanno. The mixed capacitated general routing problem under uncertainty. *European Journal of Operational Research*, 240(2):382–392, 2015. [10](#), [12](#), [198](#)
- [41] Patrizia Beraldi, Maria Elena Bruni, and Antonio Violi. Capital rationing problems under uncertainty and risk. *Computational Optimization and Applications*, 51(3):1375–1396, 2012. [12](#)
- [42] Patrizia Beraldi, Antonio Violi, Gianluca Carrozzino, and Maria Elena Bruni. The optimal energy procurement problem: A stochastic programming approach. In *International Conference on Optimization and Decision Science*, pages 357–365. Springer, 2017. [12](#)
- [43] Patrizia Beraldi, Antonio Violi, Gianluca Carrozzino, and Maria Elena Bruni. A stochastic programming approach for the optimal management of aggregated distributed energy resources. *Computers & Operations Research*, 96:200–212, 2018. [12](#)
- [44] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004. [65](#)
- [45] Lucio Bianco, Aristide Mingozzi, and Salvatore Ricciardelli. The traveling salesman problem with cumulative costs. *Networks*, 23(2):81–91, 1993. [5](#)
- [46] Louis-Philippe Bigras, Michel Gamache, and Gilles Savard. The time-dependent traveling salesman problem and single machine scheduling problems with sequence dependent setup times. *Discrete Optimization*, 5(4):685–699, 2008. [5](#), [113](#)

- [47] Roger Bilham. Lessons from the haiti earthquake. *Nature*, 463(7283):878, 2010. [77](#)
- [48] Johannes Bisschop and Marcel Roelofs. AIMMS. 3.7 User’s guide. Paragon Decision Technology B.V., The Netherlands. 2006. [149](#), [200](#)
- [49] Avrim Blum, Prasad Chalasani, Don Coppersmith, Bill Pulleyblank, Prabhakar Raghavan, and Madhu Sudan. The minimum latency problem. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 163–171. ACM, 1994. [5](#)
- [50] Chawis Boonmee, Mikiharu Arimura, and Takumi Asada. Facility location optimization model for emergency humanitarian logistics. *International Journal of Disaster Risk Reduction*, 24:485–498, 2017. [53](#)
- [51] A Bozorgi-Amiri and M Khorsi. A dynamic multi-objective location–routing model for relief logistic planning under uncertainty on demand, travel time, and cost parameters. *The International Journal of Advanced Manufacturing Technology*, 85(5-8):1633–1648, 2016. [64](#)
- [52] Luce Brotcorne, Gilbert Laporte, and Frederic Semet. Ambulance location and relocation models. *European journal of operational research*, 147(3):451–463, 2003. [122](#), [129](#)
- [53] Samuel A Broverman. *Actex study manual, Course 1, Examination of the Society of Actuaries, Exam 1 of the Casualty Actuarial Society*, volume 1. Actex Publications, 2001. [126](#)
- [54] Maria Elena Bruni, Patrizia Beraldi, and Sara Khodaparasti. A heuristic approach for the k-traveling repairman problem with profits under uncertainty. *Electronic Notes in Discrete Mathematics*, 69:221–228, 2018. [2](#), [37](#), [39](#)
- [55] Maria Elena Bruni, Patrizia Beraldi, and Demetrio Laganà. The express heuristic for probabilistically constrained integer problems. *Journal of Heuristics*, 19(3):423–441, 2013. [186](#)

- [56] Maria Elena Bruni, Francesca Guerriero, and Patrizia Beraldi. Designing robust routes for demand-responsive transport systems. *Transportation research part E: logistics and transportation review*, 70:1–16, 2014. [10](#)
- [57] Susan Budge, Armann Ingolfsson, and Erhan Erkut. Approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location. *Operations Research*, 57(1):251–255, 2009. [126](#)
- [58] Teobaldo Bulhões, Minh Hoàng Hà, Rafael Martinelli, and Thibaut Vidal. The vehicle routing problem with service level constraints. *European Journal of Operational Research*, 265(2):544–558, 2018. [52](#), [72](#)
- [59] Timothy H Burwell, James P Jarvis, and Mark A McKnew. Modeling co-located servers and dispatch ties in the hypercube model. *Computers & Operations Research*, 20(2):113–119, 1993. [129](#)
- [60] Giuseppe Carlo Calafiore and Laurent El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006. [77](#), [198](#)
- [61] Ann Melissa Campbell, Dieter Vandenbussche, and William Hermann. Routing for relief efforts. *Transportation Science*, 42(2):127–145, 2008. [51](#), [72](#)
- [62] Teresa Cardoso, Mónica Duarte Oliveira, Ana Barbosa-Póvoa, and Stefan Nickel. An integrated approach for planning a long-term care network with uncertainty, strategic policy and equity considerations. *European Journal of Operational Research*, 247(1):321–334, 2015. [186](#), [190](#), [191](#)
- [63] Teresa Cardoso, Mónica Duarte Oliveira, Ana Barbosa-Póvoa, and Stefan Nickel. Moving towards an equitable long-term care network: A multi-objective and multi-period planning approach. *Omega*, 58:69–85, 2016. [161](#)

- [64] Aakil M Caunhye, Xiaofeng Nie, and Shaligram Pokharel. Optimization models in emergency logistics: A literature review. *Socio-economic planning sciences*, 46(1):4–13, 2012. [51](#)
- [65] Fatih Cavdur, Merve Kose-Kucuk, and Asli Sebatli. Allocation of temporary disaster response facilities under demand uncertainty: An earthquake case study. *International Journal of Disaster Risk Reduction*, 19:159–166, 2016. [63](#), [66](#)
- [66] Melih Çelik. Network restoration and recovery in humanitarian operations: framework, literature review, and research directions. *Surveys in Operations Research and Management Science*, 21(2):47–61, 2016. [53](#)
- [67] Zhiqi Chang, Shiji Song, Yuli Zhang, Jian-Ya Ding, Rui Zhang, and Raymond Chiong. Distributionally robust single machine scheduling with risk aversion. *European Journal of Operational Research*, 256(1):261–274, 2017. [102](#), [103](#)
- [68] Sunarin Chanta, Maria E Mayorga, Mary E Kurz, and Laura A McLay. The minimum p-envy location problem: a new model for equitable distribution of emergency resources. *IIE Transactions on Healthcare Systems Engineering*, 1(2):101–115, 2011. [125](#)
- [69] Sunarin Chanta, Maria E Mayorga, and Laura A McLay. Improving emergency service in rural areas: a bi-objective covering location model for ems systems. *Annals of Operations Research*, 221(1):133–159, 2014. [124](#), [144](#)
- [70] Sunarin Chanta, Maria E Mayorga, and Laura A McLay. The minimum p-envy location problem with requirement on minimum survival rate. *Computers & Industrial Engineering*, 74:228–239, 2014. [125](#)
- [71] Abraham Charnes, William W Cooper, and Robert O Ferguson. Optimal estimation of executive compensation by linear programming. *Management science*, 1(2):138–151, 1955. [169](#)

- [72] Kamalika Chaudhuri, Brighten Godfrey, Satish Rao, and Kunal Talwar. Paths, trees, and minimum latency tours. In *FOCS*, volume 3, page 36, 2003. [6](#)
- [73] Liang-Hsuan Chen and Feng-Chou Tsai. Fuzzy goal programming with different importance and priorities. *European Journal of Operational Research*, 133(3):548–556, 2001. [169](#), [170](#)
- [74] Cheol-Joo Cho. An equity-efficiency trade-off model for the optimum location of medical care facilities. *Socio-Economic Planning Sciences*, 32(2):99–112, 1998. [138](#)
- [75] Soo-Haeng Cho, Hoon Jang, Taesik Lee, and John Turner. Simultaneous location of trauma centers and helicopters for emergency medical service planning. *Operations Research*, 62(4):751–771, 2014. [129](#)
- [76] Nicos Christofides. The vehicle routing problem. *Combinatorial optimization*, 1979. [38](#)
- [77] Nicos Christofides and Samuel Eilon. An algorithm for the vehicle-dispatching problem. *Journal of the Operational Research Society*, 20(3):309–318, 1969. [38](#)
- [78] Richard Church and Charles ReVelle. The maximal covering location problem. In *Papers of the Regional Science Association*, volume 32, pages 101–118. Springer, 1974. [131](#)
- [79] Lucas Dias Condeixa, Adriana Leiras, Fabricio Oliveira, and Irineu de Brito Jr. Disaster relief supply pre-positioning optimization: A risk analysis via shortage mitigation. *International Journal of Disaster Risk Reduction*, 25:238–247, 2017. [66](#)
- [80] Robert A Cook and Emmett J Lodree. Dispatching policies for last-mile distribution with stochastic supply and demand. *Transportation Research Part E: Logistics and Transportation Review*, 106:353–371, 2017. [63](#), [64](#)
- [81] William W Cooper, Lawrence M Seiford, and Kaoru Tone. Data envelopment analysis: a comprehensive text with models, applications,

- references and dea-solver software. *Springer Science & Business Media*, 2007. [165](#)
- [82] Isabel Correia and Teresa Melo. Multi-period capacitated facility location under delayed demand satisfaction. *European Journal of Operational Research*, 255(3):729–746, 2016. [190](#)
- [83] Mark S Daskin. A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation science*, 17(1):48–70, 1983. [130](#)
- [84] Mark S Daskin. *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons, 2011. [125](#)
- [85] M Davari, A Haycox, and T Walley. The Iranian health insurance system; past experiences, present challenges and future strategies. *Iranian journal of public health*, 41(9):1, 2012. [157](#)
- [86] Soheil Davari, Kemal Kilic, and Gurdal Ertek. Fuzzy bi-objective preventive health care network design. *Health care management science*, 18(3):303–317, 2015. [161](#)
- [87] Thijs Dewilde, Dirk Cattrysse, Sofie Coene, Frits CR Spijksma, and Pieter Vansteenwegen. Heuristics for the traveling repairman problem with profits. *Computers & Operations Research*, 40(7):1700–1707, 2013. [1](#), [7](#), [8](#), [10](#), [14](#), [23](#)
- [88] Tammy Drezner, Zvi Drezner, and Jeffery Guyse. Equitable service by a facility: Minimizing the Gini coefficient. *Computers & Operations Research*, 36(12):3240–3246, 2009. [125](#), [148](#)
- [89] Christophe Duhamel, Andréa Cynthia Santos, Daniel Brasil, Eric Châtelet, and Babiga Birregah. Connecting a population dynamic model with a multi-period location-allocation problem for post-disaster relief operations. *Annals of Operations Research*, 247(2):693–713, 2016. [59](#), [61](#)

- [90] Pablo A Maya Duque, Irina S Dolinskaya, and Kenneth Sørensen. Network repair crew scheduling and routing for emergency relief distribution problem. *European Journal of Operational Research*, 248(1):272–285, 2016. [62](#), [69](#)
- [91] Özgün Elçi and Nilay Noyan. A chance-constrained two-stage stochastic programming model for humanitarian relief network design. *Transportation research part B: methodological*, 108:55–83, 2018. [66](#), [73](#)
- [92] Maria Elena Bruni, Patrizia Beraldi, and Sara Khodaparasti. A fast heuristic for routing in post-disaster humanitarian relief logistics. *Transportation Research Procedia*, 30:304–313, 2018. [1](#)
- [93] Gunes Erdogun, Erhan Erkut, Armann Ingolfsson, and Gilbert Laporte. Scheduling ambulance crews for maximum coverage. *Journal of the Operational Research Society*, 61(4):543–550, 2010. [147](#)
- [94] Erhan Erkut, Armann Ingolfsson, and Güneş Erdoğan. Ambulance location for maximum survival. *Naval Research Logistics (NRL)*, 55(1):42–58, 2008. [124](#), [131](#), [143](#), [147](#)
- [95] Erhan Erkut, Armann Ingolfsson, Thaddeus Sim, and Güneş Erdoğan. Computational comparison of five maximal covering models for locating ambulances. *Geographical Analysis*, 41(1):43–65, 2009. [129](#), [131](#), [142](#), [143](#), [147](#)
- [96] Inmaculada Espejo, Alfredo Marín, Justo Puerto, and Antonio M Rodríguez-Chía. A comparison of formulations and solution methods for the minimum-envy location problem. *Computers & Operations Research*, 36(6):1966–1981, 2009. [125](#)
- [97] Imen Ome Ezzine, Frederic Semet, and Habib Chabchoub. New formulations for the traveling repairman problem. In *Proceedings of the 8th International conference of modeling and simulation*. Citeseer, 2010. [5](#)
- [98] Jittat Fakcharoenphol, Chris Harrelson, and Satish Rao. The k-traveling repairmen problem. *ACM Transactions on Algorithms (TALG)*, 3(4):40, 2007. [6](#)

- [99] Rolf Färe and Shawna Grosskopf. Modeling undesirable factors in efficiency evaluation: comment. *European Journal of Operational Research*, 157(1):242–245, 2004. [143](#)
- [100] Thomas A Feo and Jonathan F Bard. Flight scheduling and maintenance base planning. *Management Science*, 35(12):1415–1432, 1989. [28](#)
- [101] Meysam Fereiduni and Kamran Shahanaghi. A robust optimization model for distribution and evacuation in the disaster response phase. *Journal of Industrial Engineering International*, 13(1):117–141, 2017. [64](#), [68](#)
- [102] Christian Fikar, Manfred Gronalt, and Patrick Hirsch. A decision support system for coordinated disaster relief distribution. *Expert Systems with Applications*, 57:104–116, 2016. [59](#), [62](#), [67](#), [68](#), [70](#)
- [103] Matteo Fischetti, Gilbert Laporte, and Silvano Martello. The delivery man problem and cumulative matroids. *Operations Research*, 41(6):1055–1064, 1993. [5](#)
- [104] Vito Fragnelli and Stefano Gagliardo. Open problems in cooperative location games. *International Game Theory Review*, 15(03):1340015, 2013. [126](#), [127](#)
- [105] Lori S Franz, Terry R Rakes, and A James Wynne. A chance-constrained multiobjective model for mental health services planning. *Socio-economic planning sciences*, 18(2):89–95, 1984. [186](#)
- [106] Roberto D Galvao, Luis Gonzalo Acosta Espejo, and Brian Boffey. A hierarchical model for the location of perinatal facilities in the municipality of rio de janeiro. *European Journal of Operational Research*, 138(3):495–517, 2002. [160](#), [162](#)
- [107] Roberto D Galvão, Luis Gonzalo Acosta Espejo, Brian Boffey, and Derek Yates. Load balancing and capacity constraints in a hierarchical location model. *European Journal of Operational Research*, 172(2):631–646, 2006. [160](#), [162](#)

- [108] Michel Gendreau, Gianpaolo Ghiani, and Emanuela Guerriero. Time-dependent routing problems: A review. *Computers & operations research*, 64:189–197, 2015. [10](#)
- [109] Michel Gendreau, Gilbert Laporte, and Frédéric Semet. Solving an ambulance location model by tabu search. *Location science*, 5(2):75–88, 1997. [135](#)
- [110] Abdolsalam Ghaderi and Mohammad Saeed Jabalameli. Modeling the budget-constrained dynamic uncapacitated facility location–network design problem and solving it via two efficient heuristics: a case study of health care. *Mathematical and Computer Modelling*, 57(3-4):382–400, 2013. [189](#), [191](#)
- [111] Jeffrey B Goldberg. Operations research models for the deployment of emergency services vehicles. *EMS management Journal*, 1(1):20–39, 2004. [122](#)
- [112] Anna Graber-Naidich, Michael W Carter, and Vedat Verter. Primary care network development: the regulator’s perspective. *Journal of the Operational Research Society*, 66(9):1519–1532, 2015. [161](#)
- [113] Emilia Grass and Kathrin Fischer. Two-stage stochastic programming in disaster management: A literature survey. *Surveys in Operations Research and Management Science*, 21(2):85–100, 2016. [53](#)
- [114] Linda V Green and Peter J Kolesar. Anniversary article: Improving emergency responsiveness with management science. *Management Science*, 50(8):1001–1014, 2004. [122](#)
- [115] Francesca Guerriero, Maria Elena Bruni, and Francesca Greco. A hybrid greedy randomized adaptive search heuristic to solve the dial-a-ride problem. *Asia-Pacific Journal of Operational Research*, 30(01):1250046, 2013. [28](#)
- [116] Evrim Didem Güneş, Hande Yaman, Bora Çekyay, and Vedat Verter. Matching patient and physician preferences in designing a primary care facility network. *Journal of the Operational Research Society*, 65(4):483–496, 2014. [159](#), [160](#)

- [117] Mengyu Guo, Binfeng Li, Zhihai Zhang, Su Wu, and Jie Song. Efficiency evaluation for allocating community-based health services. *Computers & Industrial Engineering*, 65(3):395–401, 2013. [161](#)
- [118] Walter J Gutjahr and Nada Dzubur. Bi-objective bilevel optimization of distribution center locations considering user equilibria. *Transportation Research Part E: Logistics and Transportation Review*, 85:1–22, 2016. [59](#), [63](#), [69](#)
- [119] Mahbubeh Haghi, Seyed Mohammad Taghi Fatemi Ghomi, and Farihorz Jolai. Developing a robust multi-objective model for pre/post disaster times under uncertainty in demand and resource. *Journal of cleaner production*, 154:188–202, 2017. [65](#)
- [120] Pierre Hansen, Nenad Mladenović, and José A Moreno Pérez. Variable neighbourhood search: methods and applications. *Annals of Operations Research*, 175(1):367–407, 2010. [93](#)
- [121] Shane G Henderson. Operations research tools for addressing current challenges in emergency medical services, 2011. [122](#)
- [122] Patricio Hernandez, Antonio Alonso-Ayuso, Fernanda Bravo, Laureano F Escudero, Monique Guignard, Vladimir Marianov, and Andres Weintraub. A branch-and-cluster coordination scheme for selecting prison facility sites under uncertainty. *Computers & Operations Research*, 39(9):2232–2241, 2012. [188](#), [196](#)
- [123] José Holguín-Veras, Miguel Jaller, and Tricia Wachtendorf. Comparative performance of alternative humanitarian logistic structures after the port-au-prince earthquake: ACEs, PIEs, and CANs. *Transportation research part A: policy and practice*, 46(10):1623–1640, 2012. [72](#), [77](#)
- [124] José Holguín-Veras, Noel Pérez, Miguel Jaller, Luk N Van Wassenhove, and Felipe Aros-Vera. On the appropriate objective function for post-disaster humanitarian logistics models. *Journal of Operations Management*, 31(5):262–280, 2013. [60](#)

- [125] Maria Camila Hoyos, Ridley S Morales, and Raha Akhavan-Tabatabaei. Or models with stochastic components in disaster operations management: A literature survey. *Computers & Industrial Engineering*, 82:183–197, 2015. [53](#)
- [126] Shao-Long Hu, Chuan-Feng Han, and Ling-Peng Meng. Stochastic optimization for investment in facilities in emergency prevention. *Transportation research part E: logistics and transportation review*, 89:14–31, 2016. [66](#)
- [127] Shao-Long Hu, Chuan-Feng Han, and Ling-Peng Meng. Stochastic optimization for joint decision making of inventory and procurement in humanitarian relief. *Computers & Industrial Engineering*, 111:39–49, 2017. [61](#), [63](#)
- [128] Zhi-Hua Hu, Jiuh-Biing Sheu, Yu-Qi Yin, and Chen Wei. Post-disaster relief operations considering psychological costs of waiting for evacuation and relief resources. *Transportmetrica A: Transport Science*, 13(2):108–138, 2017. [59](#)
- [129] Michael Huang, Karen Smilowitz, and Burcu Balcik. Models for relief routing: Equity, efficiency and efficacy. *Transportation research part E: logistics and transportation review*, 48(1):2–18, 2012. [62](#)
- [130] Xiaoxia Huang and Liying Song. An emergency logistics distribution routing model for unexpected events. *Annals of Operations Research*, 269(1-2):223–239, 2018. [66](#), [70](#), [73](#)
- [131] Ana Paula Iannoni, Reinaldo Morabito, and Cem Saydam. An optimization approach for ambulance location and the districting of the response segments on highways. *European Journal of Operational Research*, 195(2):528–542, 2009. [129](#)
- [132] Armann Ingolfsson. Ems planning and management. In *Operations Research and Health Care Policy*, pages 105–128. Springer, 2013. [122](#)
- [133] Armann Ingolfsson, Susan Budge, and Erhan Erkut. Optimal ambulance location with random delays and travel times. *Health Care management science*, 11(3):262–274, 2008. [131](#)

- [134] Charbel José Chiappetta Jabbour, Vinicius Amorim Sobreiro, Ana Beatriz Lopes de Sousa Jabbour, Lucila Maria de Souza Campos, Enzo Barberio Mariano, and Douglas William Scott Renwick. An analysis of the literature on humanitarian logistics and supply chain management: paving the way for future studies. *Annals of Operations Research*, pages 1–19, 2017. [53](#)
- [135] James P Jarvis. Approximating the equilibrium behavior of multi-server loss systems. *Management Science*, 31(2):235–239, 1985. [129](#)
- [136] Afshin Kamyabniya, MM Lotfi, Mohsen Naderpour, and Yuehwern Yih. Robust platelet logistics planning in disaster relief operations under uncertainty: A coordinated approach. *Information Systems Frontiers*, 20(4):759–782, 2018. [60](#), [63](#)
- [137] İmdat Kara, Bahar Yetiş Kara, and M Kadri Yetiş. Cumulative vehicle routing problems. In *Vehicle Routing Problem*. InTech, 2008. [6](#)
- [138] Soheyl Khalilpourazari and Alireza Arshadi Khamseh. Bi-objective emergency blood supply chain network design in earthquake considering earthquake magnitude: a comprehensive study with real world application. *Annals of Operations Research*, pages 1–39, 2017. [59](#)
- [139] Sara Khodaparasti, Maria Elena Bruni, Patrizia Beraldi, Maleki, and Sedigheh Jahedi. A multi-period location-allocation model for nursing home network planning under uncertainty. *Operations Research for Health Care*, 2018. [4](#)
- [140] Sara Khodaparasti, Hamid Reza Maleki, Maria Elena Bruni, Sedigheh Jahedi, Patrizia Beraldi, and Domenico Conforti. Balancing efficiency and equity in location-allocation models with an application to strategic ems design. *Optimization Letters*, 10(5):1053–1070, 2016. [3](#), [124](#), [126](#), [162](#), [187](#)
- [141] Sara Khodaparasti, Hamid Reza Maleki, Sedigheh Jahedi, Maria Elena Bruni, and Patrizia Beraldi. Enhancing community based health programs in iran: a multi-objective location-allocation model. *Health care management science*, 20(4):485–499, 2017. [3](#), [187](#)

- [142] Dong-Guen Kim and Yeong-Dae Kim. A lagrangian heuristic algorithm for a public healthcare facility location problem. *Annals of Operations Research*, 206(1):221–240, 2013. [159](#)
- [143] Ronald K Klimberg and Samuel J Ratick. Modeling data envelopment analysis (dea) efficient location/allocation decisions. *Computers & Operations Research*, 35(2):457–474, 2008. [139](#), [142](#), [174](#)
- [144] Roger Knott. The logistics of bulk relief supplies. *Disasters*, 11(2):113–115, 1987. [50](#)
- [145] RP Knott. Vehicle scheduling for emergency relief management: A knowledge-based approach. *Disasters*, 12(4):285–293, 1988. [50](#)
- [146] Mehmet Levent Koc and M Bostancioglu. A reliability based solution to an ambulance location problem using fuzzy set theory. *International Journal of Natural and Engineering Sciences*, 5(1):13–17, 2011. [135](#)
- [147] Pavlo Krokhmal, Michael Zabaranin, and Stan Uryasev. Modeling and optimization of risk. In *HANDBOOK OF THE FUNDAMENTALS OF FINANCIAL DECISION MAKING: Part II*, pages 555–600. World Scientific, 2013. [11](#), [76](#)
- [148] Kamran Bagheri Lankarani, Seyed Moyed Alavian, and Payam Peymani. Health in the islamic republic of iran, challenges and progresses. *Medical journal of the Islamic Republic of Iran*, 27(1):42, 2013. [157](#)
- [149] Mary P Larsen, Mickey S Eisenberg, Richard O Cummins, and Alfred P Hallstrom. Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Annals of emergency medicine*, 22(11):1652–1658, 1993. [123](#)
- [150] Richard C Larson. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, 1(1):67–95, 1974. [128](#)
- [151] Richard C Larson. Approximating the performance of urban emergency service systems. *Operations Research*, 23(5):845–868, 1975. [129](#)

- [152] Christophe Lecluyse, Tom Van Woensel, and Herbert Peremans. Vehicle routing with stochastic time-dependent travel times. *4OR*, 7(4):363, 2009. [12](#)
- [153] Xueping Li, Zhaoxia Zhao, Xiaoyan Zhu, and Tami Wyatt. Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research*, 74(3):281–310, 2011. [122](#), [123](#)
- [154] Federico Liberatore, M Teresa Ortuño, Gregorio Tirado, Begona Victoriano, and Maria Paola Scaparra. A hierarchical compromise model for the joint optimization of recovery operations and distribution of emergency goods in humanitarian logistics. *Computers & Operations Research*, 42:3–13, 2014. [72](#)
- [155] Chung-Cheng Lu, Kuo-Ching Ying, and Hui-Ju Chen. Real-time relief distribution in the aftermath of disasters—a rolling horizon approach. *Transportation research part E: logistics and transportation review*, 93:1–20, 2016. [68](#)
- [156] Abilio Lucena. Time-dependent traveling salesman problem—the deliveryman case. *Networks*, 20(6):753–763, 1990. [5](#)
- [157] E Luis, Irina S Dolinskaya, and Karen R Smilowitz. Disaster relief routing: Integrating research and practice. *Socio-economic planning sciences*, 46(1):88–97, 2012. [51](#)
- [158] Masoud Mahootchi and Sajjad Golmohammadi. Developing a new stochastic model considering bi-directional relations in a natural disaster: a possible earthquake in tehran (the capital of islamic republic of iran). *Annals of Operations Research*, 269(1-2):439–473, 2018. [64](#)
- [159] Marvin B Mandell. Modelling effectiveness-equity trade-offs in public service delivery systems. *Management Science*, 37(4):467–482, 1991. [137](#)
- [160] Rhoda Margesson and Maureen Taft-Morales. Haiti earthquake: Crisis and response. Library of Congress Washington DC Congressional Research Service, 2010. [77](#)

- [161] Vladimir Marianov and Charles ReVelle. The queueing maximal availability location problem: a model for the siting of emergency vehicles. *European Journal of Operational Research*, 93(1):110–120, 1996. [130](#), [143](#)
- [162] Nikola Marković, Ilya O Ryzhov, and Paul Schonfeld. Evasive flow capture: A multi-period stochastic facility location problem with independent demand. *European Journal of Operational Research*, 257(2):687–703, 2017. [187](#)
- [163] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952. [11](#), [76](#)
- [164] Michael T Marsh and David A Schilling. Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research*, 74(1):1–17, 1994. [137](#)
- [165] Albert W Marshall and Ingram Olkin. Multivariate chebyshev inequalities. *The Annals of Mathematical Statistics*, pages 1001–1014, 1960. [198](#)
- [166] George Mavrotas. Effective implementation of the ε -constraint method in multi-objective mathematical programming problems. *Applied mathematics and computation*, 213(2):455–465, 2009. [145](#)
- [167] Leslie D Mayhew and Giorgio Leonardi. Equity, efficiency, and accessibility in urban and regional health-care systems. *Environment and Planning A*, 14(11):1479–1507, 1982. [137](#)
- [168] Donald M McAllister. Equity and efficiency in public facility location. *Geographical Analysis*, 8(1):47–63, 1976. [137](#)
- [169] Garth P McCormick. Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976. [144](#)
- [170] Laura A McLay and Maria E Mayorga. Evaluating emergency medical service performance measures. *Health Care Management Science*, 13(2):124–136, 2010. [123](#), [124](#), [125](#), [131](#), [143](#)

- [171] Ramin Mehrdad. Health system in iran. *International Medical Community Journal*, 52:69–73, 2009. [157](#)
- [172] Isabel Méndez-Díaz, Paula Zabala, and Abilio Lucena. A new formulation for the traveling deliveryman problem. *Discrete applied mathematics*, 156(17):3223–3237, 2008. [5](#)
- [173] Juan A Mesa, Justo Puerto, and Arie Tamir. Improved algorithms for several network location problems with equality measures. *Discrete applied mathematics*, 130(3):437–448, 2003. [123](#)
- [174] Ana Maria Mestre, Mónica Duarte Oliveira, and Ana Barbosa-Póvoa. Organizing hospitals into networks: a hierarchical and multiservice model to define location, supply and referrals in planned hospital systems. *OR spectrum*, 34(2):319–348, 2012. [160](#), [162](#)
- [175] Ana Maria Mestre, Mónica Duarte Oliveira, and Ana Paula Barbosa-Póvoa. Location–allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, 240(3):791–806, 2015. [162](#), [186](#), [190](#), [191](#)
- [176] Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 2012. [124](#), [147](#)
- [177] Panagiotis Mitropoulos, Ioannis Mitropoulos, and Ioannis Giannikos. Combining dea with location analysis for the effective consolidation of services in the health sector. *Computers & Operations Research*, 40(9):2241–2250, 2013. [138](#), [139](#), [161](#), [164](#), [165](#), [196](#)
- [178] Nenad Mladenović, Dragan Urošević, and Saïd Hanafi. Variable neighborhood search for the travelling deliveryman problem. *4OR*, 11(1):57–73, 2013. [6](#)
- [179] Ahmad Mohamadi, Saeed Yaghoubi, and Mir Saman Pishvae. Fuzzy multi-objective stochastic programming model for disaster relief logistics considering telecommunication infrastructures: a case study. *Operational Research*, pages 1–41. [65](#), [66](#)

- [180] Mehrdad Mohammadi, Saleh Dehbari, and Behnam Vahdani. Design of a bi-objective reliable healthcare network with finite capacity queue under service covering uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 72:15–41, 2014. [160](#)
- [181] Reza Mohammadi, SMT Fatemi Ghomi, and F Jolai. Prepositioning emergency earthquake response supplies: a new multi-objective particle swarm optimization algorithm. *Applied Mathematical Modelling*, 40(9-10):5183–5199, 2016. [68](#)
- [182] Alfredo Moreno, Douglas Alem, Deisemara Ferreira, and Alistair Clark. An effective two-stage stochastic multi-trip location-transportation model with social concerns in relief supply chains. *European Journal of Operational Research*, 269(3):1050–1071, 2018. [61](#), [66](#)
- [183] M Moshfegh and G Hosseini. The futures studies for iran demographic changes during the period 2011 to 2041. *Ma’rifat-i Farhangi Ejtemaii Farhangi Ejtemaii*, 4(1):21–41, 2013. [201](#)
- [184] John M Mulvey, Robert J Vanderbei, and Stavros A Zenios. Robust optimization of large-scale systems. *Operations research*, 43(2):264–281, 1995. [65](#)
- [185] Stéphane Mussard, María Noel Pi Alperin, Françoise Seyte, and Michel Terraza. Extensions of Dagum’s Gini decomposition. In *Siena: International Conference in Memory of Gini and Lorenz*, 2005. [148](#)
- [186] Ram Narasimhan, Srinivas Talluri, Joseph Sarkis, and Anthony Ross. Efficient service location design in government services: a decision support system framework. *Journal of operations management*, 23(2):163–178, 2005. [139](#)
- [187] Malick Ndiaye and Hesham Alfares. Modeling health care facility location for moving population groups. *Computers & Operations Research*, 35(7):2154–2161, 2008. [189](#)

- [188] Sandra Ulrich Ngueveu, Christian Prins, and Roberto Wolfler Calvo. An effective memetic algorithm for the cumulative capacitated vehicle routing problem. *Computers & Operations Research*, 37(11):1877–1885, 2010. [6](#), [27](#), [38](#)
- [189] Stefan Nickel, Francisco Saldanha-da Gama, and Hans-Peter Ziegler. A multi-stage stochastic supply network design problem with financial decisions and risk management. *Omega*, 40(5):511–524, 2012. [188](#)
- [190] Reut Noham and Michal Tzur. Designing humanitarian supply chains by incorporating actual post-disaster decisions. *European Journal of Operational Research*, 265(3):1064–1077, 2018. [64](#)
- [191] Nilay Noyan. Alternate risk measures for emergency medical service system design. *Annals of Operations Research*, 181(1):559–589, 2010. [133](#)
- [192] Nilay Noyan, Burcu Balcik, and Semih Atakan. A stochastic optimization model for designing last mile relief networks. *Transportation Science*, 50(3):1092–1113, 2015. [61](#), [70](#)
- [193] Nilay Noyan and Gökçe Kahvecioğlu. Stochastic last mile relief network design with resource reallocation. *OR Spectrum*, 40(1):187–231, 2018. [61](#)
- [194] Samuel Nucamendi-Guillén, Iris Martínez-Salazar, Francisco Angel-Bello, and J Marcos Moreno-Vega. A mixed integer formulation and an efficient metaheuristic procedure for the k-travelling repairmen problem. *Journal of the Operational Research Society*, 67(8):1121–1134, 2016. [2](#), [22](#), [25](#), [74](#)
- [195] Ceyda Og, F Sibel Salman, Zehra Bilgintürk Yalçın, et al. Order acceptance and scheduling decisions in make-to-order systems. *International Journal of Production Economics*, 125(1):200–211, 2010. [114](#), [118](#)
- [196] Ibrahim Hassan Osman. Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem. *Annals of operations research*, 41(4):421–451, 1993. [38](#)

- [197] Susan Hesse Owen and Mark S Daskin. Strategic facility location: A review. *European journal of operational research*, 111(3):423–447, 1998. [187](#)
- [198] Linet Özdamar and Mustafa Alp Ertem. Models, solutions and enabling technologies in humanitarian logistics. *European Journal of Operational Research*, 244(1):55–65, 2015. [51](#)
- [199] Jean-Claude Picard and Maurice Queyranne. The time-dependent traveling salesman problem and its application to the tardiness problem in one-machine scheduling. *Operations Research*, 26(1):86–110, 1978. [8](#)
- [200] Elham Pourrahmani, Mahmoud Reza Delavar, and Mir Abolfazl Mostafavi. Optimization of an evacuation plan with uncertain demands using fuzzy credibility theory and genetic algorithm. *International Journal of Disaster Risk Reduction*, 14:357–372, 2015. [71](#)
- [201] Elham Pourrahmani, Mahmoud Reza Delavar, Parham Pahlavani, and Mir Abolfazl Mostafavi. Dynamic evacuation routing plan after an earthquake. *Natural Hazards Review*, 16(4):04015006, 2015. [71](#)
- [202] Marcelo Prais and Celso C Ribeiro. Reactive grasp: An application to a matrix decomposition problem in tdma traffic assignment. *INFORMS Journal on Computing*, 12(3):164–176, 2000. [28](#)
- [203] Fatemeh Rahimi, Maryam Esmaeili, Abolghasem Nouri, and Alireza Mahdavi. The investigation social capital with emphasize to non-governmental organization role’s in prevention of drug dependency. 2012. [163](#)
- [204] Abdur Rais and Ana Viana. Operations research in healthcare: a survey. *International transactions in operational research*, 18(1):1–31, 2011. [159](#)
- [205] Stefan Rath, Michel Gendreau, and Walter J Gutjahr. Bi-objective stochastic programming models for determining depot locations in disaster relief operations. *International Transactions in Operational Research*, 23(6):997–1023, 2016. [63](#), [64](#), [68](#)

- [206] Panagiotis P Repoussis, Dimitris C Paraskevopoulos, Alkiviadis Vazacopoulos, and Nathaniel Hupert. Optimizing emergency preparedness and resource utilization in mass-casualty incidents. *European Journal of Operational Research*, 255(2):531–544, 2016. [68](#)
- [207] Charles ReVelle, David Bigman, David Schilling, Jared Cohon, and Richard Church. Facility location: a review of context-free and ems models. *Health Services Research*, 12(2):129, 1977. [122](#)
- [208] Charles ReVelle and Kathleen Hogan. The maximum availability location problem. *Transportation science*, 23(3):192–200, 1989. [130](#)
- [209] Mohammad Rezaei-Malek, Reza Tavakkoli-Moghaddam, Behzad Zahiri, and Ali Bozorgi-Amiri. An interactive approach for designing a robust disaster relief logistics network with perishable commodities. *Computers & Industrial Engineering*, 94:201–215, 2016. [61](#), [65](#)
- [210] Glaydston Mattos Ribeiro and Gilbert Laporte. An adaptive large neighborhood search heuristic for the cumulative capacitated vehicle routing problem. *Computers & operations research*, 39(3):728–735, 2012. [6](#), [103](#)
- [211] Daniel Rivera-Royero, Gina Galindo, and Ruben Yie-Pinedo. A dynamic model for disaster response considering prioritized demand points. *Socio-economic planning sciences*, 55:59–75, 2016. [59](#)
- [212] R Tyrrell Rockafellar. Coherent approaches to risk in optimization under uncertainty. In *OR Tools and Applications: Glimpses of Future Technologies*, pages 38–61. Informs, 2007. [11](#)
- [213] Oscar Rodríguez-Espíndola, Pavel Albores, and Christopher Brewster. Disaster preparedness in humanitarian logistics: A collaborative approach for resource management in floods. *European Journal of Operational Research*, 264(3):978–993, 2018. [61](#), [67](#)
- [214] Carlos Rodriguez-Verjan, Vincent Augusto, Thierry Garaix, Xiaolan Xie, and Valerie Buthion. Healthcare at home facility location-allocation problem. In *Automation Science and Engineering (CASE)*,

- 2012 *IEEE International Conference on*, pages 150–155. IEEE, 2012. [189](#), [191](#)
- [215] Stefan Ropke and David Pisinger. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation science*, 40(4):455–472, 2006. [103](#), [104](#), [105](#), [107](#)
- [216] Güvenç Şahin, Haldun Süral, and Sedef Meral. Locational analysis for regionalization of turkish red crescent blood services. *Computers & Operations Research*, 34(3):692–704, 2007. [139](#), [160](#), [162](#)
- [217] Faraz Salehi, Masoud Mahootchi, and Seyed Mohammad Moattar Husseini. Developing a robust stochastic model for designing a blood supply chain network in a crisis: A possible earthquake in tehran. *Annals of Operations Research*, pages 1–25, 2017. [65](#), [66](#), [68](#)
- [218] Amir Salehipour, Kenneth Sörensen, Peter Goos, and Olli Bräysy. Efficient GRASP + VND and GRASP + VNS metaheuristics for the traveling repairman problem. *4or*, 9(2):189–209, 2011. [6](#)
- [219] Javier Salmerón and Aruna Apte. Stochastic optimization for natural disaster asset prepositioning. *Production and operations management*, 19(5):561–574, 2010. [51](#)
- [220] Mohammad Reza Ghatreh Samani, S Ali Torabi, and Seyyed-Mahdi Hosseini-Motlagh. Integrated blood supply chain planning for disaster relief. *International journal of disaster risk reduction*, 27:168–188, 2018. [64](#), [65](#)
- [221] David A Schilling. Dynamic location modeling for public-sector facilities: A multicriteria approach. *Decision Sciences*, 11(4):714–724, 1980. [187](#)
- [222] Verena Schmid. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European journal of operational research*, 219(3):611–621, 2012. [154](#)
- [223] Yue Sha and Jun Huang. The multi-period location-allocation problem of engineering emergency blood supply systems. *Systems Engineering Procedia*, 5:21–28, 2012. [190](#)

- [224] Mehrdad Shahabi, Avinash Unnikrishnan, and Stephen D Boyles. An outer approximation algorithm for the robust shortest path problem. *Transportation Research Part E: Logistics and Transportation Review*, 58:52–66, 2013. [199](#)
- [225] Shahrooz Shahparvari and Babak Abbasi. Robust stochastic vehicle routing and scheduling for bushfire emergency evacuation: An australian case study. *Transportation Research Part A: Policy and Practice*, 104:32–49, 2017. [62](#), [64](#), [65](#)
- [226] Shahrooz Shahparvari, Babak Abbasi, and Prem Chhetri. Possibilistic scheduling routing for short-notice bushfire emergency evacuation under uncertainties: An australian case study. *Omega*, 72:96–117, 2017. [62](#), [64](#), [65](#), [66](#)
- [227] Shahrooz Shahparvari, Babak Abbasi, Prem Chhetri, and Ahmad Abareshi. Fleet routing and scheduling in bushfire emergency evacuation: A regional case study of the black saturday bushfires in australia. *Transportation Research Part D: Transport and Environment*, 2017. [62](#), [63](#), [64](#), [65](#)
- [228] Shahrooz Shahparvari, Prem Chhetri, Babak Abbasi, and Ahmad Abareshi. Enhancing emergency evacuation response of late evacuees: Revisiting the case of australian black saturday bushfire. *Transportation research part E: logistics and transportation review*, 93:148–176, 2016. [56](#), [62](#), [65](#)
- [229] Lloyd S. Shapley and Martin Shubik. On market games. *Journal of Economic Theory*, 1:9–25, 1969. [127](#)
- [230] Afshin Shariat-Mohaymany, Mohsen Babaei, Saeed Moadi, and Sayyed Mahdi Amiripour. Linear upper-bound unavailability set covering models for locating ambulances: Application to tehran rural roads. *European Journal of Operational Research*, 221(1):263–272, 2012. [130](#)
- [231] Mojtaba Talebian Sharif and Majid Salari. A grasp algorithm for a humanitarian relief transportation problem. *Engineering Applications of Artificial Intelligence*, 41:259–269, 2015. [71](#)

- [232] Paul Shaw. A new local search algorithm providing high quality solutions to vehicle routing problems. *APES Group, Dept of Computer Science, University of Strathclyde, Glasgow, Scotland, UK*, 1997. [107](#)
- [233] Holger Sheel. Undesirable outputs in efficiency evaluation. *European Journal of Operational Research*, 132:400–410, 2001. [143](#)
- [234] Ehsan Shekarian. A novel application of the vikor method for investigating the effect of education on housing choice. *International Journal of Operational Research*, 24(2):161–183, 2015. [184](#)
- [235] Marcos Melo Silva, Anand Subramanian, Thibaut Vidal, and Luiz Satoru Ochi. A simple and effective metaheuristic for the minimum latency problem. *European Journal of Operational Research*, 221(3):513–520, 2012. [6](#)
- [236] Honora K Smith, Paul R Harper, and Chris N Potts. Bicriteria efficiency/equity hierarchical location models for public service application. *Journal of the Operational Research Society*, 64(4):500–512, 2013. [124](#), [138](#), [162](#)
- [237] Honora K Smith, Paul R Harper, Chris N Potts, and Ann Thyle. Planning sustainable community health schemes in rural areas of developing countries. *European Journal of Operational Research*, 193(3):768–777, 2009. [160](#), [162](#)
- [238] Byung Duk Song, Young Dae Ko, and Hark Hwang. The design of capacitated facility networks for long term care service. *Computers & Industrial Engineering*, 89:177–185, 2015. [159](#), [189](#), [191](#)
- [239] Paul Sorensen and Richard Church. Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences*, 44(1):8–18, 2010. [130](#)
- [240] Maria Teresinha Arns Steiner, Dilip Datta, Pedro José Steiner Neto, Cassius Tadeu Scarpin, and José Rui Figueira. Multi-objective optimization in partitioning the healthcare system of parana state in brazil. *Omega*, 52:53–64, 2015. [161](#)

- [241] Ralph E Steuer. *Multiple criteria optimization: Theory, Computation, and application*. Willy, 1986. [169](#)
- [242] Li Sun, Gail W DePuy, and Gerald W Evans. Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. *Computers & Operations Research*, 51:350–359, 2014. [160](#)
- [243] Siddhartha S Syam and Murray J Côté. A location–allocation model for service providers with application to not-for-profit health care organizations. *Omega*, 38(3-4):157–166, 2010. [161](#)
- [244] Siddhartha S Syam and Murray J Côté. A comprehensive location-allocation method for specialized healthcare services. *Operations Research for Health Care*, 1(4):73–83, 2012. [159](#)
- [245] Jeeu Fong Sze, Said Salhi, and Niaz Wassan. The cumulative capacitated vehicle routing problem with min-sum and min-max objectives: An effective hybridisation of adaptive variable neighbourhood search and large neighbourhood search. *Transportation Research Part B: Methodological*, 101:162–184, 2017. [6](#)
- [246] M Taghvaii, A Zarrabi, and B Moghany Rahimi. Investigating effective factors of crimes in various areas. (48):1–41, 2010. (in Persian). [172](#)
- [247] Renata Algisi Takeda, João A Widmer, and Reinaldo Morabito. Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. *Computers & Operations Research*, 34(3):727–741, 2007. [129](#)
- [248] Peter Thomas, Yupu Chan, Lee Lehmkuhl, and William Nixon. Obnoxious-facility location and data-envelopment analysis: A combined distance-based formulation. *European Journal of Operational Research*, 141(3):495–514, 2002. [138](#)
- [249] RN Tiwari, S Dharmar, and JR Rao. Fuzzy goal programming—an additive model. *Fuzzy sets and systems*, 24(1):27–34, 1987. [169](#), [170](#)
- [250] S Tofghi, S Ali Torabi, and S Afshin Mansouri. Humanitarian logistics network design under mixed uncertainty. *European Journal of Operational Research*, 250(1):239–250, 2016. [60](#), [65](#)

- [251] Hector Toro-Díaz, Maria E Mayorga, Laura A McLay, Hari K Rajagopalan, and Cem Saydam. Reducing disparities in large-scale emergency medical service systems. *Journal of the Operational Research Society*, 66(7):1169–1181, 2015. [126](#)
- [252] Takeshi Uno, Kosuke Kato, and Hideki Katagiri. An application of interactive fuzzy satisficing approach with particle swarm optimization for multiobjective emergency facility location problem with a-distance. In *Computational Intelligence in Multicriteria Decision Making, IEEE Symposium on*, pages 368–373. IEEE, 2007. [135](#)
- [253] Behnam Vahdani, D Veysmoradi, F Noori, and F Mansour. Two-stage multi-objective location-routing-inventory model for humanitarian logistics network design under uncertainty. *International journal of disaster risk reduction*, 27:290–306, 2018. [65](#)
- [254] Behnam Vahdani, D Veysmoradi, N Shekari, and S Meysam Mousavi. Multi-objective, multi-period location-routing model to distribute relief after earthquake by considering emergency roadway repair. *Neural Computing and Applications*, 30(3):835–854, 2018. [59](#), [60](#), [69](#)
- [255] David Valinsky. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. *Journal of the Operations Research Society of America*, 3(4):494–512, 1955. [50](#)
- [256] Martijn van Ee and René Sitters. The a priori traveling repairman problem. *Algorithmica*, 80(10):2818–2833, 2018. [10](#)
- [257] CA Van Eijl. *A polyhedral approach to the delivery man problem*. Department of Math. and Computing Science, University of Technology, 1995. [5](#)
- [258] Sofia Vaz, Pedro Ramos, and Paula Santana. Distance effects on the accessibility to emergency departments in portugal. *Saúde e Sociedade*, 23:1154–1161, 2014. [186](#)

- [259] Vedat Verter and Sophie D Lapiere. Location of preventive health care facilities. *Annals of Operations Research*, 110(1-4):123–132, 2002. [164](#), [186](#), [196](#)
- [260] Jorge F Victoria, H Murat Afsar, and Christian Prins. Vehicle routing problem with time-dependent demand in humanitarian logistics. In *Industrial Engineering and Systems Management (IESM), 2015 International Conference on*, pages 686–694. IEEE, 2015. [72](#)
- [261] Begoña Vitoriano, M Teresa Ortuño, Gregorio Tirado, and Javier Montero. A multi-criteria optimization model for humanitarian aid distribution. *Journal of Global Optimization*, 51(2):189–208, 2011. [72](#)
- [262] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 2007. [11](#)
- [263] Xinyu Wang, Tsan-Ming Choi, Haikuo Liu, and Xiaohang Yue. A novel hybrid ant colony optimization algorithm for emergency transportation problems during post-disaster scenarios. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(4):545–556, 2018. [59](#), [62](#), [70](#), [72](#)
- [264] Sascha Wohlgemuth, Richard Oloruntoba, and Uwe Clausen. Dynamic vehicle routing with anticipation in disaster relief. *Socio-Economic Planning Sciences*, 46(4):261–271, 2012. [71](#)
- [265] Wenqi Yi, Linda Nozick, Rachel Davidson, Brian Blanton, and Brian Colle. Optimization of the issuance of evacuation orders under evolving hurricane conditions. *Transportation Research Part B: Methodological*, 95:285–304, 2017. [63](#), [64](#), [70](#)
- [266] Chian-Son Yu and Han-Lin Li. A robust optimization model for stochastic logistic problems. *International journal of production economics*, 64(1-3):385–397, 2000. [65](#)
- [267] Behzad Zahiri, Reza Tavakkoli-Moghaddam, and Mir Saman Pishvaei. A robust possibilistic programming approach to multi-period location–allocation of organ transplant centers under uncertainty. *Computers & Industrial Engineering*, 74:139–148, 2014. [189](#), [191](#)

- [268] Behzad Zahiri, Seyed Ali Torabi, and Reza Tavakkoli-Moghaddam. A novel multi-stage possibilistic stochastic programming approach (with an application in relief distribution planning). *Information Sciences*, 385:225–249, 2017. [63](#), [64](#), [65](#), [66](#)
- [269] Qiang-Lin Zeng, Dan-Dan Li, and Yi-Bin Yang. Vikor method with enhanced accuracy for multiple criteria decision making in healthcare management. *Journal of medical systems*, 37(2):9908, 2013. [184](#)
- [270] Sha-lei Zhan and Nan Liu. Determining the optimal decision time of relief allocation in response to disaster via relief demand updates. *International Journal of Systems Science*, 47(3):509–520, 2016. [62](#), [66](#)
- [271] Jun Zhang, Jia-fu Tang, Zhen-dong Pan, and Yuan Kong. Scatter search algorithm for solving weighted vehicle routing problem. *Journal of Systems Engineering*, 1:91–97, 2010. [6](#)
- [272] Jun Zhang, Jiafu Tang, and Richard YK Fung. A scatter search for multi-depot vehicle routing problem with weight-related cost. *Asia-Pacific Journal of Operational Research*, 28(03):323–348, 2011. [6](#)
- [273] Yue Zhang, Oded Berman, and Vedat Verter. The impact of client choice on preventive healthcare facility network design. *OR spectrum*, 34(2):349–370, 2012. [159](#)
- [274] Zhi-Hai Zhang and Hai Jiang. A robust counterpart approach to the bi-objective emergency medical service design problem. *Applied Mathematical Modelling*, 38(3):1033–1040, 2014. [134](#)
- [275] Zhi-Hai Zhang and Kang Li. A novel probabilistic formulation for locating and sizing emergency medical service stations. *Annals of Operations Research*, 229(1):813–835, 2015. [132](#)
- [276] Ming Zhao and Xiang Liu. Development of decision support tool for optimizing urban emergency rescue facility locations to improve humanitarian logistics management. *Safety science*, 102:110–117, 2018. [67](#)

- [277] H-J Zimmermann. Fuzzy programming and linear programming with several objective functions. *Fuzzy sets and systems*, 1(1):45–55, 1978. [169](#)
- [278] Shiva Zokaei, Ali Bozorgi-Amiri, and Seyed Jafar Sadjadi. A robust optimization model for humanitarian relief chain design under uncertainty. *Applied Mathematical Modelling*, 40(17):7996–8016, 2016. [65](#), [66](#)