

**SELECTIVE ROUTE PLANNING FOR POST-DISASTER
RESILIENCE ASSESSMENT UNDER UNCERTAIN KEY
INFORMANT PRESENCE**

ASENA AYŞE KAPLAN

ÖZYEGİN UNIVERSITY

JULY, 2025

SELECTIVE ROUTE PLANNING FOR POST-DISASTER RESILIENCE ASSESSMENT UNDER UNCERTAIN KEY INFORMANT PRESENCE

by
Asena Ayşe Kaplan

Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

in
Industrial Engineering

Advisor: Prof. Burcu Balçık Koyuncu

Graduate School of Science and Engineering
Özyeğin University
İstanbul

July, 2025

SELECTIVE ROUTE PLANNING FOR POST-DISASTER RESILIENCE ASSESSMENT UNDER UNCERTAIN KEY INFORMANT PRESENCE

Approved by:

Prof. Burcu Balçık Koyuncu, Advisor
Department of Industrial Engineering
Özyeğin University

Assoc. Prof. Dilek Günneç Danış
Department of Industrial Engineering
Özyeğin University

Prof. Raha Akhavan Tabatabaei
Department of Operations Management and
Information Systems
Sabancı University

Approval Date: June 30, 2025

To my mom and to my best friend...



DECLARATION OF ORIGINALITY

I hereby declare that I am the sole author of this thesis and that this is the true copy of my thesis, including the final revisions, approved by my thesis committee. All data and information have been obtained, produced, and presented in accordance with the rules of research ethics and principles of academic honesty. As required by these rules, to the best of my knowledge I have acknowledged ideas, thoughts, and any copyrighted material in accordance with the standard referencing rules. I certify that any part of this thesis has not been submitted for a degree or diploma in another educational institution.

Asena Ayşe Kaplan

ABSTRACT

After a disaster, rapid assessment reports provide invaluable insights into immediate needs during the first response, while long-term needs assessment studies are essential for understanding the evolving vulnerabilities of populations during the recovery and reconstruction phases. In this study, we focus on a selective routing problem under uncertainty, where field teams visit earthquake-affected sites to conduct long-term needs assessment studies with key informants with uncertain presence, about the resilience of agricultural and educational activities to disasters. Specifically, we select a sample of sites for visits and prepare daily routes for the field teams, considering the possibility of absent key informants. There are two key challenges in this problem: (1) ensuring sample representativeness in site selection, and (2) accounting for the uncertain presence of key informants in preparing efficient routing plans. To address representativeness concerns, we use a stratified sampling approach leveraging a clustering algorithm to define groups of similar sites and ensure a proportional number of visits to each. The uncertainty in the key informant's presence is tackled by a novel main-backup assignment strategy, where each selected main site is paired with a backup site from the same cluster, enabling field teams to maintain representativeness. We develop a 2-phase matheuristic approach to solve this problem. It combines two enumeration strategies with a set partitioning model to generate feasible routes and select efficient ones, ensuring representativeness. The developed solution approach is used on a large dataset of 11 provinces in Türkiye to support a real-life assessment project following the February 2023 earthquakes. The findings show that different enumeration strategies can be more advantageous in terms of solution times and expected routing costs when the probability of the key informant's presence varies.

Keywords: needs assessment, humanitarian assessment, post-disaster selective routing, routing under uncertainty

ÖZET

Hızlı değerlendirme raporları afetlerin hemen ardından acil ihtiyaçları öğrenmek için çok önemlidirler. Bunun yanında, zamanla değişen ve gelişen ihtiyaçları ve kırılganlıkları anlamak amacıyla, uzun vadeli ihtiyaç değerlendirmeleri yapılmaktadır. Bu çalışmada, saha ekiplerinin depremde etkilenen kırsal köylere, tarım aktivitelerinin depremlere olan dirençliliklerini ve felaketlerden nasıl etkilendiklerini öğrenmek için anket yapmak üzere gerçekleştireceği ziyaretlerin planlanması için belirsizlik altında seçici rotalama problemine odaklanılmaktadır. Özellikle, ziyaret edilecek bölgeden bir örneklem seçilmesi, ve anket yapılacak yetkili kişilerin köyde bulunmama olasılıkları altında günlük rotalar planlanmalıdır. Bu problemde iki temel zorluk bulunmaktadır: (1) örneklem temsil edilebilirliğinin sağlanması, ve (2) belirsizlik altında verimli rotaların hazırlanması. Temsil edilebilirliği sağlamak için, bir kümeleme yöntemiyle benzer köyleri grupluyor, ve her kümeden o kümenin boyutuna orantılı köy seçilmesini sağlayan katmanlı bir örneklem yaklaşımı kullanıyoruz. Bunun yanında, belirsizliği çözüme katmak için ziyaret edilmek üzere seçilen her ana köye bir yedek köy atıyoruz. Yalnızca aynı kümeden atanan yedek köylere, ana köydeki yetkili bulunmadığında gidilmektedir. Bu problemi çözmek için iki adımlı bir matematik-sezgisel yaklaşım geliştirilmiştir: Aşama 1: 2 sınırlı sayımlama stratejisi ile sınırlı sayıda rotaların oluşturulması, ve Aşama 2: bir küme bölümleme modeliyle bu rotalardan, örneklem temsiline dikkat ederek, en düşük ücretli rotaların seçilmesi. Çözüm yaklaşımı, Şubat 2023 depremlerinden etkilenen 11 ilde planlanan bir saha projesini desteklemek amacıyla kullanılmıştır. Bulgular gösteriyor ki, yetkililerin köyde bulunma olasılığına göre, çözüm süresi ve beklenen rota maliyetleri açısından farklı sınırlı sayımlama stratejileri avantajlı olabilmektedir.

Anahtar Kelimeler: ihtiyaç değerlendirmesi, insani yardım değerlendirmesi, afet sonrası seçici rotalama, belirsizlik altında rotalama

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who supported me throughout this academic journey. First and foremost, I am deeply grateful to my advisor, Prof. Burcu Balçık. Her support, mentorship, and insights taught me invaluable lessons: that hard and efficient work requires rest, accountability paired with compassion creates growth, and making mistakes is a stepping stone toward learning and progress. I also extend my heartfelt appreciation to Prof. Marie-Eve Rancourt, Prof. Jessica Rodriguez-Pereira, and Prof. Murat Öztürk, with whom I had the chance to collaborate on this study. Their collaborative spirit, support, and knowledge greatly enriched my M.Sc. years. Moreover, I want to thank Assoc. Prof. Elvin Çoban, Prof. Julie Paquette, and Assoc. Prof. İhsan Yanıkoğlu for their support and guidance. Finally, I would like to thank all the faculty members of the Department of Industrial Engineering for supporting me throughout my B.Sc. and M.Sc. studies.

I am also forever indebted to my beloved family. To my mom who taught me the value of independence, to my little brother who gave me the best childhood filled with laughter, to my father who showed me not to take life too seriously, and to my aunt and uncle who always supported me- thank you so much for everything. Demirhan, who has been my constant companion, I cannot express how grateful I am for your love and support. Thank you for always being there for me in the most challenging times and for helping me in every way possible.

I also want to extend my thanks to my dearest friend and colleague, Ayşe Çoban, for her love and encouragement since the first day of high school. Thanks to my dearest friends Irmak, Özge, and Naz, for their support. Also I want to extend my appreciation to Ecem, Kian, Cem, and Berk for their guidance during my M.Sc. and B.Sc. years.

Finally, this project is conducted in coordination with the 1001-BICABA (Birlikte Çalışıp Birlikte Başaracağız- A Scholarship Program) Project (No: 123K891) and is supported by a grant under the 2219 Project (“Food Security In Times Of Disasters: Data-Driven Decision-Making Models And Strategies”), both funded by the Scientific and Technical Research Council of Turkey.



TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	iv
ABSTRACT	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ACRONYMS AND ABBREVIATIONS	xiii
1 INTRODUCTION	1
2 LITERATURE REVIEW	6
2.1 Humanitarian Assessment Problems	6
2.2 Vehicle Routing Problems Literature	8
2.3 Contribution of Our Study	12
3 PROBLEM STATEMENT	13
3.1 System Background and Decision Needs	13
3.2 Problem Aspects	16
3.2.1 Sampling	17
3.2.2 Logistical Planning Under Uncertainty	18
3.3 Problem Formulation	20
4 SOLUTION METHOD	22
4.1 Clustering	23
4.1.1 Relevant Attributes	23
4.1.2 Sample Size	24
4.1.3 Sampling Approach	24
4.2 Set Partitioning	28
4.2.1 Phase I: Enumerating Subsets of Villages	29

4.2.2	Phase II: Set partitioning model	31
5	DATA ANALYSIS AND COMPUTATIONAL RESULTS	35
5.1	Data Collection and Processing	35
5.1.1	Clustering	37
5.2	Computational Analysis of the Matheuristic.....	38
5.2.1	Base Results	39
5.2.2	Parameter Analysis	43
5.2.3	Result Visualizations	46
6	CONCLUSION.....	49
	REFERENCES	51
	APPENDICES.....	57
	APPENDIX A — SARP MATHEMATICAL MODEL FORMULATION	58

LIST OF TABLES

Table 2.1: Humanitarian Assessment Literature Comparison.	9
Table 4.1: Province Data for Sample Sizes.	25
Table 5.1: Sources for Attributes Used in the Field Study (V/D : Village/District, N/B : Numeric/Binary, I : Importance Score, * <i>explained in the text</i>).	37
Table 5.2: Clustering Input and Outputs for Each Province.	38
Table 5.3: Expected Costs and Total Solution Times (mins) with Base Parameters. .	39
Table 5.4: Decrease in Expected Cost from Main-Backup to AllMain-AllBackup (%)	40
Table 5.5: Total Solution Times (min).	41
Table 5.6: Correlation of Enumeration and Set Partitioning Times (min) with Province Properties.	41
Table 5.7: Detailed Enumeration and Set Partitioning Times (min).	42
Table 5.8: Expected Cost Difference (%) from Main-Backup to AllMain-AllBackup Enumeration Algorithms.....	45
Table 5.9: Average Expected Cost and Set Partitioning Solution Times (min) Required for Different p^i Values (* <i>some instances are solved with a gap due to 30 minute time limit</i>).....	46

LIST OF FIGURES

Figure 1.1: Village Locations in 11 Provinces of Türkiye (Prepared using [1, 2]) ...	2
Figure 3.1: Türkiye's affected area and major epicenters (Reprinted from [3]).	14
Figure 3.2: Example Network for SARP	19
Figure 4.1: Solution Methodology Steps	22
Figure 5.2: Expected Costs with Base Parameters	40
Figure 5.1: Clustering Results	47
Figure 5.3: Example Solution Representation for Adana.....	48

LIST OF ACRONYMS AND ABBREVIATIONS

CVRP	Capacitated Vehicle Routing Problem
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
GVRP	Generalized Vehicle Routing Problem
SARP	Stochastic Assessment Routing Problem
SRP	Selective Routing Problem
TSP	Travelling Salesperson Problem
TSPSC	Travelling Salesperson Problem with Stochastic Customers
VRP	Vehicle Routing Problem
VRPSC	Vehicle Routing Problem with Stochastic Customers
VRPSDC	Vehicle Routing Problem with Stochastic Demand and Customers
VRPSR	Vehicle Routing Problem with Stochastic Requests

1. INTRODUCTION

In February 2023, southern Türkiye was hit by multiple earthquakes affecting 9.1 million people, of whom 35% residing in rural areas, causing more than 45,000 deaths and 110 thousand injuries in 11 provinces of Türkiye [4]. The affected region is also referred to as “Türkiye’s fertile crescent”, housing 14% of the country’s farmers to contribute 15.3% of the national agricultural gross domestic product and 20% of the agrifood exports [5]. The total economic damage to the country’s agricultural sector is estimated to be \$6.3 billion [4]. To recover from such severe impacts on agricultural activities and support the food supply chain in Türkiye, it is crucial to understand the long-term effects of the earthquakes, as well as study the factors contributing to the vulnerability and resilience of agricultural activities in rural areas to disasters.

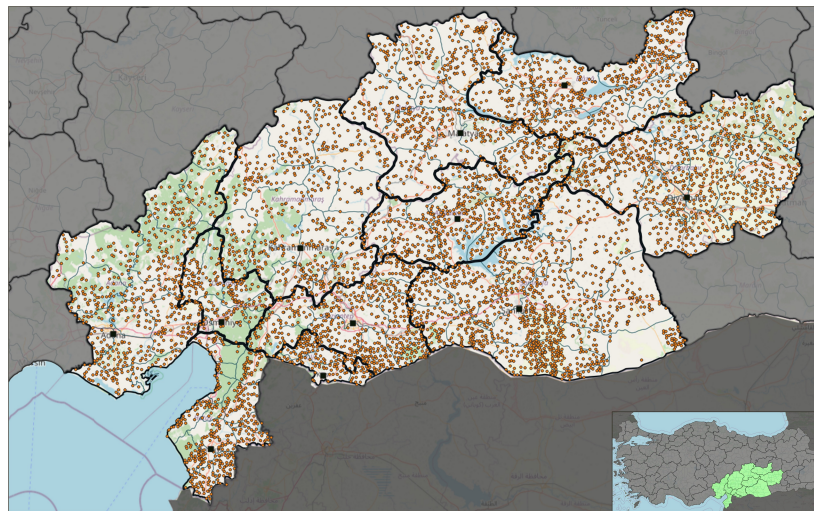
Our study is motivated by the need for logistical planning to conduct field assessments in rural areas following the February 2023 earthquake in Türkiye. Scheduled approximately 2 years after the earthquake, this detailed field assessment aims to evaluate the long-term impacts on food security and agricultural livelihoods, and understand the factors affecting the resilience of agricultural activities in rural areas to disasters by collecting essential information from villages in the region. Interdisciplinary field teams, supported by a survey company, visit selected villages within a predetermined budget and time frame to conduct detailed assessments through surveys with key informants, such as village heads, council members, and teachers. Specifically, until they complete a pre-specified number of assessment surveys in the villages of 11 provinces, they will visit villages daily and try to conduct field assessments with key informants.

The planning of these assessments presents several key challenges in this unique context. First, ensuring that the chosen sample of villages for visits accurately represents the target population is crucial in needs assessment problems [6]. Moreover, a dataset of relevant attributes for 11 provinces was not present, necessitating a large-scale data

collection and processing to define “similar” and “different” villages for ensuring representativeness. The second challenge involves uncertainties related to the potential unavailability of key informants, learned only upon the team’s arrival in the selected villages. If the key informant is not available, then the assessment survey cannot be completed in that village.

We collaborate with the field study team to support their assessment planning decisions and address these key challenges, specifically focusing on selecting a representative sample of approximately 2000 villages from a total of 6296 villages spread over 11 provinces in a large area (Figure 1.1) and planning the daily routes under uncertainty for teams to follow during the assessment study. Over a 2-year collaboration, we develop a data analysis framework and solution method that offers a systematic approach for humanitarian organizations to optimize their assessment strategies, specifically sampling and efficient route planning, ultimately contributing to more targeted and effective needs assessment and aid delivery in crises. Moreover, our close collaboration ensures that our study is grounded in practical needs and field insights, ensuring the relevance and applicability of our proposed solutions.

Figure 1.1: *Village Locations in 11 Provinces of Türkiye (Prepared using [1, 2])*



It is observed that most studies in the needs assessment literature concentrate on rapid needs assessments and sample selection based on community characteristics, using various methods to ensure representative coverage. To address representativeness concerns, our research introduces a novel approach by applying clustering techniques to large-scale datasets and adopting a stratified sampling framework to group villages based on similarities. Specifically, we collect and combine open-access data from both governmental and non-governmental sources to build a dataset of relevant attributes for 6296 villages from 126 districts, for the 11 affected provinces. The final set of relevant attributes (such as population, altitude, agricultural products, etc.) is decided by the field team, and is consistent with the attributes considered in previous literature of needs assessment studies in the agricultural context [7, 8]. To integrate data from multiple sources using different naming conventions and Turkish characters in the village names, we employed a basic natural language processing technique, regular expressions for pattern matching. The dataset for each province is then used in a weighted k-means clustering algorithm to form groups of similar villages. One major challenge in clustering was the high number of district attributes, over a few village-specific attributes. Specifically, districts are already pre-determined groups of villages. Therefore, adjusting the weights for different attributes to distinguish villages belonging to the same district, while allowing villages from different districts to be clustered together, required defining and analyzing a “diversity” metric. Finally, upon forming the final clusters, a proportional number of villages from each cluster is selected for the final sample to ensure representativeness in each province.

Moreover, to incorporate the potential key informant unavailability, our approach involves creating a priori routes that include backup villages for each selected main village. Specifically, to tackle the uncertainties related to key informant availability, we develop a new problem formulation that selects primary sites (main villages) based on

cluster features and includes backup villages within the same clusters. This enhances the accuracy of field assessments by ensuring alternative sites are available if the main villages are inaccessible. To incorporate this main-backup logic, we employ a scenario-based approach. That is, for a selected subset of villages (a possible set of main and backup pairs selected for a daily plan), we can generate scenarios based on the key informant's availability at each main village. We use an enumeration-based set partitioning model to select a set of efficient subsets (daily plans) from a feasible set of subsets generated by using different enumeration algorithms. Note that the large scale of this problem and the high number of subsets and scenarios do not allow us to solve a mathematical optimization problem to optimality. Additionally, while existing studies primarily focus on immediate post-disaster assessments, our research tackles long-term, detailed evaluations aimed at supporting recovery programs, specifically in the context of food security.

Therefore, the goal of this work is to solve a Stochastic Assessment Routing Problem (SARP) for a post-disaster field assessment, which integrates both site selection for assessments (sampling) and the routing of field teams while accounting for uncertainties in key informant unavailability. Specifically, we develop a set-partitioning-based matheuristic solution methodology capable of solving the SARP in realistic instances. Moreover, to ensure the practicality of delivered solutions to field teams, we prepare spreadsheets for the final sampling decisions and represent each daily routing plan in a decision-tree format to make sure that routes generated under uncertainty can be followed intuitively.

The remainder of this study is organized as follows: Chapter 2 reviews the relevant literature, positioning our work within the context of needs assessment in humanitarian logistics, selective routing, and vehicle routing problems under uncertainty. Chapter 3 provides a detailed problem description of the SARP. Chapter 4 presents the developed set-partitioning-based matheuristic approach, leveraging different enumeration strategies.

Chapter 5 explains the details on data collection and analysis, with a discussion of clustering and sample size distribution across provinces, and presents the computational results demonstrating the application of our set-partitioning-based matheuristic to the Türkiye data. Finally, Chapter 6 concludes this study with a discussion of implications, limitations, and directions for future research.



2. LITERATURE REVIEW

The SARP determines a set of sites to visit for conducting needs assessment surveys while ensuring representativeness across various site characteristics, and constructs efficient routes that account for uncertainty in key informant (site) availability. This problem is rooted in the context of humanitarian assessment problems. Humanitarian needs assessment problems in which field teams must be strategically deployed to perform on-site assessment surveys that evaluate both the immediate and long-term needs of disaster-affected regions are specifically related to our study. Moreover, the SARP combines elements from the Selective Routing Problem (SRP) literature to address representativeness concerns, and Vehicle Routing Problem (VRP) literature involving stochastic customers to tackle the possible unavailability of key informants. In the following sections, we review the relevant literature in these domains. First, Section 2.1 explains similar studies in humanitarian assessment problems. Then, the literature regarding the routing problems is presented in Section 2.2.

2.1 Humanitarian Assessment Problems

Following a disaster, needs assessment studies are conducted to provide initial insights into the urgent needs during the initial response, and support development activities in later stages [9]. These assessment surveys can be conducted by analyzing satellite or aerial images [10], online questionnaires [7], or by selecting and visiting affected sites to conduct in-person surveys. The most relevant class of studies is in logistical planning for selective assessment routing problems. We now review the related literature.

Specifically, the context of SARP is closest to the Selective Assessment Routing Problem formally defined in [6]. This closely related study integrates location selection with routing decisions to support purposive sampling by selecting sites based on balanced coverage for specific characteristics. This approach ensures that the sampled sites

adequately represent diverse community attributes. Building on this foundation, [11] introduce a robust optimization model addressing uncertainties in travel times within post-disaster networks, which helps manage logistical variability in disaster scenarios. [12] proposed a model designed to achieve predetermined coverage targets, focusing on rapid and effective coverage of community groups. [13] define a multi-cover routing problem similar to [12] and propose a decomposition-based tabu search algorithm to achieve quality solutions within a short amount of time. [14] develop heuristic algorithms for site selection and routing, tested through simulations to address uncertainties such as site accessibility and the presence of target communities. This study provides practical insights into managing various uncertainties encountered in post-disaster settings. Moreover, the terminology they use is very similar to our study. Specifically, the authors discuss whether a community group will be available upon a visit to a site, and consider clusters of similar villages. They use the assumed availability and cluster information to select villages in their developed heuristic. Although using clusters to define similar sites is common in both studies, we introduce a novel main-backup strategy to tackle the possible unavailability of key informants (sites), where they propose updating the route accordingly or skipping that site. To balance total assessment time and coverage of different groups [15], develop a bi-objective optimization model and define “leximin selective assessment problem” to select sites and routes in a post-disaster immediate needs assessment setting. [16] introduce a collaborative truck-and-drone system that can support or replace traditional fieldwork, particularly in scenarios where direct interviews are impractical or technology can supplement data collection.

Other notable contributions in this field include [17], who explored routing strategies for assessment teams tasked with visiting affected communities. Their model prioritizes efficient routing but does not address the complexities of site selection, assuming that all targeted sites must be visited. [18] introduce a bi-objective optimization model for

assessing damage to population centers and road segments using a combination of aerial and ground vehicles. This model is particularly useful for detailed damage evaluations but may be less applicable to initial rapid needs assessments. Recent studies focus on exact solution algorithms [19] for the selective assessment routing problem and heuristic methods (constructive heuristic and tabu search algorithm) for ensuring coverage and accuracy for assessment studies [20].

Table 2.1 presents a summary of related work in humanitarian assessment literature to SARP. Note that all studies are focused on rapid needs-assessment studies that require shorter solution times and do not necessarily have time for large-scale data collection and analysis. This additional time in our long-term field needs assessment study enables us to define clusters using a large number of attributes for representativeness, instead of focusing on a few key attributes.

Therefore, we contribute to the humanitarian assessment literature by proposing a large-scale data analysis framework to group villages according to similarities, forming clusters, and developing a new problem definition for site selection and routing based on backup site assignments to main sites under uncertain key informant presence, while ensuring sample representativeness concerning defined clusters.

2.2 Vehicle Routing Problems Literature

A closely related vehicle routing research domain is the SRP, which is solved to determine efficient routes to visit a selected subset from possible locations while ensuring some coverage over another set [21]. Our coverage definition, based on clustering and stratified sampling, is similar to the Generalized Vehicle Routing Problem (GVRP). To the best of our knowledge, the GVRP was introduced by [22], who observed that the undirected version of this problem can be transformed into a Capacitated Vehicle Routing Problem (CVRP), while [23] provided a transformation for the directed case. [24]

Table 2.1: *Humanitarian Assessment Literature Comparison.*

Ref.	Site Sel.	Routing	Coverage	Uncertainty
[17]	-	+	Visit all sites	-
[6]	+	+	Community groups with distinct characteristics- Maximize the minimum coverage ratio across all characteristics	-
[11]	+	+	Community groups with distinct characteristics- Maximize the minimum excess over targets	Travel times: uncertainty set
[13]	+	+	Community groups with distinct characteristics - Meet target coverages	-
[12]	+	+	Community groups with distinct characteristics - Meet target coverages	-
[14]	+	+	Clusters - Target number for each cluster	Travel times, assessment times, site availability: updating the routes
[15]	+	+	Community groups with distinct characteristics - Maximize coverage ratios vector	-
[19]	+	+	Community groups with distinct characteristics	-
[20]	+	+	Population sampling	Links the assessment study with uncertainty in relief distribution
Our study	+	+	Clusters - Target number for each cluster	Key informant (site) availability: main-backup assignments

proved that the reverse transformation is also valid, i.e., the CVRP can be transformed into a GVRP. These authors also provided several applications of the GVRP, including the Travelling Salesperson Problem (TSP) with profits, several vehicle routing extensions, arc routing, and the design of tandem configurations for automated guided vehicles. Other GVRP applications include the design of bus services [25], arc routing with turn penalties [26], military aircraft mission planning [27], and solid waste collection [28]. Several variants of the GVRP have also been studied, including problems with backhauling [29], with multiple pickup and delivery requests [30], with stochastic demands [31], with overlapping clusters [32], and with time windows [33]. For a recent survey, see [34].

Although using the pre-defined exhaustive and exclusive clusters to differentiate locations is similar to the GVRP literature, this definition of “coverage” for each cluster is modified to visit a proportion of each cluster in our problem. Specifically, we define target coverage levels proportional to each cluster’s size, which is similar to [12].

The uncertainty in site availability defined for the SARP stems from a fundamental variant of the vehicle routing problems, namely Vehicle Routing Problem with Stochastic Customers (VRPSC) or Vehicle Routing Problem with Stochastic Requests (VRPSR). In VRPSC/VRPSR, the set of all sites is known a priori; however, whether they are available or present for service (in SARP context, assessment surveys) is learned upon visit. In our problem, the presence of the key informant at a visited location (customer presence) is uncertain. To tackle this uncertainty, aligned with our collaboration with the field team, we assign backup locations to be visited when the main site is unavailable. The VRPSC belongs to the generic class of stochastic vehicle routing. These are VRPs that contain some sources of uncertainty, such as customer demands, travel times, service times, and the presence or absence of customers. In this study, we are only interested in the latter case. Moreover, we restrict our attention to contexts in which a set of potential customers is known a priori, but some of them may be absent when the solution is implemented. We

refer the reader to surveys [35, 36] for more details on stochastic VRPs.

Relatively few researchers have investigated the VRPSC. [37] introduced the problem for the single-vehicle case, i.e., defined the Travelling Salesperson Problem with Stochastic Customers (TSPSC) and developed several models and properties. The solution paradigm followed is called “a priori optimization”, a concept that was later formalized by [38]. In this framework, an a priori solution is first computed assuming all customers are present. The set of absent customers is then revealed, and the planned route is followed, except that the absent customers are skipped. VRPSC is especially relevant in same-day delivery problems [39] or stochastic service requests from customers [40]. Recently, [41] solved the VRPSC by column generation. [42] developed a tabu search heuristic for a version of the VRPSC with loading constraints. [43] solved the Vehicle Routing Problem with Stochastic Demand and Customers (VRPSDC) exactly by means of the integer L-shaped algorithm, and [44] solved it heuristically by tabu search.

A second solution paradigm for stochastic routing problems is reoptimization [45]. In the context of the VRPSC, vehicle routes are first computed, but no information on the set of absent customers is provided at the start of routing. The presence or absence of the customers only becomes known upon reaching them. At this point, a reoptimization procedure may be applied. The simplest one is to follow the route as planned, but one may also re-optimize the remaining portion of the route starting from the location of an absent customer, or reschedule a visit to that customer at a later time, as is sometimes done in attended home delivery [46]. Examples of route re-optimization in the context of stochastic routing can be found in [47, 48, 49].

We employ an a priori routing approach based on defined scenarios for site availability. Specifically, the routing decisions for each combination of available main villages are decided a priori. Teams are then expected to follow these plans accordingly without any re-optimization strategy.

2.3 Contribution of Our Study

The SARP contributes to the literature by uniquely combining selective assessment routing problems with vehicle routing problems with uncertain customers under the humanitarian assessment context. Clusters based on site attributes are formed upon extensive data analysis, and coverage targets are defined to ensure representativeness. Backup sites are assigned to main sites from the same clusters to maintain representativeness in case of key informant unavailability. Specifically, our research makes the following key contributions to the humanitarian logistics and vehicle routing literature:

- We introduce a novel stochastic assessment routing problem variant that integrates sample selection with logistical planning under uncertainty in key informant presence (uncertain customers-uncertainty in site availability), incorporating unique features such as cluster-based sampling and backup location selection.
- We develop efficient heuristic algorithms capable of handling realistic problem instances, addressing the computational challenges posed by the problem's complexity.
- We demonstrate the practical applicability of our approach through a case study based on a large real-world dataset from the 2023 Türkiye earthquakes, providing extensive analysis and insights into the potential benefits of integrating data analytics with field assessment logistics planning to support long-term recovery.

3. PROBLEM STATEMENT

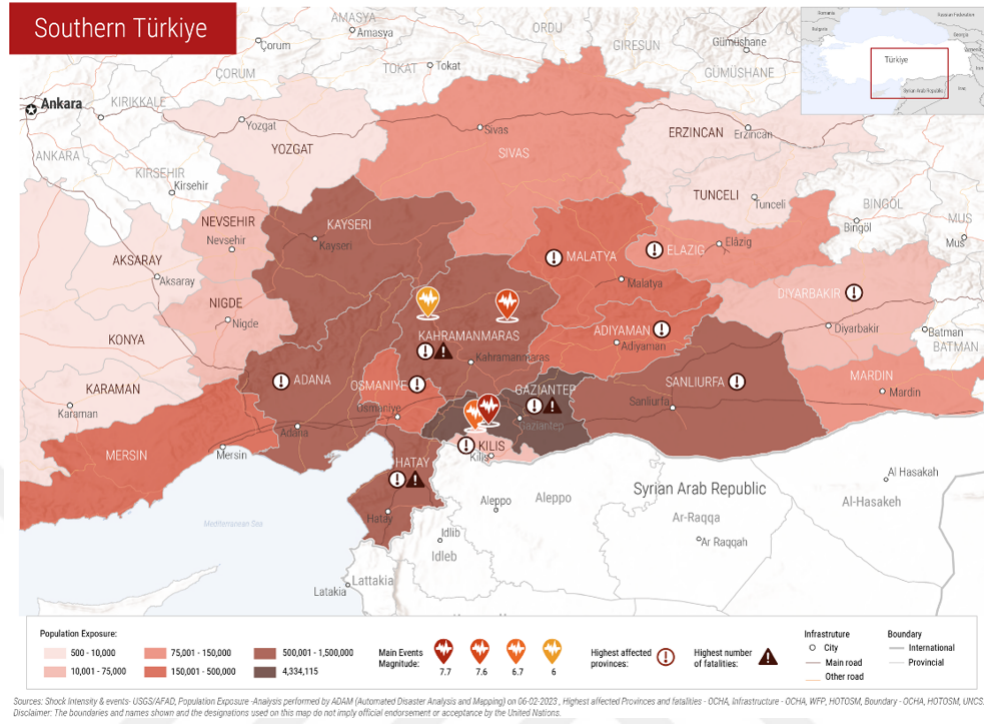
This chapter first provides a detailed system description of the field study's setting in Section 3.1. Section 3.2 elaborates on the decision support needs of the field study and presents major challenges. Finally, Section 3.3 provides a formal problem definition for the SARP and explains the mathematical complexity of the defined problem.

3.1 System Background and Decision Needs

On February 6th, 2023, southern Türkiye was hit by two major earthquakes with magnitudes 7.7 Mw and 7.6 Mw on the Richter Scale at 4:17 am and 1:24 pm [50]. The same area was hit by another earthquake of magnitude 6.4 Mw on February 20, 2023 [50]. In the following three months of the February 6th earthquakes, this region was hit with more than 33,500 aftershocks, 600 having a magnitude between 4.0-7.0 Mw [50]. Out of 81 provinces (administrative level 1 subdivisions) of Türkiye, 11 were (namely Adana, Adıyaman, Diyarbakır, Elazığ, Gaziantep, Hatay, Kahramanmaraş, Kilis, Malatya, Osmaniye, and Şanlıurfa) declared as the earthquake zone following the large-scale destruction.

The affected area and major earthquake epicenters are shown in Figure 3.1. 16.2% of Türkiye's population was located in the earthquake area, which covered $108,745 km^2$ (13.9% of the country's total area) [51, 52]. It is estimated that 9.1 million people were directly affected by the earthquakes in this hardest hit 11 provinces, and 2.7 million people were displaced [4]. With more than 50,000 lost lives and 115,000 injuries, the February 2023 earthquakes had higher casualties and therefore were more devastating than previous earthquakes in Türkiye [50]. Following the immediate Level 4 alarm calling for international help, more than 8,300 search and rescue personnel and other forms of help were rapidly mobilized globally [3].

Figure 3.1: Türkiye's affected area and major epicenters (Reprinted from [3]).



Immediately after the disaster, rapid needs assessment studies are conducted to inform decision-makers in the first response phases regarding the initial situation and urgent needs [50], provide insights into the affected population and damage caused by the earthquakes [10], or to understand the needs of the affected population in specific areas such as agricultural production [7]. While these rapid needs assessment studies provide invaluable information about the immediate needs and are crucial for first response, conducting long-term needs and impact assessment studies is also necessary to understand the situation in the long term. Moreover, analyzing factors contributing to resilience to earthquakes is essential to support preparedness activities in the future.

Two years after the 2023 earthquakes, field teams visit affected rural villages and conduct surveys to understand and analyze factors contributing to the resilience of rural populations and agricultural activities to the disaster. Specifically, to understand the

resilience of agricultural activities in rural villages to earthquakes, field teams conduct assessment surveys with key informants (heads of villages, teachers, etc.) of these villages in Türkiye's 11 affected provinces. These formal assessment surveys are planned to be combined with informal "chats" and personal observations that will provide further insight into the impact of earthquakes and the factors contributing to the resilience of local rural populations and agricultural activities. Therefore, the field team values in-person visits as more information could be gathered that way.

The team decided that a sample of 2000 villages must be selected for on-site visits from more than 6000 villages in the area. However, there are two major challenges: (1) ensuring sample representativeness while selecting the 2000 villages and (2) accounting for the possible unavailability of the key informant at a village, which is only discovered upon arrival while constructing efficient routes. Specifically, in case of an unavailable key informant, the survey cannot be completed at the visited village, and another "similar enough" village has to be visited to reach the required number of surveys based on the requests of the field team and ensuring representativeness.

We support the field study with efficient logistical plans and representative sampling decisions to minimize the necessary budget and human resources to conduct on-site surveys, under uncertain key informant presence. Specifically, we use a clustering algorithm to group villages based on similarities, using public data regarding villages and districts of the villages from secondary sources. We then construct efficient routes under key informant uncertainty by selecting a sample of main villages (and their corresponding backups) that ensures representativeness across these constructed clusters. To integrate the key informant uncertainty in our methodology, we define "scenarios" based on the main village key informant availabilities for each daily route, solve the corresponding routing problems, and calculate the expected cost of each daily route. Moreover, to ensure practical solutions for the field team in a decision-tree visual format, when a main village

is unavailable, its backup is visited immediately.

3.2 Problem Aspects

In the field study, a sample of 2000 villages must be selected from and distributed among the affected 11 provinces. Moreover, routing plans are conducted on a provincial level. Specifically, after distributing the total sample size of 2000 to 11 provinces, villages in each province must be clustered, and a sampling and routing problem must be solved separately. We assume that the following information is known.

- **Total sample size** is given by the project team.
- **Sample size in each province** is assigned based on the number of villages, population, and distance of villages to the earthquake centers (see Section 4.1.2).
- **Number of daily surveys** is constant for each team and day and is given by the project team directly.
- **Survey duration** is deterministic and known. It is based on a pilot study conducted in the area with a few selected villages to ensure survey quality.
- **Village clusters** are known, and villages from the same clusters can be interchanged without seeing any differences in the village profile or covered attributes.
- **Cluster coverage** is ensured by including each cluster in each province's sample proportional to the cluster size.
- **Probability of key informants to be absent** is known for all villages. This value is based on the project lead's expert opinion.
- **Backup presence** is guaranteed. Specifically, when the survey cannot be completed at a main village, the survey is always completed upon visiting the assigned backup village.

- **Distances between the villages** are deterministic and known.

While preparing sampling and routing plans, there are two major decision aspects: (1) site selection: selecting a representative sample of all villages, and (2) routing decisions: constructing efficient routes under key informant uncertainty. This section explains the decision-making aspects and challenges faced in the field project.

3.2.1 Sampling

Coverage over different site characteristics (attributes) is crucial for ensuring representativeness, i.e., capturing information from a sample that can be generalized to the population. In post-disaster assessment studies, various attributes, such as population, livelihood, and agro-ecological zones, geographical attributes (altitudes, location of the sites, e.g., coastal, mountainous, etc.), the impact of the disaster, and known vulnerabilities can be considered to differentiate the sites [8]. A recent example set of related attributes can be found from the agriculture sector needs assessment conducted immediately following the 2023 earthquakes [7]. In addition to information on specific needs for crop production (e.g., seeds, fertilizers, etc.), livestock (e.g., vaccines, feeds, etc.), and fisheries and aquaculture (e.g., feeds, machinery, etc.), they include information regarding the province population, and the number of registered farmers, producers, and food enterprises in the province. Therefore, attributes related to demographic information about the affected population, geographic information on the affected area, and pre-existing information on the area of interest are relevant to use while selecting a representative sample.

Given a large set of such relevant attributes, covering them in a balanced way while maintaining efficient routes is a challenge. We address this sampling concern to ensure the representativeness of the final set of selected sites.

3.2.2 *Logistical Planning Under Uncertainty*

Field teams will visit villages to conduct on-site assessment surveys with key informants. However, one major challenge in logistical planning is the uncertainty of key informants' presence in the villages, which is only discovered upon visit. Specifically, the field team could discover that they cannot complete the survey at an intended village due to various factors such as the key informant's absence as a result of them attending a funeral, using time off, visiting another village, etc., or the key informant's refusal to contribute to the survey. In such cases, to ensure the representativeness of the selected sample and complete the planned number of surveys, another village with the same (or very similar) set of attributes, namely a village from the same cluster, needs to be visited for conducting the survey.

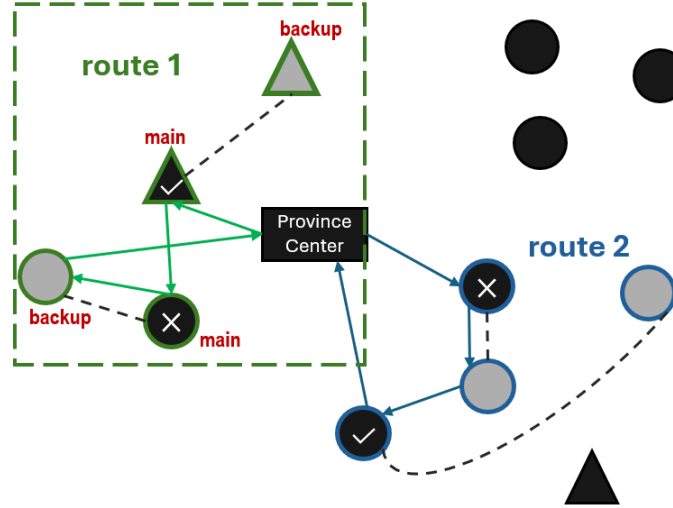
To tackle this uncertainty, we propose a dynamic "main-backup" village pairing policy, which is also requested by the field study team. Specifically, in a daily route of main villages to visit, each main village is assigned a corresponding backup village from the same cluster. The backup is only visited if the survey cannot be completed in the main village. This approach ensures completing a decided number of surveys, while visiting different village profiles, even under the uncertainty of the key informant's presence.

Figure 3.2 illustrates an example network consisting of 15 nodes across two distinct clusters represented by spheres and triangles. Four surveys are planned to be completed; when distributed proportionally to their sizes, one and three main villages (s) from the triangle and sphere clusters must be visited, respectively. The network demonstrates two survey routes: routes 1 and 2, indicated by green and blue outlines. Focusing on route 1, the routing logic operates as follows: when a main village (the triangle node) is available, its corresponding backup node (connected via dashed nodes) is not visited, and the team proceeds directly to the subsequent main node. Conversely, when a main location is unavailable, as observed in the sphere node in route 1, the team must visit its backup

immediately after the main location. Similarly, these steps can be observed in route 2. Notice that both routes start and end at the province center.

For each route, different scenarios based on each main village's availability can be constructed. Specifically, for a route with 2 main villages, there are $2^2 = 4$ scenarios. Moreover, the field team assumes a probability p^i for the key informant at a village i to be unavailable. Therefore, the probability of each scenario can be calculated using this probability. This probability is decided by the field team based on their past experience. However, we conduct a sensitivity analysis on the p^i in the following sections.

Figure 3.2: *Example Network for SARP*



It is important to note that the assignment and logistical planning of main-backup pairs is a challenging task, especially given the large size of the affected area. Moreover, routing with various scenarios of main village availability presents another challenge.

Given these aspects, we support the described field study by determining a representative sample for on-site visits and proposing effective routing plans addressing the uncertainty in key informant presence with backup village assignments. The formal definition of SARP is given in Section 3.3.

3.3 Problem Formulation

We define the SARP to minimize the total expected distance of field teams while ensuring sampling constraints, i.e., covering each cluster. Coverage is defined as representing each cluster in the final sample with a proportional number of villages to the cluster size. Specifically, the SARP is defined over a set V of villages, each belonging to one cluster (profile) from the set K of clusters. To ensure sample representativeness, N^k many main locations must be selected from each cluster $k \in K$, where binary parameter a_v^k takes the value 1 if village $v \in V$ belongs to cluster $k \in K$ and 0 otherwise. Distances d_{uv} between villages $u, v \in V$ are deterministic and known in advance. Each selected route must have n main (and therefore backup) villages. This is the number of surveys teams plan to complete in a day. Note that to complete n surveys, $2n$ villages could be visited, considering the backups. To account for the uncertainty in key informant presence, we define the set of possible scenarios S , where probability p^s is associated with each scenario $s \in S$. The presence or absence of the key informant at each village $v \in V$ in each scenario $s \in S$ is given by the binary parameter β_v^s . Note that $\beta_v^s \sim \text{Bernoulli}(p_v)$ given p_v is the probability of completing the survey at village $v \in V$, and for n mains there are 2^n many scenarios for each route. Finally, each main village must have exactly one backup, and each backup must be assigned to exactly one main village. Problem assumptions based on the collaborative meetings and discussions with the project team are given below.

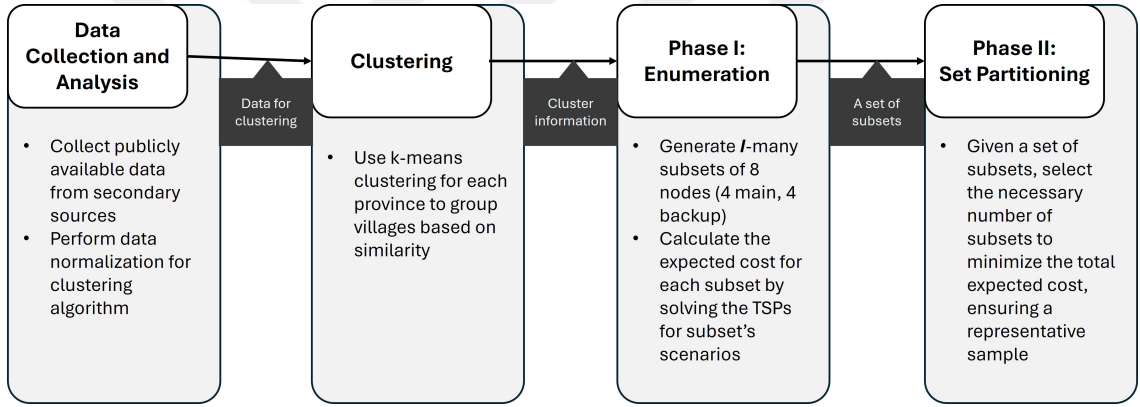
The total sample size and its distribution to provinces and clusters are known. Each team will leave the province center to visit at least n at most $2n$ villages to complete n surveys daily and return to the center. When a backup village b is visited, regardless of the β_b^s value, the survey is assumed to be completed. The mathematical model formulation of the SARP is proposed in Appendix A.

This problem cannot be solved with commercial solvers due to the high complexity and the large size of the routing problem. Specifically, when there are $|V|$ villages in a province, and R routes are to be designed, there are $\binom{|V|}{R \times 2n}$ many ways of selecting the villages for the routes. Moreover, there are at most $\binom{2n}{n}$ many main selections among the selected routes of size $2n$. For each route and main-backup configuration, 2^n scenarios must be constructed. The corresponding traveling salesperson problem (TSP) must be solved with visitation constraints since the backups must be visited after the corresponding mains. Moreover, the number of nodes from each cluster in each route must be even to ensure main-backup assignments, which adds even more complexity to the model. Therefore, this problem is intractable due to the high complexity of integrated sampling and routing decisions. Instead, we develop a matheuristic and use a set partitioning approach to select an efficient subset from a set of enumerated routes.

4. SOLUTION METHOD

Since the mathematical model cannot be solved due to its high complexity and the large size of the problem, as explained in Chapter 3, we develop a matheuristic that consists of two phases: Phase I: an enumeration process to construct feasible routes, followed by Phase II: a set partitioning model. This matheuristic is performed after data analysis and clustering studies; therefore, cluster information for each village is assumed to be known in the matheuristic. An overview of the proposed solution methodology's steps can be seen in Figure 4.1.

Figure 4.1: *Solution Methodology Steps*



Since complete enumeration, i.e., enumerating all possible subsets from each village, is not possible due to the problem size, we use different strategies to generate feasible subsets. Each subset consists of $n = 4$ main and $n = 4$ backup villages. Since each backup must have the same cluster as its assigned main, the number of villages from each cluster must be an even number in each subset. Also, for each subset, there are $2^4 = 16$ scenarios that correspond to the availability of each main village. In other words, each subset represents all possible routes of a team for a day, including the main and backup villages for any combination of available main villages. Therefore, a feasible subset must have 8 villages and an even number of villages from the included clusters.

Section 4.1 explains the clustering approach employed in detail. The proposed matheuristic with different enumeration options is presented in Section 4.2.

4.1 Clustering

It is important to select a sample that represents the whole area to generalize the final findings of the field assessment. Therefore, we use relevant attributes of villages and form clusters to group similar villages together. We assume that teams can visit one village instead of the other without information loss if they belong to the same cluster. In other words, visiting a target number of villages from a cluster will ensure “coverage” over the cluster. After covering each cluster, the information gathered from the sample can be generalized to the whole cluster.

Some of the crucial decisions in sampling are (a) relevant attributes to consider, (b) sample size, and (c) selecting the appropriate approach from different sampling techniques [53]. These decisions are explained in detail in this section.

4.1.1 *Relevant Attributes*

Villages have specific demographic (population, population change) and geographic (altitude, distance to the city center) attributes. Moreover, villages from the same district share a common socio-economic development index, land price, land cover properties, agricultural production (both used area and specific products), and total damaged area in the district. The final list of attributes is decided in coordination with the field project lead. Note that these attributes coincide with relevant characteristics for purposive sampling in the literature (see [8, 7]). It is important to note that each one of the many relevant attributes was deemed important to the field study; therefore, we could not eliminate any to reduce the dimensionality. Moreover, since some attributes belong to districts and pre-decided groups of villages, it is crucial to balance these shared attributes with village-specific ones to show the similarity among villages from different districts and the variety

of villages within the same district.

4.1.2 Sample Size

The total sample size of 2,000 villages across the 11 provinces was determined by the field team at the beginning of the project. In our study, we decide on the distribution of this sample among provinces. It is based on a simple proportional allocation considering three key factors: (1) the number of villages in each province, (2) each province's total rural population, and (3) the average distance from villages to the closest epicenter among four earthquake centers decided by the field experts. The allocation is directly proportional to (1) and (2), and is inversely proportional to (3) to prioritize provinces that are larger and are closer to epicenters. Table 4.1 shows the data used in this process.

Since there are three factors to consider, different weights are given according to their importance by the field team's opinion. Specifically, a weighted sum approach was employed with the respective weights of 15%, 15%, and 70% provided by the field team. Moreover, to ensure each main village has a backup village, the sample size of each province must be less than 50% of the number of villages it has; otherwise, the total number of villages would not be enough to ensure main-backup pairs. Moreover, in some provinces, there were small and scattered clusters. In those cases, even 50% of the province's village count could be difficult to solve. Therefore, we imposed an upper limit of 80% of the total village count for each province's sample size. The iterative algorithm is followed to allocate the total sample size of 2,000 among provinces using the three factors while ensuring the 80% upper limit is given in Algorithm 1.

4.1.3 Sampling Approach

Some of the most used sampling techniques are simple random sampling where each member of the population has an equal chance of being selected for the sample, stratified sampling that first divides the population into smaller groups (called strata) that

Table 4.1: *Province Data for Sample Sizes.*

Province	# Villages	Total Village Population	Avg. Distance to Kahramanmaraş Epicenter	Avg. Distance to Gaziantep Epicenter	Avg. Distance to Malatya Epicenter	Avg. Distance to Hatay Epicenter	Avg. Distance to Closest Epicenter	Sample Size
Kilis	139	33,376	142	43	186	122	43	56
Osmaniye	189	128,606	126	80	206	132	80	76
Hatay	450	738,091	207	118	271	40	39	180
Gaziantep	516	246,659	121	48	151	163	48	208
Adıyaman	525	189,705	109	129	64	273	61	212
Kahramanmaraş	538	311,423	57	81	126	207	43	216
Adana	588	289,961	169	143	255	145	119	224
Elazığ	593	147,176	209	267	122	412	122	208
Malatya	597	220,104	108	184	54	330	53	240
Diyarbakır	922	492,489	275	302	187	435	187	188
Şanlıurfa	1,239	775,592	192	172	146	287	140	260

Algorithm 1 Sample Size Allocation Algorithm

Require: Total sample size S , province village counts v_i , rural population p_i , and average distance of villages to the closest epicenter d_i for each province i

Ensure: Feasible sample size allocation s_i for each province i

- 1: Calculate the inverse of the average distance to the closest epicenters for each province i : $D_i = \frac{1}{d_i}$
- 2: Sort provinces in decreasing order concerning D_i
- 3: **repeat**
- 4: Calculate sample size for each province i using weighted allocation:

$$s_i = S \times \frac{w_1 \cdot v_i + w_2 \cdot p_i + w_3 \cdot D_i}{\sum_{j=1}^n (w_1 \cdot v_j + w_2 \cdot p_j + w_3 \cdot D_j)}$$

where $w_1 = 0.15$, $w_2 = 0.15$, $w_3 = 0.70$

- 5: Check feasibility constraint for each province i :
 - 6: **if** $s_i > 0.8 \times \text{villages}_i$ **then**
 - 7: Set $s_i = 0.8 \times v_i$
 - 8: Update remaining sample size: $S = S - s_i$
 - 9: Remove province i from further allocation
 - 10: **end if**
 - 11: **until** All provinces satisfy the 80% constraint
 - 12: **return** Sample sizes s_i for each province i
-

are similar in some attributes and then selects a sample from each strata independently, or cluster and systematic sampling [54]. Using selected attributes for the villages, we implement a weighted k-means algorithm to group similar villages, forming clusters, for each province. Although there are many clustering approaches, such as hierarchical clustering, k-modes, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), the widely used k-means algorithm is easy to implement and interpret. It can also be used with a large number of numeric and binary attributes and corresponding weights. In order to implement the k-means algorithm, we must first ensure the same scale for each attribute, then decide on the weights for attributes, and select the optimal cluster number (k value). These steps are now explained in detail.

Scaling the Attributes: Since k-means is based on the distances to cluster centers, attributes with a larger scale (e.g., distance to the city center in meters) would dominate any impact a binary variable with the maximum value of 1 would have. Hence, it is necessary to scale all attributes to the same range so that they have the same chance to contribute to the final clustering results. If we could use some random distributions to represent certain attributes, we could leverage distribution-specific properties for normalization. Since many of the attributes did not fit into any theoretical distribution (normal, uniform, exponential) with 5% significance using the Kolmogorov-Smirnov test, we considered two highly implemented scaling approaches: z-score normalization and MinMax scaling. The z-score normalization is sensitive to the variances and the original domain of the data; it does not guarantee a fixed range for the normalized data. Therefore, to ensure the same scale, we used MinMax scaling (see Equation 4.1) for numeric attributes using the sci-kit library in Python [55]. Each data point (x) is scaled in the interval $[0,1]$ by subtracting the minimum value ($X.min$) from x , and dividing the range ($X.max - X.min$) of the data array. There was no need to scale the binary attributes that already take either

0 or 1 as their value.

$$x_{scaled} = \frac{x - X.min}{X.max - X.min} \quad (4.1)$$

Attribute Weights: The field expert provided us with a classification of high, moderate, and low importance for all attributes. Note that the final list of attributes was already decided, and we could not omit the least important features. We decide on the weights in the k-means algorithm to reflect different levels of importance. A major challenge is that many of the attributes are shared, i.e., they are at the district level. Since the districts are already pre-determined geographic clusters of villages, we must adjust the weights for said attributes to see the effect of a few, but village-specific attributes. To limit the contribution of shared attributes, we decide on a total weight w_p^* for each province p from a set W of possible values that will be distributed among district-level attributes. The highest value w can take is equal to the number of village-specific attributes we have. This is because the value “1” as weight is given to each village-specific attribute to emphasize its importance. We test on different w^* values to find a value that allows district-level attributes to be included in the clustering results without overshadowing village-specific attributes. We define a simple “diversity” measure to count how many different clusters are included in each district. We normalize this measure with the village number in each district to compare districts of different sizes. Specifically, the diversity score $diversity_{p,d}^w$ is defined as the sum of different clusters in district d of province p , for a total weight of $w \in W$ over the total number of villages n_p in city p as represented in Equation 4.2. For each province, we select the w_p^* that results in the maximum diversity score for the highest number of districts for each province p .

$$diversity_{p,d}^w = \frac{\sum (\text{number of unique clusters in } d)}{n_p} \quad (4.2)$$

$$w_d^* = \arg \max_{w \in W} \sum_d \text{diversity}_{p,d}^w \quad (4.3)$$

Optimal Number of Clusters: The optimal number of clusters (k) should be decided by the user in the k-means algorithm. We run the weighted k-means algorithm with different k values and use the elbow method to find the optimal k . We chose the optimal k as the number of clusters where adding one additional cluster no longer decreases the inertia significantly. It is a simple and inexpensive method to decide on the optimal number of clusters [56] based on the sum of squared distances of data points to their assigned cluster centers [57]. The pseudo-code for our clustering approach is presented in Algorithm 2. This code takes the initial attribute values and village lists as input, and returns the clustering information for each province.

Algorithm 2 Clustering Pseudo-Code

Require: Village list and relevant attribute values for each province

Ensure: Cluster information for each village for each province

- 1: Use Equation 4.1 to scale each numeric attribute
 - 2: **for** w in W **do**
 - 3: Adjust district-attribute values so that their sum is w
 - 4: Solve weighted k-means, use the elbow method to decide on the optimal k
 - 5: Calculate diversity score using Equation 4.2
 - 6: **end for**
 - 7: Decide on w^* using Equation 4.3
-

4.2 Set Partitioning

We describe the two-phase matheuristic designed for solving the SARP. The first phase enumerates a manageable set of subsets of $2n$ villages, and the second phase selects the necessary number of subsets that minimize the expected cost using a set partitioning model.

4.2.1 Phase I: Enumerating Subsets of Villages

As mentioned earlier, the first phase involves identifying subsets of $2n$ sites, where n is the number of surveys to be completed in a day. Each subset actually represents the villages that should be (mains) and can be (backups) visited by a team in a day. Therefore, each subset consists of 2^4 scenarios, corresponding to the routing of the included villages (all mains and necessary backups).

Given the large number of villages, the immense number of possible site combinations makes complete enumeration computationally intractable. Therefore, we consider a smaller set of subsets of $2n$ locations. In particular, in our approach, we identify a number of feasible subsets of $2n$ sites that satisfy cluster duplication, i.e., the number of locations belonging to each cluster in a subset must be even to ensure cluster concordance between main and backup sites.

To minimize the logistical costs, selecting villages situated within a given maximum distance to the province center (Δ) is desirable. However, to ensure the feasibility of the solution in terms of sampling representation or cluster duplication, it may be necessary to allow the enumeration of subsets with villages further from the depot. To facilitate the selection of subsets in the second phase, we introduce a penalty for including villages further from the depot in the subset. Therefore, when there are enough villages for its cluster within the maximum distance to the depot Δ , those villages are more likely to be selected. Otherwise, it will be necessary to exceed this limit to satisfy sampling representation. Hence, to identify these subsets of $2n$ sites, we propose two different strategies.

The first strategy, referred to as *Main-Backup*, is based on the requirements set by the leading expert overseeing the field teams. To facilitate the execution of the routes, which are generally carried out by volunteers without logistic planning knowledge, the expert desires routes that are easy to explain and follow a consistent pattern. This pattern involves always selecting a main site first and proceeding to its designated backup site.

This approach ensures that volunteers can easily understand and navigate the routes, even though it might not be the most adequate strategy in terms of logistical cost. This is especially true when the probability of main site availability is high, as the most likely routes will either visit only the main sites or incorporate a few backup sites as needed.

Algorithm 3 details the pseudocode for this enumeration strategy, where a given number of subsets is created from each village. Specifically, a total of $I * |V|$ subsets are generated. Aiming to reduce the cost of the routes, villages belonging to a subset ω are selected based on closeness. However, to differentiate subsets from the same village, we incorporate a random aspect when selecting the main villages while keeping a limitation on the distance (δ) between villages. Note that villages can only be selected once, as either main or backup sites. Hence, if there is no suitable backup for a given main site, another village will be selected until a pair of main and backup sites is found. After generating the complete subset ω of $2n$ sites, the expected cost l_ω is computed using equation 4.4. To this end, we first solve $\binom{2n}{n}$ TSP problems, one for each scenario $s \in \hat{S}$ based on the combinatorial availability of the main villages. Then, we sum the multiplications of the cost of each scenario $s \in \hat{S}$ with its probability p_s to achieve the expected cost of selecting subset ω . The different TSPs are solved using the nearest-neighborhood algorithm. We then add a penalty cost P_ω to the expected TSP cost to penalize those subsets with villages that are further away than the maximum distance Δ to the depot. As are the cases of subsets created from a location further than Δ from the depot or those that, by adding other site locations, move further away from the depot.

$$l_\omega = \sum_{s \in \hat{S}} (NearestNeighborhood_\omega^s * p_s) + P_\omega \quad (4.4)$$

The second strategy we propose is referred to as *AllMain-AllBackup*. To maintain easy patterns for explanation to the field teams, this strategy also follows a consistent pattern where first all main villages are selected, followed by the selection of all the backup

Algorithm 3 Enumeration *Main-Backup* ($M - B$)

Require: V set of villages; d_{uv} distance between villages $u, v \in V$; p^v probability associated with availability of village $v \in V$; α_v cluster profile of village $v \in V$; δ maximum distance; I number of iterations

Ensure: Ω set of subsets of $2n$ sites; l_ω expected cost associated to subset $\omega \in \Omega$

```
1: for all villages  $v \in V$  do
2:   iter = 1
3:   repeat
4:     Generate subset  $\omega_{iter}$ 
5:      $v := M_1$ 
6:      $B_1 :=$  closest village to  $M_1$  belonging to the same cluster ( $\alpha_{B_1} = \alpha_{M_1}$ )
7:     if iter = 1 then
8:       for all  $i = 2, \dots, n$  do
9:          $M_i :=$  closest village to  $M_{i-1}$ 
10:         $B_i :=$  closest village to  $M_i$  belonging to the same cluster ( $\alpha_{B_i} = \alpha_{M_i}$ )
11:      end for
12:    else
13:      for all  $i = 2, \dots, n$  do
14:         $M_i :=$  random within a maximum distance  $\delta$  to  $M_{i-1}$ 
15:         $B_i :=$  closest village to  $M_i$  belonging to the same cluster ( $\alpha_{B_i} = \alpha_{M_i}$ )
16:      end for
17:    end if
18:    Compute expected cost  $l_{\omega_{iter}}$  using Equation 4.4
19:    iter++
20:  until iter =  $I$ 
21: end for
```

sites. This strategy aims to obtain subsets with lower logistic costs when the probability of main site availability is high. Note that since backups (which are always chosen based on closeness) are selected at the end, this allows for the possibility of selection of the closest sites as main villages. Algorithm 4 details the pseudocode for this second strategy, which follows a similar structure to the previous strategy, but as mentioned, changes the order in which backup sites are selected.

4.2.2 Phase II: Set partitioning model

The second phase of our approach involves solving a set partitioning model designed to select the $|R|$ subsets that minimize the total expected cost, taking into account

the availability of the villages. Below, we present the notation and the formulation for the set partitioning model.

Sets

Ω : set of subsets of $2n$ villages, $\omega \in \Omega$

Parameters

l_ω : expected cost of subsets $\omega \in \Omega$

a_ω^k : number of main villages from cluster $k \in K$ in subset $\omega \in \Omega$

m_v^ω : binary parameter indicating whether or not village $v \in V$ belongs to subset $\omega \in \Omega$ as main

b_v^ω : binary parameter indicating whether or not village $v \in V$ belongs to subset $\omega \in \Omega$ as backup

Decision variables

x_ω : binary variable indicating if subset $\omega \in \Omega$ is selected for the assessment

Set Partitioning Formulation

$$\text{minimize} \quad \sum_{\omega \in \Omega} l_\omega x_\omega \quad (4.5)$$

subject to

$$\sum_{\omega \in \Omega} (m_v^\omega + b_v^\omega) x_\omega \leq 1 \quad v \in V \quad (4.6)$$

$$\sum_{\omega \in \Omega} a_\omega^k x_\omega = N^k \quad k \in K \quad (4.7)$$

$$\sum_{\omega \in \Omega} x_\omega = |R| \quad (4.8)$$

$$x_\omega \in \{0, 1\} \quad \omega \in \Omega. \quad (4.9)$$

The objective function 4.5 minimizes the expected cost associated with the selected subsets. Constraints 4.6 guarantee that each village is chosen at most once. Additionally, these constraints prevent a village from being selected as both a main and a backup village. Constraints 4.7 ensure representative sampling by specifying the total number of main villages to be selected from each cluster. Constraints 4.8 guarantee that exactly $|R|$ subsets are chosen according to the size of the sampling. Finally, constraints 4.9 define the domains of the variables.

We run the set partitioning model with the subsets generated using Enumeration algorithms Main-Backup (Algorithm 3) and AllMain-AllBackup (Algorithm 4). Chapter 5 presents the case study, computational results, and parameter analysis.

Algorithm 4 Enumeration *AllMain-AllBackup* ($AM - AB$)

Require: V set of villages; d_{uv} distance between villages $u, v \in V$; p^v probability associated with availability of village $v \in V$; α_v cluster profile of village $v \in V$; δ maximum distance; I number of iterations

Ensure: Ω set of subsets of $2n$ sites; l_ω expected cost associated to subset $\omega \in \Omega$

```
1: for all villages  $v \in V$  do
2:   iter = 1
3:   repeat
4:     Generate subset  $\omega_{iter}$ 
5:     if iter = 1 then
6:        $v := M_1$ 
7:       for all  $i = 2, \dots, n$  do
8:          $M_i :=$  closest village to  $M_{i-1}$ 
9:       end for
10:      for all  $i = 1, \dots, n$  do
11:         $B_i :=$  closest village to  $M_i$  belonging to the same cluster ( $\alpha_{B_i} = \alpha_{M_i}$ )
12:      end for
13:    else
14:       $v := M_1$ 
15:      for all  $i = 2, \dots, n$  do
16:         $M_i :=$  random within a maximum distance  $\delta$  to  $M_{i-1}$ 
17:      end for
18:      for all  $i = 1, \dots, n$  do
19:         $B_i :=$  closest village to  $M_i$  belonging to the same cluster ( $\alpha_{B_i} = \alpha_{M_i}$ )
20:      end for
21:    end if
22:    Compute expected cost  $l_{\omega_{iter}}$  using Equation 4.4
23:    iter++
24:  until iter =  $I$ 
25: end for
```

5. DATA ANALYSIS AND COMPUTATIONAL RESULTS

This chapter presents the data collection and computational analysis for the proposed matheuristic approach. Section 5.1 details the data collection and pre-processing procedures for the 11 provinces in Türkiye, including clustering results. Section 5.2 follows with an extensive computational analysis evaluating the performance of the proposed solution methodology, specifically the two proposed enumeration algorithms for 11 provinces of Türkiye.

5.1 Data Collection and Processing

The affected villages are scattered across 11 provinces with distinct geographic properties over a large area. Figure 1.1 shows the locations of villages on the Türkiye map. Each province is divided into districts containing urban neighborhoods and/or rural villages. We are only interested in rural villages for this study. These villages are situated in both plain and mountainous areas, with varying altitudes. Moreover, some villages are positioned relatively close to the provincial and/or district centers, while others are in harder-to-reach areas. The team highlights that both ease of access and altitude could point to differences in infrastructure, agricultural production, and vulnerability to disasters; therefore, it is important to visit villages from different areas to see the effect of geographic location on resilience to disasters.

In addition to geographic attributes, villages have distinct demographics, including population size and population growth. The number of people and the speed of population increase/decrease could be important in comparing villages of different sizes in terms of resilience and vulnerability to earthquakes, according to the field project team. Moreover, the socio-economic development index and the price of a unit of land in districts are deemed crucial attributes to consider in the study.

The damaged area caused by the earthquakes, the area allocated to agriculture, and

the agricultural products grown in each district are other important attributes to consider when defining coverage for different village attributes in this study.

The final list of relevant attributes to the resilience assessment problem is decided in coordination with the field expert according to available data and is consistent with the current literature [7, 8]. A dataset consisting of such shared district-level and village-specific attributes is collected.

The list of rural sites is taken from [51] and is filtered according to the directions from the field expert. Current and past population values are also sourced from [51]. Additionally, [58] is used to fill in any missing values resulting from changes in village names or assigned districts. Moreover, distance and elevation information are gathered for each village using the Open Route Services Application Programming Interface (API) of [59]. Some missing altitude values are completed manually using [60]. District-level properties include the soil types and damaged area taken from the earthquake from [10], and socio-economic development indices from [61]. Moreover, the amount of land used for agriculture is taken from [62], and the production of specific agricultural products is taken from [51]. Lastly, the land price information is taken from [63]. The sources for each attribute are presented in Table 5.1. Table 5.1 also indicates in column **V/D** and **N/B** whether each attribute is village-specific or is at the district level, and whether it is a numeric or binary attribute, respectively. Lastly, column **I** shows the importance class (H-high, M-middle, L-low) for each attribute. There were 11 numeric land cover attributes and 10 binary attributes that show whether or not some agricultural products are grown. To have a concise table, the **I** column is marked with *, and the details for those attributes are explained here. For the land cover, the area covered by irrigated cropland, road, water body, and greenhouse attributes had high importance. Moderate importance was given to the area covered by soil, grassland, orchard, tree cover, and built-up area. Finally, the area covered with shrubland was given low importance by the field team. Moreover, olives,

legumes, fruits, vegetables, tubers, medical aromatic plants, oils, textile plants, tobacco, all this moderate importance; and finally, grains with low importance.

Table 5.1: *Sources for Attributes Used in the Field Study (V/D: Village/District, N/B: Numeric/Binary, I: Importance Score, * explained in the text).*

Attribute	Explanation	V/D	N/B	I	Source
Population	2023 and 2013 populations of each village	V	N	H	[51]
Distance to province center	Distance in kilometers to their province center	V	N	H	[59]
Altitude	Altitude of each location in meters	V	N	H	[59]
School	Binary attribute for school availability	V	B	H	Field expert
Socio-economic score	Development score for each district	D	N	H	[61]
High damage instensity	Area in km ² with high damage	D	B	H	[10]
Moderate damage instensity	Area in km ² with moderate damage	D	B	L	[10]
Low damage instensity	Area in km ² with high low	D	B	M	[10]
Land cover (10)	Land cover structure and distribution	D	N	*	[10]
Cropland area	Total agricultural area in decares	D	N	H	[62]
Land price	Price of land per m ² for each district	D	N	H	[63]
Agricultural products (10)	Binary attribute for product cultivation	D	B	*	[51]

5.1.1 Clustering

The clustering approach explained in 4.1 is run for each province's dataset. Table 5.2 presents the province size (# villages and districts) along with the selected total weight of districts' attributes decided using Equation 4.3, and the final number of clusters for each province. Notice that for some provinces, the number of clusters is higher than the district number (e.g., Kilis, Osmaniye), which indicates that some villages from the same district have varying characteristics according to our set of attributes and weights used in the k-means algorithm. Similarly, for other provinces such as Hatay or Gaziantep, villages from different districts may belong to the same clusters, indicating they are similar to each other compared to other villages. These clusters are used to ensure representativeness, i.e., coverage of different characteristics in the set partitioning model. Specifically, the number of villages to visit from each cluster is proportional to the cluster's size in each province. Figure 5.1 presents the distribution of clusters (with different colors) in each province.

Table 5.2: *Clustering Input and Outputs for Each Province.*

Province	# Villages	# Districts	Dist. Attr. Weight	# Clusters
Kilis	139	4	1.0	8
Osmaniye	189	7	1.0	8
Hatay	450	15	1.5	5
Gaziantep	516	9	1.0	7
Adıyaman	525	9	1.0	8
Kahramanmaraş	538	11	1.0	5
Adana	588	15	1.0	4
Elazığ	593	11	1.0	6
Malatya	597	14	1.5	7
Diyarbakır	922	17	1.0	8
Şanlıurfa	1,239	14	1.5	6

5.2 Computational Analysis of the Matheuristic

This chapter presents the computational analysis of the proposed matheuristic approach applied to 11 provinces in Türkiye. All computations were performed on a Windows 10 System with an Intel(R) Core(TM) i9-10980XE Processor (3.00GHz), 128 GB RAM, using IBM (International Business Machines Corporation) ILOG CPLEX Optimization Studio version 22.1.1.

The village set V , inter-village distances d_{uv} for all $u, v \in V$, cluster assignments α_v for each village $v \in V$, and sample sizes for all provinces and clusters are finalized during the data collection process (see Section 5.1) and are treated as deterministic parameters. Moreover, each route is prepared to represent a team's daily operational plan. The goal is to complete 4 resilience assessment surveys per day; therefore, each enumerated subset includes $n = 4$ main villages and 4 backup villages, totaling $2n = 8$ villages. As a result, there are $2^n = 2^4 = 16$ possible scenarios for each subset, corresponding to different combinations of key informant presence at the main villages.

The rest of the computational results section is organized as follows. Baseline parameter values and results are presented in Section 5.2.1. Sensitivity analysis on key informant presence probability is discussed in Section 5.2.2. Finally, Section 5.2.3 presents

the visual decision support tools in decision tree formats that were provided to field teams.

5.2.1 Base Results

The probability of successfully completing a survey upon visiting each village i is established as $p^i = 90\%$ based on field expert recommendations. The distance threshold parameters for penalty activation and selecting the next random main village are set to $\Delta = 100$ km and $\delta = 50$ km, respectively. In both enumeration algorithms (Algorithms 3 and 4) I subsets are generated from each village as input to the optimization model. Baseline computations are presented for $I \in \{15, 25, 50, 75, 100\}$ subsets per village to evaluate the performance of the set partitioning model under varying input sizes. Specifically, for both enumeration options, the value of generating more subsets in terms of total expected cost and computational time is evaluated. Expected costs and solution time in seconds to the set partitioning model for 11 provinces using these base parameter values are given in Table 5.3.

Table 5.3: *Expected Costs and Total Solution Times (mins) with Base Parameters.*

Province	Province Info					Expected Cost & Solution Time									
	# of Villages	# of Districts	# of Clusters	Sample Size	Routes to Select	Main-Backup					AllMain-AllBackup				
						I=15	I=25	I=50	I=75	I=100	I=15	I=25	I=50	I=75	I=100
Kilis	139	4	8	55	14	632 0.12	625 0.17	604 0.31	602 0.45	602 0.62	602 0.10	595 0.16	595 0.31	593 0.46	579 0.60
Osmaniye	189	7	8	75	19	1258 0.30	1223 0.30	1159 0.55	1149 0.81	1141 1.09	1157 0.20	1106 0.27	1074 0.50	1058 0.73	1042 0.97
Hatay	450	15	5	180	45	3595 1.23	3223 1.34	3045 2.80	2943 3.49	2907 4.57	3109 0.89	3053 2.06	2916 3.46	2851 4.15	2833 5.67
Gaziantep	516	9	7	206	52	7365 1.23	7342 1.84	7278 2.83	7264 4.22	7243 5.64	6571 1.43	6524 1.99	6450 3.54	6445 5.36	6066 7.14
Adıyaman	525	9	8	210	53	3765 1.25	3700 1.51	3667 2.74	3629 4.16	3629 5.53	3723 1.37	3685 1.84	3610 2.85	3592 5.04	3564 6.26
Kahramanmaraş	538	11	5	215	54	9933 1.45	9455 3.19	9249 3.23	9215 4.79	9191 6.39	8012 1.52	7857 2.16	7269 6.07	7227 5.89	7176 7.92
Adana	588	15	4	221	55	31612 1.47	31599 2.22	31572 4.36	31544 6.33	31517 8.43	30217 1.41	30199 2.27	30160 4.28	30140 6.49	30129 8.38
Elazığ	593	11	6	207	52	19106 1.26	19058 1.93	19006 3.81	18587 5.68	18561 7.64	17679 1.26	17603 1.93	17144 3.76	17009 5.70	16963 7.27
Malatya	597	14	7	238	60	25775 3.13	24665 2.32	23921 5.03	23796 7.33	23736 10.61	23162 4.35	23026 7.68	22265 6.76	22115 6.73	21565 9.03
Diyarbakır	922	17	8	186	47	7490 2.60	7485 4.23	7472 8.18	7470 12.17	7465 16.28	7370 2.66	7370 4.27	7363 8.21	7363 12.16	7363 16.33
Şanlıurfa	1239	14	6	258	65	12743 5.49	12668 8.97	12577 17.42	12518 26.05	12503 34.30	12183 5.54	12082 9.17	11961 17.87	11841 25.87	11839 34.63

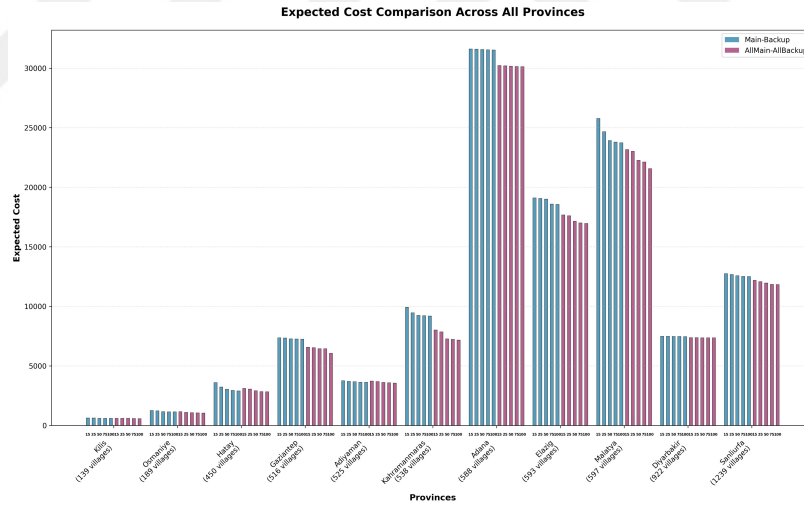
In terms of expected costs, AllMain-AllBackup (Algorithm 4) outperforms Main-Backup (Algorithm 3) in most cases (Table 5.4). These trends are visible in Figure 5.2.

However, this improvement in expected costs is usually at the cost of slightly longer, yet still reasonable, solution times (Table 5.5).

Table 5.4: *Decrease in Expected Cost from Main-Backup to AllMain-AllBackup (%)*.

Province	I=15	I=25	I=50	I=75	I=100
Kilis	4.90%	4.90%	1.40%	1.50%	3.80%
Osmaniye	8.00%	9.60%	7.30%	7.90%	8.70%
Hatay	13.50%	5.30%	4.20%	3.10%	2.50%
Gaziantep	10.80%	11.10%	11.40%	11.30%	16.30%
Adıyaman	1.10%	0.40%	1.50%	1.00%	1.80%
Kahramanmaraş	19.30%	16.90%	21.40%	21.60%	21.90%
Adana	4.40%	4.40%	4.50%	4.50%	4.40%
Elazığ	7.50%	7.60%	9.80%	8.50%	8.60%
Malatya	10.10%	6.60%	6.90%	7.10%	9.10%
Diyarbakır	1.60%	1.50%	1.50%	1.40%	1.40%
Şanlıurfa	4.40%	4.60%	4.90%	5.40%	5.30%

Figure 5.2: *Expected Costs with Base Parameters*



The enumeration times are highly correlated with the number of villages in the province, as expected. There is also a strong correlation between the enumeration times and the total province area. However, neither the village count nor the province area seems to be correlated with the set partitioning solution times (Table 5.6).

Set partitioning times (Table 5.7) are mostly affected by the density of villages,

the geographic distribution of clusters, and the province center. Notice that even though Şanlıurfa is the largest province and has the largest sample size, its computational times are less than those of Hatay, Kahramanmaraş, and Malatya when other parameters are constant. In Şanlıurfa, there are mostly at most two clusters that are close, whereas in the other three, we can see that multiple clusters are often overlapping (Figure 5.1). Therefore, we can conclude that the set partitioning times increase as the subset number per village is increased only for provinces that have clusters more spread out. However, a linear change cannot be observed in Malatya, Kahramanmaraş, Hatay, and Adıyaman, i.e., Provinces with multiple overlapping clusters take a longer time to solve due to increased computational complexity (Figure 5.1).

Table 5.5: *Total Solution Times (min).*

I	Main-Backup			AllMain-AllBackup		
	min.	avg.	max.	min.	avg.	max.
15	0.12	1.78	5.49	0.10	1.88	5.54
25	0.17	2.55	8.97	0.16	3.07	9.17
50	0.31	4.66	17.42	0.31	5.24	17.87
75	0.45	6.86	26.05	0.46	7.14	25.87
100	0.62	9.19	34.30	0.60	9.47	34.63

Table 5.6: *Correlation of Enumeration and Set Partitioning Times (min) with Province Properties.*

I	R with Enumeration Time				R with Set Partitioning Time			
	# Villages		Province Area		# Villages		Province Area	
	Main-Backup	AllMain-AllBackup	Main-Backup	AllMain-AllBackup	Main-Backup	AllMain-AllBackup	Main-Backup	AllMain-AllBackup
15	0.95	0.95	0.81	0.81	-0.01	0.04	0.1	0.16
25	0.94	0.94	0.8	0.8	0.03	0.04	0.32	0.14
50	0.94	0.94	0.8	0.79	0.22	0.05	0.27	0.29
75	0.94	0.94	0.79	0.8	0.25	0.16	0.34	0.24
100	0.94	0.94	0.8	0.79	0.21	0.16	0.31	0.27

Table 5.7: Detailed Enumeration and Set Partitioning Times (min).

Province	Main-Backup					AllMain-AllBackup				
	I=15	I=25	I=50	I=75	I=100	I=15	I=25	I=50	I=75	I=100
	Enumeration Time (min)									
Kilis	0.10	0.15	0.29	0.42	0.57	0.10	0.15	0.29	0.42	0.57
Osmaniye	0.15	0.23	0.44	0.65	0.87	0.15	0.23	0.44	0.65	0.86
Hatay	0.73	1.12	2.21	3.24	4.33	0.72	1.13	2.18	3.26	4.31
Gaziantep	0.88	1.37	2.66	3.94	5.16	0.88	1.36	2.65	3.93	5.19
Adıyaman	0.84	1.33	2.53	3.78	5.01	0.83	1.30	2.55	3.73	5.00
Kahramanmaraş	0.96	1.52	2.95	4.38	5.76	1.00	1.51	2.95	4.34	5.69
Adana	1.43	2.16	4.21	6.12	8.12	1.36	2.16	4.16	6.20	8.12
Elazığ	1.17	1.84	3.58	5.30	7.07	1.17	1.84	3.56	5.23	6.88
Malatya	1.10	1.68	3.24	4.70	6.31	1.08	1.69	3.18	4.88	6.13
Diyarbakır	2.55	4.13	7.98	11.78	15.74	2.60	4.17	7.99	11.76	15.82
Şanlıurfa	5.35	8.75	16.90	25.32	33.32	5.33	8.84	17.29	25.20	33.63
	Set Partitioning Time (min)									
Kilis	0.02	0.02	0.02	0.03	0.05	0.01	0.01	0.02	0.03	0.03
Osmaniye	0.15	0.07	0.11	0.16	0.22	0.05	0.04	0.06	0.08	0.10
Hatay	0.51	0.22	0.59	0.25	0.23	0.17	0.93	1.28	0.89	1.36
Gaziantep	0.34	0.48	0.17	0.28	0.48	0.55	0.62	0.89	1.43	1.94
Adıyaman	0.41	0.18	0.22	0.38	0.52	0.55	0.53	0.30	1.32	1.26
Kahramanmaraş	0.49	1.66	0.28	0.41	0.64	0.52	0.64	3.13	1.55	2.23
Adana	0.04	0.06	0.15	0.22	0.31	0.06	0.12	0.11	0.29	0.26
Elazığ	0.09	0.09	0.23	0.38	0.57	0.09	0.09	0.19	0.47	0.39
Malatya	2.03	0.64	1.78	2.62	4.30	3.26	5.99	3.57	1.85	2.90
Diyarbakır	0.05	0.10	0.21	0.38	0.54	0.06	0.10	0.23	0.41	0.51
Şanlıurfa	0.14	0.22	0.52	0.73	0.98	0.20	0.33	0.58	0.67	1.00

5.2.2 Parameter Analysis

The probability of completing a survey (p^i) is decided as 90% by the field expert, believing that the chance of any village's key informant is high. If this probability is lower, however, the expected cost of the route for the same subset changes since p_s changes in Equation 4.4. To analyze the performance of the proposed matheuristic under different p^i values, we generate scenarios of 50 subsets from $p^i = 0.8$ to $p^i = 0.1$ with increments of 0.1 and compare the expected costs and solution times with the base results from 5.2.1 with $p^i = 0.9$ and between the two enumeration algorithms. Using 50 subsets is selected for this analysis because the average decrease in expected cost stops after 50 subsets for almost all provinces, while the solution times start to increase rapidly.

Table 5.8 shows the expected costs for both enumeration algorithms for each p^i value and province. Specifically, the improvement part calculates the percentage decrease in expected cost when AllMain-AllBackup is used instead of Main-Backup. Therefore, positive values indicate a lower expected cost for the AllMain-AllBackup algorithm, indicating it performs better than Main-Backup for that p^i value and province. In half of the provinces (Kilis, Hatay, Gaziantep, Adıyaman, Diyarbakır, and Şanlıurfa), Algorithm 3 Main-Backup starts to perform better than Algorithm 4 as the probability of key informant presence decreases (red cells with negative improvement when AllMain-AllBackup is used).

However, in the remaining provinces (Osmaniye, Kahramanmaraş, Adana, Elazığ, Malatya), the AllMain-AllBackup algorithm always performs better than Main-Backup with the given p^i values. The advantage of the Main-Backup algorithm is ensuring the closest village from the same cluster as the backup to a selected main before selecting another main village, which could be useful when the probability of visiting that backup is high, i.e., the probability p^i of key informant presence is low.

When the villages from the same cluster are spread across the province, ensuring

the closest backup assignment before selecting another main seems to be more valuable, as it can be seen in Kilis, Hatay, Gaziantep, Adıyaman, Diyarbakır, and Şanlıurfa in Figure 5.1. In contrast, when all villages that belong to one cluster are closer to each other, or cluster distributions are spatially dense, assigning the backup after selecting all mains does not add extra distance to travel for the other provinces (Osmaniye, Kahramanmaraş, Adana, Elazığ, Malatya (Figure 5.1)) since all villages from the same cluster are fairly close to each other.

Although the Main-Backup algorithm performs better for some provinces in low p^i values, notice that the average expected cost for all provinces is lower for all p^i values when the AllMain-AllBackup algorithm is used (Table 5.9). Moreover, Table 5.9 shows that the time required to solve the set partitioning model increases when AllMain-AllBackup enumeration is used, especially as p^i values decrease. Specifically, in 4 instances using AllMain-AllBackup with $p^i \in \{0.1, 0.2, 0.3\}$, the set partitioning model stopped at the pre-defined 30-minute solution time and reported an optimality gap between 0.01-0.002.

This observation is consistent with the base results, where AllMain-AllBackup required longer solution times to provide better expected costs on average. However, as the probability of key informant presence decreases, the improvement in the expected cost of using AllMain-AllBackup at the expense of longer solution times becomes less preferable. Therefore, using AllMain-AllBackup when the probability of key informant presence is high is recommended. As this probability decreases, switching to the Main-Backup enumeration algorithm would be beneficial for the computational time and the expected cost for some provinces, even though there is a slight increase in the overall average expected cost.

Table 5.8: Expected Cost Difference (%) from Main-Backup to AllMain-AllBackup Enumeration Algorithms.

Main-Backup Algorithm									
Province	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Kilis	892	861	829	797	763	724	685	644	604
Osmaniye	1633	1588	1527	1467	1406	1345	1284	1222	1159
Hatay	3937	3842	3739	3632	3523	3408	3292	3169	3045
Gaziantep	11961	11396	10825	10248	9669	9085	8486	7882	7278
Adiyaman	5003	4859	4710	4546	4382	4215	4038	3853	3667
Kahramanmaras	12793	12378	11950	11519	11090	10633	10174	9713	9249
Adana	55175	52257	49329	46391	43451	40497	37530	34556	31572
Elazig	31674	30115	28556	26985	25412	23828	22234	20631	19006
Malatya	44582	42043	39497	36935	34371	31800	29210	26595	23921
Diyarbakir	12441	11859	11269	10654	10036	9408	8773	8128	7472
Sanliurfa	19284	18467	17646	16818	15984	15146	14300	13446	12577
AllMain-AllBackup Algorithm									
Province	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Kilis	932	905	878	846	810	767	715	656	595
Osmaniye	1626	1578	1523	1459	1396	1321	1243	1166	1074
Hatay	4081	3967	3850	3716	3570	3420	3265	3106	2916
Gaziantep	12084	11445	10770	10082	9384	8671	7940	7199	6450
Adiyaman	5218	5087	4894	4725	4540	4333	4110	3870	3610
Kahramanmaras	12388	11852	11250	10628	9998	9352	8681	8000	7269
Adana	54783	51782	48759	45720	42652	39569	36465	33338	30160
Elazig	31013	29364	27657	25943	24204	22456	20691	18921	17144
Malatya	43181	40677	38071	35463	32848	30229	27599	24952	22265
Diyarbakir	12644	12051	11441	10809	10158	9485	8794	8092	7363
Sanliurfa	19497	18592	17686	16761	15824	14877	13918	12946	11961
Improvement from Main-Backup to AllMain-AllBackup									
Province	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Kilis	-4.5%	-5.1%	-5.9%	-6.1%	-6.2%	-5.9%	-4.4%	-1.9%	1.5%
Osmaniye	0.4%	0.6%	0.3%	0.5%	0.7%	1.8%	3.2%	4.6%	7.3%
Hatay	-3.7%	-3.3%	-3.0%	-2.3%	-1.3%	-0.4%	0.8%	2.0%	4.2%
Gaziantep	-1.0%	-0.4%	0.5%	1.6%	2.9%	4.6%	6.4%	8.7%	11.4%
Adiyaman	-4.3%	-4.7%	-3.9%	-3.9%	-3.6%	-2.8%	-1.8%	-0.4%	1.6%
Kahramanmaras	3.2%	4.2%	5.9%	7.7%	9.8%	12.0%	14.7%	17.6%	21.4%
Adana	0.7%	0.9%	1.2%	1.4%	1.8%	2.3%	2.8%	3.5%	4.5%
Elazig	2.1%	2.5%	3.1%	3.9%	4.8%	5.8%	6.9%	8.3%	9.8%
Malatya	3.1%	3.2%	3.6%	4.0%	4.4%	4.9%	5.5%	6.2%	6.9%
Diyarbakir	-1.6%	-1.6%	-1.5%	-1.5%	-1.2%	-0.8%	-0.2%	0.4%	1.5%
Sanliurfa	-1.1%	-0.7%	-0.2%	0.3%	1.0%	1.8%	2.7%	3.7%	4.9%

Table 5.9: Average Expected Cost and Set Partitioning Solution Times (min) Required for Different p^i Values (* some instances are solved with a gap due to 30 minute time limit).

Average Expected Costs									
Enumeration Algorithm	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Main-Backup	18125	17242	16352	15454	14553	13644	12727.82	11804	10868
AllMain-AllBackup	17950	17027	16071	15105	14126	13135	12129	11113	10073
Improvement	0.97%	1.25%	1.72%	2.26%	2.94%	3.74%	4.70%	5.85%	7.31%
Average Solution Times (min)									
Enumeration Algorithm	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Main-Backup	3.71	2.16	1.60	0.92	0.97	0.71	0.55	0.50	0.39
AllMain-AllBackup	9.12*	7.56*	5.13*	2.47	1.77	1.11	0.91	0.87	0.94

5.2.3 Result Visualizations

A major challenge was to present the solutions in an intuitive way that would be easy to follow for the field teams. We used a decision-tree structure to show the next site to go, combining all scenarios as branches of a tree. An example route on the map and corresponding decision tree for one of the provinces (Adana) is given in Figure 5.3. The first node shows the starting site, and then at each step, teams follow the arcs marked with 1 for completed surveys and marked with 0 for failed surveys at the main sites. From each backup site, the teams should go to the next main site directly. In addition to decision trees, we provided spreadsheets that show the main and backup villages as requested by the field expert.

Figure 5.1: *Clustering Results*

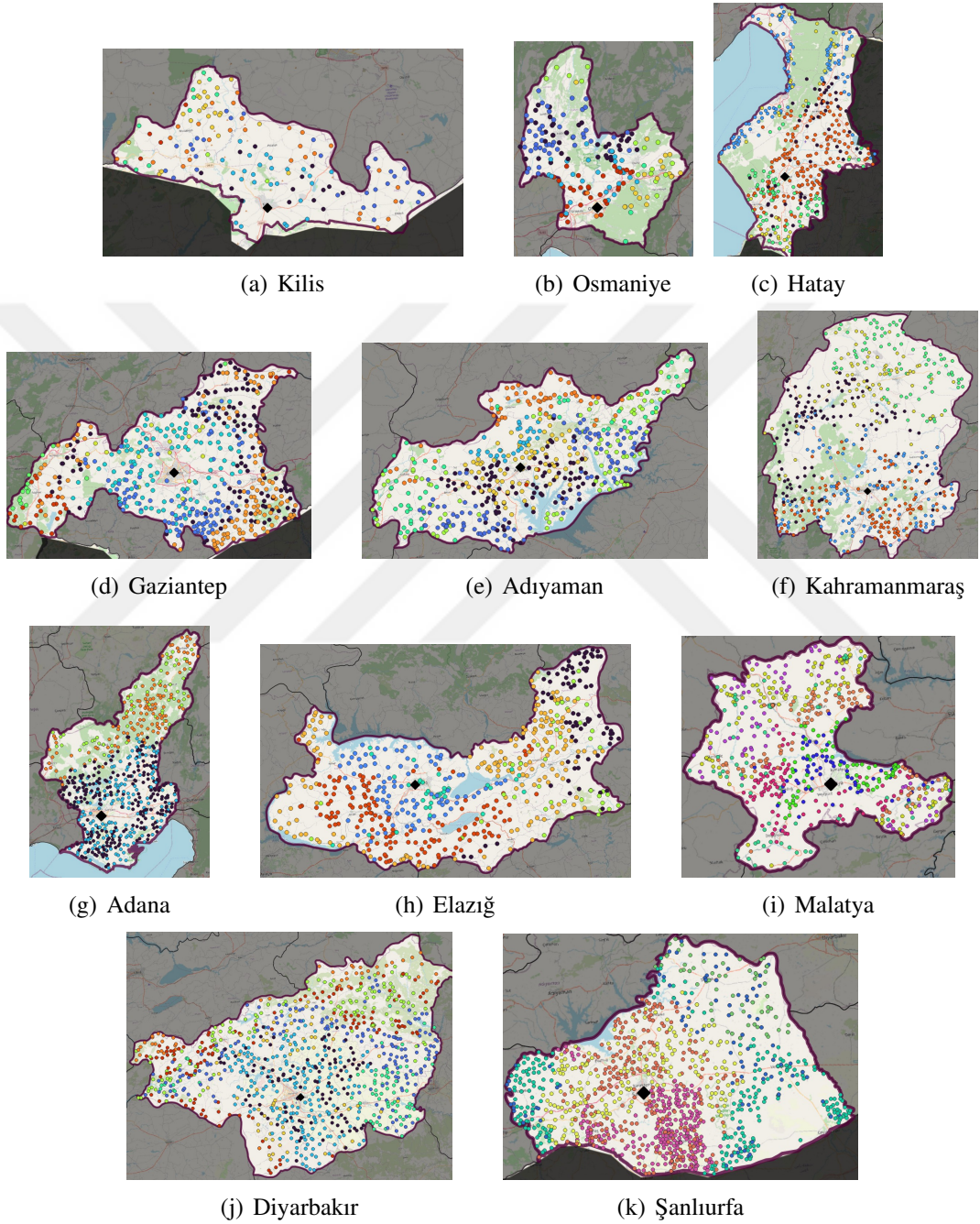
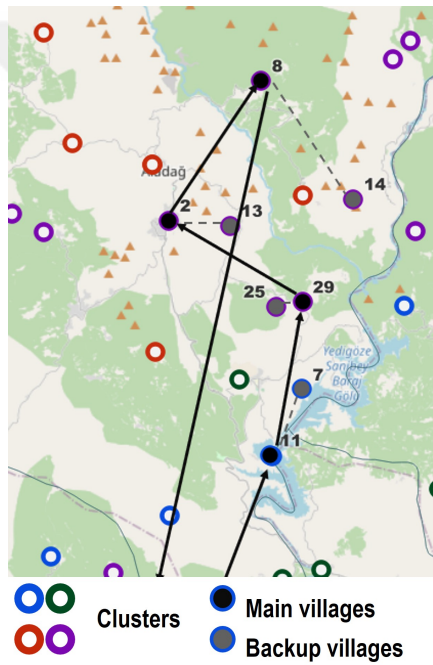
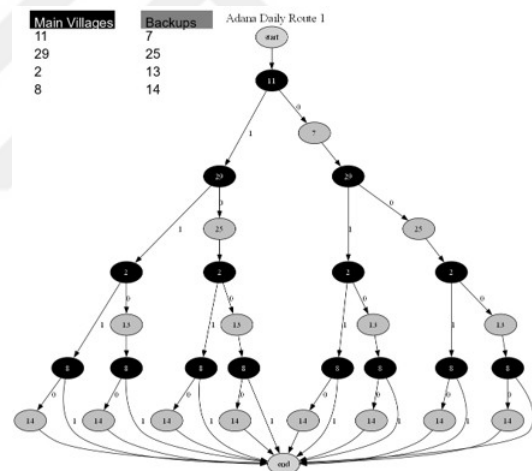


Figure 5.3: *Example Solution Representation for Adana*



(a) Example Route on Map



(b) Example Decision Tree

6. CONCLUSION

We develop a matheuristic that can provide routing decisions under uncertain key informant presence, while ensuring sample representativeness. Specifically, we propose two enumeration algorithms to generate subsets, calculate the expected cost of each scenario in the subset, and use a set partitioning model for selecting efficient routes. The results are analyzed on a large dataset of 11 provinces affected by the earthquakes. A sensitivity analysis for the probability of key informant presence is performed, and insights are drawn to compare two enumeration algorithms in terms of expected cost and solution times. Specifically, using Main-Backup enumeration as the probability of key informant presence decreases is found to be more advantageous for expected cost and/or solution time metrics, depending on the province.

As future research, a multi-depot variant where field teams can leave and return to locations other than the province center could be explored. Moreover, developing benchmarks to evaluate the proposed solution's performance is another future research direction. One notable limitation of our study is that the key informants at the backup villages are always assumed to be available. This might not be the case in the field application. Therefore, our solution methodology could be extended or modified to include a set of villages from the same cluster that are close to each selected backup village, for the teams to consider if the visited backup was not available. Another interesting variant could incorporate a decision on the number of backup villages, instead of a 1-1 assignment with mains, especially when the probability of the key informant being present varies.

Finally, the field implementation showed us several possible improvement points. Upon our discussions and meetings with the field team, we make several observations. Firstly, the funding problems affected the planned field assessment study as the team had to do some surveys online instead of in-person visits, and needed to get help from volunteers with different academic/non-academic backgrounds to conduct surveys. With

this change, the decision trees we provided were not used in planning in-person visits. After discussions with the field team, we realized that a simple phone application that integrates an online map with the decision trees and selected village lists could facilitate the adaptation of our proposed plans. However, without such a tool, it was not possible in the field to follow the decision trees, thus, routing decisions. However, teams mostly followed the sampling decisions presented in a spreadsheet format. In some cases, close villages that were not selected for the sample were visited to collect more information in a shorter amount of time.



REFERENCES

- [1] OpenStreetMap Contributors, “Planet Dump Retrieved from <https://planet.osm.org>.” <https://www.openstreetmap.org>, 2017. Accessed: 2024-09.
- [2] QGIS Development Team, “QGIS Geographic Information System.” <https://www.qgis.org>, 2024. Open Source Geospatial Foundation Project.
- [3] United Nations Office for the Coordination of Humanitarian Affairs, “Türkiye Earthquake Flash Appeal 2023,” Report, United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), 2023. Accessed: 2025-04.
- [4] Food and Agriculture Organization of the United Nations, “Türkiye: Response Overview, December 2023,” Report, Food and Agriculture Organization (FAO), 2024. Accessed: 2025-05.
- [5] Food and Agriculture Organization of the United Nations, “Türkiye: Rebuilding rural communities through cooperative movement – project profile,” Report, Food and Agriculture Organization (FAO), 2024. Accessed: 2025-04.
- [6] B. Balcik, “Site selection and vehicle routing for post-disaster rapid needs assessment,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 101, pp. 30–58, 2017.
- [7] Food and Agriculture Organization of the United Nations, “Türkiye: Agricultural sector needs assessment for earthquake-impacted provinces,” Report, Food and Agriculture Organization (FAO), Ankara, 2023. Accessed: 2025-05.
- [8] The Assessment Capacities Project, “Technical brief: purposive sampling and site selection in phase 2,” Report, ACAPS, 2011. Accessed: 2025-04.
- [9] United Nations Office for the Coordination of Humanitarian Affairs, “We coordinate.” <https://www.unocha.org/we-coordinate>, 2024.
- [10] Food and Agriculture Organization of the United Nations, “Rapid geospatial assessment after the earthquake in türkiye in 2023: Impacts on infrastructure and farming community during the period february–march 2023,” Report, Food and Agriculture Organization of the United Nations (FAO), 2023. Accessed: 2025-06.
- [11] B. Balcik and İ. Yanıkoğlu, “A robust optimization approach for humanitarian needs assessment planning under travel time uncertainty,” *European Journal of Operational Research*, vol. 282, no. 1, pp. 40–57, 2020.
- [12] D. Pamukcu and B. Balcik, “A multi-cover routing problem for planning rapid needs assessment under different information-sharing settings,” *OR Spectrum*, vol. 42, no. 1, pp. 1–42, 2020.

- [13] Y. Mıhçıoğlu and E. Albey, “A decomposition based metaheuristic approach for solving rapid needs assessment routing problem,” *Computers and Industrial Engineering*, vol. 159, 2021.
- [14] M. Hakimifar, B. Balcik, C. Fikar, V. Hemmelmayr, and T. Wakolbinger, “Evaluation of field visit planning heuristics during rapid needs assessment in an uncertain post-disaster environment,” *Annals of Operations Research*, pp. 1–42, 2022.
- [15] M. Hakimifar, V. Hemmelmayr, and F. Tricoire, “A lexicographic maximin approach to the selective assessment routing problem,” *OR Spectrum*, vol. 45, no. 1, pp. 205–249, 2023.
- [16] G. Zhang, N. Zhu, S. Ma, and J. Xia, “Humanitarian relief network assessment using collaborative truck-and-drone system,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 152, p. 102417, 2021.
- [17] M. Huang, K. R. Smilowitz, and B. Balcik, “A continuous approximation approach for assessment routing in disaster relief,” *Transportation Research Part B: Methodological*, vol. 50, pp. 20–41, 2013.
- [18] B. E. Oruc and B. Y. Kara, “Post-disaster assessment routing problem,” *Transportation Research Part B: Methodological*, vol. 116, pp. 76–102, 2018.
- [19] J. Salvà Soler, V. Hemmelmayr, and G. Raidl, “Exact methods for the Selective Assessment Routing Problem,” *Central European Journal of Operations Research*, 2024.
- [20] D. Pamukcu, M. Çelik, and B. Balçık, “Integrating logistics planning and population sampling decisions for humanitarian needs assessment,” *Socio-Economic Planning Sciences*, vol. 99, 2025.
- [21] C. F. Dursunoglu, O. Arslan, S. M. Demir, B. Y. Kara, and G. Laporte, “A unifying framework for selective routing problems,” *European Journal of Operational Research*, vol. 320, no. 1, pp. 1–19, 2025.
- [22] G. Ghiani and G. Improta, “An efficient transformation of the generalized vehicle routing problem,” *European Journal of Operational Research*, vol. 122, no. 1, pp. 11–17, 2000.
- [23] O. Bräysy, E. Martínez, Y. Nagata, and D. Soler, “The mixed capacitated general routing problem with turn penalties,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 12954–12966, 2011.
- [24] R. Baldacci, E. Bartolini, and G. Laporte, “Some applications of the generalized vehicle routing problem,” *Journal of the Operational Research Society*, vol. 61, no. 7, pp. 1072–1077, 2010.

- [25] A. Perugia, L. Moccia, Cordeau, J.-F., and G. Laporte, “Designing a home-to-work bus service in a metropolitan area,” *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1710–1726, 2011.
- [26] J. C. Micó and D. Soler, “The capacitated general windy routing problem with turn penalties,” *Operations Research Letters*, vol. 39, no. 4, pp. 265–271, 2011.
- [27] N.-H. Quttineh, T. Larsson, K. Lundberg, and K. Holmberg, “Military aircraft mission planning: a generalized vehicle routing model with synchronization and precedence,” *EURO Journal on Transportation and Logistics*, vol. 2, pp. 109–127, 2013.
- [28] L. H. Son and A. Louati, “Modeling municipal solid waste collection: A generalized vehicle routing model with multiple transfer stations, gather sites and inhomogeneous vehicles in time windows,” *Waste Management*, vol. 52, pp. 34–49, 2016.
- [29] S. Mitra, “An algorithm for the generalized vehicle routing problem with backhauling,” *Asia-Pacific Journal of Operational Research*, vol. 22, no. 2, pp. 153–169, 2005.
- [30] V. Kachitvichyanukul, P. Sombuntham, and S. Kunnapapdeelert, “Two solution representations for solving multi-depot vehicle routing problem with multiple pickup and delivery requests via pso,” *Computers & Industrial Engineering*, vol. 89, pp. 125–136, 2015.
- [31] B. Biesinger, B. Hu, and G. Raidl, “An integer L-shaped method for the generalized vehicle routing problem with stochastic demands,” *Electronic Notes in Discrete Mathematics*, vol. 52, pp. 245–252, 2016.
- [32] C. Sabo, P. C. Pop, and A. Horvat-Marc, “On the selective vehicle routing problem,” *Mathematics*, vol. 8, no. 5, pp. 771–780, 2020.
- [33] Y. Yuan, D. Cattaruzza, M. Ogier, F. Semet, and D. Vigo, “A column generation based heuristic for the generalized vehicle routing problem with time windows,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 152, 2021.
- [34] A. A. Jolfaei, M. Alinaghian, R. Bahrami, and E. B. Tirkolaee, “Generalized vehicle routing problem: Contemporary trends and research directions,” *Heliyon*, vol. 9, no. 12, 2023.
- [35] J. Oyola, H. Arntzen, and D. L. Woodruff, “The stochastic vehicle routing problem, a literature review, part II: solution methods,” *EURO Journal on Transportation and Logistics*, vol. 6, no. 4, pp. 349–388, 2017.
- [36] J. Oyola, H. Arntzen, and D. L. Woodruff, “The stochastic vehicle routing problem, a literature review, part I: models,” *EURO Journal on Transportation and Logistics*, vol. 7, no. 3, pp. 193–221, 2018.

- [37] P. Jaillet, *Probabilistic traveling salesman problems*. PhD thesis, Massachusetts Institute of Technology, 1985.
- [38] D. J. Bertsimas, P. Jaillet, and A. R. Odoni, “A priori optimization,” *Operations Research*, vol. 38, no. 6, pp. 1019–1033, 1990.
- [39] J.-F. Côté, T. Alves de Queiroz, F. Gallesi, and M. Iori, “A branch-and-regret algorithm for the same-day delivery problem,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 177, 2023.
- [40] M. W. Ulmer *et al.*, “Offline–online approximate dynamic programming for dynamic vehicle routing with stochastic requests,” *Transportation Science*, vol. 53, no. 1, pp. 185–202, 2019.
- [41] F. Lagos, M. A. Klapp, and A. Toriello, “Branch-and-price for routing with probabilistic customers,” *Computers & Industrial Engineering*, vol. 183, p. 109429, 2023.
- [42] J. Zhang, K. Luo, A. M. Florio, and T. Van Woensel, “Solving large-scale dynamic vehicle routing problems with stochastic requests,” *European Journal of Operational Research*, vol. 306, no. 2, pp. 596–614, 2023.
- [43] M. Gendreau, G. Laporte, and R. Séguin, “An exact algorithm for the vehicle routing problem with stochastic demands and customers,” *Transportation Science*, vol. 29, no. 2, pp. 143–155, 1995.
- [44] M. Gendreau, G. Laporte, and R. Séguin, “A tabu search heuristic for the vehicle routing problem with stochastic demands and customers,” *Operations Research*, vol. 44, no. 3, pp. 469–477, 1996.
- [45] M. Dror, G. Laporte, and P. Trudeau, “Vehicle routing with stochastic demands: Properties and solution frameworks,” *Transportation Science*, vol. 23, no. 3, pp. 166–176, 1989.
- [46] S. S. Özarik, V. Lurkin, L. P. Veelenturf, T. Van Woensel, and G. Laporte, “An adaptive large neighborhood search heuristic for last-mile deliveries under stochastic customer availability and multiple visits,” *Transportation Research Part B: Methodological*, vol. 170, pp. 194–220, 2023.
- [47] C. D. J. Waters, “Vehicle-scheduling problems with uncertainty and omitted customers,” *Journal of the Operational Research Society*, vol. 40, no. 12, pp. 1099–1108, 1989.
- [48] W. C. Benton and M. D. Rossetti, “The vehicle scheduling problem with intermittent customer demands,” *Computers & Operations Research*, vol. 19, no. 6, pp. 521–531, 1992.

- [49] N. Secomandi and F. Margot, “Reoptimization approaches for the vehicle-routing problem with stochastic demands,” *Operations Research*, vol. 57, no. 1, pp. 214–230, 2009.
- [50] AFAD (Disaster and Emergency Management Presidency), “06 Şubat 2023 Pazarcık-Elbistan (Kahramanmaraş) Mw: 7.7 – Mw: 7.6 Depremleri Raporu,” Report, AFAD, 2023. Accessed: 2025-04.
- [51] Turkish Statistics Institute, “Merkezi dağıtım sistemi.” <https://biruni.tuik.gov.tr/medas/?locale=tr>, 2024. Accessed: 2025-04.
- [52] Republic of Turkey Ministry of National Defense General Directorate of Mapping, “İl ve İlçe yüz Ölçümleri.” <https://www.harita.gov.tr/urun/il-ve-ilce-yuzolcumleri/176>, 2025. Accessed: 2025-04.
- [53] S. K. Thompson, *Introduction*, ch. 1, pp. 1–8. John Wiley Sons, Ltd, 2012.
- [54] S. K. Thompson, *Simple Random Sampling*, ch. 2, pp. 9–37. John Wiley Sons, Ltd, 2012.
- [55] Sci-Kit Learn, “Minmax scaler.” <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>. Accessed: 2024-09.
- [56] T. Kodinariya and P. Makwana, “Review on determining of cluster in k-means clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, pp. 90–95, 01 2013.
- [57] A. Ng, “Cs229 lecture notes: The k-means clustering algorithm.” <https://cs229.stanford.edu/notes2020spring/cs229-notes7a.pdf>, 2020. Accessed: 2024-09.
- [58] Nüfusu Ne, “Nüfusu Ne.” <https://www.nufusune.com/>, 2024. Accessed: 2025-04.
- [59] Heidelberg Institute for Geoinformation Technology, “Open route services.” <https://openrouteservice.org/>, 2024. Accessed: 2024-09.
- [60] Free Map Tools, “Elevation finder.” <https://www.freemaptools.com/elevation-finder.htm>. Accessed: 2024-09.
- [61] S. Acar, D. Şahin Cinoğlu, T. Karagöz, G. Kaygısız, M. C. Meydab, and M. Işık, *Socio-Economic Development Index of Districts*. Republic of Türkiye Ministry of Industry and Technology, 2023.
- [62] Turkish Ministry of Agriculture and Forestry, “Income-Sufficient Agricultural Land Sizes by Province and District in Türkiye,” 2024. Accessed: 2025-04.

- [63] Turkish Revenue Administration, “Arsa ve Arazi Asgari Metrekare Birim Değerleri.” <https://www.gib.gov.tr/yardim-ve-kaynaklar/yayinlar/arsa-ve-arazi-asgari-metrekare-birim-degerleri>, 2024. Accessed: 2024-09.





APPENDICES

APPENDIX A. SARP MATHEMATICAL MODEL FORMULATION

To decide on main-backup pairings and the routes to minimize the total expected distance, the following decision variables are defined:

First-stage decision variables

y_v^r : binary variable indicating if village $v \in V$ is selected as a main village in route $r \in R$;

$\hat{y}_{vu}^r$: binary variable indicating if village $u \in V$ is selected as a backup village for the main village $v \in V$ in route $r \in R$.

Second-stage decision variables

x_{vu}^{rs} : binary variable indicating if route $r \in R$ goes from village $v \in V$ to village $u \in V$ under scenario $s \in S$,

w_v^s : visiting position of village $v \in V$ under scenario $s \in S$.

Specifically, the SARP can be defined as selecting a set of routes and specifying the main and corresponding backup villages while ensuring sample representativeness and completing a specified number of surveys to minimize the expected travel distance of the assessment under key informant uncertainty. The formulation of the SARP using these sets, parameters, and decision variables is presented below.

$$\text{minimize} \quad \sum_{s \in S} \sum_{r \in R} \sum_{v \in V} \sum_{u \in V} p_s d_{vu} x_{vu}^r \quad (\text{A.1})$$

subject to

$$\sum_{r \in R} y_v^r \leq 1 \quad v \in V \quad (\text{A.2})$$

$$\sum_{u \in V} \alpha_v^k \alpha_u^k \hat{y}_{vu}^r = y_v^r \quad v \in V, k \in K, r \in R \quad (\text{A.3})$$

$$\hat{y}_{vv}^r = 0 \quad v \in V, r \in R \quad (\text{A.4})$$

$$\sum_{v \in V} y_v^r = n \quad r \in R \quad (\text{A.5})$$

$$\sum_{r \in R} \sum_{v \in V} \alpha_v^k y_v^r = N^k \quad k \in K \quad (\text{A.6})$$

$$\sum_{v \in V} x_{v0}^{rs} = 1 \quad r \in R, s \in S \quad (\text{A.7})$$

$$\sum_{v \in V} x_{0v}^{rs} = 1 \quad r \in R, s \in S \quad (\text{A.8})$$

$$\sum_{u \in V \setminus \{u\}} x_{uv}^{rs} = y_v^r + \sum_{p \in V \setminus \{p\}} (1 - \beta_p^s) \hat{y}_{pv}^r \quad v \in V, r \in R, s \in S \quad (\text{A.9})$$

$$\sum_{u \in V \setminus \{u\}} x_{vu}^{rs} = y_v^r + \sum_{p \in V \setminus \{p\}} (1 - \beta_p^s) \hat{y}_{pv}^r \quad v \in V, r \in R, s \in S \quad (\text{A.10})$$

$$w_0^s = 0 \quad s \in S \quad (\text{A.11})$$

$$w_u^s \geq w_v^s + 1 - 2n(1 - x_{vu}^{rs}) \quad v, u \in V, r \in R, s \in S \quad (\text{A.12})$$

$$w_v^s \leq 2n \quad v, u \in V, r \in R, s \in S \quad (\text{A.13})$$

$$w_u^s \geq w_v^s + 1 - 2n(1 - (1 - \beta_v^s) \hat{y}_{vu}^r) \quad v, u \in V, r \in R, s \in S \quad (\text{A.14})$$

$$y_v^r \in \{0, 1\} \quad v \in V, r \in R \quad (\text{A.15})$$

$$\hat{y}_{vu}^r \in \{0, 1\} \quad v, u \in V, r \in R \quad (\text{A.16})$$

$$x_{vu}^{rs} \in \{0, 1\} \quad v, u \in V, r \in R, s \in S \quad (\text{A.17})$$

$$w_v^s \geq 0 \quad v \in V, s \in S. \quad (\text{A.18})$$

The objective function A.1 minimizes the expected cost associated with the route design. Constraints A.2 ensure that each village is selected as the main site at most in one route. Equalities A.3 guarantee that all selected main villages have exactly one backup site, ensuring that the backup village belongs to the same cluster as the mains. Constraints A.4 prohibit a village from being selected as both a primary and backup site simultaneously. Constraints A.5 limit the number of main villages selected in each route to n . Constraints A.6 ensure representative sampling by setting the total number of main villages for each cluster. Equalities A.7 and A.8 ensure that all routes, under any scenario,

depart from and return to the province center (node 0). Similarly, equalities A.9 and A.10 guarantee that the main selected villages are visited on their corresponding route under any scenario, and backup villages when necessary. Constraints A.11 – A.13 are the well-known Miller-Tucker-Zemlin (MTZ) constraints that ensure well-defined routes and prevent the formation of subtours. Specifically, constraints A.11 set the initial position to zero at the province center (node 0); constraints A.12 control the visiting position of villages; and constraints A.13 limit the visiting position to the maximum number of visits each route can have. Note that since backup sites may be needed, a route can have up to twice the number of main villages ($2n$) per route. Constraints A.14 ensure that backup villages are visited, if necessary, sometime after their respective main village. Finally, constraints A.15 - A.18 define the domains of the variables.

