Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems 307
M. Shafik (Ed.)

doi:10.3233/ATDE240045

A Robust Optimization Method for Coordinated Optimization of Train Flow and Passenger Flow in Regional Rail Transit Systems

Jie WANG[a], Wei DONG[b,1], Xinya SUN[a] and Yindong JI[a]
[a] *Department of Automation, Tsinghua University, Beijing, China*
[b] *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China*

Abstract. Existing deterministic optimization methods suffer from decreased performance and increased risks under conditions of passenger flow uncertainty. To address these issues, a multi-scenario robust optimization method for coordinated optimization of train flow and passenger flow in regional rail transit systems is proposed. Firstly, samples of passenger travel demand, which are sampled from the potential passenger demand distribution, is used as multiple scenarios to characterize the uncertainty and diversity of passenger flow. Secondly, the mean-variance theory is employed as the foundation to establish the robust optimization model and the model's performance is discussed using variance or integration of deviations as indicator of robustness. Finally, a genetic algorithm is applied to solve the model, and the data from the Chongqing regional rail transit system is used as a case study for validation. Experimental results demonstrate that the proposed robust optimization model outperforms the deterministic model in uncertain conditions, providing better optimization performance.

Keywords. robust optimization, regional rail transit system, mean-variance model, coordinated optimization

1. Introduction

The regional rail transit system is a comprehensive rail transit system composed of various standard rail transportation modes within a city cluster or metropolitan area, serving the needs of the demands of regional economic integration. It is characterized by heterogeneity, integrity, interaction, and collaboration[1]. Regional rail transit typically includes different types (standards) of rail transit subsystems, such as high-speed or conventional railway systems, metro systems, monorail systems, maglev systems, and light rail transit (tram) systems.

Compared to single-mode transportation systems, large-scale integrated transportation systems like regional rail transit exhibit a larger scale, higher complexity, and relatively greater risks. With further socio-economic development, changes in

---

[1] Corresponding Author, Wei Dong, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China; E-mail: weidong@mail.tsinghua.edu.cn

population distribution, and evolving consumer demands, the current regional rail transit systems gradually fail to meet the public's increasing and diverse travel demands. To better meet this demand, we need to further enhance the performance of the optimization methods for regional rail transit systems.

In the practical operation of rail transit, passenger flow demand is a significant factor that needs to be considered. The actual passenger demand is an uncertain variable until it occurs. Most existing studies on rail transit system optimization do not consider the uncertainty of passenger demand and rely solely on deterministic historical data, which may result in decreased optimization performance and even pose risks when applied in scenarios with uncertain passenger flow. In this paper, the uncertainty of passenger flow demand is characterized through multi-scenario sampling, where multiple scenarios are generated by randomly sampling from assumed distributions based on historical data. This paper aims to establish a robust optimization model for regional rail transit by adopting the mean-variance theory to improve the robustness and practical application utility of the optimization model under uncertain passenger flow.

In summary, this paper contributes by using the mean-variance theory to establish a robust optimization model for train scheduling and passenger flow guidance in the regional rail transit system, addressing the impact of uncertain passenger flow and achieving better optimization performance than deterministic models.

Section 2 provides a comprehensive review of the relevant literature. Section 3 and Section 4 analyze the problem and establish the deterministic optimization model and then the robust optimization model. Section 5 validates the models using the Chongqing regional rail transit system as a case study. Finally, Section 6 presents a summary for this paper.

2. Literature Review

2.1. Uncertainty Optimization for Complex Engineering Systems

In the field of rail transit optimization, most of the literature focus on optimization under deterministic conditions[2]-[6]. The optimization models established considering system uncertainty are referred to as uncertainty optimization models. Existing methods for addressing uncertainty optimization problems mainly fall into three categories: traditional robust optimization, stochastic optimization, and distributionally robust optimization.

Traditional robust optimization typically addresses uncertainty by reformulating the model as a deterministic problem solvable by optimization solvers. This reformulation often involves the introduction of additional constraints or variables that capture the worst-case scenarios of the uncertainties, ensuring that the solution is feasible for all possible realizations within the predefined uncertainty set. The uncertainty is often described by a continuous random variable with no distribution. Valentina et al.[7] considered the uncertainty of passenger demand, established a robust optimization model with train timetables and train stopping plans as decision variables and train running time as the objective, and solved it using mixed-integer linear programming methods. Similarly, Qi[8] considered the uncertainty of passenger demand, established a robust optimization model with train timetables and train stopping plans as decision variables and transportation capacity risk as the objective, and solved it using mixed-integer linear programming methods.

In stochastic optimization methods, uncertainty is described by a unique distribution, and the optimization objective is the mean performance in a probabilistic sense. Gong et al.[9] considered the dynamic and stochastic nature of passenger demand, established a stochastic optimization model with the total number of service trains, headway settings, and speed curves as decision variables, and solved it using mixed-integer nonlinear programming and domain search algorithms. Dantzig[10] applied stochastic optimization to study modeling and solution of examples like minimum food consumption, factory shipment tasks, and three-stage problems. Kall et al.[11] proposed a method considering common stochastic variations of technical or economic parameters during the planning stage to obtain more reliable optimal solutions for engineering and economic problems.

Compared to traditional robust optimization methods, distributionally robust optimization methods use a fuzzy set containing multiple probability distributions to prevent the solutions from being overly conservative. The solutions in distributionally robust optimization consider the worst-case distributions within the fuzzy set without being excessively conservative. Lu et al.[12] considered passenger flow uncertainty, established a distributionally robust optimization model with train timetables and passenger flow control as decision variables, and used local search algorithms and mixed-integer linear programming methods to solve it. Qu et al.[13] took into account uncertainties in station and boarding passenger data, established a distributionally robust optimization model with train departure intervals and stopping timetables as decision variables, and solved it using nested heuristic genetic algorithms. Hao et al.[14] proposed a scenario-based distributionally robust optimization method for taxi pre-allocation models, addressing the problem of spatial and temporal mismatch between taxi supply and passenger demand. Cheng et al.[15] aimed to improve the efficiency of drone delivery by considering weather factors and other covariate information. They established a distributionally robust optimization model, modeling the drone's flight time as a fuzzy set, resulting in enhanced drone delivery efficiency.

Under the uncertainty conditions, most of the robust optimization methods mentioned above suffer from significant computational complexity, making it challenging to efficiently solve large-scale and highly constrained problems. When one incorporates coordinated optimization of both train flow and passenger flow within regional rail transit systems, the computational burden becomes even more pronounced.

2.2. The Railway Optimization Method Based on Multi-scenario Robust Optimization and Mean-variance Model

In large-scale and complex scenarios, robust optimization models built with distributions may become computationally intensive or difficult to solve accurately. The multi-scenario robust optimization method is a sampling-based optimization approach based on stochastic optimization, providing an approximation for dealing with uncertainties.

Several scholars have researched the multi-scenario robust optimization method for railway systems. Zhou et al.[16] investigated a robust optimization approach for high-speed railway train scheduling based on multiple demand scenarios. Each operational day's passenger demand is represented as a demand scenario, and the objective is to minimize the maximum regret value. They used train scheduling as decision variables and applied the min-max approach for optimization. Sun et al.[17] focused on the robustness optimization of railway network design under uncertain demand. They formulated three decision objectives: minimizing the total length of railway lines, minimizing the total passenger travel distance, and minimizing the total passenger transfer times. They established both the scenario model and min-max model and utilized a genetic algorithm for solving. Yang et al.[18] proposed an optimization model for the last train timetable based on the mean-variance theory. The model aimed to maximize the number of successfully transferring passengers and minimize the last train's running time. The departure schedule of the last train was considered as decision variables, and they employed a taboo search algorithm for optimization.

Based on the literature review above, it can be observed that although there have been a series of studies on robust optimization models for train scheduling and passenger flow control under uncertainty in transportation systems, there are hardly any studies that consider passenger guidance as an optimization variable, and although the computational load of multi-scenario robust optimization methods is relatively small, they still haven't been applied to the coordinated optimization of vehicle flow and passenger flow. Therefore, this paper attempts to establish a multi-scenario robust optimization model for train scheduling and passenger guidance in regional rail transit systems based on the mean-variance theory and uses an efficient genetic algorithm for solving the problem.

3. Problem Analysis and Model Construction

The deterministic model is based on the safety evaluation index system and optimization model structure proposed in reference[1]. Below, an explanation of the model is provided.

3.1. Parameter Description

The parameters used in this model are shown in Table 1[1].

Table 1. The parameter description of the optimization model.[1]

Parameter	Description
r_{ij}	The set of all paths for Origin-Destination (OD) pairs from Station i to Station j
r_{ij}^m	The m-th path of the OD pair from Station i to Station j
ω_k	Consequences of Interval k capacity risks
y_k	Passenger demand volume on Interval k (people per hour)
b_k	Transport capacity of Interval k
ω_p	Consequences of Station k capacity risks
C_p	The throughput at Station p (people per hour)
z_p	Passenger demand volume at Station p
δ_{ijk}^m	Whether r_{ij}^m passes through Station k: 1 if yes, 0 if no
λ_l	The seating capacity of the train on Line l
ξ_{lk}	Whether Interval k is on Line l: 1 if yes, 0 if no
T_l	One-way travel time on Line l (minutes)
t_{lmin}	Minimum train interval between consecutive departures of vehicles on Line l
t_{lmax}	Maximum train interval between consecutive departures of vehicles on Line l
n	Total number of vehicles in the rail network

q_{ij}	Passenger demand from Station i to Station j
θ_{ijp}^{m}	Whether Station p is the origin station, destination station, or transfer station of path r_{ij}^{m}: 1 if yes, 0 if no
d_{ij}^{m}	The length of the m-th path from the origin station i to the destination station j
d_{ijmin}	The minimum length of the paths from the origin station i to the destination station j
β	The proportion of allowable deviation from the shortest path length for every path.

3.2. Establishment of Deterministic Optimization Model

(1) Decision variables

The decision variables in this model can be mainly divided into two parts. The first part includes the departure interval of trains on each route and the number of vehicles on each route. The second part is the allocation proportions of each OD demand to different paths. The mathematical expression is shown in Table 2[1].

Table 2. The description of the decision variables.[1]

Parameter	Description
n_l	The number of vehicles on Line l
t_l	The train departure interval of Line l (minute)
x_{ij}^{m}	The proportion of passenger demand allocated to path r_{ij}^{m} out of the total demand to r_{ij}

(2) Optimization objective

The optimization objective of this model is the global transportation capacity risk of the regional rail transit system, which is calculated by summing up the capacity risks of all stations and intervals in the network, as shown in Eq. (1).

(3) Constraints

The model's constraints mainly include the constraints on the number of vehicles per line, the constraints on the minimum and maximum intervals between train departures on each line, the constraints on positive allocation proportions for the passenger demand on each path and the constraints on the difference between the shortest and longest paths to prevent significant variations in their lengths.

(4) Mathematical model[1]

The deterministic optimization model is formulated as Eqs. (1) to (10).

$$\min\ g(x_{ij}^{m}, t_l, n_l) = \sum_k \omega_k \times f(\frac{y_k}{b_k}) + \sum_p \omega_p \times f(\frac{z_p}{C_p}) \tag{1}$$

$$\text{s.t.}\ \ y_k = \sum_i \sum_j \sum_m x_{ij}^{m} \times q_{ij} \times \delta_{ijk}^{m} \quad i,j = 1,2,\dots,S \quad k = 1,2,\dots,K \tag{2}$$

$$b_k = \sum_l \frac{60\times\lambda_l}{t_l} \times \xi_{lk} \quad l = 1,2,\dots,W \tag{3}$$

$$t_l \ge \frac{T_l\times 2}{n_l} \tag{4}$$

$$t_{l\,max} \ge t_l \ge t_{l\,min} \tag{5}$$

$$\sum_l n_l \leq n \tag{6}$$

$$z_p = \sum_i \sum_j \sum_m x_{ij}^m \times q_{ij} \times \theta_{ijp}^m \quad p = 1,2,\dots,S \tag{7}$$

$$1 = \sum_m x_{ij}^m \tag{8}$$

$$0 \leq x_{ij}^m \leq 1 \tag{9}$$

$$\frac{d_{ij}^m - d_{ij\,min}}{d_{ij\,min}} \leq \beta \tag{10}$$

Eq. (1) represents the objective function, i.e., the global transportation capacity risk of the regional rail transit system. Here, the risk function is represented by $f(x) = \frac{1}{1+e^{-6x+7}}$, and the risk consequence is denoted as $\omega_k = \min(y_k, b_k)$, $\omega_p = \min(z_p, C_p)$[1]. Eq. (2) is used to calculate the passenger demand on a particular interval. Eq. (3) calculates the capacity of the interval. Eqs. (4) and (5) are constraints on the departure interval of trains. Eq. (6) ensures that the total number of vehicles in the network does not exceed the available number of vehicles. Eq. (7) calculates the passenger demand at each station. Eqs. (8) and (9) are constraints on passenger flow distribution. Eq. (10) represents the constraint on path length.

4. Robust Optimization Mathematical Model

Existing decisions in regional rail transit rely on historical data as a basis. However, in practical optimization, there is a trade-off between the robustness of the model and the degree of optimization of the objective. To enhance the robustness of the aforementioned deterministic optimization model, specifically to optimize the robustness of long-term scale decisions, this paper applies the mean-variance theory to establish a robust optimization model for train scheduling and passenger flow guidance to obtain solutions.

4.1. Uncertain Passenger Demand

This study adopts the method of multiple-scenario sampling to characterize the uncertainty in passenger demand in real-world scenarios. For the OD matrix, its three dimensions represent different time periods, origin stations (O), and destination stations (D). It is assumed that each element in the passenger OD demand matrix follows the same distribution. Multiple scenarios describing uncertain passenger demand are generated through random sampling. For instance, considering a historical OD demand data matrix element, assuming it follows a uniform distribution with a fluctuation of plus or minus thirty percent, multiple scenarios can be obtained by sampling from this distribution.

4.2. Mean-Variance Theory

The Mean-Variance Theory is a theory used to describe and analyze investment portfolios in financial markets. It is the core of modern portfolio theory, proposed by the American economist Harry Markowitz[19] in the 1950s.

According to this theory, the return of an investment portfolio can be calculated as the weighted average of the returns of various assets in the portfolio. The mean represents the average return, and the variance represents the volatility of returns. Therefore, the Mean-Variance Theory can help investors strike a balance between risk and return to achieve the optimal investment portfolio. This allows them to choose the portfolio that best fits their risk preferences and return requirements.

In this paper, the Mean-Variance Theory is utilized to balance robustness and efficiency.

4.3. Mean-Variance Theory Based Robust Optimization Model

For all N sampled scenarios, let's assume that the passenger demand OD matrix for scenario n is denoted as q^n, the probability of Scenario n occurring is denoted as $1/n$. Therefore, the objective function for Scenario n (i.e., the global transportation capacity risk under this scenario) can be transformed from equation (1) to equation (11), where for different Scenario n, the passenger demand volume on Interval k is represented as y_k^n, and the passenger demand volume on Station p is represented as $\mathrm{z}_\mathrm{p}^\mathrm{n}$.

$$\mathrm{g}^\mathrm{n}(x_{ij}^m, t_l, n_l) = \sum\nolimits_\mathrm{k} \omega_k \times f(\frac{y_k^n}{b_k}) + \sum\nolimits_\mathrm{p} \omega_p \times f(\frac{z_p^n}{C_p}) \tag{11}$$

Constraint (2) can be transformed into Constraint (12) and Constraint (7) can be transformed into Constraint (13).

$$y_k^n = \sum_i \sum_j \sum_m x_{ij}^m \times q_{ij}^n \times \delta_{ijk}^m \quad i,j = 1,2,\dots,S \quad k = 1,2,\dots,K \tag{12}$$

$$z_k^n = \sum_i \sum_j \sum_m x_{ij}^m \times q_{ij}^n \times \theta_{ijp}^m \quad p = 1,2,\dots,S \tag{13}$$

The overall robust optimization model's objective is composed of the mean and variance of N scenarios. The mean calculation is as Eq. (14):

$$\begin{aligned}
\mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l)) &= \mathrm{E}[\sum_\mathrm{k} \omega_k \times f(\frac{y_k}{b_k}) + \sum_\mathrm{p} \omega_p \times f(\frac{z_p}{C_p})] \\
&= \mathrm{E}[\sum_\mathrm{k} \omega_k \times f(\frac{y_k}{b_k})] + \mathrm{E}[\sum_\mathrm{p} \omega_p \times f(\frac{z_p}{C_p})] \\
&= \sum_{\mathrm{n}=1}^{\mathrm{N}} \frac{1}{\mathrm{N}} \sum_\mathrm{k} \omega_k \times f(\frac{y_k^n}{b_k}) + \sum_{\mathrm{n}=1}^{\mathrm{N}} \frac{1}{\mathrm{N}} \sum_\mathrm{p} \omega_p \times f(\frac{z_p^n}{C_p})
\end{aligned}$$

$$= \frac{1}{\mathrm{N}} \sum\nolimits_{\mathrm{n}=1}^{\mathrm{N}} (\sum\nolimits_\mathrm{k} \omega_k \times f(\frac{y_k^n}{b_k}) + \sum\nolimits_\mathrm{p} \omega_p \times f(\frac{z_p^n}{C_p})) \tag{14}$$

The variance calculation is as Eq. (15):

$$\mathrm{Var}(\mathrm{g}(x_{ij}^m, t_l, n_l)) = \mathrm{E}[(\mathrm{g}(x_{ij}^m, t_l, n_l) - \mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l)))^2]$$

$$= \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} ((\sum_{\mathrm{k}} \omega_k \times f(\frac{y_k^n}{b_k}) + \sum_{\mathrm{p}} \omega_p \times f(\frac{z_p^n}{C_p}) - \mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l)))^2)$$

$$= \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} ((\sum_{\mathrm{k}} \omega_k \times f(\frac{y_k^n}{b_k}) + \sum_{\mathrm{p}} \omega_p \times f(\frac{z_p^n}{C_p}) - \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\sum_{\mathrm{k}} \omega_k \times f(\frac{y_k^i}{b_k}) + \sum_{\mathrm{p}} \omega_p \times f(\frac{z_p^i}{C_p})))^2) \tag{15}$$

In the literature (e.g., [18]), the mean and the variance are often linearly combined into a single objective for optimization. However, in this paper, the mean and the variance are separated into two objectives, establishing a multi-objective robust optimization model, expressed mathematically as Eqs. (16) and (17). The reason for this separation is that addressing them as two distinct objectives provides a more intuitive understanding and representation compared to a single-objective approach.

$$\min \ \mathrm{g}_1(x_{ij}^m, t_l, n_l) = \mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l)) \tag{16}$$

$$\min \ \mathrm{g}_2(x_{ij}^m, t_l, n_l) = \mathrm{Var}(\mathrm{g}(x_{ij}^m, t_l, n_l)) \tag{17}$$

$$\mathrm{s.t.\,Eqs.}\,(12) - (13)$$

$$\mathrm{Eqs.}\,(3) - (6)$$

$$\mathrm{Eqs.}\,(8) - (10)$$

Taking into account the specificity of the problem and the directionality of the optimization objectives, since the main goal of the optimization model is to minimize the global capacity risk, it is acceptable for some scenarios to have risks smaller than the average risk. Therefore, only the deviation of scenarios with risks greater than the average risk needs to be focused on. As a result, the variance objective in Eq. (17) can be further optimized as a one-sided deviation integral, which is expressed as Eq. (18):

$$\min \mathrm{g}_3(x_{ij}^m, t_l, n_l) = \mathrm{E}[(\mathrm{g}(x_{ij}^m, t_l, n_l) - \mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l))) \times 1_{\mathrm{g}(x_{ij}^m, t_l, n_l) > \mathrm{E}(\mathrm{g}(x_{ij}^m, t_l, n_l))}] \tag{18}$$

The term $1_{(\cdot)}$ represents the indicator function, which outputs 1 when the input is true and 0 when the input is false. It can be observed that in the robust optimization model, due to the large problem scale, numerous constraints, and the complexity of the objective functions sampled from multiple scenarios, conventional optimization algorithms are not suitable for efficient solution.

In contrast, heuristic algorithms, such as genetic algorithms, are more suitable for such problems as they can often find relatively good solutions in a shorter time.

4.4. GA and NSGA-II

Genetic Algorithm is an evolutionary computation technique used to solve optimization problems based on natural selection and genetics. NSGA-II[20] is a multi-objective genetic algorithm proposed by Deb et al. in 2002. It can optimize multiple objective functions simultaneously and generates Pareto-optimal solutions. NSGA-II improves non-dominated sorting and diversity preservation. It uses a fast non-dominated sorting algorithm for classification and incorporates crowding distance to maintain diversity. NSGA-II improves convergence speed and accuracy through various measures.

5. Case Study and Result Analysis

The data for this numerical example is sourced from the experimental data presented in reference[1]. The implementation of the experimental algorithms and program development were conducted using Python 3.8 in this study.

5.1. Case Description of Chongqing Regional Rail Transit

This case study considers the Chongqing regional rail transit system, which includes nine rail lines. A set of OD passenger flow demand historical data that has been de-identified based on real data is used as input to the model. The operational period considered is from 18:00 to 21:00. For this case, only 17 OD pairs with significant passenger demand and multiple intermediate station paths are taken into account. The detailed data is listed in[1].

Since there is only one set of three-hour OD passenger flow demand data for the scenario, random sampling is employed to generate multiple scenarios of OD passenger flow demand data for robust optimization. Based on the description of passenger flow demand in Section 4.1, multiple scenarios are generated by randomly sampling from the assumed passenger flow distribution.

5.2. Design of Experiments

The proposed robust optimization model is compared with the deterministic model established in Section 3.2, which uses a single set of historical data as input to obtain a unique solution or decision.

In addition to comparing the results with the deterministic model, the robust optimization model is further analyzed through three categories of control variable experiments:

(1) Comparison of Different Robust Indicators: In measuring the robustness of the model, two indicators are used: variance and one-sided deviation integral. Variance considers the extent to which actual objectives deviate from the mean in both directions, while one-sided deviation integral only considers deviations in the direction of minimizing the objective.

(2) Comparison of Different Sampled Scenario Numbers: The number of sample scenarios used to generate the multi-scenario data for the robust optimization model input affects the performance of the model output and the computational time.

(3) Comparison of Different Distribution Assumptions: Since the actual OD demand is unknown, different distribution assumptions are considered to validate the model:

- Uniform Distribution: The passenger demand departing at time i from station j to station k in any scenario is assumed to follow a uniform distribution with the historical demand $M_{his}[i,j,k]$ as the midpoint, $0.7M_{his}[i,j,k]$ and $1.3M_{his}[i,j,k]$ as the upper and lower bounds, respectively. $M[i,j,k] \sim U(0.7M_{his}[i,j,k], 1.3M_{his}[i,j,k])$, which means that the passenger demand fluctuates within $\pm 30\%$ of the historical data.
- Normal Distribution: For the passenger demand departing at time i from station j to station k in any scenario, referring to the 3-sigma rule, assume $M[i,j,k]$ is likely to fall between $[0.7M_{his}[i,j,k], 1.3M_{his}[i,j,k]]$. Thus, the demand for any scenario is assumed to follow a normal distribution with the historical demand $M_{his}[i,j,k]$ as the mean and $0.1M_{his}[i,j,k]$ as the standard deviation, $M[i,j,k] \sim N(M_{his}[i,j,k], (0.1M_{his}[i,j,k])^2)$.
- Poisson Distribution: For any scenario, assume the passenger demand departing at time i from station j to station k follow a Poisson distribution with the historical demand $M_{his}[i,j,k]$ as the mean, $M[i,j,k] \sim P(M_{his}[i,j,k])$.

5.3. Experimental Results and Robustness Analysis

For each distribution assumption, the robustness analysis is performed by using a randomly sampled test set of 50 scenarios, following the sampling method described in Section 4.1. The objective is to validate the solutions obtained from the optimization model and assess the transportation capacity risk magnitude of the proposed schemes on the test set, as well as the variance across the 50 scenarios.

(1) Comparison between Deterministic Model and Robust Optimization Model

The deterministic model takes as input a single set of historical OD passenger demand data, representing the actual passenger flow demand within a three-hour time segment. On the other hand, the input for the robust optimization model consists of 50 sets of OD passenger demand data, randomly sampled from a uniform distribution. Each set of data also covers a three-hour duration. The robust optimization model is established using mean and variance metrics and is referred to as the baseline robust optimization model.

For both the deterministic model and the robust optimization model, the algorithm is employed with 3000 iterations. The encoding method is real-integer encoding, with a population size of 5000. The simulated binary crossover operator and polynomial mutation operator are used.

In one of the experiments, the results of the passenger flow distribution obtained by the genetic algorithm for the deterministic model are shown in Table 3, while the results for the train interval and the number of vehicles are presented in Table 4. The global capacity risk value for the non-optimized approach is 539421.38. The genetic algorithm's solution yielded an capacity risk value of 471701.77, resulting in a reduction of 12.55% compared to the non-optimized approach.

Table 3. The passenger flow allocation results obtained by the deterministic model.

OD Number	Path1	Path2	Path3	Path4	Path5
1	35.09%	12.94%	28.26%	11.89%	11.82%
2	39.81%	2.83%	0.00%	56.56%	0.80%
3	31.24%	2.24%	24.32%	13.44%	28.75%
4	0.00%	39.50%	38.26%	21.82%	0.42%
5	0.00%	0.00%	25.89%	38.21%	35.90%

6	0.00%	56.94%	1.18%	3.14%	38.74%
7	100.00%	-	-	-	-
8	100.00%	-	-	-	-
9	48.97%	6.56%	40.33%	4.14%	0.00%
10	49.40%	6.06%	12.27%	32.27%	-
11	18.81%	16.99%	0.00%	21.24%	42.95%
12	72.05%	26.71%	12.40%	-	-
13	31.29%	0.00%	37.80%	30.91%	0.00%
14	0.00%	34.71%	26.97%	37.38%	0.93%
15	11.91%	0.00%	35.09%	0.00%	53.00%
16	39.52%	0.00%	23.24%	29.58%	7.67%
17	49.15%	0.01%	0.00%	50.85%	0.00%

Table 4. The train departure interval and train amount results obtained by the deterministic model.

Line Name	Departure Interval	Train Amount
Chengdu-Chongqing High-Speed Railway	15min	32
Line 1	6min22s	40
Line 2	2min	66
Line 3	3min3s	65
Line 5	10min	18
Line 6	4min43s	46
Line 10	8min35s	15
the Loop Line	5min6s	24
Chongqing-Wanzhou Suburban Railway	50min	15

The robustness analysis results of the deterministic model solutions and the solutions on the Pareto front obtained by applying the NSGA-II algorithm to solve the robust optimization model are presented in Figure 1. The red points represent the robustness analysis results of the unique solution obtained from the deterministic model, while the blue points represent the robustness analysis results of the solutions on the Pareto front obtained by the robust optimization model. It can be observed that there exist solutions on the Pareto front that perform better than the deterministic model in terms of both the mean and variance of the testing results. Moreover, all the solutions on the Pareto front exhibit better performance in terms of testing variance than the deterministic model. This indicates that these solutions not only achieve lower capacity risk but also possess stronger robustness in practical applications.

One of the solutions on the Pareto front is highlighted in Figure 1 (red circled), and its variable values are shown in Tables 5 and 6. The comparison between this solution and the deterministic model in terms of objective function values for the 50 testing scenarios is presented in Figure 2. It can be observed that the robust optimization model achieves a lower average level and a smaller fluctuation compared to the deterministic model, indicating a superior performance.

(2) Comparison between Different Robust Metrics

In this section, the experiment adopts the one-sided deviation integral as the robustness metric and compares the results with those using variance. The other parameters of the algorithm remain the same as before. The obtained results are shown in Figure 3. The light blue Pareto front represents the results obtained by using the one-sided deviation integral as the robustness metric in the robust optimization model. It can be observed that most of the points in the light blue set are better than the blue set, indicating that the solutions obtained by using the one-sided deviation integral as the robustness metric generally perform better in the test.

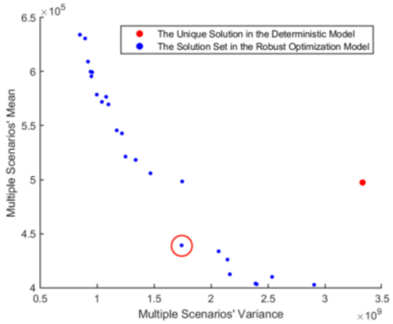

Figure 1. The robustness analysis of the robust optimization model sloved by NSGA-II with uniform distribution assumption and 50 sampled scenarios.

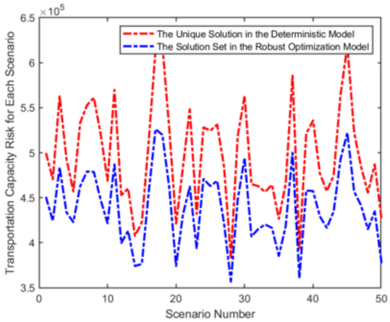

Figure 2. Comparison of robustness analysis results for one of the solution obtained from the robust optimization model with uniform distribution assumption and the deterministic model solution across multiple scenarios.

Table 5. One of the passenger flow allocation results obtained by the robust model.

OD Number	Path1	Path2	Path3	Path4	Path5
1	46.24%	0.70%	8.91%	43.46%	0.68%
2	47.38%	10.66%	0.02%	32.23%	9.71%
3	21.32%	13.50%	19.88%	40.61%	4.69%
4	5.41%	7.27%	16.05%	39.39%	31.88%
5	38.63%	7.27%	3.13%	8.77%	42.20%
6	25.53%	25.55%	0.03%	0.45%	48.44%
7	100.00%	-	-	-	-
8	100.00%	-	-	-	-
9	38.21%	22.03%	6.15%	23.67%	9.94%
10	44.49%	4.78%	44.71%	6.02%	-
11	21.70%	13.72%	14.59%	9.29%	40.70%
12	76.68%	22.49%	0.83%	-	-
13	29.93%	33.01%	27.66%	8.16%	1.25%
14	23.13%	25.67%	17.89%	0.00%	33.31%
15	23.89%	6.69%	37.83%	25.73%	5.86%
16	32.67%	14.34%	19.42%	31.52%	2.05%
17	6.86%	9.50%	29.90%	49.61%	4.13%

Table 6. One of the train departure interval and train amount results obtained by the robust model.

Line Name	Departure Interval	Train Amount
Chengdu-Chongqing High-Speed Railway	15min	32
Line 1	2min39s	45
Line 2	3min37s	35
Line 3	2min3s	100
Line 5	6min54s	11
Line 6	8min12s	23
Line 10	6min3s	23
the Loop Line	7min41s	41
Chongqing-Wanzhou Suburban Railway	50min	15

(3) Comparison between Different Sampled Scenario Numbers

This section conducted experiments with additional settings of 25, 100 and 150 sampling scenarios, in addition to the original 50 sampling scenarios used before. The other parameters of the algorithm remain the same as before. The results obtained are presented in Figure 4. From the figure, it can be observed that the solutions obtained with 50 and 100 sampling scenarios exhibit better dominance compared to using only 25 or 150 sampling scenarios. Using too few sampling scenarios can lead to underfitting, while using too many can result in overfitting, both of which lead to suboptimal performance in the optimization process.

The comparison in Figure 4 reveals that the optimal number of sampling scenarios lies between the extremes of 25 and 150, and utilizing 50 or 100 sampling scenarios seems to strike the right balance between model complexity and robustness. These findings emphasize the importance of selecting an appropriate number of sampling scenarios to achieve an effective and robust optimization result.

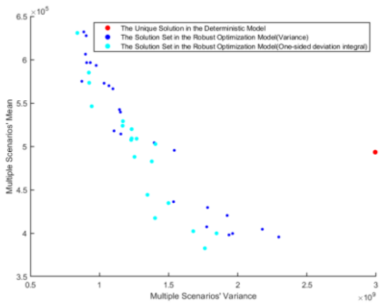

Figure 3. Robustness analysis and comparison between two different robust metrics.

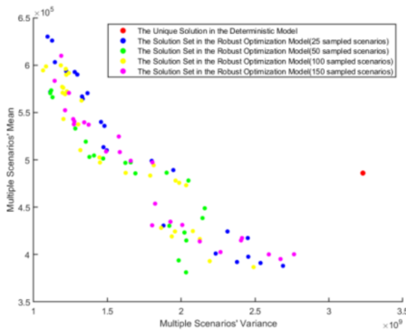

Figure 4. Robustness analysis and comparison among different number of sampled scenarios.

(4) Validation of the Robust Optimization Model under Different Distribution Assumptions

For the assumption of passenger OD demand distribution, this section validates the effectiveness of the robust optimization model using normal and poisson distributions, in addition to the uniform distribution. The robustness analysis results for the normal and poisson distributions are shown in Figures 5-6 and Figures 7-8, respectively. Both the solution and testing phases employ the same distribution assumption.

From the results, it can be observed that under the assumptions of normal and poisson distributions, the robust optimization model presented in this study, solved using a genetic algorithm with the same number of iterations and other parameters, can still obtain solutions that outperform the deterministic model in terms of average performance and robustness.

6. Conclusions

In conclusion, this research introduces a multi-scenario robust optimization method for coordinating train flow and passenger flow in regional rail transit systems. By considering passenger flow uncertainty through multiple sampled scenarios, our approach effectively captures diversity and uncertainty in demand.The robust optimization model, based on the mean-variance theory, is established, and a genetic algorithm efficiently solves it. Experimental results using the Chongqing regional rail transit system validate our model's superiority over deterministic methods under uncertain conditions, providing improved optimization performance and risk reduction.

Overall, this study has achieved certain results, yet there are areas for improvement. Specifically, two main directions for further research enhancement are identified . First, the passenger flow data considered in this paper is static OD data. Future research should delve into dynamic passenger flow data. Second, the robust optimization method used in this paper simplifies the problem. While this method reduces computational effort, it does come at the expense of precision. Subsequent studies should explore more superior robust optimization methods.

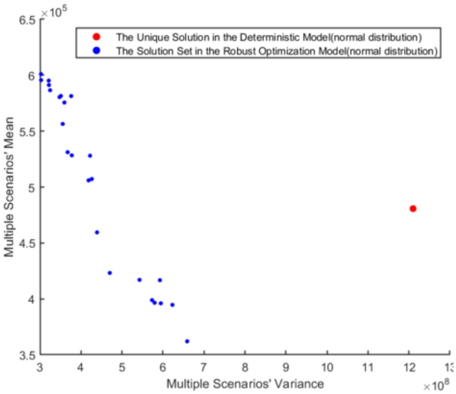

Figure 5. The robustness analysis of the robust optimization model sloved by NSGA-II with normal distribution assumption and 50 sampled scenarios.

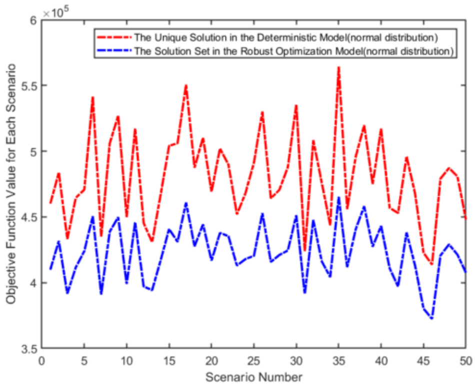

Figure 6. Comparison of robustness analysis results for one of the solution obtained from the robust optimization model with normal distribution assumption and the deterministic model solution across multiple scenarios.

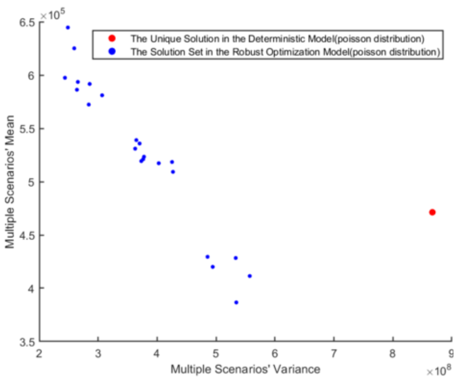

Figure 7. The robustness analysis of the robust optimization model sloved by NSGA-II with poisson distribution assumption and 50 sampled scenarios.

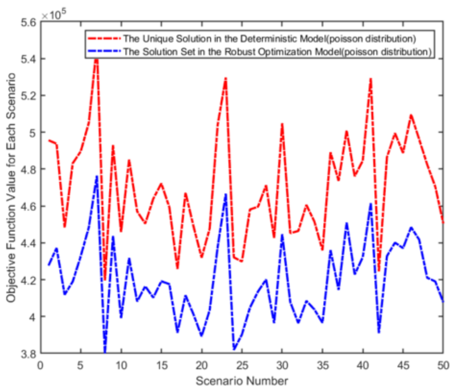

Figure 8. Comparison of robustness analysis results for one of the solution obtained from the robust optimization model with poisson distribution assumption and the deterministic model solution across multiple scenarios.

Acknowledgement

This work was supported by the National Key Research and Development Program of China under Grant 2022YFB4300502, the Research and Development Project of Qingdao National Innovation Center of High Speed Train under Grant CX/KJ-2020-0006 and the Research and Development Project of CRSC Research & Design Institute Group Co., Ltd.

References

[1] Mengyu Z. Research on Global RAMS Evaluation and Enhancement Method of Regional Rail Transit[Master's thesis]. Tsinghua University; 2020.
[2] Pan H, Yang L, Liang Z. Demand-oriented integration optimization of train timetabling and rolling stock circulation planning with flexible train compositions A column-generation-based approach[J]. European Journal of Operational Research, 2023, 305(1)184-206.
[3] Wang Y, Andrea D'Ariano, Tang B N, et al. Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line[J]. Transportation Research Part B, 2018(118).
[4] Xu X, Li K, Li X. A multi-objective subway timetable optimization approach with minimum passenger time and energy consumption[J]. Journal of Advanced Transportation, 2015, 50(1):69-95.
[5] Su S, Tang T, Li X, et al. Optimization of Multitrain Operations in a Subway System[J]. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2014, 15(2):673-684.
[6] Dechang Z. Key Technology Research on Collaborative Optimization and Simulation Platform of New Urban Rail Transit System[Master's thesis]. Tsinghua University; 2022.
[7] Cacchiani V, Qi J, Yang L. Robust optimization models for integrated train stop planning and timetabling with passenger demand uncertainty[J]. Transportation Research Part B, 2020(136):1-29.
[8] Qi J. Robust Train Timetabling and Stop Planning with Uncertain Passenger Demand[J]. Electronic Notes in Discrete Mathematics, 2018(69):213-220.
[9] Gong C, Shi J, Wang Y, et al. Train timetabling with dynamic and random passenger demand: A stochastic optimization method[J]. Transportation Research Part C, 2021(123).
[10] Dantzig, George B. Linear programming under uncertainty[J]. ManagementScience, 1955, 1(3-4):197-206.
[11] Kall, Peter, Wallace S W. Stochastic programming[M]. 2. Springer, 1994.
[12] Lu Y, Yang L, Yang K, et al. A Distributionally Robust Optimization Method for Passenger Flow Control Strategy and Train Scheduling on an Urban Rail Transit Line[J]. Engineering, 2022(12):202-220.
[13] Qu Y, Wang H, Wu J, et al. Robust optimization of train timetable and energy efficiency in urban rail transit: A two-stage approach[J]. Computers & Industrial Engineering, 2020(146).
[14] Hao, Zhaowei, Long He, Zhenyu Hu, Jun Jiang. Robust vehicle pre-allocation with uncertain covariates[J]. Production and Operations Management, 2020, 29(4): 955–972.
[15] Chun Cheng, Yossiri Adulyasak, Louis-Martin Rousseau, Melvyn Sim. Robust Drone Delivery with Weather Information[EB/OL]. 2020. https://optimization-online.org/?p=16564.
[16] Zhou W, Kang J, Qin J, et al. Robust Optimization of High-Speed Railway Train Plan Based on Multiple Demand Scenarios[J]. Mathematics, 2022(10).
[17] Sun, Y, Liu, X, Jing, C, et al. Robust Optimization of Rail Transit Network under Uncertain Demand[J]. Journal of Transportation Systems Engineering and Information Technology. 2015; 15(4): 181-186.
[18] Yang S, Yang K, Gao Z, et al. Last-Train Timetabling under Transfer Demand Uncertainty: Mean-Variance Model and Heuristic Solution[J]. Journal of Advanced Transportation, 2017.
[19] Markowitz H. Portfolio Selection[J]. The Journal of Finance, 1952(7):77-91.
[20] Deb K, Pratap A, Agarwal T M S, et al. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II[J]. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 2002, 6(2):182-197.