Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(6): 16187–16202.
DOI: 10.3934/math.2024783
Received: 21 February 2024
Revised: 22 March 2024
Accepted: 29 March 2024
Published: 08 May 2024

Research article

Optimization study of tourism total revenue prediction model based on the Grey Markov chain: a case study of Macau

Xiaolong Chen, Hongfeng Zhang and Cora Un In Wong*

Faculty of Humanities and Social Sciences, Macao Polytechnic University, Rua de Luís Gonzaga Gomes, Macau, China

* **Correspondence:** Email: corawong@mpu.edu.mo; Tel: +85385993263.

Abstract: The GM (1,1) model, grounded in gray system theory, utilizes first-order cumulative data for forecasting. While offering simplicity and efficiency, its applicability is confined to such data. In light of the constraints inherent in the conventional gray GM (1,1) prediction model when confronted with stochastic data fluctuations, the residual correction methodology was deployed to enhance the predictive efficacy of the GM (1,1) model. Subsequently, an augmented model underwent refinement through the application of the Markov chain, giving rise to a sophisticated and optimized gray Markov chain prediction model. The efficacy of this novel model was substantiated through a case study involving the prediction of Macao's aggregate tourism revenue. A comparative analysis was conducted between the outcomes generated by the traditional gray prediction model, those of the refined prediction model, and the empirical data pertaining to tourism. This scrutiny validated the proficiency and precision of the optimized prediction model. The process of model optimization manifested a discernible enhancement in both predictive accuracy and stability, thereby broadening the prospective applications of gray prediction models. This endeavor aspired to furnish a scientifically grounded point of reference for the advancement of tourism within the Guangdong-Hong Kong-Macao Greater Bay Area and, indeed, throughout China. Moreover, it introduced a fresh methodology that held promise as a decision-making support mechanism for the developmental trajectory of Macao's tourism industry.

Keywords: stochastic fluctuation; residual correction; Markov; gray prediction; Macao
Mathematics Subject Classification: 62M05, 62M10, 62P20

1. Introduction

As the financial prosperity of Chinese citizens undergoes rapid augmentation, the landscape of tourism consumption simultaneously evolves in a continuous process of refinement [1–3]. Capitalizing on its industry-induced propulsive influence, tourism has emerged as a nascent focal point for China's investment and consumption, assuming a pivotal role as a significant economic growth-supporting sector [4–6]. Based on statistical data, it is projected that in the upcoming year of 2023, the tourism sector's contribution to China's gross domestic product will rise significantly to an impressive 11.04%. To elaborate further, domestic tourism within the country is forecasted to reach an astonishing 5.539 billion visits, accompanied by a total tourism revenue anticipated to exceed 5 trillion yuan. This manifests a year-on-year surge of 10.76% and 12.3%, respectively. Of particular note, the inbound tourism sector is envisaged to host 141 million travelers, with associated consumption projected to reach an impressive 127.1 billion US dollars. This signifies incremental increments of 1.2% and 3.0%, correspondingly. Notwithstanding these promising indicators, the trajectory of tourism consumption enhancement and the swift evolution of the tourism sector engender a succession of novel scenarios. The developmental milieu is characterized by a constant state of flux, perpetually reshaping the landscape of tourism. An expanding array of factors increasingly influences its unfolding narrative.

Forecasting encapsulates the systematic analysis, estimation, and inference of future developmental trajectories of phenomena, employing judiciously curated historical data and relevant methodologies [7,8]. Forecasting is omnipresent in governmental decision-making, corporate governance, industrial safety protocols, and the mitigation of natural disasters, among other spheres. The evolution of the tourism sector is similarly intertwined with the discipline of prediction. The enhancement of predictive models holds profound implications for the advancement of regional economies [9]. It has garnered considerable scholarly attention and yielded discernible outcomes. Predominantly employed techniques for prognosticating tourism revenue primarily encompass the application of multiple linear regression [10], gray system forecasting model [11], BP neural network model [12], and time series forecasting model [13]. For example, Internet search data was introduced to predict tourism revenue [14]; based on the gray system theory, a gray model for tourist source prediction was constructed [15], and the prediction accuracy was compared with commonly used linear models; based on the BP neural network, a BP neural network model of tourism demand was established [16]. Tourism foreign exchange revenue and the number of inbound tourists were predicted and analyzed. There are also studies that analyze and predict the international tourism source market and its foreign exchange revenue based on certain statistical data [17]. The aforementioned inquiries and models have generated certain predictive insights regarding foreign exchange income from tourism. Nevertheless, there is a pressing imperative for further refinement to augment the precision and accuracy of predictions.

In summary, drawing upon the research efforts delineated by the aforementioned scholars and integrating insights gleaned from the unique stochastic fluctuations evident in historical tourism data, a nuanced enhancement is proposed. Grounded in the traditional gray GM (1,1) prediction model, this enhancement aspires to fortify the stability and precision inherent in the predictive framework. To accomplish this, an expansion of the predictive dataset is envisaged, coupled with a meticulous rectification of the residual sequence within the GM (1,1) model prediction process. A further layer of refinement is introduced through the application of the Markov chain, whereupon the residual sequence undergoes additional correction. This iterative process aims to optimize both the accuracy and stability of the prediction model. Substantiation of this optimization unfolds through meticulous example verification, attesting to the model's efficacy in navigating the intricate dynamics of tourism prediction. Ultimately, utilizing the Macao tourism market as a paradigmatic exemplar, we juxtapose the prognostications generated by the traditional gray prediction model with those produced by the enhanced predictive model. This comparative analysis encompasses empirical data pertaining to tourism, thereby validating the efficacy and precision inherent in the optimized predictive framework. The objective is to provide a refined scientific reference that aims to enrich the intricate development of tourism within the expansive fabric of the Guangdong-Hong Kong-Macao Greater Bay Area, and across the broader expanse of China.

2. Literature review

In the domain of regional analysis, an exploration of tourism income is conducted through a panoramic lens, encompassing perspectives that range from the national level to specific provinces, individual cities, or distinctive scenic locales. The taxonomy of tourism income, meticulously dissected by scholarly luminaries, delves into the categories of international (inbound) revenue, domestic tourism income, and cumulative earnings derived from tourism. Within the scholarly discourse, the trajectory of investigation traverses the multifaceted terrain of factors influencing tourism revenue, temporal and spatial variations, and the methodologies that encompass prediction and statistical analyses.

2.1. Research on factors affecting tourism income

Through the prism of regression analysis and allied methodologies, esteemed scholars posit that the tapestry of tourism income intricately weaves itself with the developmental tapestry of transportation, the accommodation industry, the culinary domain, and the broader commodity economy. Delving into the realm of the gray correlation analysis, research contends that the predominant determinant influencing tourism income resides in the per capita disposable income of urban residents' households [18]. Furthermore, within the related survey domains and drawing upon tourism-related data amassed from numerous provinces across China, the predominant determinants of tourism income are posited to lie within the economic sphere, with household income level serving as a pivotal metric [19]. Furthermore, a plethora of scholarly works have delved into the determinants shaping tourism income within specific regions, employing the sophisticated methodology of gray relational analysis. Through the utilization of the partial least squares (PLS) regression analysis model and other advanced techniques, it has been discerned that the robustness of transportation infrastructure exerts a pivotal influence on tourism income. Notably, heightened transportation convenience is identified as a catalyst for the augmentation of local tourism revenue [20]. Clearly, due to the subtle variations in research subjects, geographical contexts, and methodological frameworks, disparate conclusions emerge regarding the factors influencing the intricate fabric of tourism income.

2.2. Research on spatial and temporal differences in tourism economy based on tourism income

Certain erudite scholars have embarked upon an investigation into the spatiotemporal differentials inherent in the tourism economy, focusing exclusively on the realm of international tourism revenue. Employing the nuanced frameworks of gravity models and two-dimensional combination matrices, these scholars have meticulously delved into the spatial dispositions of China's tourism resources, their geographic locations, and the concomitant international tourism revenue [21]. Certain pertinent scholars have employed methodological tools such as standard deviation, coefficient of variation, and relative development rate to meticulously scrutinize the intricate spatiotemporal differentials within the domain of international tourism revenue across the provincial-level administrative regions in mainland China. Through the construction of a spatial effect model specifically tailored to the contours of China's inbound tourism, a discerning revelation has surfaced: There are marked and steadfast regional variances, coupled with spatial agglomeration, in the panorama of international tourism revenue among diverse provinces and cities within the nation [22]. Related investigations also reveal that, based on the dynamics encapsulated within the change index delineating the ratio of regional international tourism revenue to the Chinese average, discernible differentials in regional competition pertaining to international tourism revenue emerge across mainland China [23].

The aforementioned literature delves into the intricacies of spatiotemporal disparities within the realm of tourism revenue and economic dynamics. It predominantly traverses the temporal spectrum, elucidating the dynamic shifts in scale, structure, and regional variations of tourism revenue. This comprehensive exploration seeks to discern alterations and patterns within the tourism economy during both off-peak and peak seasons. The narrative extends further, proposing strategic interventions for the advancement of the tourism economy, encompassing considerations of developmental trajectories, regional collaborations, and more. In the spatial dimension, the inquiry spans the entirety of the nation or specific regions, utilizing tourism revenue and associated data to scrutinize variances in regional tourism economic development, spatial dislocations, and other phenomena. Causative factors, including resource endowment, geographical location, transportation infrastructure, and developmental influences, contribute to the pervasive spatiotemporal variations in tourism revenue and economy. While existing literature predominantly employs quantitative methodologies to analyze and compare spatiotemporal disparities in regional tourism revenue, a noticeable gap exists in discussions concerning the theoretical underpinnings for the coordinated and balanced development of regional tourism economies. Few studies engage in a comprehensive quantitative examination of the root causes behind these spatiotemporal differences, and some neglect to elaborate on the theoretical implications of their quantitative analyses. Moreover, the practical applicability of these insights is often obscured, diminishing their operational efficacy

2.3. Research on tourism revenue forecast analysis

The discussion pertaining to tourism income and its myriad influencing factors has increasingly become a focal point of scholarly investigation, producing a continuum of novel research findings. In the realm of tourism revenue prediction, a plethora of regression models and artificial neural network architectures have been meticulously crafted. Deliberations within this domain demand meticulous consideration of the numerous factors that influence tourism revenue, including the discerning selection of explanatory variables and a thorough assessment of data availability and relevance. Notably, the establishment of a time series model is not an omnipresent prerequisite in such inquiries. In the pursuit of advanced methodologies for forecasting tourism revenue, the calibration of predictive efficacy is often measured through error rates, embracing metrics such as mean absolute error and root mean square error. A noteworthy example lies in research employing the partial least squares method and the gray correlation method, wherein the congruence of the correlation rankings between multiple influencing factors and tourism revenue remains consistently aligned under both methodologies [24]. Certain learned scholars, exploring the temporal dimension, have undertaken projections of regional tourism income for the forthcoming five years using gray prediction models. Furthermore, they have conducted a penetrating analysis of the major factors influencing tourism income through the methodological prism of principal component analysis, yielding notable results [25]. Certain scholars have employed sophisticated statistical tools such as SPSS and EVIEWS software, harnessing the potency of multiple linear regression methods for the meticulous curation of cross-sectional data. These studies have revealed that pivotal determinants of tourism income encompass the influx of tourists, the abundance of dining establishments, pricing dynamics of scenic spot tickets, and the expanse of urban green spaces. Furthermore, these scholars have adeptly leveraged these insights to prognosticate future trajectories of tourism income through nuanced analysis [26,27]. Furthermore, there are relevant studies that propose the integration of neural networks with the Grey-Markov model to forecast tourism demand. Grounded in a comprehensive collection of tourism revenue data from recent years, these studies illustrate the ability to extrapolate projections for the upcoming years. Concurrently, a comparative analysis of the primary factors influencing tourism revenue is conducted, elucidating the complex correlation between each factor and the overarching indicators of tourism revenue [28].

3. Materials and methods

3.1. Data collection

The Macau Special Administrative Region, abbreviated as Macau, graces the southeastern fringes of mainland China, nestled on the western shores of the Pearl River Delta. It shares proximate borders with Guangdong Province, situated 42 nautical miles from Hong Kong, 145 kilometers distant from Guangzhou, and in close adjacency to the cities of Zhuhai and Zhongshan [29]. The Macau region encompasses the Macau Peninsula and two peripheral islands, namely Taipa and Coloane, constituting a collective expanse covering 30 square kilometers. Connectivity between the Macau Peninsula and Taipa is facilitated by three meticulously designed Macau-Taipa Bridges. The topography is predominantly undulating, with reclaimed land occupying the flat expanses. Situated within the subtropical monsoon domain, the climatic milieu is characterized by warmth, humidity, and copious rainfall. The strategic vision outlined in the 2019 "Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area" aspires for Macau to evolve into a globally renowned tourist destination. This commitment underscores dedication to fostering high-quality development and facilitating seamless regional integration within the tourism industry [30,31]. This facilitates the fluidity of tourism elements within the interstices of internal cities, catalyzing the profound maturation of the networked attributes inherent in the Bay Area urban agglomeration. This evolution unfolds into a multi-faceted and comprehensive new paradigm.

The "Outline of the Reform and Development Plan for the Pearl River Delta Region (2008–2020)" envisions the transformation of Macao into a globally distinguished hub for tourism and leisure, seamlessly integrated into the broader urban tourism fabric of Guangdong, Hong Kong, and Macao [32]. In a concerted effort, Guangdong, Hong Kong, and Macao ardently propel the journey of tourism integration, orchestrating multifaceted route operations under the banner of "one-journey multi-station." This endeavor seeks to seamlessly amalgamate the exceptional tourism and cultural wealth strewn across the landscapes of Guangdong, Hong Kong, and Macao, thereby birthing and championing top-tier tourism itineraries, both domestically and globally.

3.2. Data source

The total tourism revenue indicator data mainly comes from the "Macao Statistical Yearbook" from 2009 to 2020. Please also refer to the relevant information on the website of the Macau Statistics and Census Bureau, the Macau Tourism Bureau and the Macau Environmental Protection Bureau.

3.3. Statistical methods

3.3.1. GM (1,1) model theory

When employing grey theory for predictive modeling, the most frequently utilized model is the grey GM (1,1) model. Its principle involves initially accumulating the original sequence, followed by processing the accumulated sequence with a first-order exponential equation to derive a differential equation. Subsequently, the least squares method is employed to obtain the parameters of the differential equation. Upon conducting first-order accumulation to formulate a prediction model, the GM (1,1) model of the original data column is ultimately derived through the first-order accumulation (see [33,34]). The steps are as follows:

(1) First-order accumulation

Let the original sequence X^0,

$$X^0 = \left\{X_1^0, X_2^0, \cdots, X_n^0\right\}. \tag{1}$$

Perform first-order accumulation on X^0 to get X^1,

$$X^1 = \left\{X_1^1, X_2^1, \cdots, X_n^1\right\}, \tag{2}$$

where the value of X_k^1 is:

$$X_k^1 = \sum_{i=1}^{k} X_i^0, \quad k = 1, 2, \cdots, n. \tag{3}$$

(2) Establish a gray prediction model

The least squares method is used to establish a GM (1,1) gray prediction model, and the result is:

$$\hat{X}_{k+1}^1 = (X_1^0 - \frac{b}{a}) \times e^{-ak} + \frac{b}{a}, \tag{4}$$

where k=1,2,…,n.

(3) Get the predicted value

Restore the original sequence and get the predicted value:

$$\hat{X}_{k+1}^0 = \hat{X}_{k+1}^1 - \hat{X}_k^1. \tag{5}$$

3.3.2. Residual correction

The above GM (1,1) prediction results are not stable enough and the accuracy is not high enough. The residuals generated by the model need to be further corrected [35]. The specific process is as follows:

(1) Calculate the residual sequence

The residual sequence is calculated by the formula, represented by E_k^0:

$$E_k^0 = X_k^0 - \hat{X}_k^0, \tag{6}$$

where k=1,2,…,n.

Then the normalized residual sequence is obtained, represented by E_k^1:

$$E_k^1 = | E_k^0 | = | X_k^0 - \hat{X}_k^0 |, \tag{7}$$

where k=1,2,…,n.

(2) Calculate the modified residual sequence

Fit the normalized residual sequence through initial conditions and background values:

$$\begin{cases} \hat{E}_{k+1}^1 = (m - \frac{a}{b}) \times E^{-ak} + \frac{a}{b}, \\ \hat{E}_{k+1}^0 = (1 - e^{-1}) \times (m - \frac{a}{b}) \times E^{-ak}, \end{cases} \tag{8}$$

where k=1,2,…,n.

In the formula, m is the value of the leading term calculated using the optimal squares method of the normalized residual sequence. After fitting, the residual correction sequence is obtained:

$$\hat{E}_k^0 = I(K) \times \hat{E}_k^0. \tag{9}$$

Among them, when E_k^0>0, $I(K)$=1; when E_k^0<0, $I(K)$=-1.

(3) Get the predicted value

The prediction equation is obtained using the residual correction sequence:

$$\hat{X}_k^1 = \hat{X}_{k+1}^1 + \hat{E}_{k+1}^1, \tag{10}$$

where k=1,2,…,n.

3.3.3. Gray Markov chain prediction model

Based on the aforementioned residual correction sequence and in conjunction with Markov chain theory, an optimized prediction model of the Grey Markov chain is formulated to further enhance the stability and accuracy of predictions [36,37]. The specific steps are as follows:

(1) Divide the residual state

According to the residual size in the formula, it is divided into m states, each state is represented by Si, among them: $S_i \in [L_i, H_i]$. L_i and H_i represent the upper and lower boundaries of the i-th state:

$$\begin{cases} L_i = \min E_k^i + (\max E_k^i - \min E_k^i), \\ H_i = \min E_k^i + (\max E_k^i - E_k^i), \end{cases} \tag{11}$$

where k=1,2,…,n; i=1,2,…,m.

(2) Construct a state transition probability matrix

The residual sequence E is divided into m states. To construct a turntable transition probability matrix, the matrix elements are represented by P_i^j. The calculation process is as follows:

$$P_i^j = \frac{M_i^j}{M_i}. \tag{12}$$

Among them, M_i^j is the number of transitions from state E_i to state E_j in one step, and M_i is the number of data at the beginning of the turntable. After obtaining the values of the state transition matrix elements, generate a one-step transition probability matrix P:

$$P = \begin{bmatrix} P_{11}^{1} \; P_{12}^{1} \cdots P_{1m}^{1} \\ P_{21}^{1} \; P_{22}^{1} \cdots P_{2m}^{1} \\ \vdots \quad \vdots \quad \vdots \; \vdots \\ P_{m1}^{1} \; P_{m2}^{1} \cdots P_{mm}^{1} \end{bmatrix}. \tag{13}$$

(3) Get the predicted value

According to the state transition matrix, the predicted value is obtained:

$$\hat{E}_{t+1}^{0} = \hat{E}_{t}^{0} + \sum_{i=1}^{m} u_i(t) \times v_i . \tag{14}$$

Among them, $u_i(t)$ represents the probability of transforming from the previous state to the next stage in the gray residual sequence, and the parameter t is the transition time. v_i represents the midpoint of the interval.

4. Example verification

The retrospective decade of Macao's tourism revenue serves as the corpus for our analysis, with data meticulously sourced from the esteemed Macao Statistics and Census Bureau. Utilizing Macao's comprehensive tourism revenue data as our illustrative exemplar, we undertake a modeling endeavor that leverages the historical revenue dataset spanning the last ten years. The pristine historical data is elegantly presented in Table 1.

Table 1. Input data of Macao's total tourism revenue from 2010 to 2019 (unit: US$ million).

Years	Total tourism revenue
2010	4833
2011	5674
2012	30126
2013	51632
2014	43412
2015	31620
2016	31155
2017	36595
2018	41478
2019	41166

Based on the empirical data presented in Table 1, both the conventional GM (1,1) prediction model and the residual-corrected grey prediction model were employed to forecast Macao's total tourism revenue post-2011. The ensuing predictions, residual analyses, and comparisons with actual revenue are delineated in Table 2.

Table 2. Predicted values and residuals of the two prediction models.

Years	GM (1,1) model			Prediction model after correction of residual values			Actual value
	Predictive value	Residual	Fluctuation range	Predictive value	Residual	Fluctuation range	
2011	5674.23	0.23	0.7356	5674.91	0.91	0.92	5674
2012	30119.44	-6.56	-7.6397	30119.77	-6.23	-6.87	30126
2013	51638.63	6.63	7.8362	51638.88	6.88	8.92	51632
2014	43416.26	4.26	6.1156	43416.77	4.77	5.79	43412
2015	31625.38	5.38	7.0388	31625.81	5.81	5.93	31620
2016	31161.89	6.89	7.6397	31162.29	7.29	10.65	31155
2017	36592.44	-2.56	3.5851	36591.07	-3.93	-3.66	36595
2018	41474.24	-3.76	1.3376	41475.29	-2.71	-0.75	41478

As discerned from Table 2, the prediction outcomes following residual optimization surpass those of the conventional GM (1,1) prediction model in accuracy, exhibiting diminished residuals and reduced fluctuation amplitude values [38,39]. The prediction results of the two prediction models for Macao's total tourism revenue in 2019 are 41116.538 billion yuan and 4116847 million yuan respectively, while the total tourism revenue value of Macao in 2019 is 4116.6 billion yuan. It can be seen that with the increase of data, the fluctuation value of the prediction results is basically a decreasing trend, but when the fluctuation range changes, the residual-corrected prediction model has higher accuracy and stability than the prediction results of the original GM (1,1) prediction model.

Next, the Markov modified gray prediction model is used to further predict Macao's total tourism revenue [40,41]. The process is as follows.

4.1. Divide the residual state

According to the residual values in Table 2, the number of residual states is divided into 3, which represent the three states of trend slowdown, normal growth and rapid growth. S1, S2 and S3 are used respectively, and their intervals are:

$$S1 \in [-6.23, 0.91],$$

$$S2 \in [0.91, 4.77],$$

$$S3 \in [4.77, 7.29].$$

4.2. Construct a state transition matrix

Based on the residual-corrected data elucidated in Table 2 and further informed by the residual state division, the residuals therein are categorized into three distinct states. This categorization is meticulously detailed in Table 3.

Table 3. Macao's total tourism revenue forecast and residual status division.

Years	Predictive value	Residual	Status division
2011	5674.42	0.42	S1
2012	300119.12	-6.88	S1
2013	51638.73	6.73	S3
2014	43416.13	4.13	S2
2015	31625.66	5.66	S3
2016	31161.87	6.87	S3
2017	36592.47	-2.53	S1
2018	41474.28	-3.72	S1

Compute the state transition matrix and integrate it with the state division presented in Table 3 to derive the state transition matrix P for Macao's total tourism revenue:

$$P=\begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{15}$$

4.3. Get the predicted value

Combined with the state transition matrix, the predicted value is obtained. The prediction results are shown in Table 4.

Table 4. Gray Markov model prediction results.

Years	Predictive value	Actual value
2011	5674.16	5674
2012	300125.8	30126
2013	51635.13	51632
2014	43413.76	43412
2015	31623.47	31620
2016	31158.32	31155
2017	36593.65	36595
2018	41476.33	41478

4.4. Prediction of prediction model

Upon juxtaposing the predicted and actual data presented in Tables 2 and 4, a comprehensive assessment of the three prediction models in relation to the factual outcomes is undertaken. The varied results of this comparative analysis are meticulously delineated in Table 5.

Table 5. Prediction results and actual values of the three prediction models.

Years	GM (1,1) model	Residual correction	Gray Markov chain	Actual value
2011	5674.23	5674.91	5674.16	5674
2012	30119.44	30119.77	30125.8	30126
2013	51638.63	51638.88	51635.13	51632
2014	43416.26	43416.77	43413.76	43412
2015	31625.38	31625.81	31623.47	31620
2016	31161.89	31162.29	31158.32	31155
2017	36592.44	36591.07	36593.65	36595
2018	41474.24	41475.29	41476.33	41478
2019	41165.38	41168.47	41166.06	41166

Table 5 reveals that the forecasts generated by the grey Markov chain demonstrate an impressive fidelity to the actual data. Particularly noteworthy are the predictions for the years 2011 and 2018, which, despite significant data volatility, exhibit a striking alignment with the observed actual values. This emphasizes the enhanced accuracy and precision of the optimized model, especially when dealing with data predictions marked by substantial fluctuations.

4.5. Evaluation of predictive models

In the evaluation of the economic forecasting model, we embarked on a comparative analysis delineating the disparities between the predictive outcomes of the Macao tourism economic forecast model and the actual values. This investigation incorporates the concept of mean square error (MSE) to elucidate the magnitude of deviation between the projected results and the authentic values.

The mean square error serves as an expedient metric for quantifying the "average error", providing insights into the magnitude of data fluctuations. Categorized within the domains of prediction evaluation and prediction amalgamation, its nomenclature can be deconstructed thus: The "mean" denotes the arithmetic average, encapsulating the collective value; "variance" quantifies the disparity between random variables and their probabilistic estimates, representing a measure of deviation from the mean; and "error" signifies the discrepancy between the measured value and its true counterpart.

The mean square error (MSE) stands as a ubiquitous metric for evaluating errors and quantifying disparities between predicted and actual values. Its calculation method unfolds thus:

Suppose there are n samples, yi represents the true value of the i-th sample and represents the predicted value of the i-th sample. Then the calculation formula for mean square error MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{16}$$

The resultant value represents MSE, epitomizing the average of squared discrepancies between predicted and actual values. In an ideal scenario where predicted values align perfectly with actual values, the MSE would equate to zero, signifying impeccable accuracy in the model's predictions. A diminutive MSE signifies closer alignment of the model's predictive outcomes with actual values, whereas an elevated MSE indicates a greater divergence between the model's projections and the true values.

Upon juxtaposing the predictive outcomes with the actual data, as delineated in Tables 2 and 4, the test MSE results for the prediction model are elucidated in Table 6.

Table 6. Comparison of three prediction models prediction models.

Type	GM (1,1) model	Residual correction	Gray Markov chain
Mean square error (MSE)	0.68	0.47	0.31

Analysis of Table 6 reveals that the optimized gray Markov prediction model manifests a diminished MSE, indicative of reduced volatility in the predictive data [42,43]. Through this refined gray Markov prediction model, traditional challenges can be more effectively addressed. The GM (1,1) prediction model is inherently constrained, particularly in scenarios where time series exhibit pronounced volatility, leading to potentially subpar predictive outcomes [44,45]. The optimized model exhibits greater efficacy in forecasting Macao's total tourism expenditure, thereby broadening the applicability of the gray prediction model. Concurrently, it enhances both the precision and stability of predictions, offering novel insights for prognosticating and fostering the development of the Macao tourism market.

5. Conclusions

Following the assessment of the predictive outcomes derived from the trio of methodologies applied to the tourism economy, a nuanced evaluation of their respective strengths and weaknesses emerges:

(1) GM (1,1) model: Rooted in the principles of gray system theory, this model employs first-order ordinary differential equations to delineate and forecast the trajectory of systems characterized by incomplete information. Relative to its counterparts, the gray prediction model is characterized by its simplicity of implementation and operational ease. It excels particularly when confronted with limited data sets, yielding expeditious results. However, it presupposes a linear relationship among the data, potentially overlooking intricate nonlinear trends and consequently compromising its precision.

(2) Residual prediction: This approach entails a meticulous examination of the discrepancy or error between projected and actual values to refine and rectify the predictive model. It augments the fidelity of initial forecasts and facilitates iterative enhancements to the predictive framework. By assimilating prediction errors, it bolsters the model's robustness. Nonetheless, this method is not devoid of drawbacks. It may necessitate augmented computational resources during the forecasting phase. Moreover, its efficacy is contingent upon the initial forecast model; a flawed foundational model may limit the efficacy of residual corrections. If misapplied, this method can potentially amplify noise within the data, leading to overfitting.

(3) Gray Markov chain prediction: This approach synergistically amalgamates gray system theory with Markov chain theory to prognosticate future values predicated on prevailing states and transition probabilities. Its salient strength lies in its adeptness at assimilating historical data and transition probabilities for forecasting. It exhibits proficiency in navigating systems riddled with incomplete or nebulous data. Moreover, it tends to furnish more precise predictions for systems characterized by irregular or fluctuating patterns.

In summary, the GM (1,1) model, while straightforward, grapples with limitations in encapsulating nonlinear trends. Residual forecasting, although capable of refining predictions, is reliant on the integrity of the initial model [46]. On the other hand, gray Markov chain forecasting proffers heightened accuracy, albeit necessitating a deeper theoretical acumen and potentially imposing elevated computational demands.

Venturing beyond the confines of conventional economic development paradigms, this study delves into the intrinsic limitations of the classical gray prediction GM (1,1) model. Through a refined approach employing residual correction techniques and Markov processes, we have meticulously optimized the GM (1,1) model, rectifying its inherent constraints and addressing the instability in predictive outcomes. This enhanced model was subsequently deployed to forecast Macao's aggregate tourism revenue. Comparative analyses were conducted between the predictions derived from our refined GM (1,1) model, the conventional GM (1,1) model, and actual revenue values. Such rigorous validation substantiates the efficacy and precision of our novel model. The distinct advantages of this refined predictive framework equip Macao's tourism management authorities with more robust forecasting data. This empowers them to judiciously allocate the region's finite tourism resources, facilitating the provision of unparalleled tourism experiences for both domestic and international visitors. Concurrently, this initiative reinforces Macao's aspiration to evolve into a preeminent global hub for tourism and leisure. In doing so, we aim to catalyze the judicious diversification of Macao's economy. By adopting this differentiated developmental trajectory, Macao can bolster its tourism sector, thereby laying a foundation for sustainable economic diversification.

The deployment of this innovative methodology introduces a fresh perspective to the simulation and forecasting research concerning tourism revenue. It furnishes a more scientifically rigorous forecasting mechanism, laying a robust foundation for the sustainable evolution of the tourism sector. This encompasses the strategic development of tourism infrastructure and the diversification of tourism products. Concurrently, this method holds considerable significance as a guiding framework for tourism management agencies and governmental departments, offering valuable insights and references for their respective operational and policy-making endeavors.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

No potential conflicts of interest were reported by the authors.

References

1. H. He, S. Tuo, K. Lei, A. Gao, Assessing quality tourism development in China: an analysis based on the degree of mismatch and its influencing factors, *Environ. Dev. Sustain.*, **26** (2023), 9525–9552. https://doi.org/10.1007/s10668-023-03107-1
2. N. Tan, S. Anwar, W. Jiang, Intangible cultural heritage listing and tourism growth in China, *J. Tour. Cult. Change*, **21** (2023), 188–206. https://doi.org/10.1080/14766825.2022.2068373
3. Q. Chen, The impact of economic and environmental factors and tourism policies on the sustainability of tourism growth in China: evidence using novel NARDL model, *Environ. Sci. Pollut. Res.*, **30** (2023), 19326–19341. https://doi.org/10.1007/s11356-022-22925-w

4. T. Deng, W. Zhao, Y. Hu, Retirement and household tourism consumption-a case study in China, *Tour. Econ.*, **29** (2023), 1055–1073. https://doi.org/10.1177/13548166221090177
5. D. M. Gonzalez-Perez, J. M. M. Martín, J. M. G. Martínez, A. M. Pachón, Analyzing the real size of the tourism industry on the basis of an assessment of water consumption patterns, *J. Bus. Res.*, **157** (2023), 113601. https://doi.org/10.1016/j.jbusres.2022.113601
6. D. Sanjaya, M. Arief, N. J. Setiadi, P. Heriyati, Research on green tourism intention: a bibliometric analysis, *J. Syst. Manage. Sci.*, **13** (2023), 159–185. https://doi.org/10.33168/JSMS.2023.0610
7. S. C. Haw, K. Ong, L. J. Chew, K. W. Ng, P. Naveen, E. A. Anaam, Improving the prediction resolution time for customer support ticket system, *J. Syst. Manage. Sci.*, **12** (2022), 1–16. https://doi.org/10.33168/JSMS.2022.0601
8. O. Al-Jamili, H. Ibrahim, R. Ahmad, An integrated model for predicting the user continuance intention towards utilizing open government data, *J. Syst. Manage. Sci.*, **12** (2022), 295–323. https://doi.org/10.33168/JSMS.2022.0419
9. W. G. Hiyab, T. H. Hassan, M. A. Hassanin, M. Y. Almakhayitah, The epistemological values of travel & tourism competitiveness index and its predictive powers on tourist arrivals in Africa; pls-sem approach, *Geo J. Tour. Geosites*, **49** (2023), 1046–1055. https://doi.org/10.30892/gtg.49320-1104
10. M. Khairi, D. Darmawan, The relationship between destination attractiveness, location, tourism facilities, and revisit intentions, *J. Mark. Bus. Res.*, **1** (2021), 39–50. https://doi.org/10.56348/mark.v1i1.32
11. Y. C. Hu, P. Jiang, Fuzzified grey prediction models using neural networks for tourism demand forecasting, *Comput. Appl. Math.*, **39** (2020), 145. https://doi.org/10.1007/s40314-020-01188-6
12. X. Yang, J. Zhou, D. Wen, An optimized BP neural network model for teaching management evaluation, *J. Intell. Fuzzy Syst.*, **40** (2021), 3215–3221. https://doi.org/10.3233/JIFS-189361
13. J. W. Bi, H. Li, Z. P. Fan, Tourism demand forecasting with time series imaging: a deep learning model, *Ann. Tour. Res.*, **90** (2021), 103255. https://doi.org/10.1016/j.annals.2021.103255
14. A. S. Rashad, The power of travel search data in forecasting the tourism demand in Dubai, *Forecasting*, **4** (2022), 674–684. https://doi.org/10.3390/forecast4030036
15. X. Ma, Tourism demand forecasting based on grey model and BP neural network, *Complexity*, **2021** (2021), 5528383. https://doi.org/10.1155/2021/5528383
16. L. Wang, B. Wu, Q. Zhu, Y. R. Zeng, Forecasting monthly tourism demand using enhanced backpropagation neural network, *Neural Process. Lett.*, **52** (2020), 2607–2636. https://doi.org/10.1007/s11063-020-10363-z
17. W. Shi, Y. Gong, L. Wang, N. Nikolova, Heterogeneity of inbound tourism driven by exchange rate fluctuations: implications for tourism business recovery and resilience in Australia, *Curr. Issues Tour.*, **26** (2023), 450–467. https://doi.org/10.1080/13683500.2021.2023478
18. R. Brauer, M. Dymitrow, J. Tribe, The impact of tourism research, *Ann. Tour. Res.*, **77** (2019), 64–78. https://doi.org/10.1016/j.annals.2019.05.006
19. S. B. Hojeghan, A. N. Esfangareh, Digital economy and tourism impacts, influences and challenges, *Proc. Soc. Behav. Sci.*, **19** (2011), 308–316. https://doi.org/10.1016/j.sbspro.2011.05.136
20. M. Li, J. Chen, High-speed rail network in China: the contribution of fast trains to regional tourism and economic development, *Tour. Rev.*, **75** (2020), 414–432. https://doi.org/10.1108/TR-12-2018-0197
21. V. S. Lin, Y. Yang, G. Li, Where can tourism-led growth and economy-driven tourism growth occur? *J. Travel Res.*, **58** (2019), 760–773. https://doi.org/10.1177/0047287518773919

22. N. Mou, Y. Zheng, T. Makkonen, T. Yang, J. Tang, Y. Song, Tourists' digital footprint: the spatial patterns of tourist flows in Qingdao, China, *Tour. Manage.*, **81** (2020), 104151. https://doi.org/10.1016/j.tourman.2020.104151
23. L. Xu, S. Wang, J. Li, L. Tang, Y. Shao, Modelling international tourism flows to China: a panel data analysis with the gravity model, *Tour. Econ.*, **25** (2019), 1047–1069. https://doi.org/10.1177/1354816618816167
24. S. M. Rasoolimanesh, S. M. Noor, F. Schuberth, M. Jaafar, Investigating the effects of tourist engagement on satisfaction and loyalty, *Serv. Ind. J.*, **39** (2019), 559–574. https://doi.org/10.1080/02642069.2019.1570152
25. Y. C. Hu, Developing grey prediction with Fourier series using genetic algorithms for tourism demand forecasting, *Qual. Quant.*, **55** (2021), 315–331. https://doi.org/10.1007/s11135-020-01006-5
26. A. Saayman, J. de Klerk, Forecasting tourist arrivals using multivariate singular spectrum analysis, *Tour. Econ.*, **25** (2019), 330–354. https://doi.org/10.1177/1354816618768318
27. B. Wu, L. Wang, R. Tao, Y. Zeng, Interpretable tourism volume forecasting with multivariate time series under the impact of COVID-19, *Neural Comput. Appl.*, **35** (2023), 5437–5463. https://doi.org/10.1007/s00521-022-07967-y
28. Y. C. Hu, P. Jiang, P. C. Lee, Forecasting tourism demand by incorporating neural networks into Grey-Markov models, *J. Oper. Res. Soc.*, **70** (2019), 12–20. https://doi.org/10.1080/01605682.2017.1418150
29. G. McCartney, The impact of the coronavirus outbreak on Macao. From tourism lockdown to tourism recovery, *Curr. Issues Tour.*, **24** (2021), 2683–2692. https://doi.org/10.1080/13683500.2020.1762549
30. C. Li, M. K. Ng, Y. Tang, T. Fung, From a 'world factory' to China's Bay Area: a review of the outline of the development plan for the Guangdong-Hong Kong-Macao Greater Bay Area, *Plan. Theory Pract.*, **23** (2022), 310–314. https://doi.org/10.1080/14649357.2021.1958539
31. H. Weng, J. Kou, Q. Shao, Evaluation of urban comprehensive carrying capacity in the Guangdong-Hong Kong-Macao Greater Bay Area based on regional collaboration, *Environ. Sci. Pollut. Res.*, **27** (2020), 20025–20036. https://doi.org/10.1007/s11356-020-08517-6
32. X. Ma, J. Tao, Cross-border environmental governance in the Greater Pearl River Delta (GPRD), *Int. J. Environ. Stud.*, **67** (2010), 127–136. https://doi.org/10.1080/00207231003693282
33. W. Yuhong, L. Jie, Improvement and application of GM (1,1) model based on multivariable dynamic optimization, *J. Syst. Eng. Electron.*, **31** (2020), 593–601. https://doi.org/10.23919/JSEE.2020.000024
34. H. Wang, Y. Wang, D. Wu, A new seasonal cycle GM (1,1) model and its application in railway passenger volume forecasting, *Grey Syst. Theory Appl.*, **12** (2022), 293–317. https://doi.org/10.1108/GS-11-2020-0146
35. Z. Jia, Z. Zhou, H. Zhang, B. Li, Y. Zhang, Forecast of coal consumption in Gansu Province based on Grey-Markov chain model, *Energy*, **199** (2020), 117444. https://doi.org/10.1016/j.energy.2020.117444
36. V. Roy, Convergence diagnostics for Markov chain Monte Carlo, *Ann. Rev. Stat. Appl.*, **7** (2020), 387–412. https://doi.org/10.1146/annurev-statistics-031219-041300
37. C. Nemeth, P. Fearnhead, Stochastic gradient Markov chain Monte Carlo, *J. Amer. Stat. Assoc.*, **116** (2021), 433–450. https://doi.org/10.1080/01621459.2020.1847120
38. H. R. Beyer, M. Alcubierre, M. Megevand, Stability study of a model for the Klein-Gordon equation in Kerr space-time II, *Rep. Math. Phys.,* **88** (2021), 115–143. https://doi.org/10.1016/S0034-4877(21)00059-8

39. L. Wu, X. Guo, Y. Chen, Grey relational entropy calculation and fractional prediction of water and economy in the Beijing-Tianjin-Hebei Region, *J. Math.*, **2021** (2021), 4418260. https://doi.org/10.1155/2021/4418260
40. F. Mukhamedov, A. Al-Rawashdeh, Approximations of non-homogeneous Markov chains on abstract states spaces, *Bull. Math. Sci.*, **11** (2021), 2150002. https://doi.org/10.1142/S1664360721500028
41. G. Portillo-Ramírez, H. Cruz-Suárez, R. López-Ríos, R. Blancas-Rivera, Markov decision processes approximation with coupled dynamics via Markov deterministic control systems, *Open Math.*, **21** (2023), 20230129. https://doi.org/10.1515/math-2023-0129
42. T. Lam, L. Williams, A Markov chain on the symmetric group that is Schubert positive? *Exp. Math.*, **21** (2012), 189–192. https://doi.org/10.1080/10586458.2011.579020
43. A. Ruszczyński, Risk-averse dynamic programming for Markov decision processes, *Math. Program.*, **125** (2010), 235–261. https://doi.org/10.1007/s10107-010-0393-3
44. N. Chen, Analysis of the correlation between cross-border E-commerce and economic growth based on hierarchical multilevel gray evaluation model, *J. Math.*, **2022** (2022), 8455404. https://doi.org/10.1155/2022/8455404
45. I. Ben-Yair, G. B. Shalom, M. Eliasof, E. Treister, Quantized convolutional neural networks through the lens of partial differential equations, *Res. Math. Sci.*, **9** (2022), 58. https://doi.org/10.1007/s40687-022-00354-y
46. J. Xu, K. Liu, Isomorphism problem of China's marine industries: analysis based on the gray correlation theory, *J. Interdiscip. Math.*, **21** (2018), 479–487. https://doi.org/10.1080/09720502.2017.1420577

AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)