Journal of Industrial and Systems Engineering
Vol. 15, No. 2, pp. 109-129
Spring (April) 2023

Data-driven robust optimization for hub location-routing problem under uncertain environment

MirMohammad Musavi[1], Ali Bozorgi-Amiri[1]

[1] *School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran*

mohammad.musavi@ut.ac.ir, alibozorgi@ut.ac.ir

Abstract

This study addresses the Hub Location-Routing Problem (HLRP) in transportation networks, considering the inherent uncertainty in travel times between nodes. We employed a method centered on data-driven robust optimization, utilizing Support Vector Clustering (SVC) to form an uncertainty set grounded in empirical data. The proposed methodology is compared against traditional uncertainty sets, showcasing its superior performance in providing robust solutions. A comprehensive case study on a retail store's transportation network in Tehran is presented, demonstrating significant differences in hub locations, allocations, and vehicle routes between deterministic and robust models. The SVC-based model proves to be particularly effective, yielding substantially improved objective function values compared to polyhedral and box uncertainty sets. The study concludes by highlighting the practical significance of this research and suggesting future directions for advancing transportation network optimization under uncertainty.

Keywords: Robust optimization, hub location, machine learning, data-driven approach, support vector clustering

1- Introduction

The efficient management of transportation networks is of paramount importance in contemporary logistics and supply chain management. In the realm of network design and optimization, the Hub Location-Routing Problem (HLRP) holds a pivotal position, addressing the intricate challenge of determining optimal locations for hubs and the corresponding allocation of demand nodes to these hubs. However, the conventional HLRP formulations typically assume deterministic travel times between nodes, a simplification that might not adequately capture the complexities inherent in real-world transportation systems.

In practice, transportation networks often operate under conditions of considerable uncertainty. Variations in travel times caused by factors such as traffic congestion, weather conditions, and unexpected incidents can substantially impact the performance and cost-effectiveness of hub-based logistics solutions. To address these uncertainties and enhance the robustness of HLRP solutions, the emerging field of Data-Driven Robust Optimization (DDRO) offers a promising avenue for research and application.

*Corresponding author
ISSN: 1735-8272, Copyright c 2023 JISE. All rights reserved

This paper delves into the challenging domain of the HLRP, where travel times between nodes are inherently uncertain. Our objective is to develop a robust optimization framework that leverages data-driven techniques to enhance the resilience and reliability of hub-based logistics operations. By incorporating real-world data on travel time uncertainties into the optimization process, our approach seeks to yield solutions that are not only cost-efficient but also capable of withstanding the vagaries of an uncertain operational environment.

The contributions of this paper can be summarized as follows:

- We present a comprehensive mathematical model that integrates data-driven insights into the traditional HLRP framework, allowing for the representation and management of uncertain travel times.
- We introduce robust optimization techniques to address the inherent uncertainties, ensuring that the proposed hub location and routing solutions remain effective under a range of plausible scenarios.
- We demonstrate the practical relevance of our approach through a set of illustrative examples and case studies, showcasing its potential to improve decision-making in logistics and supply chain management.
- By advancing the understanding of the HLRP in uncertain environments and offering innovative solutions, this work contributes to the theoretical foundations of robust optimization in logistics.
- Our research provides a valuable decision support tool for practitioners seeking resilient hub-based logistics solutions that can adapt to the ever-changing dynamics of transportation networks.

The subsequent sections of this paper can be outlined as follows: Section 2 offers a comprehensive review of the relevant literature. In Section 3, we establish a clear problem definition and present the mathematical formulation. Section 4 delves into the intricacies of data-driven robust optimization. Moving on to Section 5, a thorough numerical analysis is provided. Section 6 showcases a compelling case study, and lastly, Section 7 encapsulates the concluding insights.

2- Literature review

The HLRP with uncertain travel times constitutes a variant of the classic hub location problem that incorporates uncertainty in travel times between nodes. Klincewicz (1991), Campbell (1994) and Ernst & Krishnamoorthy (1999) can be credited as pioneers in the field of classical hub locations. Their seminal papers laid the groundwork for subsequent research in this area, shaping the way we understand and analyze hub location problems. Their early contributions provided valuable insights and methodologies that continue to influence and inform current studies in the field. Interested readers seeking to delve deeper into the evolving landscape of hub location problems can find valuable insights in the following review papers: S. A. Alumur et al. (2021), S. Alumur & Kara (2008) and Campbell & O'Kelly (2012). These comprehensive works offer a nuanced exploration of the historical development, key methodologies, and emerging trends within the field, providing an invaluable resource for researchers and practitioners alike.

One notable variant of hub location problems is HLRPs, which combines the determination of hub locations with the optimization of routing decisions. A seminal contribution to this field was made by de Camargo et al. (2013), whose work introduced a novel formulation for this pivotal problem. The solution approach employed a meticulously crafted Benders decomposition algorithm, showcasing both innovation and effectiveness. Another noteworthy contribution in the domain of HLRP are the works by Catanzaro et al. (2011) and Rodríguez-Martín et al. (2014), where they introduced branch-and-cut algorithm specifically tailored for addressing the HLRP. This algorithm represents a significant advancement in the field, providing a powerful tool for efficiently solving complex instances of this intricate problem. The work by Lopes et al. (2016) presented innovative heuristics tailored for addressing the many-to-many HLRP. Subsequently, Karimi (2018) made a notable advancement by integrating capacitated hub covering location considerations into the simultaneous pickup and delivery vehicle routing problem (VRP). This integration represents a significant extension of previous research, addressing the added complexity of capacity constraints in hub locations while simultaneously managing pickup and delivery operations. Danach et al.

(2019) provided a lagrangian relaxation and a hyper-heuristic solution method for this problem. Recently, researchers including Ghaffarinasab et al., (2018), Ratli et al. (2022), and Wu et al. (2022) have made substantial strides in exploring diverse solution methodologies for the HLRP. Their investigations have encompassed a range of approaches, such as large neighborhood decomposition and continuous approximation techniques, as well as the development of highly efficient heuristics. These efforts represent a concerted endeavor to advance the state-of-the-art in solving this complex problem, offering a spectrum of tools and strategies for achieving optimal or near-optimal solutions across various real-world applications.

In the field of transportation logistics, managing uncertainty is a critical aspect of ensuring efficient and reliable operations. Traditional approaches often rely on predefined uncertainty sets and perturbation ranges, which may lead to overconservative solutions. Asefi et al. (2019) addressed the challenge of Municipal Solid Waste Management in large cities, emphasizing the need for practical decision-making tools. They introduced a two-stage stochastic optimization approach to effectively support cost-effective ISWM transportation system planning under uncertainty. Jiang et al. (2020) tackled uncertainty in a regional logistics network design problem with CO2 emission reduction goals in urban clusters. They introduced an improved adjustable robust optimization approach to address uncertainty in demands, providing a practical guide for sustainable logistics development. Russell et al. (2020) emphasized the growing uncertainties in container port logistics, exacerbated by factors like the COVID-19 pandemic. They proposed a concise framework to evaluate port logistics capacity, aiding in navigating uncertain scenarios. In recent years, data-driven approaches have emerged as a transformative paradigm in the field of optimization, offering novel solutions to address uncertainty and variability in decision-making processes. Bertsimas et al. (2018) stands as a pioneer in the realm of data-driven robust optimization, being instrumental in its introduction and early development. This groundbreaking work revolutionized the way uncertainty is handled in optimization models, paving the way for more resilient and adaptable decision-making processes in the face of real-world variability and unpredictability. Shang et al. (2017) introduced a transformative approach known as data-driven robust optimization, leveraging the principles of kernel learning. This pioneering work represents a significant paradigm shift in handling uncertainty within optimization models. Shang & You (2019) also put forth a pioneering data-driven robust optimization methodology for scenario-based stochastic model predictive control.

Numerous researchers have applied data-driven robust optimization across a diverse range of applications. Noteworthy implementations include: large-scale industrial energy systems (Shen et al., 2020), wastewater sludge-to-biodiesel supply chain (Mohseni & Pishvaee, 2020), crude oil blending (Dai et al., 2020), optimization of grinding processes (Inapakurthi et al., 2020), supply chain planning (Gumte et al., 2021), multi-objective renewable energy location (Lotfi et al., 2022), integrated network design for solar photovoltaic to microgrid systems (Gilani et al., 2022), scheduling of power to methanol processes (Zheng et al., 2022), and privacy-preserving energy trading management in networked microgrids (Mohseni et al., 2023).

Recently, Zhang et al. (2022) introduced an innovative machine learning-based data-driven robust optimization approach tailored for uncertain environments. This method harnesses the power of machine learning to enhance the robustness and adaptability of optimization models, representing a significant advancement in the field. In parallel, Goerigk & Kurtz (2023) put forth a cutting-edge data-driven robust optimization framework, leveraging deep neural networks to effectively handle uncertainty. This approach demonstrates the potential of advanced neural network architectures in bolstering the resilience of optimization models to real-world variability and unpredictability.

The current state of research in the HLRP with uncertain travel times reveals several noteworthy gaps. Existing literature predominantly focuses on deterministic scenarios, with limited attention to uncertainty in travel times. Traditional robust optimization approaches with predefined uncertainty sets are prevalent, yet there is a significant opportunity to harness available data for more tailored uncertainty characterization. Moreover, while Support Vector Clustering (SVC) has demonstrated efficacy in machine learning, its application in constructing uncertainty sets for HLRP remains relatively unexplored. Bridging these gaps would lead to a more comprehensive understanding of the impact of uncertain travel times and result in more efficient and robust solutions.

3- Problem definition

The HLRP under uncertain environment stands as a variant of the classic hub location problem, distinguished by its integration of uncertain travel times between nodes. This stochastic element introduces a layer of complexity to the optimization process, considering factors such as traffic variations, weather conditions, and unforeseen events, which significantly impact the efficiency and cost-effectiveness of transportation operations.

In this context, we consider a total of N nodes, with a subset of P nodes designated as hubs and the remainder as non-hub locations. Each hub node is associated with a deterministic number of vehicles, denoted as V, each possessing a specific load capacity, denoted as L. The primary objective is to judiciously allocate the non-hub nodes to the selected hubs, establishing an optimal network configuration that minimizes transportation costs. This entails determining the optimal hub locations and devising the most efficient routes between nodes, while effectively managing the uncertainties inherent in travel times. Each vehicle within the fleet is endowed with a specific load capacity and is dispatched to provide services to various customers, facilitating the delivery of goods. Every customer is exclusively serviced by a single vehicle, and each vehicle is mandated to return to its assigned hub upon the completion of its delivery operations.

With the integration of uncertain travel times into the HLRP, the conventional metric of minimizing total travel time is replaced by a comprehensive objective function. This function seeks to minimize the combined path length and associated costs. This comprehensive approach to cost optimization is designed to ensure the efficient allocation of resources and the timely delivery of goods, even in the face of unpredictable travel conditions.

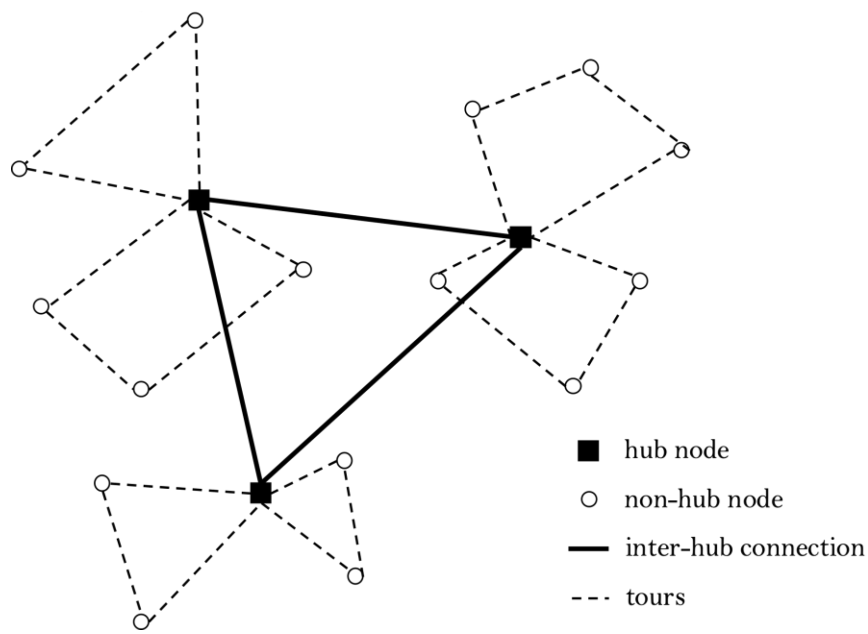


Fig. 1. Graphical representation of the hub location-routing problem

3-1- Problem assumptions

Model assumptions for the HLRP under Uncertain Environment:

- **Uncertain Travel Times**: The planning horizon is continuous, allowing for the utilization of mathematical techniques to address the uncertainties in travel times.
- **Euclidean Geometry**: Distance computations are based on a simplified Euclidean space, providing a practical approximation for real-world transportation scenarios.
- **Homogeneous Fleet**: The vehicle fleet is composed of identical vehicles in terms of capacity, speed, and operational efficiency, simplifying the optimization process.
- **Fixed Customer Demand**: Customer demands remain constant throughout the planning horizon, eliminating the need to account for demand fluctuations or uncertainties.
- **Static Hub Locations**: Hub locations are predetermined and remain fixed throughout the planning period, providing a stable foundation for the optimization process.
- **Single Allocations**: Each customer node is allocated to a single hub for servicing, streamlining the allocation process and ensuring clear responsibility for customer satisfaction.
- **Direct Routes**: Vehicles follow direct routes between their respective hubs and allocated nodes, disregarding factors like detours or multi-step routes.

3-2- Problem formulation

Indices:

$i, j, k, l \in \{0,1,2,\dots,N+1\}$	Index for nodes

Parameters:

P	Number of hubs to locate
α	Discount factor for inter-hub connections, $0 \le \alpha \le 1$
V	Number of available vehicles at each hub
L	Vehicle capacity
M	Big number
c_{ij}	Transportation cost between nodes i and j
w_{ij}	Flow of products from node i to node j
t_{ij}	Travel time between nodes i and j

Positive variables:

at_i	Arrival time of delivery at node i
$load_i$	Load of vehicle in delivery of node i
y_{ikl}	Flow of products from node i, routed through hubs k and l

Binary variables:

x_{ij}	Equal to 1 if node i is allocated to node j, otherwise 0, node i is a hub if $x_{ii} = 1$
z_{ijk}	Equal to 1 if customer i is served before customer j, which both nodes allocated to distribution center k, otherwise 0

Objective function:

$$\sum_{i=1}^{N}\left[\sum_{k=1}^{N} c_{ik}x_{ik}\left(\sum_{j=1}^{N} w_{ij}\right)\right] + \alpha\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} y_{ikl}c_{kl} + \sum_{i=1}^{N}\left[\sum_{k=1}^{N} c_{ki}x_{ik}\left(\sum_{j=1}^{N} w_{ji}\right)\right] + \sum_{i=1}^{N} at_i \qquad (1)$$

Constraints:

$$\sum_{i=1}^{N} x_{ii} = P \tag{2}$$

$$\sum_{k=1}^{N} x_{ik} = 1 \qquad \forall i = 1,2,\dots,N \tag{3}$$

$$x_{ik} \leq x_{kk} \qquad \forall i,k = 1,2,\dots,N \tag{4}$$

$$\sum_{l=1}^{N} y_{ikl} - \sum_{l=1}^{N} y_{ilk} = \sum_{j=1}^{N} w_{ij}x_{ik} - \sum_{j=1}^{N} w_{ij}x_{jk} \qquad \forall i,k = 1,2,\dots,N \tag{5}$$

$$2z_{ijk} \leq x_{jk} + x_{ik} \qquad \forall i,j,k = 1,2,\dots,N, i \neq j, i \neq k, j \neq k \tag{6}$$

$$\sum_{\substack{i=0 \\ i\neq j \\ i\neq k}}^{N} z_{ijk} = x_{jk} \qquad \forall j,k = 1,2,\dots,N, j \neq k \tag{7}$$

$$\sum_{\substack{j=1 \\ j\neq i \\ j\neq k}}^{N+1} z_{ijk} = x_{ik} \qquad \forall i,k = 1,2,\dots,N, i \neq k \tag{8}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{N} z_{0,jk} \leq V \qquad \forall k = 1,2,\dots,N \tag{9}$$

$$\sum_{\substack{i=1 \\ i\neq k}}^{N} z_{i,N+1,k} \leq V \qquad \forall k = 1,2,\dots,N \tag{10}$$

$$at_i \geq \widetilde{t_{ik}} z_{0,ik} \qquad \forall i,k = 1,2,\dots,N \tag{11}$$

$$at_j + M(1 - z_{ijk}) \geq \widetilde{t_{ij}} + at_i \qquad \forall i,j,k = 1,2,\dots,N, i \neq j, i \neq k, j \neq k \tag{12}$$

$$load_i \geq \sum_{j=1}^{N} w_{ji}\, z_{0,ik} \qquad \forall i,k = 1,2,\dots,N, i \neq k \tag{13}$$

$$load_j + M(1 - z_{ijk}) \geq \sum_{j=1}^{N} w_{ji} + load_i \qquad \forall i,j,k = 1,2,\dots,N, i \neq j, i \neq k, j \neq k \tag{14}$$

$$load_i \leq L \qquad \forall i = 1,2,\dots,N \tag{15}$$

$$at_i, load_i, y_{ikl} \geq 0 \quad z_{ijk}, x_{ik} \in \{0,1\} \qquad \forall i,j,k = 0,1,2,\dots,N+1 \tag{16}$$

The objective function (1) comprises four components. The first term computes the total transportation costs between non-hub and hub nodes. The second term calculates the total transportation costs between hub nodes. The third term assesses the total transportation costs between hub nodes and non-hub nodes. The fourth term quantifies the total travel time of vehicles.

Equation (2) enforces the constraint that only a maximum of P hubs can be located. Equation (3) mandates that each non-hub node must be allocated to exactly one hub node. Equation (4) states that node allocation to hub k is contingent on the presence of hub k.

Equation (5) delineates the threshold for inter-hub product transportation. Equation (6) stipulates those vehicles within hub k are permitted to serve non-hub nodes i and j exclusively if these non-hub nodes have been assigned to hub k. Equations (7) and (8) constitute the standard constraints of a vehicle routing problem, ensuring that each vehicle enters and exits the assigned node only once.
Equations (9) and (10) validate that the commencement and conclusion of vehicle routes within hubs hinge on the number of vehicles in those hubs. Equations (11) and (12) compute the arrival times of vehicles at nodes. Equations (13) and (14) ascertain the load of vehicles while servicing nodes.
Equation (15) mandates that the load of vehicles must not exceed their capacity. Finally, equation (16) outlines the variable types utilized in the proposed model.

4- Data-driven robust optimization

The foundation of conventional robust optimization methodologies often lies in the construction of uncertainty sets, a process that traditionally neglects the wealth of available information regarding uncertain parameters. This results in a challenging task for the user, who must select an appropriate uncertainty set, often leading to unnecessary overestimation. Significantly, methods in robust optimization induced by uncertainty sets have been developed assuming restricted knowledge about the exact values of uncertain parameters. In contrast, stochastic programming relies on past data to get the precise distribution of them. However, in practical scenarios, obtaining extensive historical data for accurate distribution estimation can be a formidable task.

For our specific application in the HLRP under Uncertain Environment, recent advancements have introduced a groundbreaking data-driven robust optimization approach. This innovative method harnesses empirical data and deploys machine learning techniques, including Support Vector Clustering (SVC), to cover data samples which results to an uncertainty.

In particular, the data-driven robust optimization methodology, as introduced by (Shang et al., 2017), extracts an uncertainty set directly from the available information on uncertain parameters. The choice of SVC as the preferred uncertainty set in our study is rooted in its unique capability to construct a data-driven uncertainty set tailored to the specific characteristics of uncertain parameters. SVC, a powerful machine learning technique, enables us to accurately identify and encompass relevant data samples that closely align with the genuine distribution of uncertainties in the HLRP. Unlike predefined uncertainty set structures, SVC allows for a flexible and adaptive approach, eliminating unnecessary overcoverage and providing a more precise representation of the uncertainty faced in real-world scenarios. By employing SVC, it accurately identifies and encompasses pertinent data samples, aligning closely with the genuine distribution of uncertainties in the HLRP. This data-driven approach represents a promising avenue for optimizing the HLRP, as it flexibly adapts to the specific characteristics of uncertain travel times. Consequently, it significantly enhances the robustness and practical applicability of the solution methodology.

4-1- Support vector clustering

Ben-Hur et al. (2001) introduced the SVC algorithm within the domain of machine learning theory, aims to identify a compact hypersphere with the least possible volume. This hypersphere is designed to tightly encompass all available data samples (Shang et al., 2017). To elucidate the algorithm, consider a collection of N data samples denoted as $S = \{s_i | i = 1,2, \dots, N\}$. Through a nonlinear transformationϕ, the original input space is mapped into a higher-dimensional feature space. Subsequently, the algorithm seeks the smallest hypersphere, characterized by a radius denoted as R, encompassing all representations of data samples within this specific feature space. This objective is addressed through the following optimization problem:

$$\begin{aligned} & \min R^2 \\ & s.t. \|\phi(s_i) - c\|^2 \le R^2 \qquad \forall i = 1,2, \dots, N \end{aligned} \tag{17}$$

Here, the center of the hypersphere's center is c and the Euclidean norm is showed by $\|.\|$. Then slack variables to permit certain data samples to deviate from the confines of the hypersphere are added:

$$\begin{aligned} &\min R^2 + H\sum_{i=1}^{N}\xi_i \\ &s.t. \|\phi(s_i) - c\|^2 \leq R^2 + \xi_i, \forall i = 1, \dots, N \\ &\xi_i \geq 0, \forall i = 1, \dots, N \end{aligned} \tag{18}$$

Here, H serves as a parameter determining the level of penalty. The resolution of this problem involves the formulation of the Lagrangian in the following manner:

$$L(R, c, \beta, \alpha, \xi) = R^2 - \sum_{i=1}^{N}(R^2 + \xi_i - \|\phi(s_i) - c\|^2)\beta_i - \sum_{i=1}^{N}\xi_i\alpha_i + H\sum_{i=1}^{N}\xi_i \tag{19}$$

Here, β_i and α_i (where $\alpha_i \geq 0$) denote the Lagrangian multipliers. Calculating the derivatives of L yields:

$$\sum_{i=1}^{N}\beta_i = 1 \tag{20}$$

$$c = \sum_{i=1}^{N}\beta_i\phi(s_i) \tag{21}$$

$$\beta_i = H - \alpha_i \tag{22}$$

By substituting the aforementioned definitions, equations (20) to (22) are incorporated into the Lagrangian, resulting in the dual form. This form takes the shape of a quadratic programming problem, expressed as follows:

$$\begin{aligned} &Max \sum_{i=1}^{N}K(s_i, s_i)\beta_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j K(s_i, s_j) \\ &s.t. \sum_{i=1}^{N}\beta_i = 1 \\ &0 \leq \beta_i \leq H, \forall i = 1, \dots, N \end{aligned} \tag{23}$$

In this context, $K(s_i, s_i)$ represents the kernel function. The kernel function is a key component of the SVC technique, influencing how data points are mapped in the high-dimensional feature space. Choosing an appropriate kernel function involves a trade-off between capturing fine-grained details and maintaining a generalized representation. A well-suited kernel function will align with the underlying distribution of uncertain parameters in the HLRP. There are several widely recognized forms of kernel functions available (Hsu et al., 2003):

1. Polynomial Kernel: $K(s_i, s_j) = \left(\gamma s_i^T s_j + 1\right)^d$
2. Radial Basis Function (RBF) Kernel: $K(s_i, s_j) = \exp\left\{-\left\|\frac{s_i - s_j}{2\sigma^2}\right\|\right\}$
3. Sigmoid Kernel: $K(s_i, s_j) = \tanh\{\gamma s_i^T s_j + r\}$

Employing these functions introduces intricate nonlinear terms into the robust optimization problem. Shang et al. (2017) have proposed the utilization of the subsequent generalized intersection kernel function:

$$K(s_i, s_j) = \sum_{i=1}^{N} l_k - \left\|F(s_i - s_j)\right\|_1 \tag{24}$$

In this function, F is determined as $F=\Sigma^{-\frac{1}{2}}$, which is the whitening matrix. Additionally, l_k serves as the width parameter, which is adjusted to guarantee the positive-definiteness of the kernel function K and make the problem convex. This parameter tuning ensures the stability and effectiveness of the robust optimization process. It is noteworthy that Σ is to minimize biases that might arise due to a limited sample size (N).

4-1- Robust counterpart formulation

Upon attaining the optimal solutions for β_i as determined by model (23), the computation of the hypersphere's radius proceeds as outlined below:

$$R^2 = K(s_{i'}, s_{i'}) - 2\sum_{i=1}^{N} K(s_{i'}, s_i)\beta_i + \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j K(s_i, s_j), i' \in SV \tag{25}$$

The area encompassed by the hypersphere constitutes the uncertainty set derived from S, and its characterization is:

$$U(S) = \left\{s \middle| K(s,s) - 2\sum_{i=1}^{N} K(s,s_i)\beta_i + \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j K(s_i,s_j) \le R^2\right\} \tag{26}$$

By substituting equation (25) into (26) we have:

$$\begin{aligned} U(S) &= \left\{s \middle| \sum_{i=1}^{N} K(s_{i'}, s_i)\beta_i \le \sum_{i=1}^{N} K(s,s_i)\beta_i, i' \in SV\right\} \\ &= \{s | \textstyle\sum_{i\notin ID} K(s_{i'}, s_i)\beta_i \le \sum_{i\notin ID} K(s,s_i)\beta_i, i' \in SV\} \end{aligned} \tag{27}$$

It's worth noting that the condition $i \notin ID$ in the second equality stems from $\beta_i = 0$ when $i \in ID$. When we substitute the kernel function (24) into equation (27), we arrive at the subsequent form of the uncertainty set:

$$U(S) = \{s | \textstyle\sum_{i\notin ID} \|F(s-s_i)\|_1\beta_i \le \sum_{i\notin ID}\|F(s_{i'}-s_i)\|_1\beta_i, i' \in SV\} \tag{28}$$

By incorporating auxiliary variables $Z=[z_1,\dots,z_N]$ and defining $\Omega=\sum_{i\notin ID}\|F(s_{i'}-s_i)\|_1\beta_i$, the expression for the uncertainty set (28) can be rephrased as:

$$U(S) = \left\{ s \middle| \begin{array}{c} \exists z_i, i \notin ID, s.t. \\ \sum_{i\notin ID} z_i\beta_i \le \Omega \\ -z_i \le F(s-s_i) \le z_i \;\; i \notin ID \end{array} \right\} \tag{29}$$

The robust adaptation within the uncertainty set $U(S)$ is delineated in the following:

$$\begin{aligned} &\min_{x\in X} c^T x \\ &s.t. \max_{a\in U(S)} a^T x \le b \end{aligned} \tag{30}$$

To derive the equivalent LP formulation of (30), we can express the internal maximizing problem in the following way:

$$\begin{aligned} &\max_{s,z_i} s^T x \\ &s.t. \sum_{i\notin ID} z_i\beta_i \le \Omega \\ &-z_i \le F(s-s_i) \le z_i \quad \forall i \notin ID \end{aligned} \tag{31}$$

The dual form of the (31) will be:

$$\min_{\gamma_i,\vartheta_i,\lambda} \sum_{i\notin ID} (\gamma_i - \vartheta_i)^T F s_i + \Omega\lambda$$
$$s.t. \sum_{i\notin ID} F(\gamma_i - \vartheta_i)^T + x = 0 \quad (32)$$
$$\gamma_i + \vartheta_i = \lambda\beta_i \quad \forall i \notin ID$$
$$\vartheta_i, \gamma_i \in R_+^n \lambda \geq 0$$

In this context, γ_i, ϑ_i, and λ denote dual variables. By incorporating (32) into (30), we arrive at the ensuing robust counterpart problem:

$$\min_{x\in X} c^T x$$
$$s.t. \sum_{i\notin ID} (\gamma_i - \vartheta_i)^T F s_i + \Omega\lambda \leq b$$
$$\sum_{i\notin ID} F(\gamma_i - \vartheta_i)^T + x = 0 \quad (33)$$
$$\gamma_i + \vartheta_i = \lambda\beta_i \quad \forall i \notin ID$$
$$\vartheta_i, \gamma_i \in R_+^n \lambda \geq 0$$

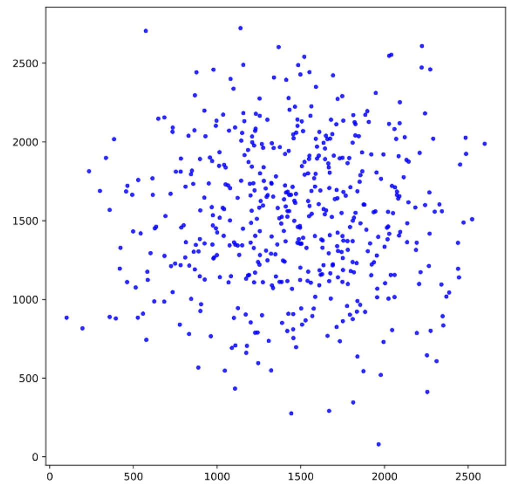
Finally, we discuss the step-by-step procedure for implementing the proposed approach. To maintain generality, let's assume the model at hand can be represented like the problem in (30). Here, the set $D = \{s_i\}_{i=1}^N$ comprises N available samples, representing realizations of uncertainties. This formulation accommodates more complex cases in a similar manner. The parameter H is the regularization factor and can be customized based on the conservatism of the decision maker.
The subsequent robust optimization (30) can be tackled through the subsequent steps:

1. Calculate Σ using the sample datapoints. Obtain the weighting matrix $F = \Sigma^{-\frac{1}{2}}$.
2. Establish the kernel parameters $\{l_k\}$.
3. Create the kernel matrix $K(s_i, s_j)$ utilizing the kernel function (24) with the N samples.
4. Resolve the SVC model by addressing the problem (23) with the kernel matrix K. Acquire the indices of support vectors and β.
5. Utilize the support vectors $\{s_i, i \in SV\}$ along with their corresponding Lagrange multipliers $\{\beta_i, i \in SV\}$ to construct the robust counterpart problem (32). Subsequently, employ it to substitute the constraint containing uncertainty in (30).
6. Resolve the modified problem.

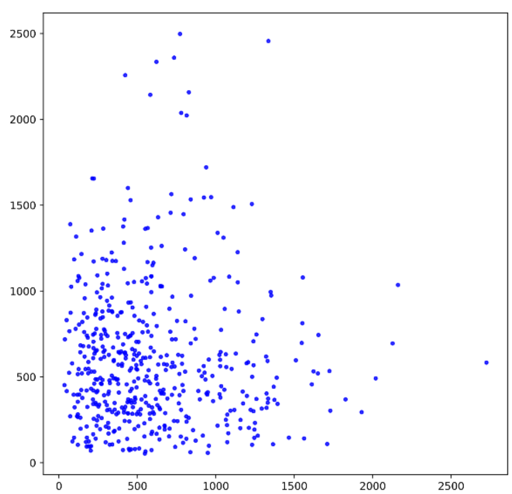
5- Numerical analysis

Our primary objective in this section is to validate and assess the efficacy of the SVC-based uncertainty set, following the methodology introduced by Shang et al. (2017).
To accomplish this, we have generated three distinct datasets for t_{ij} utilizing bivariate gamma, truncated uniform, and mixed Gaussian distributions. The remaining parameters of these datasets have been adopted from the widely utilized CAB dataset, originally presented by O'kelly (1987). This dataset is derived from airline passenger flows between 25 cities, serving as a valuable benchmark for our analysis. For the experimental setup, we have set α to 0.5, P to 3, and V to 2. Additionally, the parameter L is determined as $\frac{1.5\times\sum_{i,j} w_{ij}}{P\times V}$.

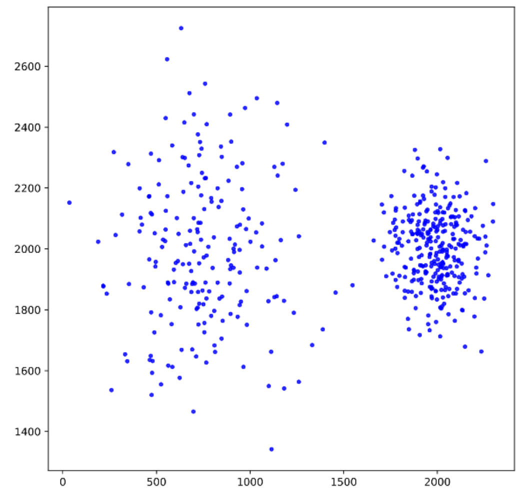
It's worth noting that all coding implementations have been executed in the Julia programming language. The experiments have been conducted on a computer system equipped with a 2 GHz Quad-Core Intel Core i5 processor and 16 GB of RAM. This robust computational environment ensures reliable and efficient execution of our experiments. Figure 2 illustrates the uncertainty data generated by different distributions for two nodes for computational experiments.



(a) Truncated Uniform



(b) Bivariate Gamma



(c) Mixed Gaussian

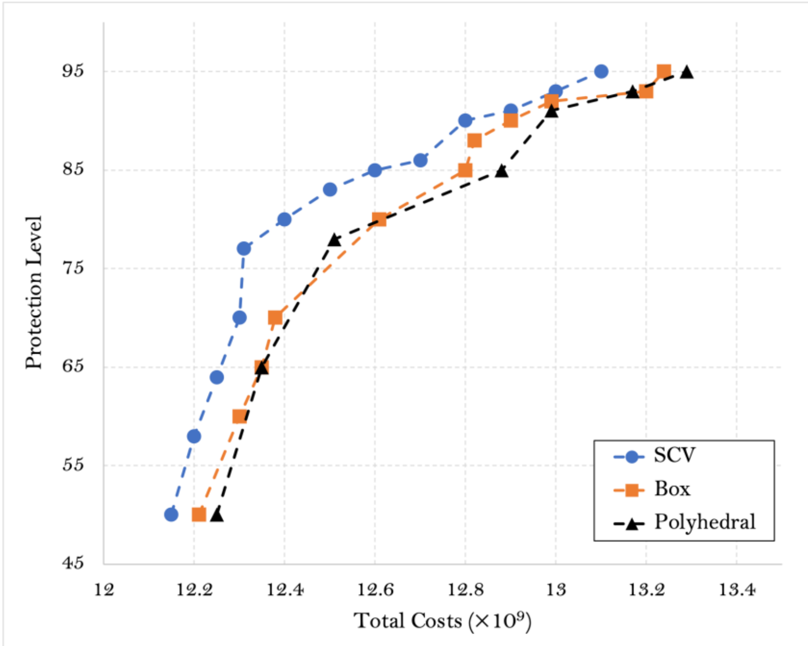
Fig. 2. Uncertainty data generated by different distributions for computational experiments

Once the uncertainty sets are defined, we formulate the robust counterpart optimization problem. Following this, a variety of robust solutions is derived for each set by modifying the size of its corresponding uncertainty set. To assess the efficacy of these robust solutions, 1500 scenarios are produced, randomly. Each scenario is carefully examined to ascertain if a robust solution results in any constraint breaches. The degree of protection provided by a solution is subsequently gauged by calculating the percentage of random scenarios that do not incur any constraint violations.

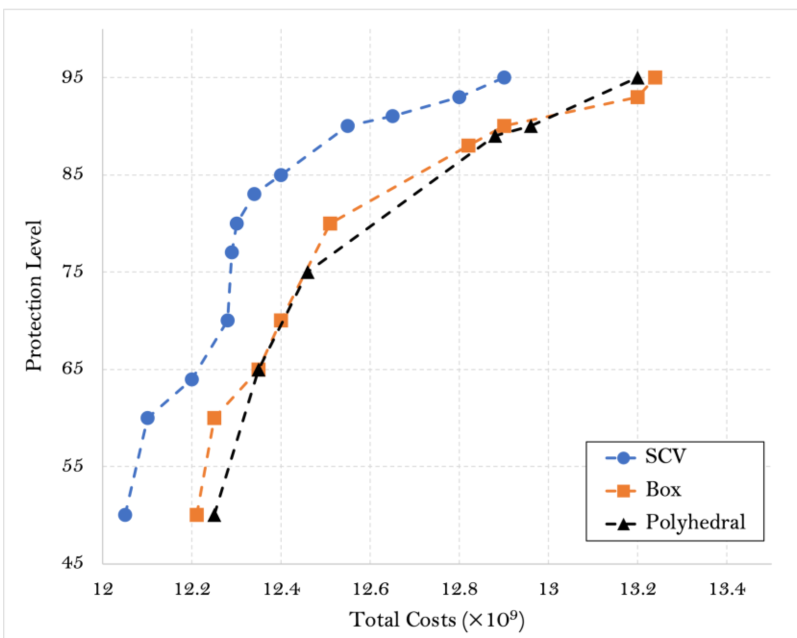
The computational analysis results, illustrated in figure 3, show the effectiveness of the SVC-based uncertainty sets in solving the minimization problem for the total costs of the hub location-routing scenario. In this figure, we represent the total costs of the network for each protection level. It's important to note that since our problem is a minimization task, a smaller objective function value under the same protection level signifies a less conservative solution.

Consequently, we aim for a performance-tradeoff curve that resides as close as possible to the lower right corner. The results indicate that the data-driven SVC-based uncertainty set substantially outperforms both the box uncertainty set and the polyhedral uncertainty set.

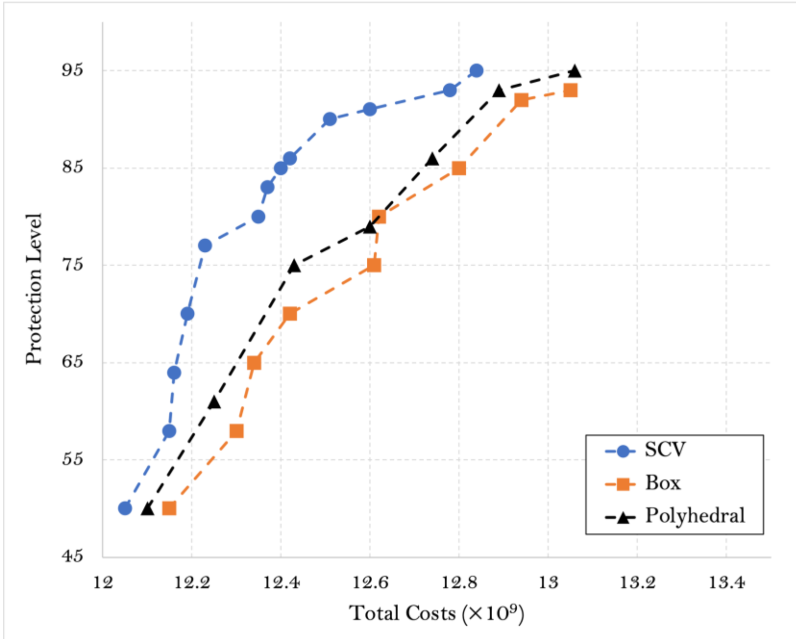
Even in cases involving other uncertainty set (mixed Gaussian and bivariate gamma), the SVC-based uncertainty set consistently demonstrates superior performance. This underscores the notion that by efficiently encapsulating the distributional characteristics of our data, the uncertainty set can effectively mitigate uncertainties and reduce the conservatism inherent in the solutions of robust optimization problems.



(a) Truncated Uniform



(b) Bivariate Gamma



(c) Mixed Gaussian

Fig. 3. Performance of the robust HLRP under different uncertainty sets

6- Case study

In this section, we delve into a comprehensive case study centered around the transportation network of a prominent retail store situated within the city of Tehran, which encompasses 22 distinct areas. Our objective is to subject the proposed model to rigorous scrutiny in four distinct configurations: the deterministic model, the robust model incorporating a box uncertainty set, the robust model utilizing a polyhedral uncertainty set, and finally, the robust model employing the data-driven SVC-based uncertainty set.

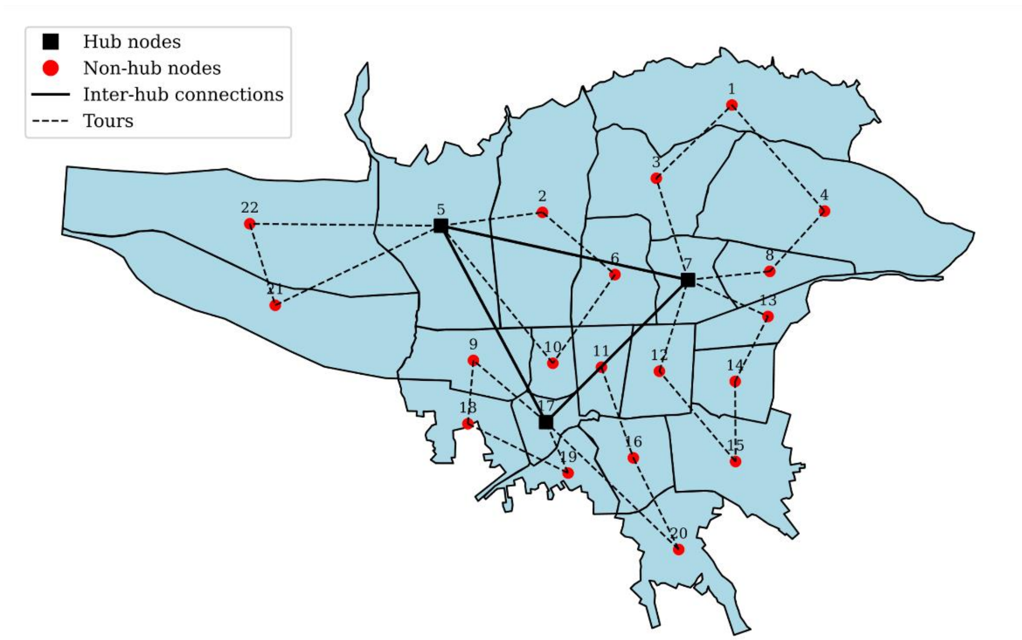
A critical facet of this case study lies in the pivotal data regarding t_{ij}, signifying the travel time between the 22 unique areas of Tehran. It is imperative to note that this data is inherently uncertain, introducing a crucial dimension to our analysis. To obtain this dataset, we harnessed Google Maps data and undertook a web scraping process utilizing the Google Maps API. The process involved several key steps:

1. **API Configuration**: We started by setting up the Google Maps API, configuring it to retrieve travel time data between specific geographic coordinates. This ensured that we could obtain precise information tailored to the 22 unique areas of Tehran.
2. **Coordinate Selection**: We identified the geographic coordinates corresponding to each of the 22 areas within Tehran. These coordinates were crucial for accurately querying travel time data from the API.
3. **Automated Querying**: We designed a script to automate the querying process. The script systematically sent requests to the API for travel time data between all possible pairs of the 22 areas. This automated approach streamlined the data collection process and minimized the potential for errors.
4. **Data Validation and Cleaning**: Upon retrieval, we implemented a validation process to ensure the integrity and reliability of the acquired data. This involved checking for outliers, anomalies, and inconsistencies. Any questionable data points were flagged for further review.

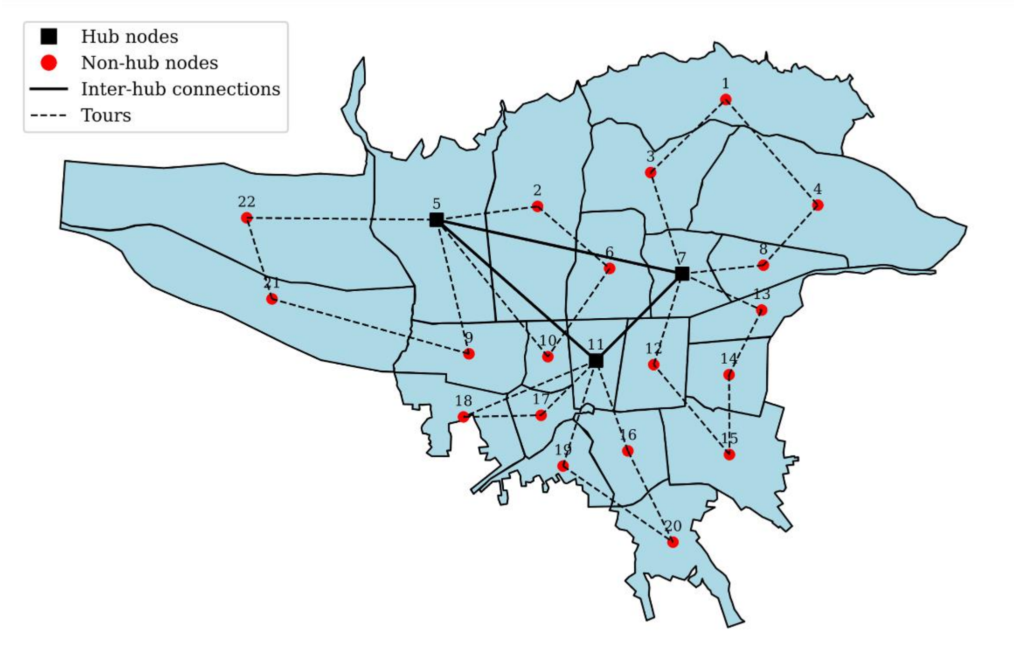
The resulting dataset comprises historical data spanning a 12-month period, capturing the travel times between each pair of the 22 areas. This comprehensive dataset forms the bedrock of our case study, facilitating a robust and thorough evaluation of the proposed model across varying uncertainty scenarios.
The transportation network derived from the robust models is established based on a protection level of 85%, a highly regarded threshold for conservative decision-makers. A comparative analysis between the deterministic and robust models reveals substantial disparities in hub locations, allocations, and vehicle routes.

In figure 4, we present the optimized transportation networks, encompassing hub locations, allocations, and vehicle routes, for each model. Notably, the robust models yield a more centralized network configuration, indicative of their enhanced resilience to uncertainties.

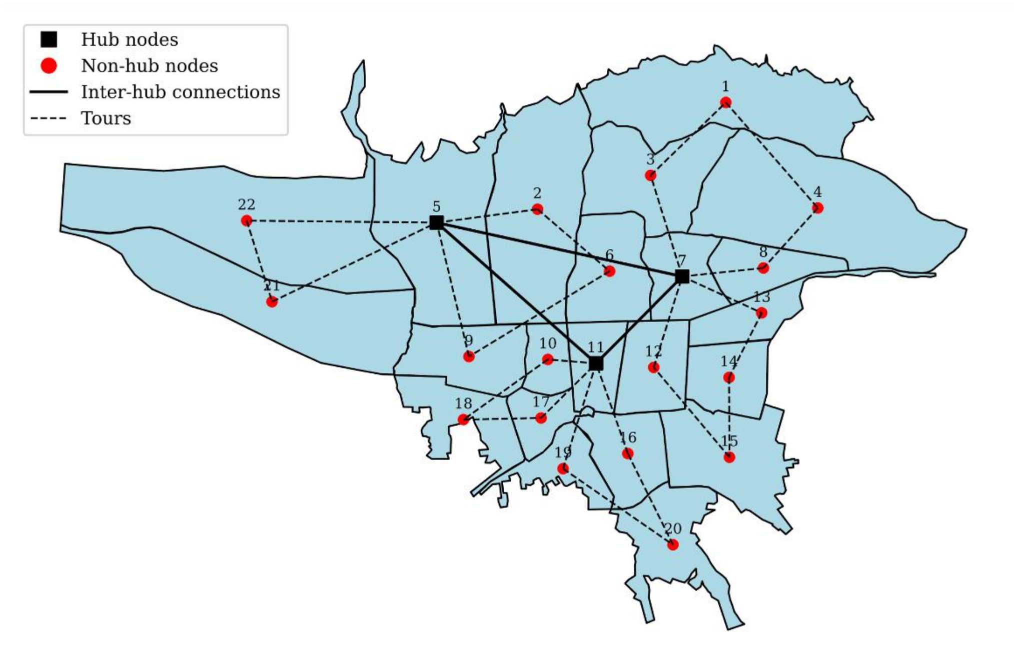
Upon evaluation, the SVC-based model demonstrates a 5.83% improvement in objective function compared to the polyhedral uncertainty set, underscoring its superior performance. Additionally, it exhibits a 2.41% enhancement in objective function when contrasted with the box uncertainty set, further solidifying its efficacy in addressing uncertainties within the transportation network.



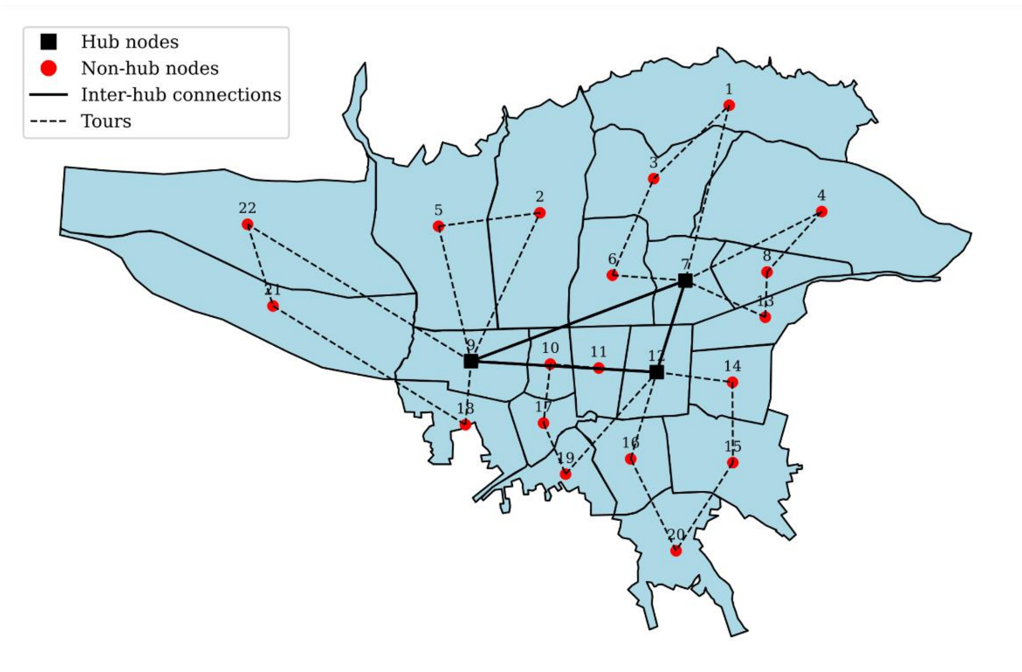
(a) Deterministic



(b) Box uncertainty set



(c) Polyhedral uncertainty set



(d) SVC-based uncertainty set

Fig. 4. Optimal transportation networks as determined by deterministic and robust models

Furthermore, the robust models exhibit a marked influence on the overall efficiency and cost-effectiveness of the transportation network. Specifically, they lead to a more streamlined allocation of resources, optimized vehicle routes, and strategically positioned hub locations, all of which collectively contribute to a more resilient and adaptable system.

In terms of practical implications, the centralized network structure resulting from the robust models may translate to improved resource utilization, reduced operational costs, and enhanced service quality. This can be particularly advantageous in scenarios where uncertainties in travel times between areas are prevalent, such as urban environments with fluctuating traffic conditions.

It's worth noting that the utilization of the SVC-based uncertainty set proves to be instrumental in achieving these improvements. By harnessing data-driven insights, this approach leverages real-world information to construct a more accurate representation of uncertainty, resulting in more reliable and effective transportation solutions.

The outcomes of our study carry substantial significance for real-world decision-making, particularly in the domain of HLRPs under uncertain environments. The implementation of robust models yields transformative effects on the efficiency and cost-effectiveness of transportation networks, fundamentally reshaping the way resources are allocated and vehicles navigate through the system.

Specifically, the optimization brought about by these models leads to a more strategic placement of hub locations. This, in turn, contributes to a centralized network structure, which has far-reaching practical implications. Firstly, it enhances resource utilization, ensuring that assets are deployed in a manner that maximizes their effectiveness. This streamlined allocation not only minimizes wastage but also optimizes the use of valuable resources, ultimately reducing operational costs.

Moreover, the optimized vehicle routes generated by the robust models have a twofold impact. On one hand, they lead to time savings, as routes are carefully selected to minimize travel times between areas. This not only improves the overall efficiency of operations but also translates to cost savings, particularly in contexts where time-sensitive deliveries or services are involved. On the other hand, these optimized routes contribute to a reduction in environmental impact, as fuel consumption and emissions are curtailed.

In urban environments characterized by unpredictable traffic conditions, the benefits are even more pronounced. The robust models provide a framework that can dynamically adapt to changing traffic patterns, ensuring that routes remain effective even as the urban landscape evolves. This adaptability is a crucial asset in real-world scenarios, where flexibility and responsiveness to changing conditions are imperative for success.

7- Conclusions

In conclusion, this study addresses the critical challenge of optimizing hub location-routing problems in the face of uncertain travel times. By introducing a data-driven robust optimization approach, we have demonstrated a significant advancement in tackling uncertainties inherent in transportation networks. Our findings showcase the superiority of the SVC-based uncertainty set in providing robust solutions, outperforming traditional uncertainty sets. The centralized network structures resulting from the robust models not only improve resource allocation and vehicle routes but also enhance overall system adaptability.

Furthermore, our case study on the transportation network of a retail store in Tehran has provided valuable insights into real-world applications. The comparison between deterministic and robust models reveals stark differences in hub locations, allocations, and vehicle routes. This highlights the necessity of considering uncertainties in decision-making processes, particularly in transportation and logistics scenarios. The SVC-based model, in particular, emerges as a standout performer, yielding significantly improved objective function values compared to both polyhedral and box uncertainty sets.

This research underscores the crucial role of data-driven robust optimization methodologies in addressing uncertainties in transportation networks. The outcomes of this study not only contribute to the theoretical advancements in the field but also hold practical significance for industries reliant on efficient logistics operations. We anticipate that our findings will inspire further exploration and adoption of robust optimization techniques in diverse applications, ultimately leading to more resilient and adaptable transportation systems.

Future research in this field could explore dynamic environments with evolving uncertainties, integrating real-time data updates and adaptive decision-making strategies. Additionally, investigating the integration of advanced machine learning techniques for more accurate uncertainty characterization holds immense potential. It is important to note that while our proposed methodology shows promising results, it may face challenges in scenarios with highly volatile and unpredictable uncertainties. Exploring multi-objective optimization frameworks, sustainability metrics, and scalability for larger and more complex networks could further advance practical implementation in metropolitan areas. It is noteworthy that the proposed model integrates two inherently complex and NP-Hard problems: hub location and vehicle routing. Solving large-scale instances of this combined problem can present significant computational challenges. While we have achieved promising results in our experiments, it is important to acknowledge that further research on efficient solution techniques, parallel computing, and heuristic algorithms tailored to our specific formulation could enhance scalability for real-world applications in even larger urban environments. These avenues promise to enhance the responsiveness and resilience of transportation networks.

References

Alumur, S. A., Campbell, J. F., Contreras, I., Kara, B. Y., Marianov, V., & O'Kelly, M. E. (2021). Perspectives on modeling hub location problems. *European Journal of Operational Research*, *291*(1), 1–17. https://doi.org/https://doi.org/10.1016/j.ejor.2020.09.039

Alumur, S., & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research*, *190*(1), 1–21. https://doi.org/https://doi.org/10.1016/j.ejor.2007.06.008

Asefi, H., Shahparvari, S., & Chhetri, P. (2019). Integrated Municipal Solid Waste Management under uncertainty: A tri-echelon city logistics and transportation context. *Sustainable Cities and Society*, *50*, 101606.

Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research*, *2*(Dec), 125–137.

Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, *167*, 235–292.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, *72*(2), 387–405.

Campbell, J. F., & O'Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science*, *46*(2), 153–169.

Catanzaro, D., Gourdin, E., Labbé, M., & Özsoy, F. A. (2011). A branch-and-cut algorithm for the partitioning-hub location-routing problem. *Computers & Operations Research*, *38*(2), 539–549. https://doi.org/https://doi.org/10.1016/j.cor.2010.07.014

Dai, X., Wang, X., He, R., Du, W., Zhong, W., Zhao, L., & Qian, F. (2020). Data-driven robust optimization for crude oil blending under uncertainty. *Computers & Chemical Engineering*, *136*, 106595.

Danach, K., Gelareh, S., & Neamatian Monemi, R. (2019). The capacitated single-allocation p-hub location routing problem: a Lagrangian relaxation and a hyper-heuristic approach. *EURO Journal on Transportation and Logistics*, *8*(5), 597–631. https://doi.org/https://doi.org/10.1007/s13676-019-00141-w

De Camargo, R. S., de Miranda, G., & Løkketangen, A. (2013). A new formulation and an exact approach for the many-to-many hub location-routing problem. *Applied Mathematical Modelling*, *37*(12), 7465–7480. https://doi.org/https://doi.org/10.1016/j.apm.2013.02.035

Ernst, A. T., & Krishnamoorthy, M. (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, *86*(0), 141–159. https://doi.org/10.1023/A:1018994432663

Ghaffarinasab, N., Van Woensel, T., & Minner, S. (2018). A continuous approximation approach to the planar hub location-routing problem: Modeling and solution algorithms. *Computers & Operations Research*, *100*, 140–154. https://doi.org/https://doi.org/10.1016/j.cor.2018.07.022

Gilani, H., Sahebi, H., & Pishvaee, M. S. (2022). A data-driven robust optimization model for integrated network design solar photovoltaic to micro grid. *Sustainable Energy, Grids and Networks*, *31*, 100714.

Goerigk, M., & Kurtz, J. (2023). Data-driven robust optimization using deep neural networks. *Computers & Operations Research*, *151*, 106087.

Gumte, K. M., Pantula, P. D., Miriyala, S. S., & Mitra, K. (2021). Data driven robust optimization for handling uncertainty in supply chain planning models. *Chemical Engineering Science*, *246*, 116889.

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A practical guide to support vector classification*. Taipei, Taiwan.

Inapakurthi, R. K., Pantula, P. D., Miriyala, S. S., & Mitra, K. (2020). Data driven robust optimization of grinding process under uncertainty. *Materials and Manufacturing Processes*, *35*(16), 1870–1876.

Jiang, J., Zhang, D., Meng, Q., & Liu, Y. (2020). Regional multimodal logistics network design considering demand uncertainty and CO2 emission reduction target: A system-optimization approach. *Journal of Cleaner Production*, *248*, 119304.

Karimi, H. (2018). The capacitated hub covering location-routing problem for simultaneous pickup and delivery systems. *Computers & Industrial Engineering*, *116*, 47–58. https://doi.org/https://doi.org/10.1016/j.cie.2017.12.020

Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research*, *53*(1), 25–37. https://doi.org/https://doi.org/10.1016/0377-2217(91)90090-I

Lopes, M. C., de Andrade, C. E., de Queiroz, T. A., Resende, M. G. C., & Miyazawa, F. K. (2016). Heuristics for a hub location-routing problem. *Networks*, *68*(1), 54–90.

Lotfi, R., Kargar, B., Gharehbaghi, A., Afshar, M., Rajabi, M. S., & Mardani, N. (2022). A data-driven robust optimization for multi-objective renewable energy location by considering risk. *Environment, Development and Sustainability*, 1–22.

Mohseni, S., & Pishvaee, M. S. (2020). Data-driven robust optimization for wastewater sludge-to-biodiesel supply chain design. *Computers & Industrial Engineering*, *139*, 105944.

Mohseni, S., Pishvaee, M. S., & Dashti, R. (2023). Privacy-preserving energy trading management in networked microgrids via data-driven robust optimization assisted by machine learning. *Sustainable Energy, Grids and Networks*, *34*, 101011.

O'kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, *32*(3), 393–404.

Ratli, M., Urošević, D., El Cadi, A. A., Brimberg, J., Mladenović, N., & Todosijević, R. (2022). An efficient heuristic for a hub location routing problem. *Optimization Letters*, 1–20.

Rodríguez-Martín, I., Salazar-González, J.-J., & Yaman, H. (2014). A branch-and-cut algorithm for the hub location and routing problem. *Computers & Operations Research*, *50*, 161–174. https://doi.org/https://doi.org/10.1016/j.cor.2014.04.014

Russell, D., Ruamsook, K., & Roso, V. (2020). Managing supply chain uncertainty by building flexibility in container port capacity: a logistics triad perspective and the COVID-19 case. *Maritime Economics & Logistics*, 1–22.

Shang, C., Huang, X., & You, F. (2017). Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering*, *106*, 464–479.

Shang, C., & You, F. (2019). A data-driven robust optimization approach to scenario-based stochastic model predictive control. *Journal of Process Control*, *75*, 24–39.

Shen, F., Zhao, L., Du, W., Zhong, W., & Qian, F. (2020). Large-scale industrial energy systems optimization under uncertainty: A data-driven robust optimization approach. *Applied Energy*, *259*, 114199.

Wu, Y., Qureshi, A. G., & Yamada, T. (2022). Adaptive large neighborhood decomposition search algorithm for multi-allocation hub location routing problem. *European Journal of Operational Research*, *302*(3), 1113–1127. https://doi.org/https://doi.org/10.1016/j.ejor.2022.02.002

Zhang, C., Wang, Z., & Wang, X. (2022). Machine learning-based data-driven robust optimization approach under uncertainty. *Journal of Process Control*, *115*, 1–11.

Zheng, Y., You, S., Li, X., Bindner, H. W., & Münster, M. (2022). Data-driven robust optimization for optimal scheduling of power to methanol. *Energy Conversion and Management*, *256*, 115338.