*Journal of Infrastructure, Policy and Development* 2024, 8(12), 6700.
https://doi.org/10.24294/jipd.v8i12.6700

Article

# Analysis of tourist flow prediction model of rural tourism on the edge of big cities in China

**Na Yin, Chonlavit Sutunyarak***

Chakrabongse Bhuvanarth International Institute for Interdisciplinary Studies (CBIS), Rajamangala University of Technology Tawan-OK, Bangkok 10400, Thailand
*** Corresponding author:** Chonlavit Sutunyarak, Chonlavit_su@rmutto.ac.th

**CITATION**

Yin N, Sutunyarak C. (2024). Analysis of tourist flow prediction model of rural tourism on the edge of big cities in China. Journal of Infrastructure, Policy and Development. 8(12): 6700. https://doi.org/10.24294/jipd.v8i12.6700

**ARTICLE INFO**

Received: 28 May 2024
Accepted: 26 June 2024
Available online: 31 October 2024



**Abstract: Introduction:** With the adoption of the rural rehabilitation strategy in recent years, China's rural tourist industry has entered a golden age of growth. Due to the lack of management and decision-support systems, many rural tourist attractions in China experience a "tourist overload" problem during minor holidays or Golden Week, an extended vacation of seven or more consecutive days in mainland China formed by transferring holidays during a specific holiday period. This poses a severe challenge to tourist attractions and relevant management departments. **Objective:** This study aims to summarize the elements influencing passenger flow by examining the features of rural tourist attractions outside China's largest cities. Additionally, the study will investigate the variations in the flow of tourists. **Method:** Grey Model (1,1) is a first-order, single-variable differential equation model used for forecasting trends in data with exponential growth or decline, particularly when dealing with small and incomplete datasets. Four prediction algorithms—the conventional GM(1,1) model, residual time series GM(1,1) model, single-element input BP neural network model, and multi-element input BP network model—were used to anticipate and assess the passenger flow of scenic sites. **Result:** The multi-input BP neural network model and residual time series GM(1,1) model have significantly higher prediction accuracy than the conventional GM(1,1) model and unit-input BP neural network model. A multi-input BP neural network model and the residual time series GM(1,1) model were used in tandem to develop a short-term passenger flow warning model for rural tourism in China's outskirts. **Conclusion:** This model can guide tourists to staggered trips and alleviate the problem of uneven allocation of tourism resources.

**Keywords:** short-term passenger flow; rural tourism in fringe areas; big cities in China; warning model of rural tourism

## 1. Introduction

As people's needs have changed and development levels have accelerated, the tourist sector has proliferated in scope and quality. Tourism has grown in importance in a relatively rich culture as a measure of rising living standards and a way of life (Ma et al., 2022). With overall revenue reaching 6.63 trillion yuan in 2019 and tourism and allied industries reaching 4.5 trillion yuan in added value, accounting for around 4.56% of GDP, China's domestic tourist industry has risen at an average annual pace of about 10.6% since 2012. There has been increasing consolidation of tourism's role as a critical pillar industry of the national economy.

At present, the pace of urbanization development in China is accelerating, and some large cities have entered the stage of suburban urbanization and will face reverse urbanization development. By comparing the relationship between urbanization and spatial patterns of tourism behavior, scholars have found that the rise of the tourism industry around cities is a common phenomenon that occurs when urbanization reaches a certain level. Large cities, urban agglomerations, and metropolitan areas have gradually become the focus of tourism industry development. Rural scenic spots on the edge of large cities are increasingly becoming essential providers of urban leisure and vacation tourism products and essential carriers of the tourism industry in metropolitan areas (Chen and Cong, 2022). Rural tourism has sprung up in the tourism market with its leisure and healthy tourism positioning, a form of tourism with fewer stops in one place, convenient transportation close to central cities, and tailored tourism resources. The government will promote rural tourism to improve services and produce high-quality goods. Through the growth of tourism, many formerly “impoverished villages” will become “beautiful villages” and “happy villages.” China’s rural tourism industry saw an average annual growth rate of 20% between 2012 and 2019.

With the continuous development of technology, intelligent scenic spots supported by information and Internet of Things technology have become a new trend in developing tourist attractions and even the tourism industry. The construction of intelligent scenic spots is a complex system engineering, among which tourism passenger flow prediction is one of the core contents of the intelligent scenic spot information management platform prediction system (Xie et al., 2021). Achieving passenger flow prediction in different periods is crucial for tourist attractions. However, in China, due to various external factors such as natural climate, unique vacation systems, and tourism emergencies, the short-term passenger flow of tourism exhibits complex characteristics such as nonlinearity, seasonality, and randomness. Traditional prediction methods often need help to achieve accurate predictions.

The short-term passenger flow forecast of rural tourism in China, particularly in enormous urban fringe areas, is a new subject of interdisciplinary integration. Its research scope covers geography, regional economics, management science, and other disciplines, and it is comprehensive, cutting-edge, professional, technical, and applied. The keywords of rural tourism in extensive urban fringe areas, rural tourism, and passenger flow prediction are hot to figures in geography, economics, and management, respectively (Ruan and Zhang, 2021). Many scholars have already discussed this, and the research on cross figures needs to be further deepened and explored based on previous studies.

The study of China’s tourism in extensive urban fringe areas began in the late 1980s in China and has gradually become the focus of academic attention due to the rapid development of China’s urbanization. Region is geography’s starting point and destination (Liu et al., 2022). Regional development and sustainable development are the eternal themes of geography research. Geography is the main subject of tourism research in urban fringe areas. Geography theory suggests that large cities with a certain level of economic development often have unique advantages in developing suburban tourism in their peripheral areas (also known as the “urban-rural fringe”). Rural tourism in urban fringe areas may have natural scenery, historical and cultural landscapes, large sports venues, or sightseeing agricultural parks with strong local flavor, which have the potential to transform into functional tourism resources such as sightseeing, vacation, entertainment, fitness, sports, and so on, and is a recreation space suitable for the development of “short trip.”

Rural tourism is a crucial way to develop the regional economy, an important starting point for implementing the new urbanization development strategy, and a critical direction for promoting tourism recovery after the COVID-19 epidemic (Tang, 2022). To overcome the obstacle preventing rural tourism from developing further and to better capitalize on the industry's leadership role in advancing rural economic development, it is imperative to examine how rural tourism is evolving in the context of the modern period, ecological civilization, cultural prosperity, social stability, employment, and benefiting the people, scientific guidance on tourism development planning and public service construction has essential theoretical and practical value (Wang et al., 2021). Urban-rural integration and the coordinated development of the regional system of rural human land relations are critical strategic demands of national and regional development and a new hot spot in regional economics. The challenge is to identify the pain points in the development of rural tourism and use them to improve the quality and efficiency of services. After that, qualified rural areas can be guided by tourism to take a new path of urbanization with Chinese characteristics.

Rural tourism, rural revitalization, high-quality development, consumption expansion and upgrading, and supply-side reform have become keywords in China's economic development in recent years. The State Council formally unveiled the "13th Five Year Plan for the Development of Tourism Industry" in 2016. It made clear that, between 2016 and 2020, we should accelerate supply-side structural reform and uphold work toward transforming China into a comprehensive and moderately prosperous tourism nation. The Central Committee's No. 1 Document 2018 formally advocated the rural revitalization strategy. It carried out projects for rural tourism boutiques (Chen et al., 2023). Among the 14 departments that jointly released the Action Plan for Promoting the Quality and Upgrading of Rural Tourism Development (2018–2020) in 2018 was the National Development and Reform Commission. The plan emphasized how the rural tourism business has tremendous development potential, significantly improves people's lives, and how in-demand it is. It is a crucial means of advancing the consumption of locals, putting rural regeneration plans into action, and encouraging high-caliber growth in the modern period. In the No. 1 central document (Opinions of the CPC Central Committee and the State Council on comprehensively promoting rural revitalization and accelerating agricultural and rural modernization) published in 2021, the central government proposed the creation of leisure and entertainment as well as boutique routes for rural tourism (Xie et al., 2022). In 2021, the National Rural Revitalization Administration was officially launched, which marked the beginning of a new journey of beautiful rural construction and brought new opportunities for rural tourism development.

## 2. Literature review

With the rapid growth of China's rural tourist industry, particularly in the fringe areas of large cities, the need for effective management and decision-support systems has become increasingly evident. This research aims to develop a sophisticated model for predicting tourist flow in these areas, leveraging the power of new technologies.

(1) Big urban:

Large cities are one of the official classifications used in China for city sizes. In a November 2014 notification on modifying the rules for dividing urban scale, the State Council defined major cities as those having a permanent population of one million to five million. Large cities are classified as Type I if they have a permanent population of three to five million and Type II if they have a permanent population of one to three million. The main cities discussed in this study include megacities, vast urban areas with a population exceeding ten million. They are notable for their vast scale, high population density, and significant economic, cultural, and political influence on a global scale. Megacities often present unique infrastructure, sustainability, and governance challenges while also serving as hubs for innovation and economic growth. Type I and above, and major cities with an urban population of over three million (Wang et al., 2022). These big cities cover four municipalities directly under the Central Government, major provincial capitals, regional central cities, and prefecture-level cities with relatively developed economies. The population size of urban areas is the main criterion for defining human cities (Wu et al., 2021). As shown in Table 1, large cities with large population sizes in urban areas radiate to rural areas within their city and surrounding cities, provinces, and even the whole country.

**Table 1.** List of population proportion in urban areas of large and mega cities in China (Ministry of Housing and Urban-rural Development, 2021).

| City level | Serial number | City | Urban population (unit: Ten thousand people) | Resident population in the city | Proportion of urban population |
|---|---|---|---|---|---|
| Megacities (urban population exceeding 10 million) | 1 | Shanghai | 1987 | 2487.1 | 79.89% |
| | 2 | Beijing | 1775 | 2189.3 | 81.08% |
| | 3 | Shenzhen | 1744 | 1756 | 99.32% |
| | 4 | Chongqing | 1634 | 3205.4 | 50.98% |
| | 5 | Guangzhou | 1488 | 1867.6 | 79.67% |
| | 6 | Chengdu | 1334 | 2093.8 | 63.71% |
| | 7 | Tianjin | 1093 | 1386.6 | 78.83% |
| Megacities (urban population 5–10 million) | 8 | Wuhan | 998 | 1232.65 | 80.96% |
| | 9 | Dongguan | 956 | 1046.6 | 91.34% |
| | 10 | Xi’an | 928 | 1295 | 71.66% |
| | 11 | Hangzhou | 874 | 1193.6 | 73.22% |
| | 12 | Foshan | 854 | 949.8 | 89.91% |
| | 13 | Nanjing | 791 | 931.4 | 84.93% |
| | 14 | Shenyang | 707 | 907 | 77.95% |
| | 15 | Qingdao | 601 | 1007.1 | 59.68% |
| | 16 | Jinan | 588 | 920.2 | 63.90% |
| | 17 | Changsha | 555 | 1004 | 55.28% |
| | 18 | Harbin | 550 | 1000 | 55.00% |
| | 19 | Zhengzhou | 534 | 1260 | 42.38% |
| | 20 | Kunming | 534 | 846 | 63.12% |
| | 21 | Dalian | 521 | 745 | 69.93% |

(2) Concept of big urban fringe areas:

In 1936, Harbert Louis first put forward the concept of rural tourism in urban fringe areas from the perspective of urban ecology, which refers to the urban and rural transitional zone with the nature of urban and rural land use (Harbert, 2020). However, due to its dynamic spatial nature, there has been no unified definition standard regarding spatial scope (Mou, 2022). Urban-urban fringe has become synonymous with suburb, Desakota, urban-urban ecotone, urban-urban transition zone, urban influence area, and urban outer suburbs. These terms convey the idea of nesting or confining each other in space.

From the perspective of geography's analysis of regional spatial socio-economic characteristics, conceptual divisions can be made from static phenomena and dynamic processes. For example, Wilvin defined urban fringe areas as a transition from industrial land to agricultural land (Gao et al., 2022). Pryor believes that the urban fringe areas, located between the continuously built-up areas and the pure agricultural hinterland, are zones that change in land use and social and demographic characteristics. It has both urban and rural characteristics. Russwurm divided the urban structure into five significant parts (Gao et al., 2022):

1) Urban central area: urban core built-up area;
2) Inner fringe zone: Near the urban center, the vast majority of land has been used or planned for urban construction purposes;
3) Outer fringe zone: The characteristics of rural land use are apparent, but the urban influence has infiltrated, with independent residential and commercial networks distributed along the highway.
4) Urban shadow area: Natural landscape is less affected, but non-agricultural population and non-agricultural land exist so that you can feel the impact of the city;
5) Rural hinterland: Some urban impacts, such as establishing second homes for non-agricultural populations, mostly on land, can be observed.

Chinese scholars generally accept the concept of urban fringe, but the specific description and conceptual definition of urban fringe still need to be unified. Many scholars have elaborated on this concept from different perspectives (Xu et al., 2022). GU Chaolin thinks that the urban fringe areas are situated on the outskirts of the urban built-up region based on the features of the large urban fringe areas. This region naturally becomes a transitional zone between urban and rural economies; from the perspective of community type, it is a zone that connects the city and the countryside. Urban fringe areas are divided into inner fringe areas and outer fringe areas. The inner fringe area is close to the city (Wang et al., 2021). It is strongly affected by urban Diffusion and is closely linked with the central city, which reflects the functions of the city, such as residence, production, and commercial services. There are many high-tech industrial development zones and new areas, and the municipal infrastructure and social service system are relatively complete (Qin et al., 2022). The outer zone exhibits more rural landscape characteristics, gathering relatively independent satellite small towns, agricultural industrial parks, and tourism main functional areas. Zhou Jie put forward the view of "Desakota" and believed that the evolution of urban suburbs from urban to rural environments is an integral part of the urban regional structure. The Desakota is the most intricate and constantly evolving rural and urban development area. According to experts, the suburbs of big cities have distinctions and similarities between rural and urban communities (Zhang et al., 2021). However, due to the influence and radiation of urban areas, suburban towns have formed particular land use patterns and spatial layouts, and their development characteristics and mechanisms are unique. Experts believe that the suburb of a big city is the space for the expansion of land use in the central urban area and is the transitional zone where the urban center attracts and gathers the economic hinterland and affects diffusion.

ANN (Artificial Neural Network) has strong processing power for nonlinear data, which does not rely on the relationships between variables and has no special requirements for data distribution characteristics. However, it also has its shortcomings. Firstly, as a computational method, compared to traditional time series models, it needs more theoretical support and a systematic modeling process (Zeng, 2021). The selection of parameters often needs to be obtained through repeated experiments, and the learning process usually takes time and effort and is prone to falling into local optima. Secondly, the method requires a large amount of data for training. In foreign tourism passenger flow prediction, ANN has shown superiority in prediction. However, this is often difficult to achieve in China because the information on tourism scenic spots in China started relatively late, and the genuinely complete preserved tourism data samples are small and limited. Hence, there are certain limitations in tourism passenger flow prediction. Finally, the explanatory power of ANN technology could be more substantial, and it cannot explain tourism passenger flow nicely from the perspective of economic theory (Li et al., 2022). It cannot analyze and predict short-term fluctuations and seasonal issues of passenger flow well, and due to the slow learning process of neural networks and poor adaptability to emergencies, it cannot provide more suggestions and assistance for policy evaluation and decision-making; this further limits the breadth and depth of ANN's application in predicting tourist traffic. So, it is necessary to build a professional prediction system based on the characteristics of passenger flow changes and then combine ANN technology for input and output training to improve prediction accuracy and interpretability.

## 3. Research method and design

This study addresses the challenge of developing a scientifically sound and rational algorithm for accurately predicting short-term tourist flow in rural tourism areas on the outskirts of major cities. The research is anchored in the fringe areas of large cities, with a specific focus on rural tourism. Case studies are conducted in the Chongming District of Shanghai, Huangpi District of Wuhan, and Longquanyi District of Chengdu. The methodology employs a single machine-learning model and an ensemble of machine-learning models to forecast short-term tourist flow. The main topics of this study are the combinations of theoretical exploration and empirical research, qualitative and quantitative analysis, and theoretical exploration and applied research. The particular techniques used are as follows:

(1) Literature research method

The literature review method involves systematically organizing and analyzing existing literature to identify and understand key concepts such as metropolitan fringe areas, rural tourism, and short-term tourism dynamics. This method is crucial for understanding the development context, current state, and challenges of rural tourism in urban fringe areas. The review mainly focuses on analyzing various forecasting algorithms, exploring their relationship with rural tourism promotion, and establishing a logical framework for this study.

(2) Comprehensive analysis method

The comprehensive analysis method applies a multidisciplinary approach to developing rural tourism in China's periphery. It encompasses leisure and sightseeing travel within the theoretical transition frameworks, multidisciplinary integration, and rural rejuvenation policy. The study utilizes geography, economics, scientific management, mathematics, and statistics as guiding theories to support rural tourism's growth and enhance the precision of short-term visitor flow estimates. The aim is to investigate the rise of intelligent rural tourism and contribute innovative concepts to the field.

(3) Quantitative research method

Through the statistical yearbook, tourism big data, literature review, and other methods to obtain data necessary for research, based on field investigations of primary data and the analysis of secondary data, make full use of quantitative analysis methods such as time series algorithms, grey theory algorithms, and the Backpropagation neural network algorithm, and comprehensively employ algorithm comparison, optimization, and combination. By comparing the predicted values of different algorithms of rural tourist flow in the fringe areas of big cities in China, the algorithm model with the highest accuracy is obtained, which makes the relevant research conclusions more scientific and convincing.

**Table 2.** 2022 China tourist cities ranking (China Tourism Network, 2022).

| Rank | City | District | Tier | Population | Passenger index | Scenic index | Traffic index |
|---|---|---|---|---|---|---|---|
| 1 | Shanghai | Eastern | Megacity | 2475.89 | 0.9845 | 0.9745 | 0.9745 |
| 2 | Beijing | Eastern | Megacity | 1912.80 | 0.9726 | 0.9732 | 0.9456 |
| 3 | Wuhan | Central | Megacity | 1080.64 | 0.9645 | 0.9745 | 0.9475 |
| 4 | Hangzhou | Eastern | Megacity | 1002.14 | 0.9345 | 0.9448 | 0.9177 |
| 5 | Chengdu | Western | Megacity | 1257.14 | 0.9169 | 0.9141 | 0.9085 |
| 6 | Chongqing | Western | Megacity | 1617.50 | 0.9159 | 0.9018 | 0.9061 |
| 7 | Tianjin | Eastern | Megacity | 1160.07 | 0.9049 | 0.9037 | 0.8985 |
| 8 | Guangzhou | Eastern | Megacity | 1367.78 | 0.8847 | 0.8931 | 0.8785 |
| 9 | Dongguan | Eastern | Megacity | 1080.44 | 0.8749 | 0.8847 | 0.8681 |
| 10 | Shenzhen | Eastern | Megacity | 1766.18 | 0.9011 | 0.8537 | 0.8985 |

In our study, we utilized a refined backpropagation (BP) neural network to predict tourist flows in rural areas. The model's clarity and understandability are enhanced by detailing its multilayer perceptron (MLP) architecture, parameter initialization with random weights and zero biases, and the selection of key hyperparameters like learning rate and momentum through experimentation. Our iterative training process, using mean squared error as the loss function, concluded when validation performance plateaued or met iteration limits. This concise explanation aims to deepen readers' grasp of our model's predictive capacity.

In 2023, the China Tourism Network released 2022 China's Tourism Cities list. Combined with the classification of China's big cities in the "City Yearbook 2022" of the Ministry of Housing and Urban-rural Development, the popularity of tourism in the top ten big cities was obtained. Shanghai, Wuhan, and Chengdu are megacities in eastern, central, and western China. These three cities are not only representative of regional big cities but also have the urban function of driving the development of suburbs (marginal areas), and rural tourism is representative.

Shanghai is the birthplace of the Communist Party of China, a national center city, a megacity, the center of its metropolitan area, a well-known historical and cultural city in China, and a global first-tier city. It is also a municipality directly under the control of the People's Republic of China. Shanghai has become a significant worldwide trade, finance, shipping, and economic hub. It also serves as the foundation for a scientific and technical innovation center that will impact the world. Shanghai has a total size of 6340.5 square kilometers. Shanghai has 24.7589 million permanent residents as of 2022. Located at the entrance of the Yangtze River, Chongming District covers an area of 1413 square kilometers, about 80 kilometers from the city center, and has a population of 672,900. In the 1950s, Chongming District was transferred from Jiangsu Province to Shanghai to ensure Shanghai's self-sufficiency in agricultural products. Chongming District has been positioned to develop the agricultural economy for a long time. Although it is the largest district in Shanghai, its economic aggregate has consistently ranked low. In 2015, the state put forward the strategy of poverty alleviation and rural revitalization; Chongming District began to pay attention to the use of agricultural industry advantages to vigorously develop rural tourism and gradually build Chongming Hongqiao Flower Township Scenic spot, Chongming District Natural Echo, Chongming District Yuan Yi Garden, and other 3A scenic spots, in order to improve the local economy and people's happiness.

Hubei Province's capital, Wuhan, is a megacity, a sub-provincial city, and a test region for extensive innovation and transformation. It is an essential comprehensive transportation hub of water, land, and air, the economic and geographical center of China, one of nine central cities in the country, one of the four central cities of science and education in China, and one of the three intellectually intensive areas in the country. By the end of 2021, the city will be in charge of 13 districts totaling 8569.15 square kilometers in size, with 13,648,900 people living there permanently. In 2021, the city's gross regional product reached 1.77 trillion yuan, ranking ninth in the country's urban GDP. Huangpi District (referred to as "Huangpi") is located north of Wuhan, about 40 km from the city center. It covers an area of 2256.7 km$^2$, accounting for about a quarter of Wuhan's land area. Huangpi is the largest and best ecological urban area in Wuhan and the core area of the Yangtze River New City and Wuhan Airport Economic Zone. Since the 1980s, tourism has been used as a supporting industry to create several 5A-level scenic spots, which is the highest rating given to tourist attractions in China's grading system, such as Mulan Lake, Mulan Mountain, and Mulan Grassland, which have been the industrial basis of rural tourism for many years.

The metropolitan area is a strategic tourism innovation and development platform, with many tourism development opportunities, a large tourism industry, a relatively perfect industrial system, and apparent tourism development effects. The fringe area of the big city is a critical geographical unit of the metropolitan area. On the fringes of big cities, rural scenic spots have the unique advantage of being close to tourist source markets and directly radiated by economic and social developments. Therefore, they are the critical bearing areas of rural tourism. Regarding spatial structure, product supply, and market consumption, rural tourist attractions in big cities' fringe areas have remarkable characteristics.

The research methods described in this paper are meticulously applied and reflected throughout the study, ensuring a robust and systematic approach to investigating tourist flow in rural areas on the outskirts of major cities. The paper begins with a comprehensive literature review that establishes the conceptual framework and identifies the knowledge gaps that our research intends to address. This review serves as the foundation for our theoretical exploration, which is then complemented by an empirical study focusing on the case areas of Chongming District in Shanghai, Huangpi District in Wuhan, and Longquanyi District in Chengdu.

In our study's methods section, we detail our model selection, focusing on the grey model (GM) and neural networks, and explain the exclusion of regression models due to their inability to capture data's nonlinearity and seasonal trends. The GM(1,1) model's process, from data accumulation to parameter estimation, is outlined for its effectiveness with incomplete information. Our Backpropagation (BP) neural network, a multi-layer perceptron, is described in terms of architecture, weight initialization, and training dynamics, including learning rate and momentum. These concise explanations aim to clarify each model's utility and constraints for complex data analysis.

The qualitative insights gained from the literature review are deepened through a quantitative research method that collects and analyzes data from various sources, including statistical yearbooks and big tourism data. This approach allows us to quantify trends and patterns in tourist flow, providing a numerical basis for our predictions. The paper employs time series algorithms, grey theory algorithms, and Backpropagation neural network algorithms to process the data, reflecting a multifaceted quantitative analysis that enhances the accuracy of our forecasting model.

The integration of theoretical exploration with empirical research is further exemplified by applying both a single machine-learning model and an ensemble of machine-learning models. This dual approach showcases the paper's commitment to leveraging cutting-edge techniques to improve the predictive power of our models. Comparing these models' predicted values against actual tourist flow data is a critical component of the paper, demonstrating our methods' practical application and validation.

Moreover, the paper's focus on interdisciplinary knowledge and technology integration is evident in how we harness insights from geography, economics, scientific management, mathematics, and statistics to inform our analysis. This multidisciplinary approach is reflected in the theoretical underpinnings of our study and in the practical steps we take to collect, analyze, and interpret data.

In summary, the research methods outlined in this paper are deeply embedded in every investigation stage, from the initial conceptualization of the problem to the final presentation of our findings. The paper's structure and content are designed to demonstrate the systematic application of these methods, ensuring that our conclusions are well-supported and our predictions are both scientifically sound and practically relevant.

# 4. Empirical research and results

## 4.1. An empirical study of passenger flow prediction based on residual time series-GM(1,1) model

In 1948, Wiener's seminal work cybernetics was published, symbolizing the birth of cybernetics. Wiener realized that information processing and transmission are integral to all control and communication systems (Li et al., 2022). A communication system can continuously transmit information with different ideological contents according to people's needs, and an automatic control system must adjust its movement according to changes in the surrounding environment. These points confirm the foundation of information and feedback in cybernetics. It highlights that data that communication and control systems receive is information and fits into a specific statistical distribution. In order to generate a statistically satisfactory action for a class of statistically expected inputs, the communication and control system's structure must also adjust to this statistical quality. Cybernetics usually refers to the problem or field of study as a system and uses a shade of color to describe how much information the system contains. People often refer to the system information that can be mastered entirely as a white system, which cannot be mastered entirely as a black system, providing an inspiring research idea for creating the grey system theory.

Professor Deng (1985), a Chinese academic, introduced the grey system theory in 1982. This new branch of control theory and system engineering discipline is founded on mathematical theory. It is the theoretical result of fusing operations research and control theory and applying cybernetics theories and methodologies to social and economic systems. The primary objective of this activity is to find mathematical relationships and changes in variables or between variables based on distinctive data of ecological, social, economic, and other systems. Abstract systems can be analyzed, modeled, predicted, controlled, and judgments made by stem theory. It has several uses and has emerged as a new transdisciplinary field for academics to study and modify objective systems.

The primary distinction between the white and black systems, which comprise the grey system, is whether or not the system's components have a clear link. The white system possesses all the information needed to be quantitatively defined, has clear laws governing its development and change, and its internal components are related. In contrast, the internal characteristics of the black system are entirely unknown. Some information can be obtained in a grey system, and the relationship between various factors within the system could be more precise, so it is not easy to achieve a quantitative description. Generally, original data is processed with the help of the grey system theory to obtain newly generated numbers, weaken the randomness of original data, and find rules in new data columns so that the grey system becomes clear and gradually whitening. The system has characteristics of "small sample, poor information."

Set the original number lists $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}$, the original sequence $x^{(0)}$ for an accumulator, then get a new set of sequence $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$, where $x^{(0)}(k) = \sum_{i=a}^{k} x^{(0)}(i)\ (k = a, \dots, n)$, define the grey derivative of $x^{(1)}$ as:

$$dx^{(1)}(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1) \tag{1}$$

Make $z^{(1)}$ is a sequence of $x^{(1)}$ adjacent to mean sequence, namely:

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2,3, \dots, n$$

then $z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)\}$.

The differential equation model for GM(1,1) is defined as $dx^{(1)}(k) + ax^{(1)}(k) = b$.

$$x^{(0)}(k) + ax^{(1)}(k) = b \tag{2}$$

where $x^{(0)}(k)$ is the grey derivative, a is the development coefficient, $z^{(1)}(k)$ is the whitening background value, and b is the grey action.

Set $k$ = 2, 3, $n$ generation into the equation is:

$$\begin{cases} x^{(0)}(2) + ax^{(1)}(2) = b \\ x^{(0)}(3) + ax^{(1)}(3) = b \\ \dots\dots \\ x^{(0)}(n) + ax^{(1)}(n) = b \end{cases} \tag{3}$$

Set $Y = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T$, then

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots\dots & \dots\dots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{4}$$

If $Y$ is a data vector, $B$ is a data matrix, and $u$ is a parameter vector, then the GM(1,1) model can express the matrix equation $Y = Bu$. It can be obtained by the least square method:

$$u' = (a', b')^T = (B^T B)^{-1} B^T Y \tag{5}$$

The grey differential equation of GM(1,1) is obtained for the white differential equation:

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = b \tag{6}$$

when $t = t_0$, $x^{(1)} = x^{(1)}(t_0)$ white solution of differential equation for $x^{(1)}(t) = \left[x^{(1)}(t_0) - \frac{b}{a}\right] e^{-a(t-t_0)} + \frac{b}{a}$, the discrete values of equal interval sampling is $x^{(1)}(k+1) = [x^{(1)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a}$. To estimate $a'$ and $b'$ into $x^{(1)}(k+1) = [x^{(1)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a}$, time corresponding equation is obtained as follows:

$$x'^{(1)}(k+1) = [x^{(1)}(1) - \frac{b'}{a'}]e^{-a'k} + \frac{b'}{a'}$$

when $k$ = 1, 2, 3 ... when $n - 1$, the fitting value is the $x'^{(1)}(k+1)$ calculated by the *XX* formula. When $k \geq n$, $x'^{(1)}(k+1)$ is the predicted value. This is the fit relative to the one-time cumulative sequence reduced by the subtraction operation; when $k$ = 1, 2, 3, ..., $n - 1$, the fitting value of the original sequence $x^{(0)}$ can be obtained: $x'^{(0)}(k+1)$; when $k \geq n$, the predicted value of the original sequence $x^{(0)}$ can be obtained.

## 4.2. Empirical analysis of the traditional GM(1,1) model

Shanghai, Wuhan, and Chengdu are the largest cities in the east, central, and west of China, respectively, in terms of tourist popularity, and their changes in passenger flow are very representative. Therefore, the statistical data of tourist flow during the National Day Golden Week from 2000 to 2018 of Chongming District in Shanghai, Huangpi District in Wuhan, and Longquanyi District in Chengdu were adopted. The classic GM(1,1) model was adopted to forecast and analyze the changes in tourist flow from 2019 to 2020.

**Table 3.** Statistical table of actual tourist population during the National Day Golden Week in Chongming, Huangpi, Longquanyi from 2000 to 2020.

| Year | Chongming | Huangpi | Longquanyi |
|---|---|---|---|
| 2000 | 93,725 | 284,823 | 92,311 |
| 2001 | 97,870 | 290,519 | 96,124 |
| 2002 | 102,870 | 296,330 | 98,451 |
| 2003 | 107,255 | 302,256 | 110,004 |
| 2004 | 112,205 | 311,324 | 110,463 |
| 2005 | 128,333 | 320,664 | 110,571 |
| 2006 | 133,333 | 330,284 | 119,143 |
| 2007 | 143,333 | 346,798 | 166,857 |
| 2008 | 170,100 | 364,138 | 191,714 |
| 2009 | 188,333 | 382,344 | 274,000 |
| 2010 | 220,000 | 420,579 | 331,429 |
| 2011 | 421,667 | 462,637 | 391,714 |
| 2012 | 527,083 | 508,900 | 451,429 |
| 2013 | 632,500 | 763,351 | 521,657 |
| 2014 | 674,667 | 839,686 | 547,743 |
| 2015 | 716,833 | 1,007,623 | 553,943 |
| 2016 | 759,000 | 1,229,322 | 568,943 |
| 2017 | 801,167 | 1,695,652 | 552,571 |
| 2018 | 822,250 | 1,955,467 | 592,029 |
| 2019 | 843,333 | 2,100,287 | 657,543 |
| 2020 | 905,140 | 2,168,125 | 737,114 |

We found that the overall trend of Chongming, Huangpi, and Longquanyi data was upward with apparent volatility. The National Day Golden Week tourist flow in these three regions is developed based on time series, and factors affecting are complicated. Many factors are involved in dynamic change, so it is not easy to quantify them. The National Day Golden Week tourist flow in these three regions is a grey quantity containing known and unknown information, which has obvious grayness and conforms to the general characteristics of data processed by the GM(1,1) model. The grey GM(1,1) model is a single-variable first-order linear model. GM(1,1) models are usually built using time series data of the prediction object itself. We selected the National Day Golden Week tourist flow of Shanghai Chongming District, Wuhan Huangpi District, and Chengdu Longquanyi District from 2000 to 2018 as the initial data in **Table 4**. We established a GM(1,1) model for these data.

The data in **Table 4** are accumulated yearly to form a cumulative generation sequence (**Table 4**). The cumulatively generated sequence is defined as $x^{(1)}$, and the smoothness ratio is calculated.

**Table 4.** Cumulative generating sequence (unit: person).

| Year | Chongming | | | Huangpi | | | Longquanyi | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Tourism flow** | **Sequence** | **Ratio** | **Tourism flow** | **Sequence** | **Ratio** | **Tourism flow** | **Sequence** | **Ratio** |
| 2000 | 93,725 | 93,725 | - | 284,823 | 284,823 | - | 92,311 | 92,311 | - |
| 2001 | 97,870 | 191,595 | 0.54 | 290,519 | 575,342 | 1.02 | 96,124 | 188,435 | 1.04 |
| 2002 | 102,870 | 294,465 | 0.36 | 296,330 | 871,672 | 0.52 | 98,451 | 286,886 | 0.52 |
| 2003 | 107,255 | 401,720 | 0.28 | 302,256 | 1,173,928 | 0.35 | 110,004 | 396,890 | 0.38 |
| 2004 | 112,205 | 513,925 | 0.25 | 311,324 | 1,485,252 | 0.27 | 110,463 | 507,353 | 0.28 |
| 2005 | 128,333 | 642,258.3 | 0.21 | 320,664 | 1,805,916 | 0.22 | 110,571 | 617,924.4 | 0.22 |
| 2006 | 133,333 | 775,592.7 | 0.18 | 330,284 | 2,136,200 | 0.18 | 119,143 | 737,067.3 | 0.19 |
| 2007 | 143,333 | 918,925 | 0.19 | 346,798 | 2,482,998 | 0.16 | 166,857 | 903,924.4 | 0.23 |
| 2008 | 170,100 | 1,089,025 | 0.17 | 364,138 | 2,847,136 | 0.15 | 191,714 | 1,095,639 | 0.21 |
| 2009 | 188,333 | 1,277,358 | 0.17 | 382,344 | 3,229,480 | 0.13 | 274,000 | 1,369,639 | 0.25 |
| 2010 | 220,000 | 1,497,358 | 0.28 | 420,579 | 3,650,059 | 0.13 | 331,429 | 1,701,067 | 0.24 |
| 2011 | 421,667 | 1,919,025 | 0.27 | 462,637 | 4,112,696 | 0.13 | 391,714 | 2,092,782 | 0.23 |
| 2012 | 527,083 | 2,446,108 | 0.26 | 508,900 | 4,621,596 | 0.12 | 451,429 | 2,544,210 | 0.22 |
| 2013 | 632,500 | 3,078,608 | 0.22 | 763,351 | 5,384,947 | 0.17 | 521,657 | 3,065,867 | 0.21 |
| 2014 | 674,667 | 3,753,275 | 0.19 | 839,686 | 6,224,633 | 0.16 | 547,743 | 3,613,610 | 0.18 |
| 2015 | 716,833 | 4,470,108 | 0.17 | 1,007,623 | 7,232,256 | 0.16 | 553,943 | 4,167,553 | 0.15 |
| 2016 | 759,000 | 5,229,108 | 0.15 | 1,229,322 | 8,461,578 | 0.17 | 568,943 | 4,736,496 | 0.14 |
| 2017 | 801,167 | 6,030,275 | 0.14 | 1,695,652 | 10,157,230 | 0.20 | 552,571 | 5,289,067 | 0.12 |
| 2018 | 822,250 | 6,852,525 | 0.12 | 1,955,467 | 12,112,697 | 0.19 | 592,029 | 5,881,096 | 0.11 |

As can be seen from **Table 4**, more than 90% of the cumulative generated sequence smoothness ratio is less than 2, satisfying the exponential growth law, so the GM(1,1) model can be established.

1) Construct differential equations

Build GM(1,1) differential equation for $x^{(1)}$:

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = b \tag{7}$$

where $a$ is the development parameter of the model, and $b$ is the coordination coefficient.

According to the formula $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3, \dots, n$, establish the next-mean sequence $z^{(1)}$ for $x^{(1)}$:

$$z^{(1)} = \left\{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(21)\right\} \tag{8}$$

The grey differential equation was established to obtain the matrix of Chongming, Huangpi, and Longquanyi, respectively:

$$\begin{bmatrix} 97870 \\ 102870 \\ \dots \\ 822250 \end{bmatrix} = \begin{bmatrix} -142660 & 1 \\ -243030 & 1 \\ \dots & \dots \\ -6441400 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{9}$$

$$\begin{bmatrix} 290519 \\ 296330 \\ \dots \\ 2168125 \end{bmatrix} = \begin{bmatrix} -430083 & 1 \\ -723507 & 1 \\ \dots & \dots \\ -15297047 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} 96124 \\ 98451 \\ \dots \\ 592029 \end{bmatrix} = \begin{bmatrix} -140373 & 1 \\ -237661 & 1 \\ \dots & \dots \\ -5585082 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{11}$$

2) Calculate parameters $a$ and $b$

By the least square method, we can calculate $(a, b)^T = (B^T B)^{-1} B^T Y a$ and obtain the development coefficient a and coordination coefficient $b$.

### 4.3. Empirical analysis based on residual time series-GM(1,1) model

Since the prediction model only summarizes the rules of the original data in Chongming, Huangpi, and Longquanyi, the relative error between the original data and predicted data is extensive, so it is necessary to optimize the model. In this chapter, grey modeling is planned for the residual part, and modeled data is added with the original predicted value to form a new predicted value to improve prediction accuracy.

1) Data analysis

According to **Table 5**, Chongming's residual value was positive in specific years and negative in 2013. In some years, Huangpi was positive; in others, it was negative in 2011, 2012, and 2013. Longquanyi was positive in 2012–2024 and harmful in other years. There are positive and negative residual values in three regions, indicating noise in the data. Therefore, it is necessary to analyze and model the residual sequence to make up for incomplete data and the influence of noise and then improve the prediction accuracy. In order to increase the prediction model's accuracy, this article used the grey residual GM(1,1) model to estimate the passenger flow in Chongming, Huangpi, and Longquanyi during the National Day Golden Week.

**Table 5.** Residual value of tourist flow during the National Day Golden Week in Huangpi District from 2000 to 2018.

| | | **Chongming** | | | **Huangpi** | | | **Longquanyi** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Year** | **Population** | **Estimate** | **Ratio** | **Residual** | **Estimate** | **Ratio** | **Residual** | **Estimate** | **Ratio** | **Residual** |
| 2000 | 93,725 | 93,725 | | 0 | 284,823 | | 0 | 92,311 | | 0 |
| 2001 | 97,870 | 287,390 | 193.64% | −189,520 | 203,446 | 29.97% | 87,073 | 92,710 | 3.55% | 3414 |

**Table 5.** (*Continued*).

| | | Chongming | | | Huangpi | | | Longquanyi | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Year** | **Population** | **Estimate** | **Ratio** | **Residual** | **Estimate** | **Ratio** | **Residual** | **Estimate** | **Ratio** | **Residual** |
| 2002 | 102,870 | 170,585 | 65.83% | −67,715 | 126,737 | 57.23% | 169,593 | 164,668 | 67.26% | −66,217 |
| 2003 | 107,255 | 191,464 | 78.51% | −84,209 | 147,013 | 51.36% | 155,243 | 181,379 | 64.88% | −71,375 |
| 2004 | 112,205 | 214,900 | 91.52% | −102,695 | 170,533 | 45.22% | 140,791 | 199,785 | 80.86% | −89,322 |
| 2005 | 128,333 | 241,203 | 87.95% | −112,870 | 197,816 | 38.31% | 122,848 | 220,059 | 99.02% | −109,488 |
| 2006 | 133,333 | 270,727 | 103.05% | −137,393 | 229,463 | 30.53% | 100,821 | 242,391 | 103.45% | −123,248 |
| 2007 | 143,333 | 303,864 | 112.00% | −160,531 | 266,174 | 23.25% | 80,624 | 266,988 | 60.01% | −100,131 |
| 2008 | 170,100 | 341,057 | 100.50% | −170,957 | 308,758 | 15.21% | 55,380 | 294,082 | 53.40% | −102,368 |
| 2009 | 188,333 | 382,803 | 103.26% | −194,469 | 358,155 | 6.33% | 24,189 | 323,925 | 18.22% | −49,925 |
| 2010 | 220,000 | 429,658 | 95.30% | −209,658 | 415,455 | 1.22% | 5124 | 356,797 | 7.65% | −25,368 |
| 2011 | 421,667 | 482,248 | 14.37% | −60,582 | 481,922 | 4.17% | −19,285 | 393,004 | 0.33% | −1290 |
| 2012 | 527,083 | 541,276 | 2.69% | −14,192 | 559,022 | 9.85% | −50,122 | 432,886 | 4.11% | 18,542 |
| 2013 | 632,500 | 607,528 | 3.95% | 24,972 | 648,458 | 15.05% | 114,893 | 476,815 | 8.60% | 44,842 |
| 2014 | 674,667 | 681,890 | 1.07% | 7223 | 752,202 | 10.42% | 87,484 | 525,202 | 4.12% | 22,541 |
| 2015 | 716,833 | 765,354 | 6.77% | −48,520 | 872,544 | 13.41% | 135,079 | 578,499 | 4.43% | 24,557 |
| 2016 | 759,000 | 859,033 | 13.18% | −100,033 | 1,012,138 | 17.67% | 217,184 | 637,205 | 12.00% | −68,262 |
| 2017 | 801,167 | 964,179 | 20.35% | −163,013 | 1,174,066 | 30.76% | 521,586 | 701,868 | 27.02% | −149,297 |
| 2018 | 822,250 | 1,082,195 | 31.61% | −259,945 | 1,361,900 | 30.35% | 593,567 | 773,094 | 30.58% | −181,065 |

2) Data select

We start with data converted from positive and negative numbers to take the residual value (yellow data) and again use the GM(1,1) model for modeling. For example, since 2013 in Huangpi, the residual values have been positive, and there is no data noise, so this case will select data from 2014 to carry out residual sequence analysis. The data were accumulated yearly to form the cumulative residual sequence table in **Table 6**.

**Table 6.** Residual cumulative sequence.

| | Chongming | | Huangpi | | Longquanyi | |
|---|---|---|---|---|---|---|
| **Year** | **Modelable residuals** | **Generate date** | **Modelable residuals** | **Generate date** | **Modelable residuals** | **Generate date** |
| 2013 | | | −114,893 | −114,893 | | |
| 2014 | 7223 | 7223.157113 | −87,484 | −202,377 | | |
| 2015 | 48,520 | 55,743.32726 | −135,079 | −337,456 | 24,557 | 24,557 |
| 2016 | 100,033 | 155,776.5095 | −217,184 | −554,640 | 68,262 | 92,819 |
| 2017 | 163,013 | 318,789.1468 | −521,586 | −1,076,226 | 149,297 | 242,116 |
| 2018 | 259,945 | 578,734.5141 | −593,567 | −1,669,793 | 181,065 | 423,181 |

3) Data processing

Accumulated year by year to form a cumulative generation sequence (**Table 7**). The cumulatively generated sequence is defined as $x^{(1)}$ and calculates the smoothness ratio.

**Table 7.** Residual sequence smoothness ratio.

| | Chongming | | | Huangpi | | | Longquanyi | | |
|---|---|---|---|---|---|---|---|---|---|
| **Year** | **Modelable residuals** | **Generate date** | **Ratio** | **Modelable residuals** | **Generate date** | **Ratio** | **Modelable residuals** | **Generate date** | **Ratio** |
| 2013 | - | - | - | −114,893 | −114,893 | - | - | - | - |
| 2014 | 7223 | 7223 | - | −87,484 | −202,377 | 0.76 | - | - | - |
| 2015 | 48,520 | 55,743 | 6.72 | −135,079 | −337,456 | 0.67 | 24,557 | 24,557 | - |
| 2016 | 100,033 | 155,777 | 1.49 | −217,184 | −554,640 | 0.64 | 68,262 | 92,819 | 2.78 |
| 2017 | 163,013 | 318,789 | 1.05 | −521,586 | −1,076,226 | 0.94 | 149,297 | 242,116 | 1.48 |
| 2018 | 259,945 | 578,735 | 0.82 | −593,567 | −1,669,793 | 0.55 | 181,065 | 423,181 | 0.75 |

**Table 7** shows that the smoothness ratio of most cumulatively generated sequences is less than 2, satisfying the exponential growth law, so the GM(1,1) model can be established.

**Figures 1–3** serve the purpose of visually representing the accuracy of the passenger flow predictions made using the residual GM(1,1) model for the rural tourism areas of Chongming, Huangpi, and Longquanyi, respectively.

**Figure 1** aims to illustrate the comparison between the predicted values generated by the residual GM(1,1) model and the actual passenger flow data for Chongming. This visual comparison helps to assess the model's predictive performance in Chongming by displaying how closely the predicted values align with the actual data points. See **Figure 1** as follow.

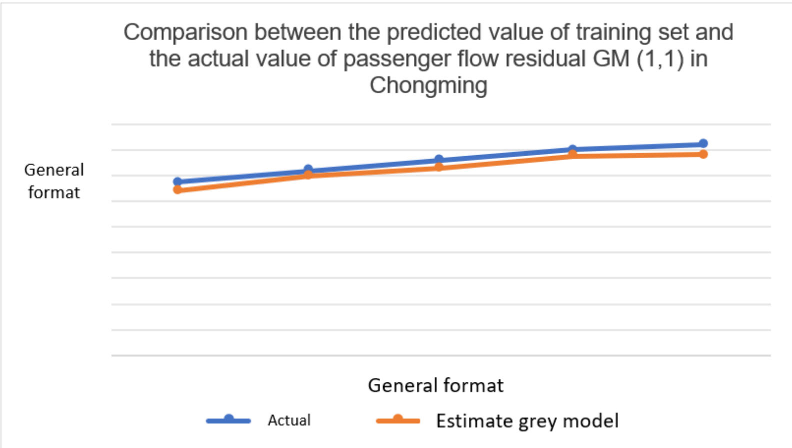


**Figure 1.** Comparison between the predicted value of the training set and the actual value of passenger flow residual GM(1,1) in Chongming.

**Figure 2** presents the same type of analysis for Huangpi, showing the model's predicted values against the actual observed data. This figure is crucial for evaluating the residual GM(1,1) model's effectiveness in accurately forecasting tourist flow for the Huangpi District. See **Figure 2** as follow.

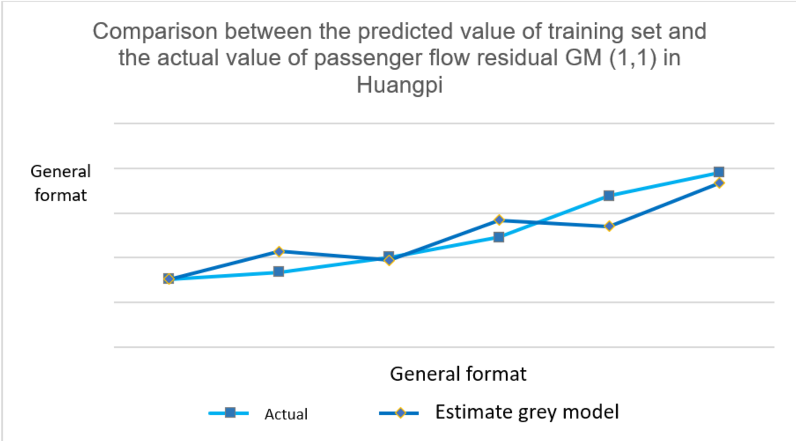


**Figure 2.** Comparison between the predicted value of training set and the actual value of passenger flow residual GM(1,1) in Huangpi.

**Figure 3** does the same for Longquanyi, providing a visual representation of the model's predictive accuracy in this district. By comparing the predicted values with the actual data, Figure 3 allows readers to gauge the model's reliability and precision in the context of Longquanyi. See **Figure 3** as follow.

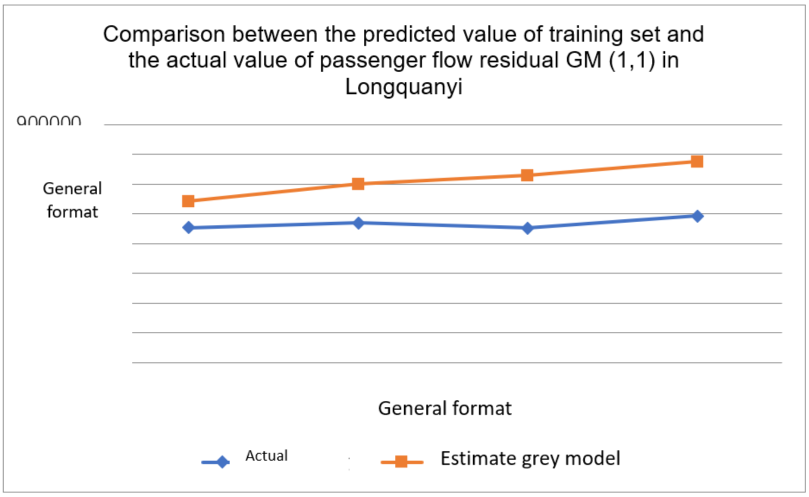


**Figure 3.** Comparison between the predicted value of the training set and the actual value of passenger flow residual GM(1,1) in Longquanyi.

## 4.4. An empirical study of passenger flow prediction based on BP neural network model

The global neural network has been an active marginal interdisciplinary field in recent years. Its research has surged again, on the one hand, due to the significant progress made in neuroscience itself. On the other hand, there is also an urgent need to develop new computers and expand new ways of artificial intelligence.

So far, in many problems that need artificial intelligence to solve, the human brain is far more intelligent than a computer. To create a new generation of intelligent computers, we must understand the human brain and study the information processing mechanism of the human brain neural network system. Besides, the artificial neural network model developed based on neuroscience research results reflects some essential characteristics of human brain function and opens up a new way for computers to use neural networks. It is a powerful challenge to traditional computer structure and artificial intelligence and has attracted significant attention from experts in various fields.

## 5. Discussion and conclusion

The findings of this study offer significant insights into the dynamics of short-term tourist flow in rural tourism areas on the outskirts of major cities. Through the application of advanced machine learning models, our research has successfully predicted tourist flow with a high degree of accuracy. The multi-input BP neural network model demonstrated the best performance, with a prediction error rate below ±5.5%, closely followed by the improved GM(1,1) model, which showed a prediction difference rate of less than ±15%. These results underscore the efficacy of employing sophisticated algorithms in tourism flow prediction, particularly in rural tourism areas that are increasingly gaining popularity.

Mean Squared Error(MSE) is a common measure used to evaluate the performance of a predictive model. We compared predictive models using MSE() and $R^2$, applied $k$-fold cross-validation, and conducted $t$-tests and $F$-tests for statistical significance. Visualizing results with charts and graphs, we integrated these analyses to assess model performance. The neural network and residual time series models excelled in accuracy over traditional models, offering valuable insights for tourism flow forecasting and management decisions.

This study is significant because it can guide strategic planning and resource allocation in rural tourism management. By providing more precise predictions, tourism authorities can better anticipate tourist influxes, thereby ensuring a higher standard of service and a more sustainable tourism environment. Moreover, the models' high predictive accuracy contributes to a more effective response to peak tourist seasons, potentially preventing overcrowding and resource strain.

However, with every research comes the opportunity for further refinement. It is recommended that future work explores the incorporation of additional variables into the models, such as economic indicators, social events, and environmental factors, which may influence tourist flow. Additionally, the models could be enhanced through continuous learning algorithms that adapt to new data over time, improving their predictive power as more information becomes available.

Furthermore, the models should be tested across rural tourism contexts to evaluate their generalizability and robustness. This would provide a comprehensive understanding of their applicability and effectiveness in various scenarios, ultimately enriching the predictive modeling framework for the broader field of tourism studies.

In summary, this study presents a significant methodological advancement in predicting rural tourist flow, with clear implications for improving tourism management practices. The suggested improvements aim to build upon these findings, ensuring that the models remain at the cutting edge of predictive analytics and continue to offer valuable insights into rural tourism.

**Author contributions:** Conceptualization, NY and CS; methodology, NY; software, NY; validation, NY formal analysis, NY; resources, CS; data curation, NY; writing—review and editing, NY; visualization, NY; supervision, CS; project administration, CS; funding acquisition, CS. All authors have read and agreed to the published version of the manuscript.

**Conflict of interest:** The authors declare no conflict of interest.

## References

Chen, X., & Cong, D. (2022). Application of improved algorithm based on four-dimensional ResNet in rural tourism passenger flow prediction. Journal of Sensors, 2022, 1–8. https://doi.org/10.1155/2022/9675647

Chen, X., Huang, Y., & Chen, Y. (2023). Spatial pattern evolution and influencing factors of tourism flow in China's Chengdu—Chongqing economic circle. ISPRS International Journal of Geo-Information, 12(3), 121. https://doi.org/10.3390/ijgi12030121

Gan, C., Voda, M., Wang, K., et al. (2021). The spatial network structure of the tourism economy in urban agglomeration: A social network analysis. Journal of Hospitality and Tourism Management, 47, 124–133. https://doi.org/10.1016/j.jhtm.2021.03.009

Gao, Y., Nan, Y., & Song, S. (2022). High-speed rail and city tourism: Evidence from Tencent migration big data on two Chinese golden weeks. Growth and Change, 53(3), 1012–1036. https://doi.org/10.1111/grow.12473

Li, Y., Gong, G., Zhang, F., et al. (2022). Network structure features and influencing factors of tourism flow in rural areas: Evidence from China. Sustainability, 14(15), 9623. https://doi.org/10.3390/su14159623

Liu, C., Qin, Y., Wang, Y., et al. (2022). Spatio-temporal distribution of tourism flows and network analysis of traditional villages in Western Hunan. Sustainability, 14(13), 7943. https://doi.org/10.3390/su14137943

Ma, X., Yang, Z., & Zheng, J. (2022). Analysis of spatial patterns and driving factors of provincial tourism demand in China. Scientific Reports, 12(1), 2260. https://doi.org/10.1038/s41598-022-04895-8

Mou, J. (2022). Extracting network patterns of tourist flows in an urban agglomeration through digital footprints: The case of the greater Bay area. IEEE Access: Practical Innovations, Open Solutions, 10, 16644–16654. https://doi.org/10.1109/access.2022.3149640

National Tourism Administration of the People's Republic of China. (2022). China Tourism Network, 2022 City Yearbook, 14-15

Qin, X., Li, X., Chen, W., et al. (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. https://doi.org/10.1080/10941665.2023.2166420

Ruan, W. Q., & Zhang, S.-N. (2021). Can tourism information flow enhance regional tourism economic linkages? Journal of Hospitality and Tourism Management, 49, 614–623. https://doi.org/10.1016/j.jhtm.2021.11.012

Tang, Y. (2022). Discrete dynamic modeling analysis of rural revitalization and ecotourism sustainable prediction based on big data. Discrete Dynamics in Nature and Society, 2022, 1–9. https://doi.org/10.1155/2022/9158905

The Ministry of Housing and Urban-rural Development. (2021). 2021 Urban Construction Statistical Yearbook. 21-23

Wang, L., Wu, X., & He, Y. (2021). Nanjing's intracity tourism flow network using cellular signaling data: A comparative analysis of residents and non-local tourists. ISPRS International Journal of Geo-Information, 10(10), 674. https://doi.org/10.3390/ijgi10100674

Wang, Y., Chen, H., & Wu, X. (2021). Spatial structure characteristics of tourist attraction cooperation networks in the Yangtze River Delta are based on tourism flow. Sustainability, 13(21), 12036. https://doi.org/10.3390/su132112036

Wu, S., Wang, L., & Liu, H. (2021). Study on tourism flow network patterns on May Day Holiday. Sustainability, 13(2), 947. https://doi.org/10.3390/su13020947

Xie, X., Zhang, L., Sun, H., et al. (2021). Spatiotemporal difference characteristics and influencing factors of urbanization in China's major tourist cities. International Journal of Environmental Research and Public Health, 18(19), 10414. https://doi.org/10.3390/ijerph181910414

Xie, Y., Meng, X., Cenci, J., & Zhang, J. (2022). Spatial pattern and formation mechanism of rural tourism resources in China: Evidence from 1470 national leisure villages. ISPRS International Journal of Geo-Information, 11(8), 455. https://doi.org/10.3390/ijgi11080455

Xu, D., Zhang, J. H., Huang, Z., et al. (2022). Tourism community detection: A space of flows perspective. Tourism Management, 93, 104577. https://doi.org/10.1016/j.tourman.2022.104577

Zeng, B. (2021). Pattern of Chinese tourist flows in Japan: A social network analysis perspective. In: Tourism Spaces. Routledge. pp. 42–64.

Zhang, H., Duan, Y., & Han, Z. (2021). Research on spatial patterns and sustainable development of rural tourism destinations in the Yellow River Basin of China. Land, 10(8), 849. https://doi.org/10.3390/land10080849