

Optimizing Rural Area Tourism Routes Using Structural Equation Modeling

Fenglian Liu

Associate Professor, College of Culture and Tourism, Urumqi Vocational University, Urumqi, 830002, China, E-mail: liufenglian2025@163.com

Project Management

Received February 10, 2026; revised April 17, 2026; September 29, 2026; accepted October 6, 2026
Available online October 20, 2026

Abstract: With the advancement of the rural revitalization strategy, rural tourism has emerged as a significant growth point for regional economies. However, current tour routes predominantly rely on empirical planning, making it difficult to meet tourists' demands for personalized and immersive experiences. To enhance the rural tourism experience and improve tour efficiency, this study is based on 421 valid tourist questionnaires collected from typical rural areas in Shandong Province, China. The study first constructs an evaluation model for the importance of rural tourism resources. This model quantifies spatial characteristics of resources across multiple dimensions and calculates node centrality using social network analysis. Subsequently, the study employs a structural equation model to analyze the impact of latent variables, such as personal attributes, on tourists' travel choices, thereby forming relevant mechanisms. A double-layer Nested Logit model is introduced, in which the upper layer selects activity chain types, and the lower layer selects travel modes, with parameters estimated via a phased approach. Finally, a dual-objective model integrating landscape service levels and travel time is constructed. This model takes node weights and mode shares as inputs and uses a genetic algorithm to find optimal solutions, yielding a generation model for rural landscape tour-route networks. The results indicate that travel preferences have the most direct impact on tourists' choices, with a path coefficient of 0.55, followed by transportation attributes and tourism activity attributes. Significant behavioral differences are observed among different groups. For instance, the younger demographic prefers immersive activity chains, accounting for 42%, while family travelers rely heavily on public transportation, at 65%. Dual-objective optimization enhances the overall efficiency of the express-arterial-branch road network by 28%. This study couples the objective characteristics of resources with tourists' subjective behaviors, providing a data-driven quantitative tool for rural tourism route planning.

Keywords: Rural tourism, structural equation modeling, genetic algorithm, nested logit, objective optimization.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).
DOI 10.32738/JEPPM-2026-227

1. Introduction

Rural landscapes, with their unique blend of nature and culture, meet the growing demand for tourism. The comprehensive implementation of the rural revitalization strategy has made rural tourism an important driver of regional economic growth. Additionally, the rapid development of rural tourism has placed higher demands on transportation network planning. However, the current rural landscape tourism routes rely heavily on experiential planning, with serious homogenization. The contradiction between tourists' demand for personalized and immersive experiences and the supply of scenic facilities is becoming increasingly prominent (Lin et al., 2024). With the surge of online Position of Interest (POI) related data, it has become difficult for humans to efficiently process massive amounts of information, and tourism research has turned to computer-automated processing (Bhosle and Musande, 2023). Scholars at home and abroad have made many research achievements in the fields of network structure, development models, and technologies of rural tourism. For example, Qin et al. (2022) established a rural tourism flow network analysis framework using Ctrip's digital tourist footprint data to study spatial patterns. The research found the network had low overall density, pronounced node centrality stratification, significant structural holes, and a distinct core-periphery structure. Internally, it exhibited development patterns centered on core nodes, linear extension along major transport routes, and clustered regional development. Chen et al. (2024) proposed a self-service tourism route-planning system using a hybrid genetic algorithm based on the Traveling Salesman Problem (TSP). Their research found that the algorithm performs more route crossover operations and yields better optimization outcomes after multiple iterations, particularly in scenarios with numerous urban attractions and large distances.

Additionally, several experts have proposed key methods for sustainable development and innovative practices in rural tourism. Li (2022) constructed a SWOT (Strengths, Weaknesses, Opportunities, Threats) structural analysis model for e-

commerce tourism strategy based on big data and applied the Analytic Hierarchy Process (AHP) to prioritize the SO, ST, WO, and WT strategic sequences for green development. Harbiankova and Gertsberg (2022) proposed a rural sustainable development information model using systematic methods, comparative analysis, and mathematical modeling to monitor sustainability at the planning level. Wang et al. (2023) introduced an immersive tourism marketing model that leverages new media and VR and AR technologies via platforms such as Weibo and WeChat to enhance brand influence. Tsigdinos et al. (2024) also examined flexible public transport in rural Greece through quantitative social research and spatial analysis, identifying cost and travel time as key factors and proposing eight recommended routes to guide adaptive transport planning. Pei et al. (2022) addressed the lack of geographical perspective, insufficient resource integration, and time-cost trade-offs in tourism route optimization by integrating the ArcGIS accessibility evaluation model with the Vehicle Routing Problem (VRP) in network analysis, producing accessibility results for Lushunqiu District and optimized one and two-day tour routes to aid governments and travel agencies. Yin and Li (2025) addressed the ineffectiveness of traditional rural tourism route optimization by combining an improved genetic algorithm with Ant Colony Optimization (ACO), achieving average absolute errors between 0.784 and 0.801 and a maximum error rate of 0.56% compared with other methods, thereby enhancing optimization accuracy. Li et al. (2025) addressed uneven accessibility between urban and rural tourism resources in Wuhan's urban fringe using a "quantity-location-capacity" principle with location-configuration and algorithmic models, revealing higher suburban accessibility, identifying eight potential new tourism sites after accounting for optimization costs, and improving overall efficiency and fairness.

Traditional methods are largely based on static road networks or single-objective optimization. They overlook the dynamic balance among tourists' landscape aesthetics, cultural identity, and time costs throughout the tourism process, thereby making it difficult to balance experience quality and spatial efficiency in route design. In addition, existing models are insufficiently responsive to the heterogeneity and distributional characteristics of rural landscapes and lack an effective framework for coupling tourists' subjective perceptions with objective path structures. Structural Equation Modeling (SEM) provides a new approach to constructing an integrated circuit generation mechanism for the "perception behavior path" by revealing complex causal relationships among latent variables. Therefore, this study addresses the following research questions: RQ1 (What): What are the key latent factors and causal paths that determine tourists' travel mode choices in rural tourism? RQ2 (What): What activity chains and travel modes do different tourist groups prefer, and how can these preferences be quantified? RQ3 (How): How can node weights and behavior shares be integrated into a dual-objective model that maximizes landscape service level and minimizes travel time? RQ4 (How): How does the proposed model perform compared with a travel-time-minimizing baseline model? Based on behavioral theory, the study proposes the following hypothesis, tested using SEM: H1: Tourists' travel preferences have a direct positive effect on transportation mode choice. H2: Individual attributes indirectly affect transportation mode choice through travel preferences. H3: Tourism activity attributes have a direct positive effect on transportation attributes. The theoretical framework of this study is mainly based on the "attitude-behavior" theory in travel behavior research and the "random utility maximization" framework in destination selection. The former provides a conceptual basis for the latent variables and their influencing paths in the SEM, to explain how psychological factors shape decisions; the latter provides the core logic for the Nested Logit model (NL) to simulate the hierarchical choice behavior of tourists in the "activity chain and travel mode", that is, it assumes that tourists tend to choose the option combination with the highest perceived utility. To this end, a bidirectional driving SEM that integrates resource and demand is developed to incorporate tourists' activity chains and travel mode choices into the path algorithm, thereby optimizing landscape and time. The research further introduces a double-layer NL model and embeds a joint selection mechanism of "activity chain type travel mode" to characterize tourist decision-making behavior. Using node weights and sharing rates as inputs, a dual objective optimization model is constructed, and a genetic algorithm is used to solve it. Finally, a tourism route network generation model that meets tourist preferences is constructed. The innovation of this study lies in integrating mature methods such as SEM, NL models, Social Network Analysis (SNA), and genetic algorithms into a unified "resource evaluation, behavior analysis, and path generation" collaborative framework for the first time. This provides methodological support for systematically resolving the complex relationship between tourists' subjective preferences and the optimization of objective paths. The research aims to develop a suitable method for planning rural landscape route networks based on tourists' travel characteristics to optimize service levels and travel efficiency.

2. Methods and Materials

This study employed a stratified random sampling method and conducted a questionnaire survey in the key rural tourism counties of Shandong Province from May to August 2023. Data collection was carried out through a combination of on-site interviews and online platforms. A total of 456 questionnaires were collected. After excluding those with incomplete answers or obvious contradictions, 421 valid samples were obtained, with an effective recovery rate of 92.3%. The samples covered tourists of different ages, incomes, and family structures to support the robustness of the subsequent behavioral analysis. The questionnaire was divided into four parts: The first part covered the social and economic attributes of the individual and the family; the second part used the Likert five-point scale to measure tourists' preferences and perceptions regarding travel costs, comfort, time, etc.; the third part collected the activity chain, type of scenic spots and travel mode of the respondents' most recent rural tourism experience; the fourth part was an objective evaluation of the transportation attributes. After the pre-test, the questionnaire's reliability was assessed using validity and reliability tests (Cronbach's $\alpha > 0.8$), confirming the scale's reliability. The factor loadings reported in this study were all based on the 421 valid tourist questionnaires collected in Shandong Province. They were calculated through the measurement model part of the SEM.

2.1. Development of Importance Evaluation Model for Rural Tourism Resources

The development of a rural landscape tourism route model integrates cultural heritage and community participation, serving as both a technical exploration and a deeper inquiry into rural socio-cultural values. Due to the heterogeneous distribution of resources, an evaluation model incorporating multi-dimensional spatial features and SNA is used to determine node

centrality and weights. Furthermore, tourism and transportation in rural areas constitute a deeply coupled system that evolves dynamically in response to tourist demand (Li et al., 2024). Fig. 1 shows the interaction mechanism between rural landscape tourism and transportation.

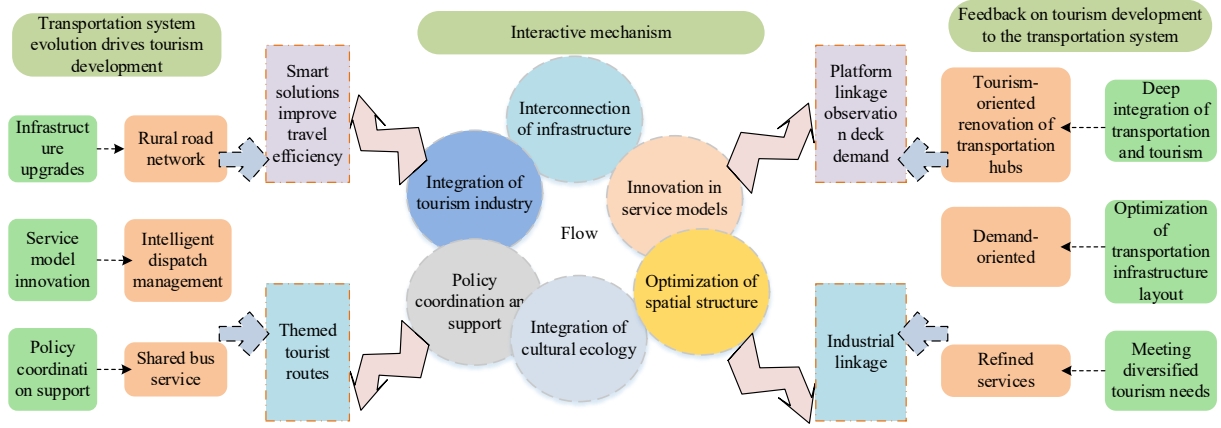


Fig. 1. Interaction mechanism between rural landscape tourism and transportation

In Fig. 1, the evolution of the transportation system is the core driver of tourism development, enhancing the accessibility of rural attractions, shortening spatiotemporal distances and connecting dispersed landscapes into tourism routes through improved road and bus networks. Optimized traffic management, like flexible scheduling and landscape road construction, further activates rural tourism resources and upgrades supporting facilities. Conversely, rural tourism development promotes upgrades to the transportation system, and personalized demands spur the development of refined services, such as customized public transportation and scenic micro-circulation routes. Increased passenger flow drives the expansion of facilities such as parking lots and transfer hubs, and landscape transformations, such as cycling greenways, integrate transport into the tourism experience. The vast rural areas and abundant ecological resources nurture unique tourism characteristics, with resources transformed into a multi-faceted product system through policy drivers (Qian and Tian, 2024). Spatially, natural resources cluster by terrain, while human resources are scattered. Insufficient transport-tourism linkages affect development patterns, and spatial clustering can be analyzed using methods such as the nearest neighbor index and kernel density analysis (Zhang et al., 2023). The nearest neighbor distance index is commonly used to evaluate the spatial distribution of point features, which calculates the average distance $\bar{d}_0$ between scenic spots within a region and their nearest neighbors. Compared with the expected distance d_e of a random distribution, the ratio is the nearest neighbor index, which is used to determine the distribution type. The calculation process is shown in Eq. (1).

$$\begin{cases} r = \frac{\bar{d}_0}{d_e} \\ \bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_i \\ d_e = \frac{1}{2} \sqrt{\frac{n}{s}} \end{cases} \quad (1)$$

In Eq. (1), r represents the degree of deviation between the measured and expected average distance of the scenic spot. n is the number of scenic spots. s is the area of the region. $r > 1$ denotes the clustered distribution, $r = 1$ for random distribution, and $r < 1$ for uniform distribution. Kernel density analysis is used to measure local density changes and explore popular areas by covering smooth surfaces above each scenic spot to form a circle with a search radius. The surface value decreases with increasing distance (Tang and Huang, 2023). The calculation formula for nuclear density is shown in Eq. (2).

$$f(x) = \frac{1}{nl^2} \sum_{i=1}^n \sigma\left(\frac{x - x_i}{l}\right) \quad (2)$$

In Eq. (2), the smaller $f(x)$, the sparser the points. σ represents the kernel function. $x - x_i$ is the distance from the measurement point to the prediction point. l indicates the prediction scale of kernel density. The Gini coefficient is calculated by the Lorenz curve to assess the equilibrium distribution of regional scenic spots: the cumulative function is calculated by arranging the number of scenic spots in ascending order, and the proportion of the area enclosed by the curve and the absolute equilibrium line is the Gini coefficient value (G) (Lucio, 2023). $G \in [0, 1]$, the larger the value, the higher the concentration of scenic spots and the greater the regional differences. $G < 0.2$ means absolute equilibrium, 0.2 to 0.4 indicates relative balance, and > 0.6 indicates a significant disparity. The scenic spot network can be regarded as a social network, with scenic spots as nodes and tourist flow as directed edges. SNA quantifies node importance through centrality

indicators: degree centrality represents the number of direct connections between nodes, while degree centrality reflects the radiation and attraction of tourist attractions, respectively (Gaddy et al., 2024). The formula for degree centrality is given by Eq. (3).

$$\begin{cases} C_i^O = \sum_{j \neq i \in G} O_{i,j} \\ C_i^T = \sum_{j \neq i \in G} T_{j,i} \end{cases} \quad (3)$$

In Eq. (3), G represents a collection of scenic spots. In the tourism network, $O_{i,j}$ and $T_{j,i}$ respectively represent the correlation of tourist flow. The centrality of intermediaries is quantified by calculating the number of shortest paths passing through the node, as shown in Eq. (4).

$$Y_k = \sum_{(i,j) \in G} \frac{V_{k,(i,j)} / V_{(i,j)}}{(n-1)(n-2) / 2} \quad (4)$$

In Eq. (4), intermediary centrality $V_{(i,j)}$ measures the media capability of node k by the frequency ($V_{k,(i,j)} / V_{(i,j)}$) of the shortest path between scenic spots. The higher the value, the stronger the node's control and centrality in the network. Proximity centrality reflects the degree of proximity as the sum of the shortest-path distances from a node to all other nodes. The smaller the sum, the higher the proximity, indicating that the node is closer to the geometric center of the network. The calculation is shown in Eq. (5).

$$D_i = 1 / \sum_{j=1}^n d(i, j) \quad (5)$$

In Eq. (5), $d(i, j)$ represents the shortest distance from the node i to j . Proximity centrality measures a scenic spot's hub position by its total spatial distance to all other nodes, with higher values indicating stronger spatial centrality as the spot is closer to the network's geometric center (Stamos, 2023). In tourism network analysis, such centrality indicators help identify core attractions. Assessing the importance of rural tourism resources helps governments, businesses, and communities make targeted investments, thereby reducing risk and improving efficiency.

2.2. Multi-Layer Nested Tour Path Optimization Method Based on SEM

While evaluating resource importance and quantifying spatial characteristics through methods like SNA are crucial, these objective measures alone are insufficient for accurate tourism route planning. Tourist choices result from a complex interplay of subjective preferences and objective conditions. SEM addresses this by integrating factor and path analysis. Through its measurement model (linking latent variables to observed indicators) and structural model (defining relationships among latent variables), SEM enables a comprehensive analysis of the factors influencing travel mode selection. The covariance of the observed variables is the "raw fuel" that drives the operation of the entire observation model. Its mathematical expression is given in Eq. (6).

$$\begin{cases} y = \Lambda_y \eta + \varepsilon \\ x = \Lambda_x \xi + \delta \end{cases} \quad (6)$$

In Eq. (6), in the SEM, y and x are the endogenous and exogenous observation variable vectors, respectively. η and ξ are endogenous and exogenous latent variable vectors, respectively. Λ_y and Λ_x are the factor loading matrices, which represent the linear influence of latent variables on observed variables. ε and δ are the measurement error or residual term. In the SEM, Eq. (7) can be used to describe the causal relationship between latent variables.

$$\eta = Z_A \eta + \theta \xi + \gamma \quad (7)$$

In Eq. (7), γ represents the interference term. Z_A represents the relationship matrix. θ is the impact of exogenous latent variables on endogenous latent variables. SEM involves steps such as model construction, identification, parameter estimation using methods such as Maximum Likelihood Estimation (MLE), and evaluation using fit indicators, including χ^2/df and GFI (Juhari, 2023). Applied to rural tourism, the model posits that tourists' travel choices (e.g., self-driving vs. public transportation) are influenced by four latent variables: personal attributes (which indirectly affect preferences and activity attributes), travel preferences, tourism activity attributes (which determine available transport modes), and transportation attributes like cost and time, collectively forming a "latent variable to observation variable to decision result" framework, with rural characteristic attractions categorized based on features such as hydrological and cultural landscapes. Among them, H represents the departure and destination points of tourists, usually the accommodation location, distribution center or transportation hub. The R represents core rural landscape nodes, including natural or cultural attractions with high resource value. The O represents general sightseeing nodes or secondary attractions, with relatively low resource weight. Based on these node types, the activity chain is divided into five structures as shown in Fig. 2.

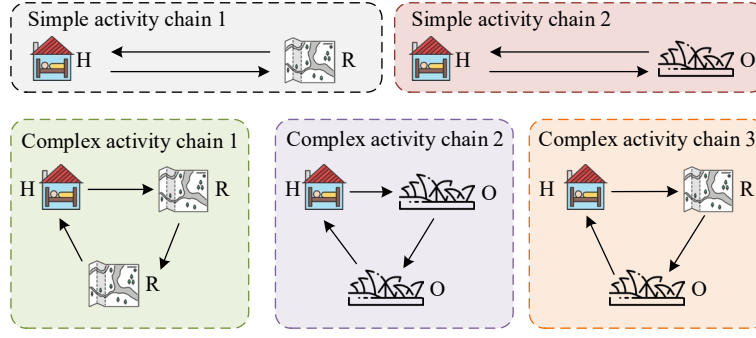


Fig. 2. Rural landscape tour activity chain

Fig. 2 classifies tourist activity chains into simple closed-loop structures (e.g., $H \rightarrow R \rightarrow H$ for deep experience and $H \rightarrow O \rightarrow H$ for standardized behavior) and complex multi-node loops (e.g., $H \rightarrow R \rightarrow R \rightarrow H$, $H \rightarrow R \rightarrow O \rightarrow H$, $H \rightarrow O \rightarrow R \rightarrow H$) reflecting thematic preferences, core resource allocation, and multi-destination flexibility. Travel decision-making is a hierarchical, nested process, simulated by an NL model which overcomes the Independence of Irrelevant Alternatives (IIA) limitation of the Multinomial Logit (MNL) model. Taking the double-layer NL model as an example, if the upper selection set $H = (h_1, h_2, \dots, h_i)$ contains i schemes and the lower selection set $U = (u_1, u_2, \dots, u_j)$ contains j schemes, the total utility S of the schemes is a combination function of the utility of each layer and the associated parameters. The calculation process is shown in Eq. (8).

$$S_{hu} = Q_u + Q_h + Q_{hu} + W_u + W_h + W_{hu} \quad (8)$$

In Eq. (8), Q_h is the fixed utility of the upper level scheme h . Q_u is the fixed utility of the lower level scheme u . Q_{hu} is the fixed utility of the combination scheme (h, u) . W_h, W_u, W_{hu} represent the random error terms when selecting the upper layer scheme h , lower layer scheme u , and a combination scheme (h, u) . Based on this, a joint hierarchical selection model for the tourism activity chain and travel mode is constructed. The upper level defines five types of activity chains: HRH simple characteristic, HOH simple non-characteristic, HRRH complex characteristic, HROH mixed type, and HOOH complex non-characteristic. The lower level selects the travel mode based on the selected activity chain. The NL parameters are estimated using MLE and calibrated with the Python Biogeme software package. The model estimation is carried out in two stages: First, the lower-level utility function is estimated using sample data that include alternative options. Second, the Inclusion Value (IV) is calculated from the lower-level estimation results, and the upper-level utility function is then estimated. The model assumes that all error terms follow the generalized extreme value distribution, and based on this, the joint selection probability of the alternatives is derived.

The phased approach was chosen for the research because it clearly reveals the model's hierarchical relationships, is easy to operate and is easy to explain. The evaluation of model results includes two aspects: significance test and overall evaluation. The likelihood ratio test and the goodness-of-fit test are standard practices in SEM. The calculation formula for the likelihood ratio test is shown in Eq. (9).

$$LR = -2 \ln \frac{L(\theta_0)}{L(\theta_1)} \quad (9)$$

In Eq. (9), $L(\theta_0)$ represents the value of the likelihood function under the null hypothesis. $L(\theta_1)$ represents the value of the likelihood function under unconstrained conditions. The larger LR The worse the fit between the null hypothesis and the data, the more likely it is that the null hypothesis will be rejected. The overall adaptability of the model is characterized by the McFadden coefficient χ^2 , which is calculated using Eq. (10).

$$\chi^2 = 1 - \frac{L(\theta_0)}{L(\theta_1)} \quad (10)$$

In Eq. (10), the McFadden coefficient $\chi^2 \in [0, 1]$, and the closer it is to 1, the stronger the explanatory power of the model. In practice, an accuracy of $\chi^2 \geq 0.2$ is considered acceptable and is commonly used for model optimization.

2.3. Model for Generating Rural Landscape Tourism Route Network Under Dual Objective Optimization

Building upon the optimization logic from SEMs and the activity chain sharing rates from the NL model, the node weights and behavior data are transformed into a quantitative basis for route planning through a dual-objective optimization framework; this framework crucially balances landscape experience quality with travel efficiency by converting node weights into landscape service level indicators and activity chain sharing rates into time cost estimates, thereby solving the coordination problem between experience and spatial efficiency in traditional planning. To evaluate the effectiveness of the model proposed in this study, a baseline model is established as a benchmark. This baseline model does not account for tourists'

behavioral preferences and only aims to minimize total travel time. It also uses a genetic algorithm to find a solution and generates a tourism route network. This study will compare the average landscape service level and average travel time of the two models in the same experimental group. The dual-objective model for maximizing landscape service level and minimizing travel time is given by Eq. (11).

$$\begin{cases} \text{Maximize } S(L) = \sum i \in L^{\omega_i \cdot I_i} \\ \text{Minimize } T(L) = \sum (i, j) \in \text{Path}(L) t_{ij} \cdot f_{ij} \end{cases} \quad (11)$$

In Eq. (11), ω_i is the node weight (i.e., resource importance). I_i represents whether the node is included in line L . t_{ij} is the travel time between nodes. f_{ij} is the activity chain-sharing rate, reflecting the probability of travel mode selection. Enforcing a closed-loop constraint for activity chains (e.g., $H \rightarrow R \rightarrow H$), route optimization uses a genetic algorithm that simulates biological evolution, starting from an initial solution and evolving populations through inheritance, crossover, and mutation over multiple generations to retain outstanding individuals and approach the optimal solution (Manosuthi, Lee, and Han, 2024); in the tourism bus network, a route is a gene and a set of routes forms a chromosome, as illustrated in the dual-objective optimization and double-layer NL model for rural landscape tourism route network generation in Fig. 3.

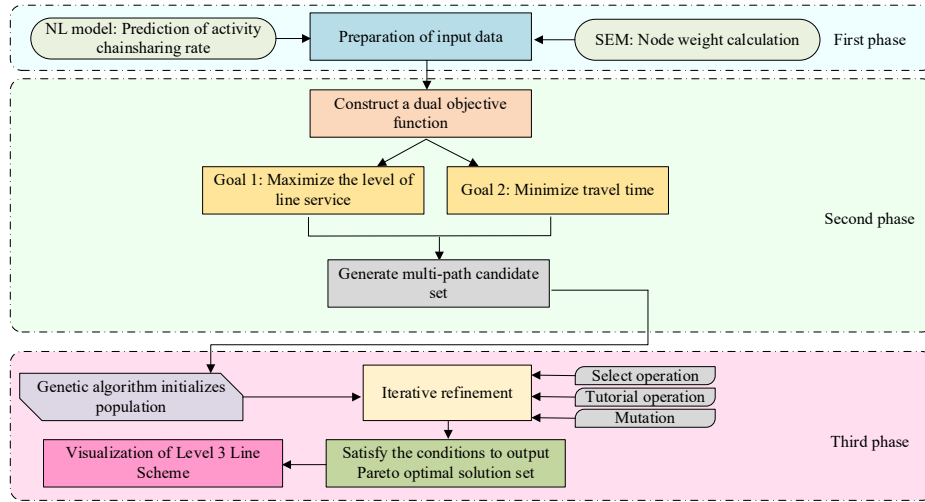


Fig. 3. Model for generating rural landscape tour route network

In Fig. 3, during the design of the rural landscape tour route network generation model, landscape node weights are first input based on resource importance derived from the SEM. Using the passenger flow distribution probabilities predicted by the NL model, constraints such as mode share and node connectivity can be derived. Subsequently, a multi-objective optimization model is constructed, with the first objective to maximize the service level for high-weight nodes and the second to minimize travel time on paths with high mode shares. To transform multi-objectives into a single objective for route generation, a three-tier network is designed: express lines connect core high-weight attractions, arterial lines link moderately clustered nodes, and branch lines cover niche, distinctive nodes. A standard permutation-based genetic algorithm is employed, with chromosomes representing complete route network schemes encoded in a mixed integer permutation and binary form, where each gene unit is structured as [Attraction ID·Transportation Mode]. Key genetic algorithm parameters, determined via sensitivity analysis, include a population size of 50, roulette wheel selection with elite retention, a crossover probability of 0.8, and a mutation probability of 0.15. The process iterates for 500 generations or until the change in fitness is less than 0.1%. Crossover exchanges segments of high-importance nodes based on their weights, while mutation increases the replacement probability for low-importance nodes according to betweenness centrality, with a mutation rate of up to 50%. The outputs include a Pareto front solution set balancing landscape and time, a GIS-visualized three-tier route network, and a flexible scheduling mechanism for dynamic weight adjustment during peak periods. The computational complexity of this integrated framework stems primarily from the serial processing of multi-stage models, SEM's covariance matrix estimation, the NL model's MLE, and GA's iterative optimization. The GA's search space grows exponentially with the number of candidate nodes (N), making it the main computational burden.

3. Results and Analyses

3.1. Empirical Analysis of Factors Influencing Tourist Behavior

To analyze the influencing factors of travel choices in rural tourism and reveal the relationships among personal attributes, travel preferences, tourism activity attributes, and transportation attributes, as well as their mechanisms of action on travel choices, the study utilizes the proposed rural landscape tour route network generation model to construct an SEM with seven influence paths between variables. The contribution of observed variables to latent variables is evaluated through factor loadings. The results are presented in Fig. 4.

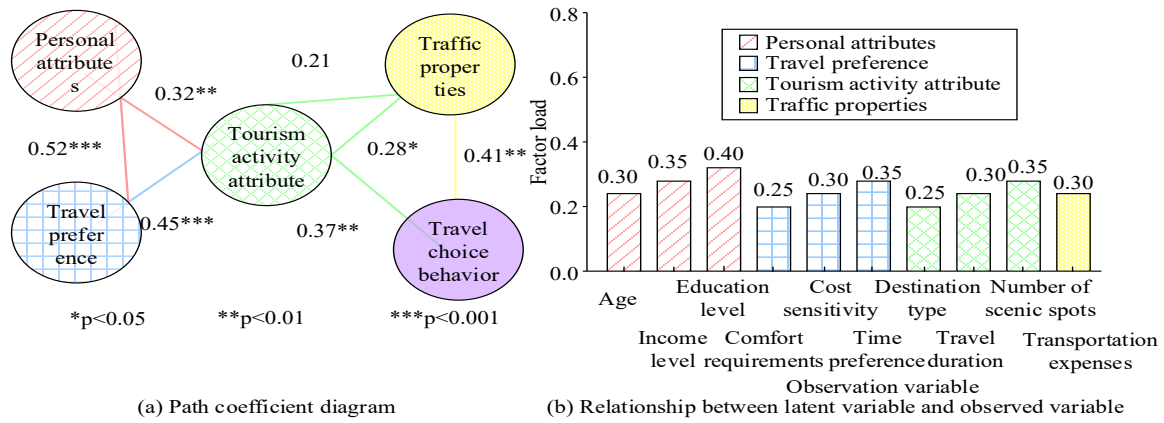


Fig. 4. Analysis of influencing factors of rural travel choice based on SEM

In Fig. 4(a), travel preferences exhibited the strongest direct impact on travel choices, with a path coefficient of 0.55 ($p < 0.001$), followed by transportation attributes (0.45, $p < 0.001$) and tourism activity attributes (0.38, $p < 0.001$). Personal attributes had significant positive effects on travel preferences (0.42, $p < 0.001$), tourism activity attributes (0.35, $p < 0.001$), and transportation attributes (0.28, $p < 0.01$). Additionally, tourism activity attributes also had a notable influence on transportation attributes (0.32, $p < 0.01$). In Fig. 4(b), all factor loadings for the observed variables exceeded 0.5. Transportation costs (0.85) and time consumption (0.82) within transportation attributes contributed the most to the latent variable, while destination type (0.82) under tourism activity attributes and income (0.78) under personal attributes also demonstrated high factor loadings. The overall goodness-of-fit of the model is excellent: $\chi^2/df = 2.15 (< 3)$, RMSEA = 0.052 (< 0.08), CFI = 0.95 (> 0.90), TLI = 0.94 (> 0.90), SRMR = 0.038 (< 0.05). These indicators indicated that the model had a high degree of compatibility with the covariance matrix of the observed data, and the hypothesized structural relationships were reasonable. These findings indicated that the observed variables were key indicators of their corresponding latent variables, and that the overall model fit was satisfactory, validating the effectiveness of the constructed SEM. Meanwhile, the study also analyzed tourist behavior flows and visualized the relationships among tourist attributes, activity chain types, and travel modes using a Sankey diagram, with the results presented in Fig. 5.

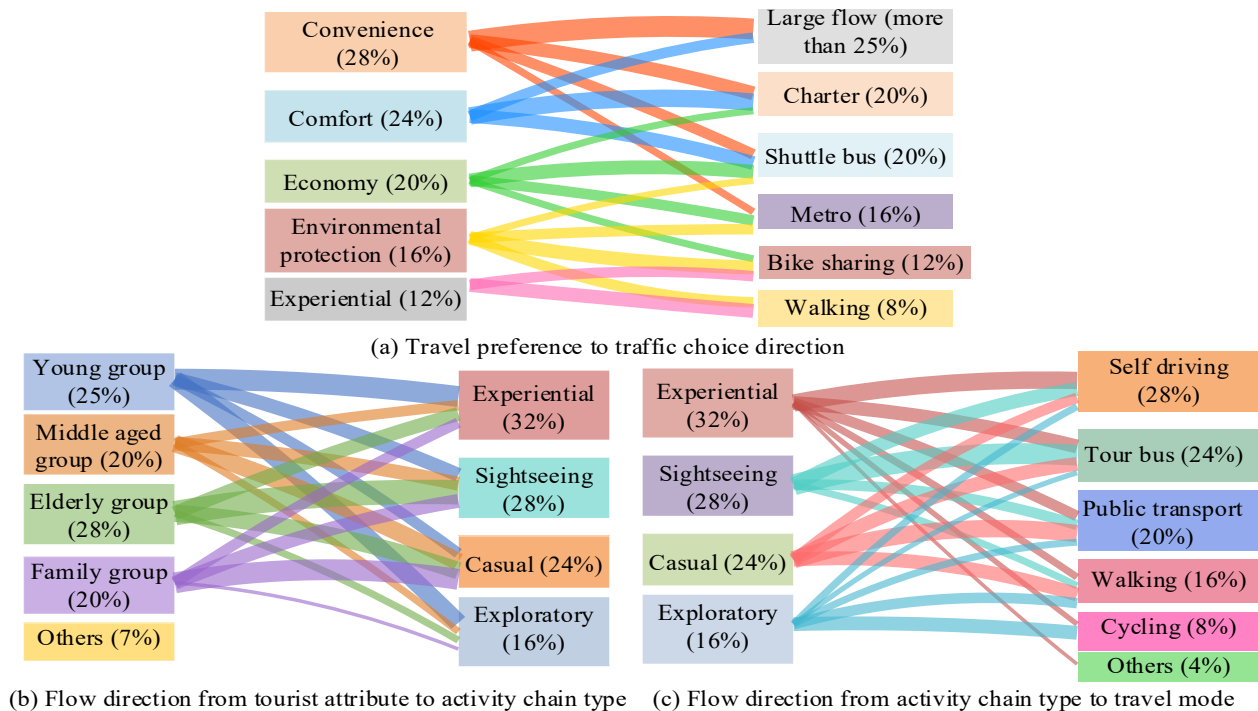


Fig. 5. Tourist behavior flow analysis by Sangit

Fig. 5 reveals distinct tourist behavior patterns: younger tourists strongly preferred (42%) immersive, interactive experiences, while the elderly group favored (38%) sightseeing at classic landmarks. Regarding transportation, family travelers showed a high reliance on public transit (65%), whereas solo travelers preferred self-driving (58%). A significant correlation was found: 78% of convenience-prioritizing tourists chose public transport. These findings uncover the logic behind tourist decision-making and provide an empirical basis for constructing the NL model. The convergence analysis results of genetic algorithm optimization are presented in Fig. 6.

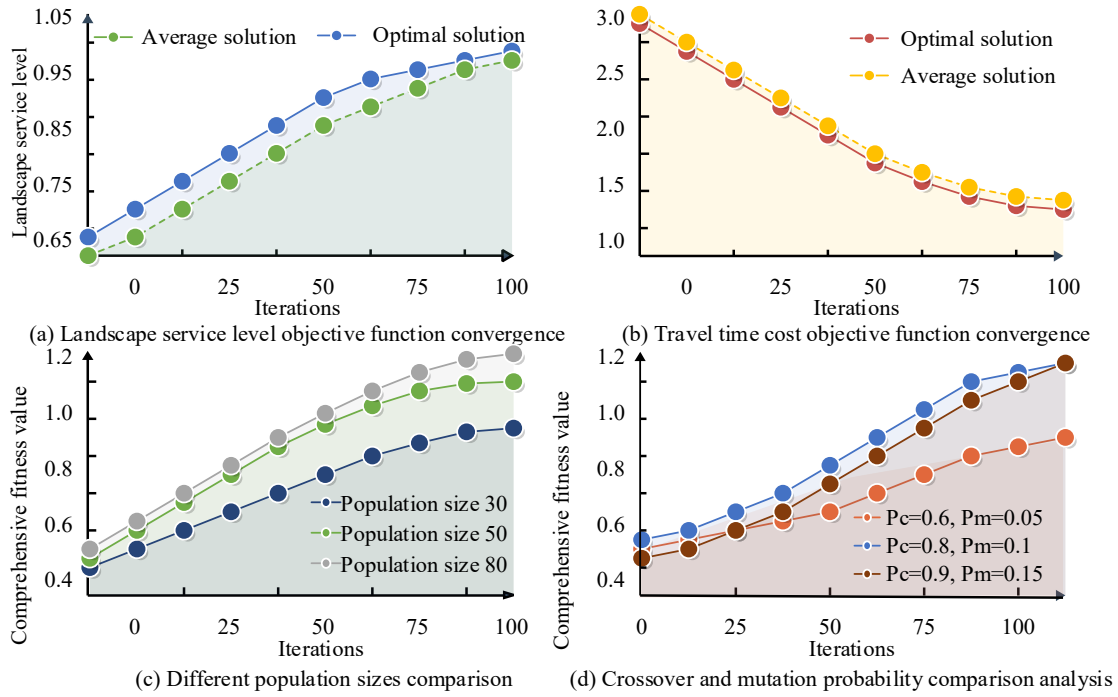


Fig. 6. Convergence analysis of genetic algorithm optimization

As observed in Fig. 6(a), the objective function for landscape service levels demonstrated rapid convergence within the first 30 generations, with the optimal solution improving from an initial value of 0.3 to 0.85, while the average solution stabilized around 0.75, indicating the algorithm's significant optimization effect on landscape quality. In Fig. 6(b), the objective function for travel exhibited a more gradual convergence, with the optimal solution stabilizing after 60 generations at a relatively low cost level of 0.2, demonstrating the algorithm's effectiveness in controlling time costs. Fig. 6(c) reveals that a population size of 50 performed best, outperforming configurations of 30 and 80 in terms of both convergence speed and solution quality, suggesting that a moderate population size can balance search efficiency and solution diversity. Fig. 6(d) shows that a combination of a crossover probability of 0.8 and a mutation probability of 0.1 achieved optimal convergence, validating the critical impact of parameter configuration on algorithm performance. A sensitivity analysis was conducted on the key parameters of the genetic algorithm. The results showed that the mutation rate had the most significant impact on the quality and diversity of the solutions. When the mutation rate was below 0.05, the algorithm was prone to getting stuck in local optima; when it was above 0.25, the convergence speed was too slow, and the solutions were unstable. Finally, the crossover probability was set at 0.8 and the mutation probability at 0.15. Comparative verification with different initial population sizes (30, 50, 80) confirmed the robustness of the current parameter settings.

3.2. Application Analysis of the Model for Generating Rural Landscape Tourism Route Network

To validate the proposed rural landscape tour route generation model, the study tested five experimental groups, each representing a distinct activity chain. Group 1 (H→R→H) assessed the balance between landscape quality and time cost in a deep-experience framework. Group 2 (H→O→H) evaluated route suitability for standardized tours. Group 3 (H→R→R→H) examined the clustering effect of themed routes. Group 4 (H→R→O→H) measured the efficiency of a core-resource prioritization model. Group 5 (H→O→R→H) tested adaptability in multi-destination interconnected scenarios. These five experimental groups were divided based on the joint analysis results of the SEM-NL model, corresponding to different tourist decision-making patterns: Group 1 (deep natural immersion type) prefers hydrological landscapes and slow-paced activities. Group 2 (efficient classic sightseeing type) prefers landmarks and group buses. Group 3 (cultural theme linking type) prefers in-depth exploration of the same resource type. Group 4 (mixed core experience type) prioritizes access to core resources. Group 5 (flexible multi-stop linked type) involves multi-destination switching and accommodation changes. The outcomes are detailed in Table 1.

Based on the results in Table 1, different activity chain structures demonstrated distinct performance trade-offs. Experimental group 3 (feature-connected closed loop) achieved the highest landscape service level (3.75) and model fit (0.51), but with the longest travel time (58.43 min). In contrast, group 2 (single non-feature round-trip) had the shortest travel time (28.12 min) but the lowest landscape value (1.56) and fit (0.28). Group 1 (single-feature round-trip) balanced service level (3.28) and time cost (32.56 min) effectively with high node coverage (92.35%), while group 4 (mixed-attraction closed loop) showed moderate performance across all metrics. Group 5 (cross-accommodation mixed loop) indicated that multi-destination links substantially increase travel time (52.31 min) for a moderate landscape gain (2.13). The correlation between the importance ratings of different types of rural tourism attractions and tourist flow volumes is presented in Fig. 7.

Table 1. Empirical analysis results of performance testing for rural tourism route optimization

Experimental group No.	Average landscape service level	Average travel time (minutes)	Model goodness of fit (McFadden coefficient)	Coverage rate of high-weight nodes (%)
1	3.28	32.56	0.42	92.35
2	1.56	28.12	0.28	65.78
3	3.75	58.43	0.51	94.62
4	2.89	45.77	0.39	82.15
5	2.13	52.31	0.35	76.89

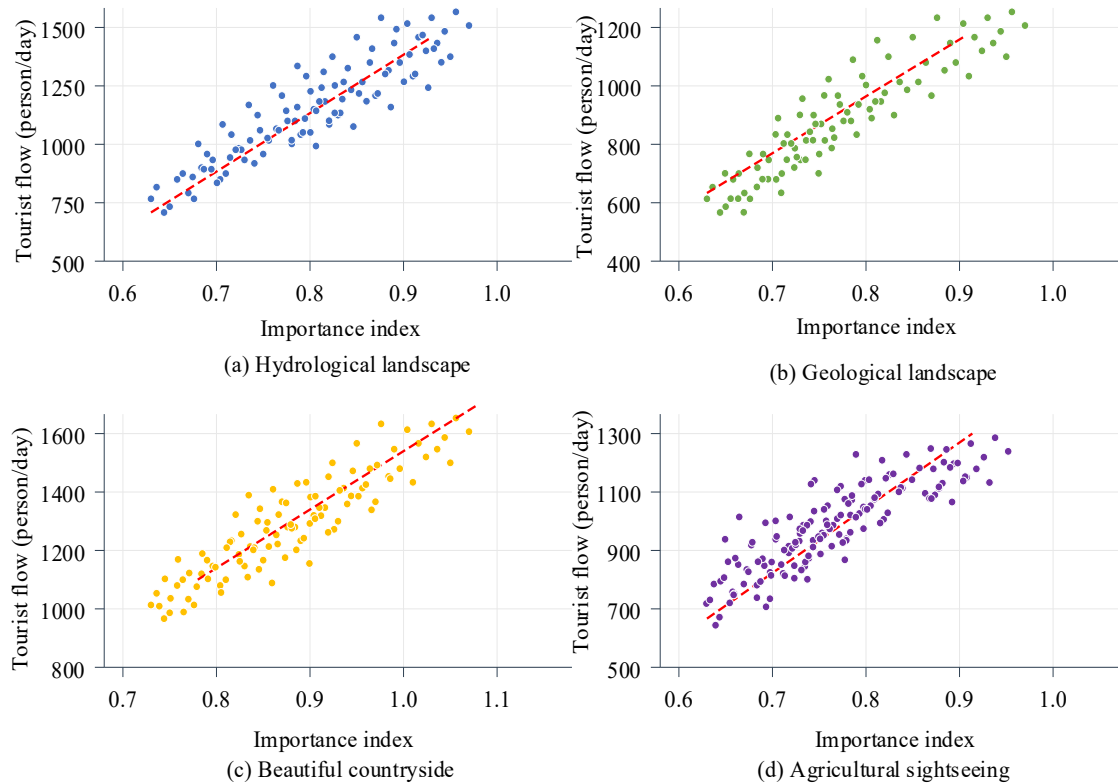
**Fig. 7.** Analysis of the relationship between the importance of rural tourism attractions and tourist flow

Fig. 7 presents the correlation between the importance ratings and tourist flow volumes for four types of rural attractions. Hydrological landscapes (Fig. 7(a)) show a strong positive correlation ($r=0.78$), with higher ratings attracting larger flows (500–1500 visitors). Geological landscapes (Fig. 7(b)) also show a strong positive trend ($r=0.72$), with visitor flows ranging from 400 to 1200. Beautiful countryside attractions (Fig. 7(c)) exhibit the strongest correlation ($r=0.85$), indicating a close match between rated importance and tourist preference, with visitor flows of 800–1600. Agricultural tourism (Fig. 7(d)) shows a relatively weaker but still positive association ($r=0.69$), with flows ranging from 500 to 1300 visitors. Overall, importance ratings positively correlate with visitor volumes across all attraction types, with beautiful countryside sites aligning most closely with tourist demand. The comparison results of this research model with the baseline model are presented in Table 2.

Table 2 compares the performance of the model proposed in this study with that of the baseline model across five experiments. It can be seen that across the five activity chain structures, the model in this study significantly outperformed the baseline model in terms of average landscape service level and high-weight node coverage, while the average travel time increased only slightly. For example, in experiment group 1, the average landscape service level of the model in this study was 3.28, which was 77.3% higher than the baseline model's 1.85. The high-weight node coverage increased from 65.40% to 92.35%, while the average travel time only increased by 2.44 minutes (8.1%). Based on the weighted objective function calculation, the efficiency of this study's model was 28% higher than that of the baseline model.

4. Discussion

The study proposed a framework for generating rural landscape tourism route networks that integrates SEM, double-layer NL models, and genetic algorithms, demonstrating advantages for analyzing tourist decision-making mechanisms and optimizing dual objectives. Key nodes were identified through SNA, revealing that core-node attractiveness was closely tied to tourist preferences for “comfort” and “convenience,” providing a micro-behavioral explanation for the core-periphery

structure noted by Qin et al. (2022). Compared to the distance optimization using a genetic algorithm based on the traveling salesman problem by Chen et al. (2024). This study's dual-objective model synergistically balanced landscape service level and time cost, increasing satisfaction for a "Beautiful Rural Route" in Shandong by 0.27–0.31. Unlike the marketing-oriented new media and VR/AR approach by Wang et al. (2023), this SEM-NL-genetic algorithm framework integrated "activity chain travel mode" choices into path optimization to better capture actual spatial behavior. At the resource evaluation level, its focus on "importance" over sustainability metrics complements the rural sustainable development information model by Harbiantkova and Gertsberg (2022), who stated that while empirical results, such as beautiful rural resources scoring 8.3 and transportation convenience 7.7, confirmed the ability to identify highly attractive and accessible core nodes, aligning with the cluster development viewpoint of Lin et al. (2024). The deep mechanism revealed by SEM showed that personal attributes indirectly influenced transport choices through travel preferences. For example, family travelers' 65% dependence on public transit. This finding both supports and extends the "cost and time" findings of Tsigdinos et al. (2024) by identifying a "preference-behavior" mediating effect. Additionally, the double-layer NL model captured nested relationships between activity chains and transport modes, with 42% of young people preferring "immersive activity chains + self-driving" versus elderly travelers opting for "sightseeing activity chains + public transportation." Furthermore, it should be noted that the effectiveness of the route network generation model proposed in this study has currently been verified mainly through internal simulations and comparisons with baseline models. Although the comparison results showed that the model significantly improved landscape service level, node coverage, and network structure efficiency, these conclusions have not yet been tested in real-world implementations. Future research can apply the route plans generated by this model in actual rural areas for pilot applications. By comparing key performance indicators such as tourist satisfaction, traffic flow distribution, and scenic spot income before and after implementation, the actual planning effect of the model can be further verified. Additionally, such practical feedback can also be used to iteratively optimize model parameters and enhance its practical adaptability.

Table 2. Comparison of model performance with baseline model

Experimental group number	Model type	Average landscape service level	Average travel time (minutes)	Coverage rate of high-weight nodes (%)
1	This research model	3.28	32.56	92.35
	Baseline model	1.85	30.12	65.4
2	This research model	1.56	28.12	65.78
	Baseline model	0.95	25.34	45.23
3	This research model	3.75	58.43	94.62
	Baseline model	2.1	55.67	70.15
4	This research model	2.89	45.77	82.15
	Baseline model	1.65	42.89	60.33
5	This research model	2.13	52.31	76.89
	Baseline model	1.2	49.87	55.47

This study provides a clear decision-making transition path for tourism and transportation planners, shifting from supply-oriented planning based on expert experience and static resource assessments to demand-responsive planning that integrates insights into tourist behavior and multi-objective collaborative optimization. Managers can prioritize investments in high-impact latent variables identified by SEM and differentiate transportation service allocation and landscape maintenance priorities based on the optimized hierarchical network of trunk, branch, and sub-branch routes, thereby achieving Pareto improvement in overall experience and efficiency within budget constraints.

The model parameters in this study were calibrated using data from a rural area in Shandong Province, making them context-dependent. Direct application to regions with different transportation infrastructure, tourist demographics, or spatial distribution of resources would require parameter adjustments. However, the integrated framework of "resource evaluation to behavior modeling to dual-objective optimization" is methodologically universal, allowing other regions to input local data and follow the same technical process to generate customized route networks suited to their unique resource endowments and tourist characteristics.

The theoretical contribution lies in the deep integration and extension of destination choice behavior theory from tourism research with discrete choice modeling and network optimization theory from transportation planning. By employing SEM, the study revealed unique path relationships among latent variables in the rural tourism context, enriching theoretical understanding of travel decision-making mechanisms in specific settings. Furthermore, nesting SEM latent variables representing psychological factors within the NL model representing rational choice provided a methodological example for constructing a more comprehensive "psychological-behavioral" integrated transportation choice model.

The study has several limitations. The case selection was limited to a rural area in Shandong, which may not be generalizable to regions dominated by hydrological landscapes. Although genetic algorithms achieved dual-objective optimization, the 50% mutation rate was set empirically and lacked an adaptive mechanism. Additionally, the quantification of “protective value” in resource importance evaluation relied on expert scoring, introducing potential subjectivity in the accuracy of node weights. Future research could introduce adaptive adjustment mechanisms to address dependence on mutation rates, incorporate dynamic data such as real-time tourist flow and weather conditions, and develop elastic scheduling models to enable dynamic updates to the route network.

5. Conclusion

In response to RQ1, travel preference is the strongest factor influencing travel choice (0.55), followed by transportation attributes (0.45) and tourism activity attributes (0.38). RQ2 identifies distinct groups: e.g., 42% of young tourists prefer immersive activity chains, and 65% of family travelers rely on public transport. RQ3, SEM-based node weights and NL-based mode shares are integrated into a dual-objective genetic algorithm. RQ4, the proposed model improves overall efficiency by 28% over the baseline model. This study developed a rural landscape tour route generation model by integrating SEM, an NL model, and a genetic algorithm within a “resource analysis to behavioral interpretation to path optimization” framework, finding that travel preference was the strongest factor influencing choices (path coefficient 0.55) with distinct group behaviors, such as 42% of youth preferring immersive activity chains and 65% of families relying on public transport, while the dual-objective genetic algorithm improved satisfaction across five activity chain types by 0.19–0.31. Spatially, kernel density and SNA identified high-weight nodes, exemplified by beautiful countryside attractions with a resource value of 8.3 and accessibility of 7.7, showing a 0.85 correlation with tourist flow. Behaviorally, the NL model quantified “activity chain-travel mode” selection probabilities to generate passenger-flow corridors, ultimately translating node weights and mode shares into landscape-service and time-cost metrics. Therefore, enabling precise coverage of a three-tier network (express, arterial, and feeder routes). This model effectively couples objective resource characteristics with subjective tourist behavior, offering a data-driven tool for route planning.

While the proposed model demonstrated theoretical and methodological innovation, several areas require improvement. The model parameters, calibrated using rural data from Shandong Province, necessitate re-calibration when applied to regions with different resource types, transportation conditions, or tourist structures. Future cross-regional comparative studies could evaluate their transferability. In evaluating resource importance, indicators such as “protection value” currently rely on expert scoring. Introducing public-participation assessments or social-media sentiment analysis could enhance objectivity and dynamism. Parameters like the mutation rate in the genetic algorithm still depend on empirical settings. Future research should explore adaptive parameter adjustment mechanisms to improve optimization efficiency and stability. The model has not yet integrated real-time data. There are future opportunities to incorporate IoT and mobile signaling data to build a dynamic, flexible route recommendation system that improves timeliness and personalization.

The contributions of this study lie in two aspects. Methodologically, it integrates SEM, NL, and GA to construct a quantifiable and replicable “psychology–behavior–space” coupling analysis framework, offering new tools for optimizing tourism routes in complex decision-making contexts. Paradigmatically, it empirically demonstrates a logical loop that directly drives spatial planning with tourist behavior data, providing a feasible path for shifting tourism planning from a static “resource supply” orientation to a dynamic “behavior demand” response. The deeper insight is that this framework implies that future rural planning must focus on dynamic adaptation to, and guidance of, complex human behaviors rather than merely on facility layout. However, because model validity depends on stable preferences and structured data, this framework serves less as a definitive solution than as a starting point for an iterative “computational laboratory,” indicating that integrating human complexity with spatial computation is key to achieving truly intelligent and humanistic rural tourism planning.

Funding

The research is supported by the Priority Research Project of Urumqi Vocational University in 2025, Study on the spatial evolution pattern and reconstruction of oasis Karez settlements in the Turpan Basin (No. 2025XZ005). Institutional Review Board (IRB) Statement. We confirm that this study involved human participants via an anonymous questionnaire. The study was reviewed by Urumqi Vocational University and determined to be exempt from full IRB review. The study used anonymous, voluntary questionnaires. It involved no intervention, no biological samples, no vulnerable populations, and no identifiable private information.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

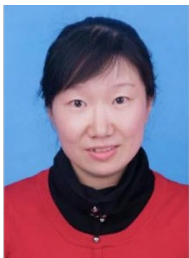
The authors confirm that no AI tools were used in the preparation of this manuscript.

References

- Vada, S., Dupre, K., and Zhang, Y. (2023). Route tourism: a narrative literature review. *Current issues in tourism*, 26(6), 879-889, DOI: 10.1080/13683500.2022.2151420.
- Bhosle, K. and Musande, V. (2023). Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. *Artif. Intell. Appl.*, 1(2), 114-118, DOI: 10.47852/bonviewAIA3202441.
- Chen, Z., Zhang, P., and Peng, L. (2024). Application of a hybrid genetic algorithm based on the travelling salesman problem in rural tourism route planning. *International Journal of Computing Science and Mathematics*, 19(1), 1-14,

DOI: 10.1504/IJCSM.2024.136816.

- Gaddy, Y., Jones, E. C., Wells, R., and others. (2024). Effects of Resource Sharing Networks on Community Anti-Drug Coalitions Outcomes: A Social Network Analysis. *Prevention Science*, 25(7), 1029-1039, DOI: 10.1007/s11121-024-01719-1.
- Harbiankova, A. and Gertsberg, L. (2022). Information Model for Sustainable Rural Development. *Energies*, 15(11), 1-15, DOI: 10.3390/en15114009.
- Izquierdo-Gascón, M. and Rubio-Gil, Á. (2023). Theoretical approach to Api-tourism routes as a paradigm of sustainable and regenerative rural development. *Journal of Apicultural Research*, 62(4), 751-766, DOI: 10.1080/00218839.2022.2079285.
- Juhari, J. (2023). Loyalty of domestic tourists to beach tourism: role of image, experiential, digital marketing and satisfaction. *IJBE (Integrated Journal of Business and Economics)*, 7(2), 473-487, DOI: 10.33019/ijbe.v7i2.715.
- Li, H., Wang, J., Wei, Q., and Yang, S. (2024). The Mechanism of Rural Tourism Development—Taking Yangshuo County as an Example. *Journal of Resources and Ecology*, 15(5), 1240-1250, DOI: 10.5814/j.issn.1674-764x.2024.05.012.
- Li, K. (2022). SWOT analysis of e-commerce development of rural tourism farmers professional cooperatives in the era of big data. *IET communications*, 16(5), 592-603, DOI: 10.1049/cmu2.12358.
- Li, Q., Ai, M., Luo, J., Sun, Y., and Tian, L. (2025). Three-step optimization based on a multi-model of rural tourism sites. *Plos one*, 20(9), e0332878, DOI: 10.1371/journal.pone.0332878.
- Lin, X., Shen, Z., Teng, X., and Mao, Q. (2024). Cultural routes as cultural tourism products for heritage conservation and regional development: A systematic review. *Heritage*, 7(5), 2399-2425, DOI: 10.3390/heritage7050114.
- Lucio, B. B. (2023). Equivalent Gini coefficient, not shape parameter! *Scientometrics*, 128(1), 867-870, DOI: 10.1007/s11192-022-04571-8.
- Manosuthi, N., Lee, J. S., and Han, H. (2024). Green behavior at work of hospitality and tourism employees: Evidence from IGSCA-SEM and fsQCA. *Journal of Sustainable Tourism*, 32(1), 85-107, DOI: 10.1080/09669582.2022.2115051.
- Pei, Q., Wang, L., Du, P., and Wang, Z. (2022). Optimization of tourism routes in Lushunkou District based on ArcGIS. *Plos one*, 17(3), e0264526, DOI: 10.1371/journal.pone.0264526.
- Qian, S. and Tian, Z. (2024). A nearest neighbor query method for searching objects with time and location informations based on spatiotemporal similarity. *Evolutionary Intelligence*, 17(4), 3031-3041, DOI: 10.1007/s12065-024-00926-7.
- Qin, X., Li, X., Chen, W., Tan, H., Luo, L., and Xu, X. (2022). Tourists' digital footprint: the spatial patterns and development models of rural tourism flows network in Guilin, China. *Asia Pacific Journal of Tourism Research*, 27(12), 1336-1354, DOI: 10.1080/10941665.2023.2166420.
- Stamos, I. (2023). Transportation networks in the face of climate change adaptation: A review of centrality measures. *Future Transportation*, 3(3), 878-900, DOI: 10.3390/futuretransp3030049.
- Tang, L. and Huang, S. (2023). Spatial distribution and influencing factors of the key villages of national rural tourism in Fujian Province. *Journal of Central South University of Forestry & Technology*, 43(9), 181-190, DOI: 10.14067/j.cnki.1673-923x.2023.09.019.
- Tsigdinos, S., Karolemeas, C., Siti, M., Papadaki, K., Athanasopoulos, K., and Tzouras, P. G. (2024). Route Planning for Flexible Bus Services in Regional Cities and Rural Areas: Combining User Preferences with Spatial Analysis. *Future Transportation*, 4(4), 1476-1500, DOI: 10.3390/futuretransp4040071.
- Wang, X., Yao, J., Zhang, C., and Yan, X. (2023). Research on immersive experience of rural tourism marketing in the Internet Era. *Journal of Social Science Humanities and Literature*, 6(5), 124-127, DOI: 10.53469/jsshl.2023.06(05).18.
- Yin, L. and Li, M. (2025). Application of improved GA in optimizing rural tourism routes. *Nonlinear Engineering*, 14(1), 20250127, DOI: 10.1515/nleng-2025-0127.
- Zhang, H., Li, J., and Liu, J. (2023). Spatial differentiation characteristics and influencing factors of key rural tourism villages in Hunan Province. *Journal of Central South University of Forestry & Technology*, 43(9), 171-180, DOI: 10.14067/j.cnki.1673-923x.2023.09.018.



Fenglian Liu obtained her bachelor's degree in Geography from Xinjiang Normal University in 2002. She received her master's degree in human geography (with a focus on tourism planning and development) from Xinjiang Normal University in 2005. She is currently an associate professor at the College of Culture and Tourism, Urumqi Vocational University. Her research interests include tourism management, tourism information technology, and scenic area development and planning.