Applied Soft Computing Journal 92 (2020) 106320

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Inbound tourism demand forecasting framework based on fuzzy time series and advanced optimization algorithm

Ping Jiang [a], Hufang Yang [b,*], Ranran Li [b], Chen Li [b]

[a] School of Statistics, Xi'an University of Finance and Economics, Xi'an, China
[b] School of Statistics, Dongbei University of Finance and Economics, Dalian, China

ARTICLE INFO

Article history:
Received 12 November 2019
Received in revised form 5 March 2020
Accepted 14 April 2020
Available online 20 April 2020

Keywords:
Inbound tourism demand
Small sample forecasting
Fuzzy time series model
Information optimization technology
Atom search optimization algorithm

ABSTRACT

The tourism industry has been integrated into the national strategic system in China. Thus, tourism demand forecasting has become a concern for the sustainable development of the tourism industry. Unfortunately, the sample size for tourism in China is always small and cannot satisfy the hypothesis test of an economic model or the data volume for a traditional time series model. In this study, a novel hybrid forecasting framework combining fuzzy time series (FTS) and an atom search optimization (ASO) algorithm is proposed for inbound tourism demand forecasting; this forecasting framework is particularly suitable for small sample sizes. Specifically, information optimization technology is applied in the FTS to improve the recognition ability of the system and effectively identify small sample information. The ASO algorithm is applied to search the optimal parameters of FTS that can further improve forecasting performance. All comparison experiments and tests verify the effectiveness and superiority of our proposed model, which provides excellent forecasting results for tourism demand and a basis for policymakers and managers to plan appropriately for the tourism market.

1. Introduction

Around the world, the development and promotion of tourism has become a source of personal income and government revenue [1]. In an increasingly competitive environment, it is important for governments and businesses to embrace demand trends in the tourism industry; this is fundamental to developing appropriate strategies for resources allocation and business investment to guarantee sustainable development in tourism [2].

Forecasting tourist volume is becoming increasingly important for predicting future economic development. However, due to the complex and evolutionary nature of the tourism market, a tourist demand series tends to be inherently noisy, conditionally non-stationary, and in some cases, deterministically chaotic. Because modeling dynamic, non-stationary demand series is challenging, a system is needed that allows for more accurate forecasts with less noise and complexity [3]. Current forecasting methods for tourism demand can be divided into econometric models and time series models.

Econometric models attempt to use related variables to explain and forecast the dependent variable based on economic theory. Econometric models include static models such as traditional regression models [4], gravity models [5], linear almost ideal demand systems [6,7], and dynamic econometric models such as vector autoregressive (VAR) models [8,9], time varying parameter models [10], and error correction models [11].

Time series models are popular in recent research because they are based only on historical tourism demand data [12]. Time series models can be divided into linear and nonlinear models. The most popular linear methods that have been successfully applied in practical applications are the exponential smoothing (ES) model [13,14] and autoregressive integrated moving averages (ARIMA) models [15,16]. In practice, linear models possess an important advantage in easy implementation and interpretation. However, when linear models perform poorly in both in-sample fitting and out-of-sample forecasting, more complex nonlinear models should be considered. Nonlinear time series models study the nonlinear relationship between historical series and predictive indicators such as artificial neural networks [17–19] and support vector regression [20–22]. It is widely believed that nonlinear methods are superior to linear methods for efficient and prudent decision-making when applied in modeling economic behavior.

Additionally, there are several combined models proposed to improve forecasting accuracy [23]. Robert R [24] combined different time length forecasts for application in tourism demand forecasting. John T. Coshall [25] combined volatility and smoothing forecasts in UK tourism demand. Kuan-Yu Chen [26] combined linear and nonlinear models for tourism demand forecasting. Haiyan Song [27] combined statistical and judgmental

* Corresponding author.
E-mail address: dufesos_1530@126.com (H. Yang).

https://doi.org/10.1016/j.asoc.2020.106320
1568-4946/

forecasts to forecast demand for Hong Kong tourism. It cannot be denied that the combined forecasting models can achieve better performance than individual models.

Econometric models always rely on economic theory. A regression equation need to tests a series of hypotheses, which is difficult to be fulfilled in actual applications. Time series models always require sufficient historical data to explore the rules and patterns in the data. Linear time series models have a poor extrapolation effect and narrow forecasting scale; nonlinear time series models are unstable, and highly dependent on data.

In practice, the sample size for tourism is always small, making it difficult to test the hypothesis. Thus, a traditional time series model may be not appropriate or accurate. Moreover, the data usually contains uncertainty and fuzziness due to limitations of resources and statistical techniques or inherent dataset characteristics. If a traditional model is applied to explain the stability and trend analysis of a time series, the model may incur causal judgment deviation and forecasting error, or increase the error between forecasting results and actual values.

Accordingly, fuzzy theory provides a powerful tool for uncertainty and incomplete or limited data. Some scholars present the concept of fuzzy time series by combining fuzzy theory and fuzzy mathematics methods that have been applied in some fields for time series forecasting [28]. Song and Chissom [29] proposed a fuzzy time series model for forecasting university enrollment. Yu [30] proposed weighted fuzzy time series models for Taiwan Stock Index forecasting. Li et al. [31] applied fuzzy time series for air quality forecasting with large fluctuations in the concentration of pollutants. Stefanakos [32] applied the fuzzy time series method to forecast non-stationary wind and wave data. There have been some advances to improve the accuracy of fuzzy time series methods. Cheng et al. [33] added the adaptive expectation model in fuzzy time series forecasting to improve forecasting performance for the Taiwan Stock Index. Sadaei et al. [34] combined fuzzy time series and a convolutional neural network to control over-fitting phenomenon for short-term load forecasting. Dincer [35] applied the fuzzy k-medoid clustering algorithm to address the outliers and abnormal observations in a fuzzy time series for air pollution forecasting.

In a fuzzy time series forecasting model, partitioning the interval length and establishing the fuzzy logic relationship are two main phases that can significantly affect forecasting performance. Huang et al. [36] applied particle swarm optimization to adjust the interval lengths in the fuzzy time series for forecasting student enrollment. Yang et al. [37] applied a multi-objective differential evolution algorithm to determine the interval lengths to balance the accuracy and stability for wind speed forecasting. Lu et al. [38] applied an interval information granules technique to optimize the interval partition in a fuzzy time series. Chen et al. [39,40] used particle swarm optimization techniques to determine the optimal partitions of intervals used for the TAIEX, NTD/USD exchange rates, and the forecasting of enrollment at the University of Alabama. Cheng and Yang [41] applied a rule-based algorithm rather than fuzzy logical relationships to establish forecast rules in a fuzzy time series for stock price forecasting. Rubio et al. [42] proposed a new weighted fuzzy-trend method to assign weights in a fuzzy time series for forecasting stock market indices. Sadaei et al. [43] used a seasonal auto-regressive fractionally integrated moving average and particle swarm optimization to establish the fuzzy logical relationship for seasonal time series forecasting.

For the tourism demand time series, the sample size is always small. In some cases, the information is insufficient to accurately understand the statistical rules of the samples. Thus, information optimization technology including information distribution and information diffusion is applied using a fuzzy time series method to discretely distribute and superpose the sample information to excavate the true structure of overall sample information and as much useful information as possible to further enhance the understanding of the sample. In most previous studies, the parameters in a fuzzy time series are selected by experience, which can significantly affect the model performance in the forecasting progress. Thus, in our study the ASO algorithm is applied to search the optimal parameters in the fuzzy time series. Accordingly, the contributions of our study are as follows:

(1) ***A novel hybrid forecasting model focusing on small sample size for tourism demand forecasting is developed.*** The sample size of most economic data for tourism is too small to meet the requirements of traditional forecasting models that use large data samples. Thus, in this study, a hybrid forecasting method combined with a fuzzy time series and an advanced optimization algorithm is proposed. It is appropriate for small sample forecasting and can provide excellent results for tourism demand forecasting.

(2) ***Information optimization technology including information distribution and diffusion is applied in a fuzzy time series method.*** Information distribution technology can allocate the information carried by sample points to fill information gaps in small sample sets. Accordingly, forecasting accuracy can be improved. Normal information diffusion can spread the information carried by sample points to multiple monitoring points, which can both simplify the operation and improve the recognition ability of the system, improving forecasting accuracy.

(3) ***An advanced optimization algorithm, ASO, is used to search the optimal parameters in the fuzzy time series model.*** In a fuzzy time series, the main parameters are the ambiguity, fuzzy intervals, and diffusion coefficient (h), which are selected by experience in most research and play a key role in the forecasting progress. The ASO algorithm is applied to determine the optimal values of these parameters, which can significantly improve the forecasting accuracy.

(4) ***Tourism demand is analyzed based on previous tourism demand and out-of-sample forecasting obtained by the optimal model.*** According to the comprehensive evaluation indexes, tests, and forecasting effectiveness, the optimal model for tourism demand is selected and applied to forecast the annual number of inbound tourists from 2017 to 2020. Combined with historical data and forecasting results, the development trend of tourism demand is analyzed. Trend analysis is a prerequisite for policymakers and managers to enact plans and adopt measures to guarantee the balance of the tourism market.

The rest of the paper is organized as follows: Section 2 presents the clustering process for different provinces. Section 3 introduces the framework of our forecasting model. Section 4 describes and compares the empirical results. Section 5 discusses statistical tests, forecasting effectiveness, reproducibility, and universality of the proposed forecasting framework. In Section 6, out-of-sample forecasting is conducted and analyzed by the forecasting model with the best performance. Conclusions are presented in Section 7. The structure of our forecasting framework is illustrated in Fig. 1.

2. The clustering of province tourism based on fuzzy c-means clustering algorithm

Due to the diversity of the resource endowment, economic level, geographical location, cultural background, and infrastructure, the development of tourism in different provinces varies

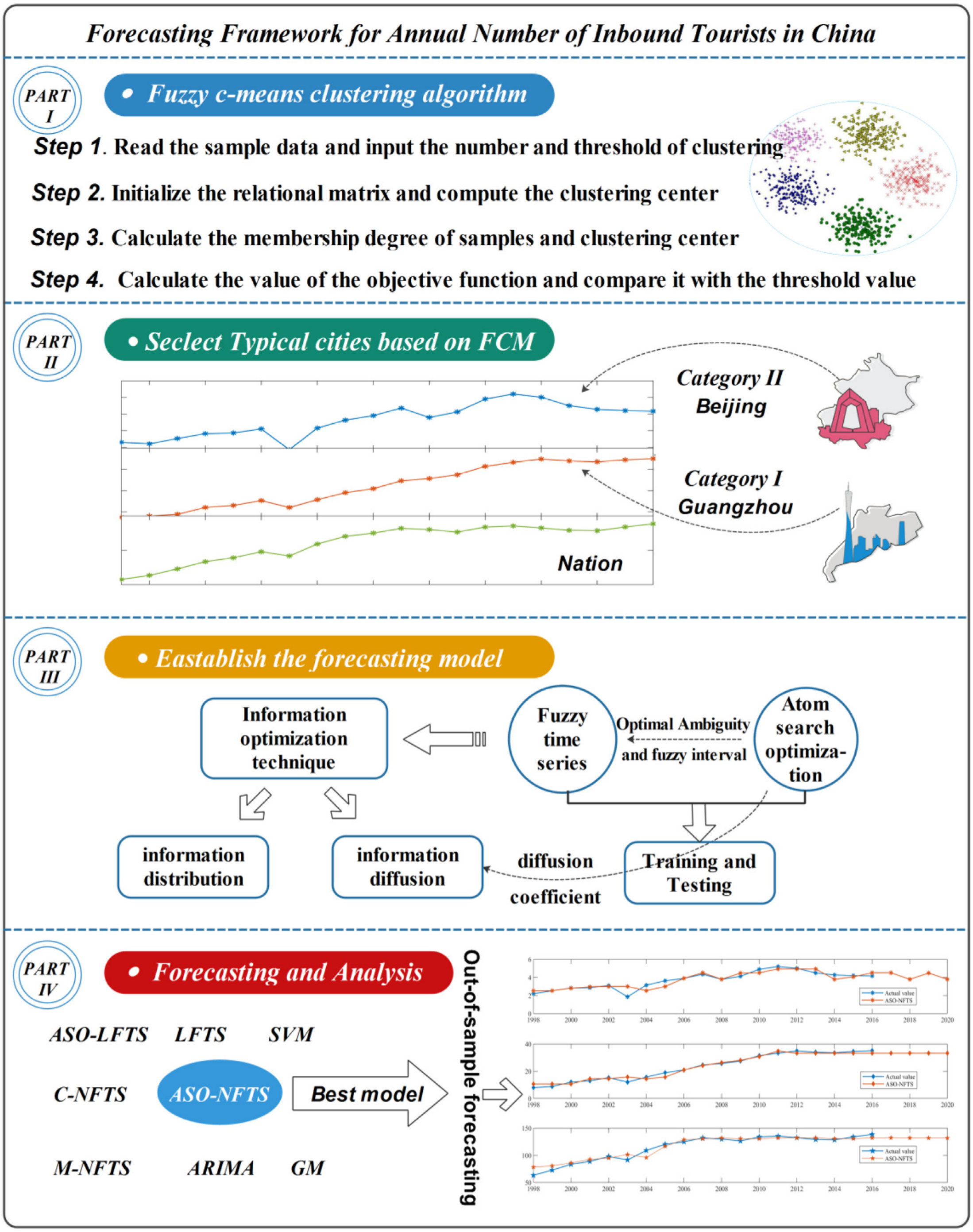

Fig. 1. Structure of our forecasting framework.

significantly. Therefore, to intuitively recognize the characteristics of tourism demand development from the perspectives of region distribution and economic development, the fuzzy *c*-means clustering (FCM) algorithm is applied to classify all 31 provinces [44] (data for Hong Kong, Macao, and Taiwan cannot be obtained).

The FCM algorithm is a flexible clustering algorithm that can partition based on an objective function. It aims to maximize the similarity between objects in the same class and minimize the similarity between different clusters. The iteration center can update the membership value corresponding to data points according to a predefined number of clusters, making it easier to achieve a global optimum [45]. FCM relies on the principle that each data point belongs to several different clusters with different membership values that range from [0, 1]. The specific calculating process of FCM is shown in [46] and [47]. The pseudo code of FCM is presented in Appendix 1.

The provinces are clustered into five categories based on FCM and the average number of inbound tourists in 2016 for each category is calculated, as presented in Table 1.

The comprehensive development of China's tourism shows obvious spatial regularity in geographical distribution. Generally, the comprehensive development of tourism is higher in eastern coastal regions than in central and western regions.

The eastern coastal regions are mainly concentrated in the first three categories and are tourism destinations. The fourth category is mainly the central region and part of the western region. The fifth category is mainly the western region and a few regions with less tourism development. The visual clustering results are presented in Fig. 2.

Although Hebei and Tianjin are located in North China near Beijing, the level of tourism development is relatively low, due to the lack of characteristic tourism resources and the impact of Beijing's tourism status. Although Liaoning, Jilin, and Heilongjiang are located in the northeastern region, Liaoning is far superior to Heilongjiang and Jilin in terms of tourism development. The main advantage is that Liaoning is located along the coast. The 'Coastal Avenue', including Dalian, Dandong, and Jinzhou, has greatly promoted the development of tourism in Liaoning.

Tourism is also an industry with strong dependence on regional economic patterns. The economic patterns correspond to tourism development in the provinces with traditionally strong tourism on the eastern coast, the provinces with high tourism development in the central region, and the provinces with low tourism development in the western and marginal regions. There is an obvious 'polarization phenomenon'; the difference in tourism and economic development between developed coastal areas and western and remote provinces has widened.

3. Introduction of the methods

In practice, the data is insufficient and not always easily accessible. To address the small sample time series and avoid model causality deviation and forecasting error caused by traditional models, a forecasting system based on fuzzy time series (FTS) and the atom search optimization (ASO) algorithm is proposed in our study to forecast the number of annual inbound tourists.

In the fuzzy time series, fuzzy information optimization technology is applied to excavate useful information from the time series and improve the system identifiable accuracy and the forecasting accuracy [48]. The core of fuzzy information optimization technology is information distribution and diffusion.

3.1. Background of the forecasting framework

In this section, the background of the basic methods applied in forecasting framework are introduced.

3.1.1. Fuzzy time series based on information distribution and diffusion

Set time series $Y(t) = \{y(1), y(2), \ldots, y(n)\}$. U is the domain of discourse, $\left\{U_j, j = 1, 2, \ldots, l, \right\}$ is the equidistant sequence segmentation set of U. The membership function is $\mu\left(y(t), u\right)$.

Thus, the fuzzy set of $y(t)$ on U is $\{\mu_1(y(t)), \mu_2(y(t)), \ldots, \mu_l(y(t))\}$, $0 \leq \mu_j \leq 1, j = 1, 2, \ldots, l$, which is defined as $\{F(t)\}$, The set of $F(t)$ is the fuzzy time series distributing on $Y(t)$, simplified as $F(t) = (\mu_1, \mu_2, \ldots \mu_l)$.

- **Information distribution**

Information distribution is a mathematical processing method extending from the processing of small sample data that contain incomplete information such that the statistical rules of the sample are insufficient to provide accurate understanding. The information distribution method fuzzifies the explicit boundary of the traditional histogram method in the process of analyzing small sample probability events to make the fuzzy transition information useable.

Set time series $Y(t) = \{y(1), y(2), \ldots, y(n)\}$. $V = \left\{V_j,\ j = 1, 2, \ldots, l\right\}$ is the median of the $U = \left\{U_j, j = 1, 2, \ldots, l\right\}$. For $y \in Y$, $v \in V$, $\Delta = \left|V_j - V_{j-1}\right|$ then

$$\mu(y, v) = \begin{cases} 1 - \frac{|y-v|}{\Delta} & |y - v| \leq \Delta \\ 0 & |y - v| \geq \Delta \end{cases} \tag{1}$$

μ is the linear information allocation [49], Δ is the ambiguity [50].

Regarding the linear information allocation as the membership function, every observation value y_i corresponds to a fuzzy set $\{F(i)\}$ generated based on the membership function and Eq. (1). The set of $F(i)$ is the fuzzy time series generated by the information distribution. Accordingly, a time series is transformed into a fuzzy time series.

- **Information diffusion**

Information diffusion is based on the membership function, and the normal diffusion function is a common and accurate function among several information diffusion functions [51]. The normal information diffusion method can divide the information carried by a sample point, to create more sample points, to fill information gaps in the small sample data and improve the accuracy of time series forecasting.

Set time series $Y(t) = \{y(1), y(2), \ldots, y(n)\}$. u_j is the monitoring point of the normal information diffusion function, and the set of u_j is the monitoring space recorded as $U = \left\{u_1, u_2, \ldots, u_j, \ldots, u_l\right\}$. As for $y \in Y$, $u \in U$,

$$\mu\left(y, u\right) = \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(u-y)^2}{2h^2}\right] \tag{2}$$

$\mu\left(y, u\right)$ is the normal information diffusion function [49], h is the window-width, also known as the information diffusion coefficient.

Regarding the normal information diffusion function as the membership function, every observation value y_i corresponds to a fuzzy set $\{F(i)\}$ generated based on the membership function and Eq. (2). The set of $F(i)$ is the fuzzy time series generated by information diffusion. Accordingly, a time series is transformed into a fuzzy time series.

Based on the information distribution and diffusion, the fuzzy set on input set (X) and output set (Y) can be obtained. Set $U \times U' = \left\{(U_j, U'_m), j = 1, 2, \ldots, l, m = 1, 2, \ldots, l'\right\}$, U and U' are the domain of discourse of the input set and output set, respectively. The information of (X, Y)can be distributed or diffused on (U, U')according to Eq. (1) and Eq. (2). Thus, the information of (x, y)falls to (U_j, U'_m) and can be expressed as

Information distribution:

$$q_{jm} = \begin{cases} \mu_j(x, u_j) \times \mu(y, u'_m) & \left|u_j - x\right| \leq \Delta_x \text{ and } \left|u'_m - y\right| \leq \Delta_y \\ 0 & others \end{cases} \tag{3}$$

Information diffusion:

$$q_{jm} = \mu_j(x, u_j) \times \mu(y, u'_m) \tag{4}$$

q_{jm} is the distribution or diffusion information of one sample observation. The total information of all observations can be calculated as

$$Q_{jm} = \sum_{i=1}^{m} q_{jm} \tag{5}$$

Then, the fuzzy information matrix is as follows:

$$\mathbf{Q} = \begin{array}{c c} & \begin{array}{cccc} u'_1 & u'_2 & \cdots & u'_{l'} \end{array} \\ \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_l \end{array} & \left\{ \begin{array}{cccc} Q_{11} & Q_{12} & \cdots & Q_{1l'} \\ Q_{21} & Q_{22} & \cdots & Q_{2l'} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{l1} & Q_{l2} & \cdots & Q_{ll'} \end{array} \right\} \end{array} \tag{6}$$

The fuzzy information matrix can be transformed into a fuzzy relationship matrix as

$$R = r_{jm_{l \times l'}} = \frac{Q_{jm}}{p_{m\ l \times l'}} \tag{7}$$

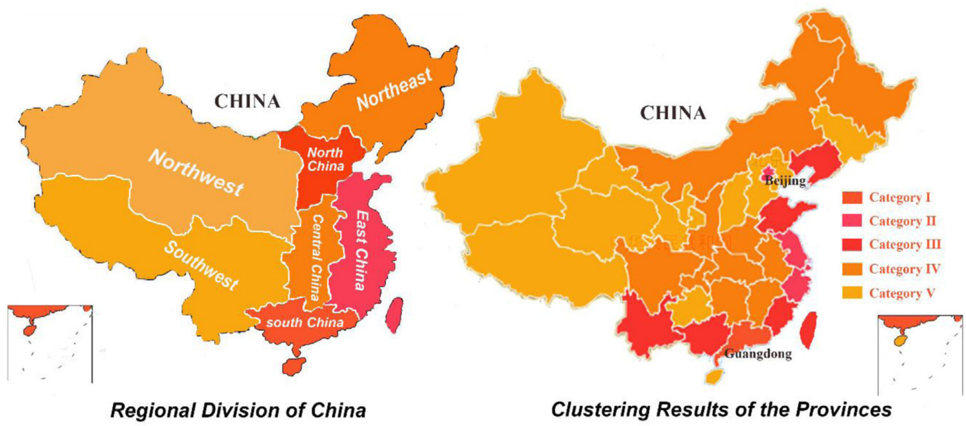

Fig. 2. Visual region division and clustering results of China.

Table 1
Clustering results of the 31 provinces and the average value of each category.

Province	Category	Mean	Region	Province	Category	Mean	Region
Guangdong	Category I	3507.21	South China	Liaoning	Category III	459.374	Northeast
Beijing	Category II	490.58	North China	Fujian	Category III		East China
Shanghai	Category II		East China	Shandong	Category III		East China
Jiangsu	Category II		East China	Guangxi	Category III		south China
Zhejiang	Category II		East China	Yunnan	Category III		Southwest
Inner Mongolia	Category IV	230.805	North	Tianjin	Category V	58.91	North China
Heilongjiang	Category IV		Northeast	Hebei	Category V		North China
Anhui	Category IV		East China	Shanxi	Category V		North China
Jiangxi	Category IV		East China	Jilin	Category V		Northeast
Henan	Category IV		Central China	Hainan	Category V		south China
Hubei	Category IV		Central China	Guizhou	Category V		Southwest
Hunan	Category IV		Central China	Xizang	Category V		Southwest
Chongqing	Category IV		Southwest	Gansu	Category V		Northwest
Sichuan	Category IV		Southwest	Qinghai	Category V		Northwest
Shaanxi	Category IV		Northwest	Ningxia	Category V		Northwest
				Xinjiang	Category V		Northwest

Table 2
Annual number of inbound tourists in different regions (in millions).

Year	Beijing	Guangzhou	Nation	Year	Beijing	Guangzhou	Nation
1997	2.2984	7.3916	57.5879	2007	4.3548	24.6087	131.8733
1998	2.2009	7.8865	63.4784	2008	3.7904	25.6797	130.0274
1999	2.5239	8.7602	72.7956	2009	4.1251	27.4780	126.4759
2000	2.8209	11.9894	83.4439	2010	4.9007	31.4093	133.7622
2001	2.8579	12.9238	89.0129	2011	5.2040	33.3163	135.4235
2002	3.1038	15.2588	97.9083	2012	5.0086	34.8943	132.4053
2003	1.8512	11.9696	91.6621	2013	4.5013	33.9790	129.0778
2004	3.1550	15.6365	109.0382	2014	4.2745	33.5543	128.4983
2005	3.6292	18.9699	120.2923	2015	4.1996	34.5035	133.8204
2006	3.9029	20.8971	124.9421	2016	4.1653	35.0721	138.4438

Table 3
Evaluation index equations.

Metric	Definition	Equation
MAE	Mean absolute error of forecasting results	$\mathbf{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left\|y_i - \hat{y}_i\right\|$
RMSE	Root mean square value of the errors	$\mathbf{RMSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}$
MAPE	Average of absolute percentage error	$\mathbf{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| \times 100\%$
TIC	Theil inequality coefficient of forecasting results	$\mathbf{TIC} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \Big/ \left(\sqrt{\frac{1}{N} \times \sum_{i=1}^{N} y_i^2} + \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \hat{y}_i^2}\right)$
VAR	Variance of the forecasting error	$\mathbf{Var} = E\left(e - E\left(e\right)\right)^2$

$$p_m = \max_{1\le j\le l} Q_{jm},\ \ m = 1, 2, \ldots, l' \tag{8}$$

According to the fuzzy approximate reasoning theory, the fuzzy time series forecasting model can be established as

$$\mu_{U'}(u'_m) = \mu_U(u_j) \circ R \tag{9}$$

In this equation, the operation '$\circ$' is selected as '$\vee - *$', and can be expressed as

$$\mu_{U'}(u'_m) = \max_{j=1}^{l} \mu_U(u_j) * r_{jm} \tag{10}$$

The final forecasting results can be obtained by the fuzzy set-gravity method.

$$F(n) = \frac{\sum_{m=1}^{l'} \mu_{U'}(u'_m) \times u'_m}{\sum_{m=1}^{l'} \mu_{U'}(u'_m)} \tag{11}$$

3.1.2. Atom search optimization algorithm

The atom search optimization algorithm was a novel physics-inspired meta-heuristic optimization algorithm based on atom dynamics proposed by Zhao [52].

ASO starts the optimization by generating a set of random solutions. The atoms update their positions and velocities in each iteration, and the position of the best atom found is also updated in each iteration. In addition, the acceleration of atoms comes from two forces. One is the interaction force caused by the Lennard-Jones potential, which is the vector sum of the attraction and repulsion exerted by other atoms. The other is the constraint force caused by the bond-length potential, which is the weighted position difference between each atom and the best atom. All updating and calculations are performed interactively until the stop criterion is satisfied. Finally, the position and the fitness value of the best atom are returned as an approximation of the global optimum. The specific process is presented in [53].

3.2. Contribution of the forecasting framework

For the macroeconomic data, the number of samples is always small. Unfortunately, most of the traditional econometric time series models require a specific probability distribution or restrictive assumptions. However, it is difficult to find an accurate probability distribution between series or find agreement under various assumptions with small sample data.

Therefore, a hybrid forecasting system is proposed that combines the ASO algorithm and the fuzzy time series based on fuzzy information optimization technology. In our forecasting system, the fuzzy time series is the main forecasting method, and fuzzy information optimization technology is applied to discretely distribute and superpose the sample information, which does not need a set probability distribution or restrictive assumptions. In this method, the raw information matrix of sample data is used to excavate the true structure of overall sample information, which can enhance the understanding of the sample. In the fuzzy information optimization process, the linear (L) information distribution and normal (N) information diffusion are indicated as LFTS and NFTS, respectively, and applied to establish the fuzzy time series forecasting model. The ASO algorithm is applied to optimize the parameters in the fuzzy time series method, and can significantly improve forecasting performance.

4. Empirical study

To verify the effectiveness and applicability of our proposed forecasting method, experiments and comparisons are presented in this section, including data description, evaluation criteria, parameter setting, experimental results, and analysis.

Table 4
Parameters setting in the Atom Search Optimization.

Parameters	Value
Number of atom population	100
Maximum of iterations.	100
Depth weight	50
Multiplier weight	0.2
Dimension of NFTS	4
Dimension of LFTS	2

4.1. Typical city selection and data analysis

As a vital constituent of the three major tourism markets (domestic tourism, inbound tourism, and outbound tourism), inbound tourism is an important index to measure the comprehensive development of tourism in a country or region. Inbound tourism can effectively earn foreign exchange, increase tourism income, and solve employment problems. According to the cluster analysis in Section 2, the top two classes for average number of tourists are selected as the analysis object.

In these two classes, the annual number of inbound tourists from 1997 to 2016 in Beijing and Guangdong are used for empirical study; the data is presented in Table 2. The number of inbound tourists in Guangzhou and Beijing in 2016 was 35.07 million and 4.17 million, respectively, as shown in Fig. 3. The annual number of national inbound tourists has the maximum absolute growth rate, but Guangzhou has the fastest average growth rate with 8.54% from 1997 to 2016. Beijing has the lowest growth rate among the three regions with 3.18%, and the national average growth rate is 4.72%. The tourism in Guangzhou has experienced great development since 1997. In our experiments, the data from 1997 to 2011 is used for training; data from 2012 to 2016 is used for testing.

4.2. Evaluation criteria

To compare the forecasting performance of different models and validate the efficiency of our proposed forecasting model, error evaluation criteria is considered in the experiments. No single index can fully measure the forecasting performance of all models. Therefore, a relative comprehensive evaluation system including various indexes is used in our study, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), Theil inequality coefficient (TIC), and forecasting error variance (VAR). The MAE, RMSE, MAPE, and TIC are calculated on the basis of the forecasting error and actual value to assess the forecasting accuracy of multiple points. Because forecasting error can be positive or negative, the MAE mitigates the inability of absolute error to reflect the error level through absolute value of the forecasting error. The RMSE is sensitive to outliers and can evaluate the forecasting accuracy efficiently. The VAR is applied to evaluate the model stability. Equations for these indexes are presented in Table 3.

4.3. Parameter setting

In our forecasting system, the main forecasting method is the fuzzy time series based on linear information distribution and normal information diffusion technology. ASO is applied to optimize the parameters in the fuzzy time series and diffusion function.

For the ASO, the parameters are set according to several experiments and literature [52,53]. Parameter values are presented in Table 4. Once the parameters are determined, the optimization process for the main forecasting system can be conducted.

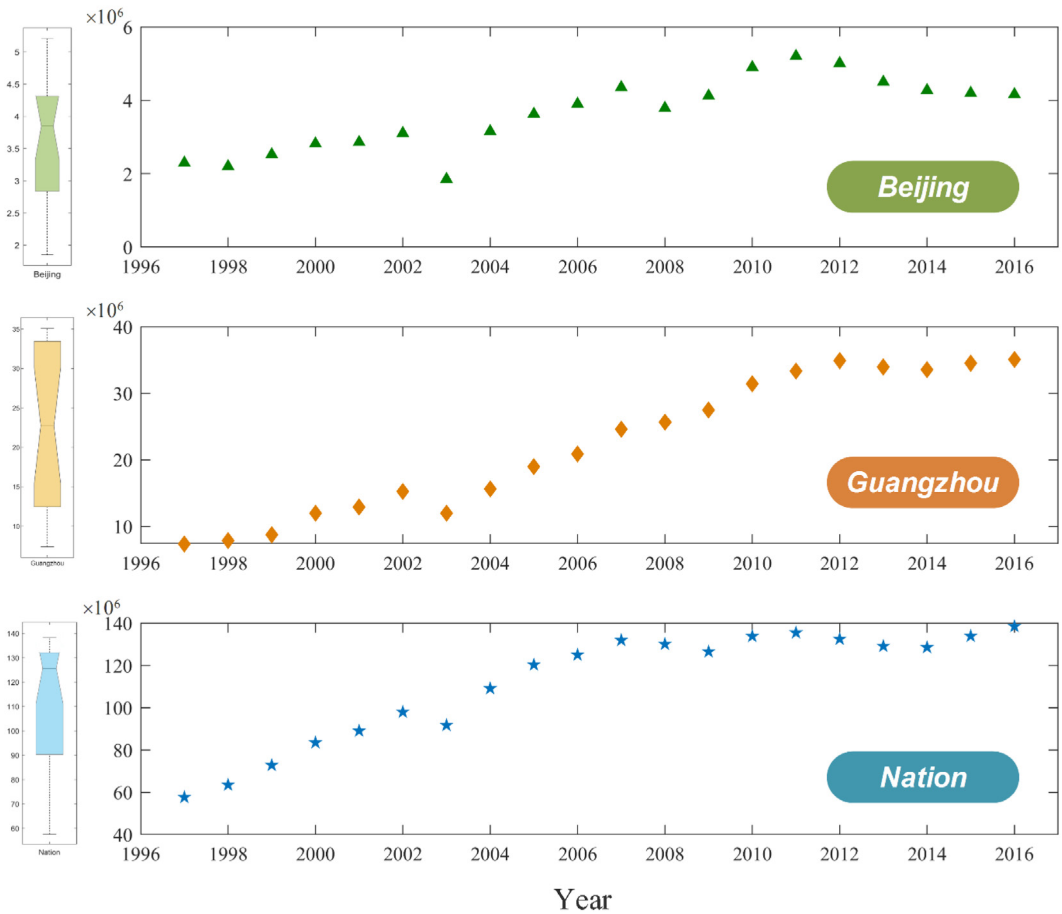

Fig. 3. Annual number of inbound tourists in three regions.

For the *fuzzy time series*, the main parameters affecting forecasting accuracy are the ambiguity and fuzzy interval. In the literature, the partition methods are varied. In our study, ASO is applied to search the optimal ambiguity and fuzzy interval in both LFTS and NFTS. The ambiguity and the number of fuzzy intervals obtained by ASO in the three time series are presented in Table 5.

For the *normal diffusion function*, the main parameter is the diffusion coefficient (h). The diffusion coefficient is a critical parameter for estimating information diffusion, and has a decisive effect on the calculation results. If it is too small, the result will be unstable; if it is too large, the resolution of the results will be too low. Thus, under the constraint of a limited number of samples, it is desirable to find the proper diffusion coefficient to achieve a balance between stability and resolution. The methods to determine the diffusion coefficient use the closest principle and the minimum mean square error of integration (MISE) principle (C-NFTS and M-NFTS).

• **Closest principle:**

$$h = \begin{cases} 0.8164 \times (max - min) & n = 5 \\ 0.5690 \times (max - min) & n = 6 \\ 0.4560 \times (max - min) & n = 7 \\ 0.3860 \times (max - min) & n = 8 \\ 0.3362 \times (max - min) & n = 9 \\ 0.2986 \times (max - min) & n = 10 \\ 2.6851 \times (max - min)/(n-1) & n \geq 11 \end{cases} \quad (12)$$

where $max = \max_{i=1}^{n} y_i$, $min = \min_{i=1}^{n} y_i$, n is the number of the sample.

• **MISE principle:**

$$h = 1.059\sigma n^{-0.2} \quad (13)$$

σ is the standard deviation of sample observations, and n is the number of the sample.

In our study, ASO is applied to search the optimal diffusion coefficient.

Accordingly, the optimization results of the ambiguity, fuzzy intervals, and information diffusion coefficient are presented in Table 5.

Table 5
Optimization results of the parameters in fuzzy time series.

Cities	Variable	ASO-NFTS	ASO-LFTS
Beijing	Ambiguity	0.4790	0.4790
	N-fuzzy interval	7	8
	diffusion coefficient (h_{in}/h_{out})	0.1000/0.2972	/
Guangdong	Ambiguity	3.0756	5.5361
	N-fuzzy interval	9	8
	diffusion coefficient (h_{in}/h_{out})	1.0034/1.4108	/
Nation	Ambiguity	8.9840	16.1712
	N-fuzzy interval	12	6
	diffusion coefficient (h_{in}/h_{out})	10.1211/18.3975	/

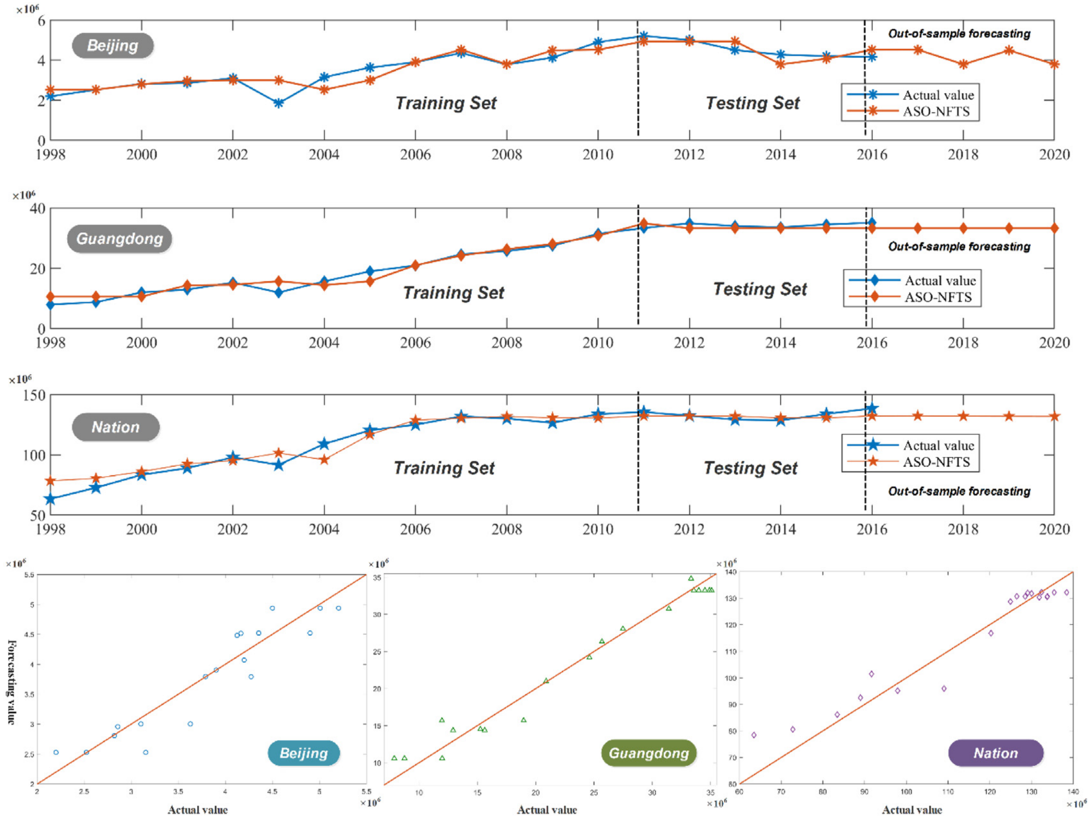

Fig. 4. Forecasting results of the annual number of inbound tourists by ASO-NFTS.

4.4. Experimental results

After the forecasting system is established, the annual inbound tourist data for Beijing and Guangdong from 1997 to 2011 are used to train the system and the data from 2012 to 2016 are used to test the performance of the system.

Based on Table 5, the fuzzy information matrix can be obtained using Eq. (5) and Eq. (6). For Guangdong, the fuzzy information matrixes of **ASO-NFTS** are calculated as given in Box I.

According to Eq. (7), the fuzzy relationship matrix can be calculated as given in Box II.

Based on the fuzzy approximate reasoning theory and the fuzzy gravity theory, the forecasting results can be calculated using Eq. (11), and are presented in Fig. 4.

4.5. Model comparison

To evaluate the effectiveness of our proposed forecasting system, the basic fuzzy time series (LFTS, C-NFTS, and M-NFTS) and other traditional small sample forecasting models (GM, ARIMA (p, d, q), and SVM) are compared. The parameters in GM and SVM adopt the default setting while the autoregressive term (p), moving average number (q), and difference times (d) in ARIMA are set based on the Akaike information criterion (AIC). To validate the universal adaptability of our proposed forecasting, three time series (Beijing, Guangzhou, and Nation) for the annual number of inbound tourists are used for experiments.

To compare the forecasting accuracy of these methods, MAE, RMSE, and MAPE are used to calculate the forecasting accuracy. The results are listed in Table 6. For Beijing, the MAPE is 6.78, 7.86, 13.24, 13.09, 14.11, 17.32, 11.81, and 29.90 for ASO-NFTS, ASO-LFTS, C-NFTS, M-NFTS, LFTS, ARIMA, GM, and SVM, respectively. It is observed that the forecasting accuracy of our proposed optimized FTS forecasting system (ASO-NFTS and ASO-LFTS) is significantly higher than the basic fuzzy time series and traditional small sample forecasting models. The MAPE of the SVM model in forecasting the annual number of national inbound tourists indicates slightly higher accuracy than ASO-LFTS. However, the MAPE of the proposed ASO-LFTS indicates much higher accuracy than SVM for Beijing and Guangzhou. Thus, the performance of the proposed forecasting system is superior to the other methods.

Comparing the information distribution and information diffusion technology in the fuzzy time series, the performance of the ASO-NFTS system demonstrates higher forecasting accuracy than the ASO-LFTS system, especially for Guangzhou. Thus, for the optimized model, information diffusion produces better results than information distribution.

To evaluate the stability of our proposed forecasting system, the percentage error of each point in the testing sets is calculated and the variance of the forecasting error is analyzed. The variance of the forecasting error of the different models is presented in Table 6 and the percentage error is list in Table 7. The minimum percentage error of our proposed forecasting system and other models is 0.11%, 0.14%, 6.80%, 6.16%, 7.83%, 2.20%, 0.41%, and 0.52%, and the maximum percentage error is 11.26%, 19.46%, 25.31%, 21.57%, 22.65%, 25.06%, 20.64%, and 51.63%, for ASO-NFTS,

$$
\begin{bmatrix}
0.1020 & 0.1169 & 0.0115 & 0.0002 & 0 & 0 & 0 & 0 & 0 \\
0.0022 & 0.1014 & 0.2123 & 0.0320 & 0.0001 & 0 & 0 & 0 & 0 \\
0.0109 & 0.1122 & 0.0241 & 0.0832 & 0.0267 & 0.0001 & 0 & 0 & 0 \\
0.0002 & 0.0017 & 0.0003 & 0.0163 & 0.0802 & 0.0070 & 0.0004 & 0 & 0 \\
0 & 0 & 0 & 0.0013 & 0.0147 & 0.1047 & 0.0156 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.0008 & 0.0708 & 0.0959 & 0.0057 & 0.0003 \\
0 & 0 & 0 & 0 & 0 & 0.0037 & 0.0296 & 0.0921 & 0.0360 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0103 & 0.0714 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0015 & 0.0118
\end{bmatrix}
$$

Box I.

$$
\begin{bmatrix}
1 & 1 & 0.0541 & 0.0030 & 0 & 0 & 0 & 0 & 0 \\
0.0214 & 0.8674 & 1 & 0.3849 & 0.0012 & 0 & 0 & 0 & 0 \\
0.1066 & 0.9602 & 0.1136 & 1 & 0.3335 & 0.0008 & 0 & 0 & 0 \\
0.0017 & 0.0149 & 0.0013 & 0.1962 & 1 & 0.0669 & 0.0041 & 0 & 0 \\
0 & 0 & 0 & 0.0162 & 0.1831 & 1 & 0.1623 & 0.0002 & 0 \\
0 & 0 & 0 & 0 & 0.0096 & 0.6768 & 1 & 0.0618 & 0.0035 \\
0 & 0 & 0 & 0 & 0.0002 & 0.0350 & 0.3090 & 1 & 0.5050 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.0004 & 0.1120 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0167 & 0.1653
\end{bmatrix}
$$

Box II.

Table 6
Comparison of results of the forecasting models.

	Criteria	*Optimized FTS*		*Basic FTS*			*Traditional small sample models*		
		ASO-NFTS	ASO-LFTS	C-NFTS	M-NFTS	LFTS	ARIMA(1,1,2)	GM	SVM
Beijing	MAE	0.2938	0.3654	0.5998	0.5958	0.6448	0.7651	0.5168	1.2739
	RMSE	0.3369	0.4796	0.6483	0.6595	0.7316	0.8004	0.5908	1.4693
	MAPE (%)	6.7750	7.8643	13.2445	13.0927	14.1092	17.3243	11.8054	29.8992
	TIC	0.0378	0.0549	0.0784	0.0797	0.0889	0.0829	0.0647	0.1456
	VAR	0.1413	0.2645	0.0757	0.1000	0.1492	0.0691	0.3520	0.9316
Guangzhou	MAE	1.1487	3.0186	4.0738	5.1292	4.5734	6.6069	3.3733	4.3622
	RMSE	1.2804	3.1516	4.1129	5.1603	4.6353	6.9973	3.8980	4.9140
	MAPE (%)	3.3130	8.7564	11.8183	14.8870	13.2774	19.2420	9.7379	12.7196
	TIC	0.0189	0.0479	0.0635	0.0810	0.0722	0.0927	0.0551	0.0672
	VAR	0.3998	1.0263	0.3998	0.3998	0.7134	6.6396	14.9918	7.1763
Nation	MAE	2.9342	4.2596	12.0142	14.3286	12.5255	4.1525	8.1141	3.8408
	RMSE	3.5342	5.5407	12.5618	14.7363	13.0192	4.2702	9.1032	4.3707
	MAPE (%)	2.1912	3.1503	9.0032	10.7592	9.3922	3.1441	6.1215	2.8888
	TIC	0.0134	0.0212	0.0497	0.0588	0.0516	0.0160	0.0333	0.0166
	VAR	14.6225	21.2289	16.8211	14.8123	15.7634	20.4015	25.6789	22.6797

Table 7
Percentage error of the testing set generated by different methods for three datasets.

	Testing year	*Optimized FTS*			*Basic FTS*		*Traditional small sample models*		
		ASO-NFTS	ASO-LFTS	C-NFTS	M-NFTS	LFTS	ARIMA(1,1,2)	GM	SVM
Beijing	2012	1.408%	19.458%	25.306%	21.569%	22.645%	12.315%	12.835%	4.240%
	2013	9.704%	6.178%	16.888%	12.820%	13.927%	25.057%	0.459%	13.901%
	2014	11.262%	2.372%	11.229%	11.124%	10.121%	21.994%	9.575%	38.824%
	2015	3.072%	5.206%	9.645%	10.064%	9.452%	15.731%	15.520%	51.632%
	2016	8.429%	6.107%	7.476%	10.646%	9.319%	11.524%	20.639%	40.899%
Guangzhou	2012	4.707%	7.004%	12.911%	13.089%	16.114%	6.661%	10.229%	0.516%
	2013	2.140%	10.312%	15.607%	10.748%	13.854%	17.765%	1.151%	11.237%
	2014	0.901%	6.421%	10.139%	9.619%	12.764%	24.247%	7.331%	17.471%
	2015	3.627%	6.996%	12.143%	12.105%	15.164%	23.625%	11.919%	18.531%
	2016	5.190%	13.049%	15.586%	13.530%	16.539%	23.913%	18.059%	15.844%
Nation	2012	0.111%	4.186%	9.554%	8.928%	10.147%	3.395%	0.414%	4.371%
	2013	2.266%	0.137%	7.222%	6.580%	7.831%	4.537%	5.339%	0.598%
	2014	1.701%	1.082%	6.804%	6.159%	9.067%	2.715%	9.115%	4.096%
	2015	2.361%	2.929%	9.882%	10.449%	12.684%	2.198%	8.044%	1.413%
	2016	4.518%	7.417%	13.499%	12.900%	14.067%	2.876%	7.694%	3.966%

ASO-LFTS, C-NFTS, M-NFTS, LFTS, ARIMA, GM, and SVM, respectively. For both the minimum and maximum percentage error, the proposed forecasting system demonstrates better performance than the other models. The minimum percentage error of GM and SVM are relatively small, but the maximum is large, which indicates the instability of these models. Overall, fuzzy time series forecasting models are more stable than traditional forecasting models.

Remarks: For forecasting accuracy and stability, our proposed optimized fuzzy time series forecasting models are generally superior to other models. Of the two optimized models, ASO-NFTS (information diffusion) demonstrates better performance than ASO-LFTS (information distribution).

5. Discussion

In this section, three parts are discussed. A no-parameters test is used to determine if there is a significant statistical difference between the models. The forecasting effectiveness is verified. The reproducibility and universality of the proposed forecasting framework is verified through repeated and supplementary experiments.

5.1. Mann–Whitney U test

The Mann–Whitney U test, also known as the Mann–Whitney rank-sum test or the Mann–Whitney–Wilcoxon (MWW) test, was proposed by H. B. Mann and D. R. Whitney in 1947 [54]. This test determines whether there is a significant difference between the mean values of two populations. The MWW test explicitly considers the ranks of each measured value in each sample, and uses more information than the Symbolic Test [55].

Assumption 1. All observations from both groups are independent of each other.

Assumption 2. The responses are ordinal (i.e., one can say which is the greater of any two observations).

Assumption 3. Under the null hypothesis H_0, the distributions of both populations are equal.

Assumption 4. The alternative hypothesis H_1 is that the distributions are not equal.

The specific steps of the test method are as follows:

Step 1: Mix the two sets of sample data and rank them in ascending order. The smallest data level is 1, the second smallest is 2, etc. (if there is equality in the mixed data, the same data should have the same level value and use the average value;

Step 2: Calculate the rank sum of two samples, $R_1$and R_2;

Step 3: Set n_1 = the sample size for sample 1, n_2 = the sample size for sample 2.

$$\begin{aligned} U_1 &= R_1 - n_1 \times (n_1 + 1)/2 \\ U_2 &= R_2 - n_2 \times (n_2 + 1)/2 \end{aligned} \tag{14}$$

$$U_1 + U_2 = R_1 + R_2 - (n_1 \times (n_1 + 1) + n_2 \times (n_2 + 1))/2 \tag{15}$$

Knowing that$R_1 + R_2 = N\,(N + 1)\,/2$ and$N = n_1 + n_2$, we find that the sum of $U_1 + U_2 = n_1 \times n_2$;

Step 4: The smaller value of U_1 and U_2 is the one used to compare with critical valueU_α. When$U < U_\alpha$, the null hypothesis H_0 is rejected, indicating that the distributions of the two samples are not equal and there are significant differences between them.

Table 8
Mann–Whitney U test results.

ASO-NFTS vs.	Beijing		Guangzhou		Nation	
	One tail	*Two tails*	*One tail*	*Two tails*	*One tail*	*Two tails*
ASO-LFTS	0.34524	0.69048	0.00397	0.00794***	0.15476	0.30952
C-NFTS	0.02778	0.05556*	0.00397	0.00794***	0.00397	0.00794***
M-NFTS	0.02778	0.05556*	0.00397	0.00794***	0.00397	0.00794***
LFTS	0.02778	0.05556*	0.00397	0.00794***	0.00397	0.00794***
ARIMA	0.00397	0.00794***	0.00397	0.00794***	0.15476	0.30952
GM	0.21032	0.42063	0.11111	0.22222	0.02778	0.05556*
SVM	0.04762	0.09524*	0.00397	0.00794***	0.50000	1.00000
ASO-LFTS vs.	Beijing		Guangzhou		Nation	
	One tail	*Two tails*	*One tail*	*Two tails*	*One tail*	*Two tails*
ASO-NFTS	0.34524	0.69048	0.00397	0.00794***	0.15476	0.30952
C-NFTS	0.04762	0.09524*	0.04762	0.09524*	0.01587	0.03175**
M-NFTS	0.04762	0.09524*	0.07540	0.15079	0.01587	0.03175**
LFTS	0.04762	0.09524*	0.00794	0.01587**	0.00794	0.01587**
ARIMA	0.00397	0.00794***	0.00397	0.00794***	0.07540	0.15079
GM	0.15476	0.30952	0.04762	0.09524*	0.01587	0.03175**
SVM	0.02778	0.05556*	0.00397	0.00794***	0.27381	0.54762

Table 9
Forecasting effectiveness of different models.

MODELS	*Beijing*		*Guangzhou*		*Nation*	
	1st-order	*2nd-order*	*1st-order*	*2nd-order*	*1st-Order*	*2nd-order*
ASO-NFTS	0.93225	0.89216	0.96687	0.94960	0.97809	0.96261
ASO-LFTS	0.92136	0.85997	0.91244	0.88644	0.96850	0.94081
C-NFTS	0.85891	0.79737	0.86723	0.84688	0.90608	0.88188
M-NFTS	0.86756	0.82621	0.88182	0.86745	0.90997	0.88454
LFTS	0.86907	0.81986	0.85113	0.83775	0.89241	0.86946
ARIMA	0.82676	0.77735	0.80758	0.74680	0.96856	0.95996
GM	0.88195	0.81556	0.90262	0.84666	0.93878	0.90616
SVM	0.70101	0.56136	0.87280	0.80848	0.97111	0.95413

The test results of our proposed forecasting methods and other models are presented in Table 8. The values in Table 8 are the p value with one tail and two tails of the MWW test between each pair of models. It is observed that the difference between ASO-NFTS and ASO-LFTS is not significant in most cases. In a few experiments, the difference between our proposed forecasting system and the traditional small sample forecasting models is verified to be not significant, but in most cases, the differences are great. Thus, based on the experimental results, our proposed model can significantly improve forecasting performance compared with traditional models.

5.2. Forecasting effectiveness

Forecasting effectiveness (FE) is applied to compare the degree of forecasting validity for different models. This indicator is based on the mean squared deviation of the forecasting accuracy and the sum of the squared errors [56]. The *1st-order* FE is based on the expected value of the forecasting accuracy sequence, and the *2nd-order* FE is based on the difference between its expected value and standard deviation. A higher FE indicates better performance.

The calculation results of the 1st- and 2nd-order forecasting effectiveness are listed in Table 9. The ASO-NFTS forecasting method demonstrates the highest forecasting effectiveness in both 1st-order and 2nd-order conditions for all three cases. The 1st-order forecasting effectiveness is 0.932, 0.967, and 0.978, and the 2nd-order forecasting effectiveness is 0892, 0950, and

Table 10
Statistical indicators of the 30 experimental results.

City	Criteria	Mean	Max	Min	Median	Std.
Beijing	MAE	0.3162	0.3684	0.2364	0.3089	0.0410
	RMSE	0.3729	0.4262	0.2777	0.3695	0.0448
	MAPE (%)	7.2226	8.3891	5.4290	7.0290	0.9065
	TIC	0.0420	0.0471	0.0312	0.0424	0.0048
	VAR	0.1448	0.2022	0.0630	0.1482	0.0487
City	Criteria	Mean	Max	Min	Median	Std.
Guangzhou	MAE	1.1806	1.2682	1.0883	1.1849	0.0391
	RMSE	1.3092	1.3886	1.2265	1.3129	0.0352
	MAPE (%)	3.4058	3.6605	3.1374	3.4180	0.1137
	TIC	0.0194	0.0206	0.0181	0.0194	0.0005
	VAR	0.3998	0.3998	0.3998	0.3998	0.0000
City	Criteria	Mean	Max	Min	Median	Std.
Nation	MAE	3.1576	3.9602	2.8511	3.0473	0.2892
	RMSE	3.7825	4.5491	3.3858	3.5936	0.4055
	MAPE (%)	2.3616	3.0232	2.1275	2.2886	0.2283
	TIC	0.0143	0.0174	0.0128	0.0136	0.0016
	VAR	14.6122	15.7459	13.8615	14.3622	0.6711

0.963 for the forecasting results of the annual number of Beijing, Guangzhou, and national inbound tourists, respectively. For the forecasting results for annual number of national inbound tourists, the forecasting effectiveness obtained by SVM is higher than ASO-LFTS, but lower in the other cases. Among the three cases, the forecasting results for annual number of national inbound tourists indicate the highest effectiveness for all methods.

5.3. Reproducibility and universality of proposed forecasting framework

To verify the reproducibility and universality of the proposed forecasting framework, some repeated experiments and supplementary experiments are conducted.

The earlier experiments are repeated 30 times to observe the variety of the forecasting results. The statistical indicators of the 30 experimental results are presented in Table 10. The maximum MAPE is 8.3891%, 3.6605%, and 3.0232%, and the minimum MAPE is 5.4290%, 3.1374%, and 2.1275, in the 30 experiments for Beijing, Guangzhou, and Nation, respectively. All results are superior compared to the small sample forecasting models presented in Section 4.5. The mean and median values are close, which indicates that the distribution of results is approximately symmetrical. The standard deviation of the results indicates the high stability of the proposed forecasting framework. The repeated experiments verified that our forecasting framework is reproducible and the experimental results are stable.

To verify that the universality of the proposed forecasting framework is greater than classical small sample forecasting models, we selected eight additional tourism demand time series for comparison. The performance of the proposed forecasting is superior to the compared models in all indicators and datasets, as shown in Table 11. It is also observed that the compared small sample forecasting models are not stable, and in some cases the evaluation criteria are abnormal. The supplementary experiments show that our proposed forecasting framework is reproducible, stable, and applicable to different small sample time series.

Remarks: The performance of our proposed forecasting framework is superior to the other models in forecasting accuracy, stability, and effectiveness. Moreover, the repeated experiments and supplementary experiments have verified the reproducibility and universality of the proposed forecasting framework.

6. Out-of-sample forecasting and discussion

The experiments and tests indicate that our proposed ASO-NFTS model can achieve higher accuracy and more stable forecasting results than other models for annual number of inbound tourists. The lower part of Fig. 4 illustrates the actual and forecasted values of ASO-NFTS. It is observed that the scatter is close to the diagonal. Thus, in this section, ASO-NFTS is applied to make out-of-sample forecasting for Beijing, Guangdong, and national annual number of inbound tourists from 2017 to 2020. The fitting value of the training set, the forecasting value of the testing set, and the out-of-sample forecasting value are presented in Fig. 4.

China's tourism industry started somewhat later than some developed countries. Before 1949, China's economy was depressed, tourism development was essentially stagnant, and the tourism industry did not exist. Since the reform and opening the borders, with the sustained development of China's economy and increased personal income, the number of tourists and tourism income continue to grow rapidly. The tourism industry has become an important part of the national economy and one of the fastest growing areas of residents' consumption. According to the China Tourism Statistics Bulletin issued by the State Tourism Administration, from 2006 to 2015 domestic tourism revenue showed a steady growth trend. China's tourism conditions have improved and the attraction of domestic and foreign tourists has been enhanced by abundant tourism resources.

It is observed in Fig. 4 that the annual number of inbound tourists in Guangzhou increased rapidly from 1998 to 2012. The growth rate was higher than in Beijing and the nation. From 1997 to 2012, the annual number of inbound tourists in Beijing increased from 2.298 million to 5.01 million, while tourists in Guangzhou increased from 7.392 million to 34.894 million; the average growth rates were 5.33% and 10.90%, respectively. The annual number of national inbound tourists in the same period increased from 57.588 million to 138.444 million and the average growth rate was 4.723%. From 2012 to 2016, the annual numbers of inbound tourists in provinces and the nation were steady and consistently high. According to the out-of-sample forecasting from 2017 to 2020, the inbound tourism markets in China will continue to grow. In support of national policy, governments at all levels seize opportunities to vigorously develop mass tourism and promote tourism consumption. China's tourism consumption scale has steadily expanded, the structure has changed rapidly, and the quality has been continuously improved. With stable macroeconomic and social environments, the demand for national tourism and tourism consumption is continuously increasing, and industrial investment and innovation are more active. The tourism market has entered a stage of steady growth.

Table 11
Experimental results comparison for eight cities.

Anhui	ASO-NFTS	ARIMA	GM	SVM	*Hebei*	ASO-NFTS	ARIMA	GM	SVM
MAE	0.2519	1.4515	0.8550	0.8545	MAE	0.1198	0.2667	0.2321	0.3123
RMSE	0.3314	1.6269	0.9367	0.9582	RMSE	0.1898	0.3238	0.2574	0.3688
MAPE (%)	7.9874	50.1460	28.5479	29.7219	MAPE (%)	11.0808	32.7907	25.2200	38.5470
TIC	0.0580	0.2183	0.1411	0.1404	TIC	0.1081	0.1562	0.1364	0.1721
VAR	0.0600	0.6751	0.6179	0.2351	VAR	0.0397	0.0586	0.0777	0.0481
Henan	ASO-NFTS	ARIMA	GM	SVM	*Hunan*	ASO-NFTS	ARIMA	GM	SVM
MAE	0.1680	0.2708	0.3868	0.2811	MAE	0.1569	0.4578	0.3512	1.0345
RMSE	0.2152	0.4557	0.4018	0.3462	RMSE	0.1726	0.4743	0.4041	1.2305
MAPE (%)	11.2485	20.7299	26.2849	21.1050	MAPE (%)	6.7826	20.2129	15.2544	45.2104
TIC	0.0731	0.1407	0.1291	0.1088	TIC	0.0391	0.0944	0.0839	0.2170
VAR	0.0578	0.1679	0.1667	0.0755	VAR	0.0064	0.0192	0.1443	0.5561
Inner Mongolia	ASO-NFTS	ARIMA	GM	SVM	*Shaanxi*	ASO-NFTS	ARIMA	GM	SVM
MAE	0.1082	0.1192	0.1681	0.5002	MAE	0.3144	0.6505	0.4083	1.7813
RMSE	0.1311	0.1587	0.2072	0.5297	RMSE	0.3468	0.9035	0.5024	1.9287
MAPE (%)	6.3774	7.0672	10.0384	29.9525	MAPE (%)	10.6815	24.1367	13.1630	62.6885
TIC	0.0410	0.0484	0.0597	0.1882	TIC	0.0582	0.1360	0.0835	0.2479
VAR	0.0069	0.0299	0.0252	0.0379	VAR	0.1503	0.4914	0.3149	0.6839
Shanghai	ASO-NFTS	ARIMA	GM	SVM	*Tianjin*	ASO-NFTS	ARIMA	GM	SVM
MAE	0.61956	0.7854	0.7436	0.8166	MAE	0.0192	0.4199	0.2128	0.5726
RMSE	0.668451	1.0239	0.8139	0.8998	RMSE	0.0258	0.4232	0.2141	0.5793
MAPE (%)	9.418299	12.3288	11.3318	12.7637	MAPE (%)	2.4093	54.5376	27.4198	74.2532
TIC	0.053985	0.0742	0.0596	0.0660	TIC	0.0169	0.2471	0.1214	0.2727
VAR	0.078716	0.5394	0.3393	0.5554	VAR	0.0005	0.2187	0.0007	0.0096

In summary, the overall growth trend for annual number of inbound tourists continues for individual provinces and the nation. Some marked decline occurred in 2003, 2008, 2009, and 2013. Along with actual economic and social environments, these structural breaks are caused by incidental events such as regional plagues, the international financial environment, and natural disasters. Thus, the handling and response of emergency events is also crucial for sustained and healthy development of the tourism industry.

7. Conclusion

Excellent forecasting performance is a prerequisite and foundation of scheduling and management in economic and industrial fields. Some traditional models, such as the autoregressive integrated moving average, artificial neural network, and regression models meet with theoretical assurance only when the sample size is large. However, in most actual circumstances, it is a formidable or impossible task for researchers to collect complete information and many samples due to limited data. Thus, forecasting techniques for small sample data should be considered.

In this study, a novel and effective small sample forecasting system using ASO-NFTS and ASO-LFTS is proposed to forecast the annual number of inbound tourists. The proposed models can be effective for small sample forecasting. The advantages of this method are: (1) the information optimization technology including linear information distribution and normal information diffusion is applied in fuzzy time series that can improve the recognition ability of the system, effectively identify small sample information, and further improve the forecasting performance of the model; (2) the ASO algorithm, which is simple and easily implemented, is applied to search the optimal ambiguity, interval number, and diffusion coefficients in information optimization technology. For fuzzy time series, optimal parameters can greatly improve forecasting performance.

To verify the effectiveness of our proposed forecasting method, three cases selected by the FCM algorithm were used to perform experiments. The main conclusions are: (1) our proposed optimized FTS forecasting method (ASO-NFTS and ASO-LFTS) is superior to basic fuzzy time series models and traditional small sample forecasting models in most cases. The MAPE of SVM and ARIMA for forecasting the annual number of national inbound tourists are slightly lower than ASO-LFTS but much higher for Beijing and Guangzhou. Thus, the forecasting accuracy of the proposed system is superior to other methods; (2) comparing information distribution and information diffusion technology for information optimization in fuzzy time series, the difference between the basic fuzzy time series models is small, but the performance of the ASO-NFTS forecasting system is superior to the ASO-LFTS for the two optimization FTS forecasting models, especially in the forecasting accuracy for Guangzhou. Therefore, for optimization models, information diffusion is more effective than information distribution; (3) the percentage error and variance of the forecasting error both indicate better stability in the proposed methods; (4) three different cases indicate that our proposed forecasting system adapts well to different data sets; (5) the no-parameters test and forecasting effectiveness verify the significant difference between our proposed forecasting method and traditional small sample forecasting models. Based on the conclusions drawn from the experimental results, tourism industry forecasting should consider the small sample characteristics of the time series and excavate as much sample information as possible to improve data identification ability and forecasting performance.

Furthermore, based on the compared experiments and tests, the optimal model is selected for the out-of-sample forecasting. Based on the actual data and forecasted values of the number of inbound tourists, the changing trend, growth rate, and structural breaks of tourism demand in typical cities are analyzed in this study. According to the analysis, the overall trend of annual number of inbound tourists is rising except during periods with incidental events. Thus, the development of the tourism industry still has tremendous potential and it is important to strengthen coping strategies and responses for emergency events that can significantly affect the sustainable development of the tourism industry.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Ping Jiang: Conceptualization, Resources, Methodology. **Hufang Yang:** Formal analysis, Writing - original draft, Writing - review & editing. **Ranran Li:** Software, Writing - review & editing. **Chen Li:** Software.

Acknowledgment

This work was supported by Major Program of National Social Science Foundation of China (Grant No. 17ZDA093).

Appendix A

The pseudo code of Fuzzy c-means algorithm.

Algorithm 1. Fuzzy c-means algorithm

Parameters:

m -Fuzzification parameter $(1 \leq m \leq \infty)$.

ε - Threshold $\varepsilon > 0$.

ξ -constant.

d_{ki} -Euclidean distance between the training pattern x_k and the class center v_i.

1 Begin

2 /* Initialize class centers $v_i = (2 \leq i \leq n)$ and fuzzy c-partition $\mathbf{U}^{(0)}$ randomly. Given m, ε and ξ */

3 /*Calculate the membership matrix $\mathbf{U} = [\mu_{ki}]$ based on

$$\mu_{ki} = \frac{\sum_{i=1}^{n} \left(1/d_{ki}^2 \right)^{1/(m-1)}}{d_{ki}^2}$$ */

4 /*Update the class centers according to

$$v_i = \sum_{k=1}^{c} \frac{x_k \sum_{i=1}^{n} \mu_{ki}^m}{\mu_{ki}^m}$$ */

5 /*Compute $\Delta = \max(|\mathbf{U}^{(t+1)} - \mathbf{U}^t|)$, if $\Delta > \varepsilon$, then go to Step 2; otherwise go to Step 6.*/

6 /*Find the results for the final class centers.*/

7 End

Appendix B

The pseudo code of ASO algorithm.

Algorithm 2: ASO algorithm

Objective functions:

Minimize: $\mathbf{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$

Input:

$x_{tr} = (x(1), x(2), \cdots, x(k)) - \textit{the training set}$

$x_{te} = (x(k+1), x(k+2), \cdots, x(n)) - \textit{the testing set}$

Output:

$\hat{x}_f = (\hat{x}(k+1), \hat{x}(k+2), \cdots, \hat{x}(n)) - \textit{the forecasting results}$

Parameters:

T -the maximum number of iterations

N -the number of Atoms

t -current iteration number

d -the number of dimension

$rand_j$ -a random number in [0, 1]

$\lambda(t)$ -the lagrangian multiplier

β -the multiplier weight

$x_{best}(t)$ -the position of the best atom at the tth iteration

$Fit_{best}(t)$ -the fitness values of the best atoms at the tth iteration

$Fit_{worst}(t)$ -the fitness values of the worst atoms at the tth iteration

$Fit_i(t)$ -the function fitness value of the ith atom at the tth iteration.

F_{ij} -the interaction force acted on the ith atom from the jth atom at the tth iteration

1 /*Set the parameters of ASO*/

2 /*Randomly initialize a set of atoms X (solutions) and velocity v, Fit_{best}=inf */

3 WHILE the stop criterion is not satisfied **DO**

4 **FOR EACH** atom X_i **DO**

/*Calculate the fitness value Fit_i*/

5 **IF** $Fit_i < Fit_{best}$ **THEN**

6 /* Fit_{best}=Fit_i */

7 /* $X_{best} = X_i$ */

8 **END IF**

9 /*Calculate the mass

$$M_i(t) = e^{-\frac{Fit(t)-Fit_{best}(t)}{Fit_{worst}(t)-Fit_{best}(t)}} \quad Fit_{best}(t) = \min_{i\in\{1,2,\ldots,N\}} Fit_i(t)$$

$$m_i(t) = \frac{M_i(t)}{\sum_{j=1}^{N} M_j(t)} \qquad Fit_{worst}(t) = \max_{i\in\{1,2,\ldots,N\}} Fit_i(t) \quad */$$

10 /*Determine its K neighbors using equation

$$K(t) = N-(N-2)\times\sqrt{\frac{t}{T}} \quad */$$

11 /*Calculate of the interaction force F_i and the constraint force G_i

$$F_i^d(t) = \sum_{j\in kbest} rand_j F_{ij}^d(t)$$

$$F_{ij}(t) = \frac{24\varepsilon(t)}{\sigma(t)}\left[2\left(\frac{\sigma(t)}{r_{ij}(t)}\right)^{13} - \left(\frac{\sigma(t)}{r_{ij}(t)}\right)^{7}\right]\frac{r_{ij}(t)}{r_{ij}^d(t)}$$

$$G_i^d(t) = \lambda(t)(x_{best}^d(t) - x_t^d(t))$$

$$\lambda(t) = \beta e^{-\frac{20t}{T}} \quad */$$

12 /*Calculate the acceleration

$$\alpha_i^d(t) = \frac{F_i^d(t)}{m_i^d(t)} + \frac{G_i^d(t)}{m_i^d(t)} \, */$$

13 /*Update the velocity

$$v_i^d(t+1) = rand_i^d v_i^d(t) + \alpha_i^d(t) \, */$$

14 /*Update the position

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad */$$

15 END FOR

16 END WHILE

17 Find the best solution so far X_{best}

References

[1] X. Sun, W. Sun, J. Wang, Y. Zhang, Y. Gao, Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China, Tour. Manag. 52 (2016) 369–379, http://dx.doi.org/10.1016/j.tourman.2015.07.005.

[2] B. Peng, H. Song, G.I. Crouch, A meta-analysis of international tourism demand forecasting and implications for practice, Tour. Manag. 45 (2014) 181–193, http://dx.doi.org/10.1016/j.tourman.2014.04.005.

[3] W. Jun, L. Yuyan, T. Lingyu, G. Peng, Modeling a combined forecast algorithm based on sequence patterns and near characteristics: an application for tourism demand forecasting, Chaos Solitons Fractals 108 (2018) 136–147.

[4] T. Iverson, The advanced econometrics of tourism demand, Tour. Manag. 31 (2010) 449.

[5] C. Morley, J. Rosselló, M. Santana-Gallego, Gravity models for tourism demand: theory and use, Ann. Tour. Res. 48 (2014) 1–10, http://dx.doi.org/10.1016/j.annals.2014.05.008.

[6] Z. Han, R. Durbarry, M.T. Sinclair, Modelling US tourism demand for European destinations, Tour. Manag. 27 (2006) 1–10, http://dx.doi.org/10.1016/j.tourman.2004.06.015.

[7] M. De Mello, A. Pack, M.T. Sinclair, A system of equations model of UK tourism demand in neighbouring countries, Appl. Econ. 34 (2002) 509–521.

[8] Z. Cao, G. Li, H. Song, Modelling the interdependence of tourism demand: the global vector autoregressive approach, Ann. Tour. Res. 67 (2017) 1–13, http://dx.doi.org/10.1016/j.annals.2017.07.019.

[9] Y.Y. Liu, F.M. Tseng, Y.H. Tseng, Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model, Technol. Forecast. Soc. Change 130 (2018) 123–134, http://dx.doi.org/10.1016/j.techfore.2018.01.018.

[10] H. Song, G. Li, S.F. Witt, G. Athanasopoulos, Forecasting tourist arrivals using time-varying parameter structural time series models, Int. J. Forecast. 27 (2011) 855–869, http://dx.doi.org/10.1016/j.ijforecast.2010.06.001.

[11] F. Halicioglu, An econometric analysis of the aggregate outbound tourism demand of Turkey, Tour. Econ. 16 (2010) 83–97, http://dx.doi.org/10.5367/000000010790872196.

[12] J. Wang, N. Zhang, H. Lu, A novel system based on neural networks with linear combination framework for wind speed forecasting, Energy Convers. Manage. 181 (2019) 425–442, S0196890418313542.

[13] J.H. Kim, M.T. Ngo, Modelling and forecasting monthly airline passenger flows among three major Australian cities, Tour. Econ. 7 (2001) 397–412, http://dx.doi.org/10.5367/000000001101297946.

[14] C.J.S.C. Burger, M. Dohnal, M. Kathrada, R. Law, A practitioners guide to time-series methods for tourism demand forecasting - a case study of Durban, South Africa, Tour. Manag. 22 (2001) 403–409, http://dx.doi.org/10.1016/S0261-5177(00)00068-6.

[15] M. Akin, A novel approach to model selection in tourism demand modeling, Tour. Manag. 48 (2015) 64–72, http://dx.doi.org/10.1016/j.tourman.2014.11.004.

[16] C. Lim, M. McAleer, Time series forecasts of international travel demand for Australia, Tour. Manag. 23 (2002) 389–396, http://dx.doi.org/10.1016/S0261-5177(01)00098-X.

[17] C.F. Chen, M.C. Lai, C.C. Yeh, Forecasting tourism demand based on empirical mode decomposition and neural network, Knowl.-Based Syst. 26 (2012) 281–287, http://dx.doi.org/10.1016/j.knosys.2011.09.002.

[18] Y. Yao, Y. Cao, X. Ding, J. Zhai, J. Liu, Y. Luo, S. Ma, K. Zou, A paired neural network model for tourist arrival forecasting, Expert Syst. Appl. 114 (2018) 588–614, http://dx.doi.org/10.1016/j.eswa.2018.08.025.

[19] S. Li, T. Chen, L. Wang, C. Ming, Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index, Tour. Manag. 68 (2018) 116–126, http://dx.doi.org/10.1016/j.tourman.2018.03.006.

[20] K.Y. Chen, C.H. Wang, Support vector regression with genetic algorithms in forecasting tourism demand, Tour. Manag. 28 (2007) 215–226, http://dx.doi.org/10.1016/j.tourman.2005.12.018.

[21] W. Lijuan, C. Guohua, Seasonal svr with foa algorithm for single-step and multi-step ahead forecasting in monthly inbound tourist flow, Knowl.-Based Syst. 110 (2016) 157–166, http://dx.doi.org/10.1016/j.knosys.2016.07.023.

[22] A. Altan, S. Karasu, The effect of kernel values in support vector machine to forecasting performance of financial time series, J. Cogn. Syst. 4 (n.d.) 17–21.

[23] A. Altan, S. Karasu, S. Bekiros, Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques, Chaos Solitons Fractals 126 (2019) 325–336, http://dx.doi.org/10.1016/j.chaos.2019.07.011.

[24] R.R. Andrawis, A.F. Atiya, H. El-Shishiny, Combination of long term and short term forecasts, with application to tourism demand forecasting, Int. J. Forecast. 27 (2011) 870–886, http://dx.doi.org/10.1016/j.ijforecast.2010.05.019.

[25] J.T. Coshall, Combining volatility and smoothing forecasts of UK demand for international tourism, Tour. Manag. 30 (2009) 495–511, http://dx.doi.org/10.1016/j.tourman.2008.10.010.

[26] K.Y. Chen, Combining linear and nonlinear model in forecasting tourism demand, Expert Syst. Appl. 38 (2011) 10368–10376, http://dx.doi.org/10.1016/j.eswa.2011.02.049.

[27] H. Song, B.Z. Gao, V.S. Lin, Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system, Int. J. Forecast. 29 (2013) 295–310, http://dx.doi.org/10.1016/j.ijforecast.2011.12.003.

[28] J. Wang, H. Li, H. Lu, Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China, Appl. Soft Comput. (2018) http://dx.doi.org/10.1016/j.asoc.2018.07.030.

[29] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series — part II, Fuzzy Sets and Systems 62 (1994) 1–8.

[30] H.-K. Yu, Weighted fuzzy time series models for TAIEX forecasting, Physica A 349 (2005) 609–624, http://dx.doi.org/10.1016/J.PHYSA.2004.11.006.

[31] J. Wang, H. Li, H. Lu, Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China, Appl. Soft Comput. J. 71 (2018) 783–799, http://dx.doi.org/10.1016/j.asoc.2018.07.030.

[32] C. Stefanakos, Fuzzy time series forecasting of nonstationary wind and wave data, Ocean Eng. 121 (2016) 1–12, http://dx.doi.org/10.1016/j.oceaneng.2016.05.018.

[33] C.-H. Cheng, T.-L. Chen, H.J. Teoh, C.-H. Chiang, Fuzzy time-series based on adaptive expectation model for TAIEX forecasting, Expert Syst. Appl. 34 (2008) 1126–1132, http://dx.doi.org/10.1016/J.ESWA.2006.12.021.
[34] H.J. Sadaei, P.C. de Lima e Silva, F.G. Guimarães, M.H. Lee, Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series, Energy 175 (2019) 365–377, http://dx.doi.org/10.1016/j.energy.2019.03.081.
[35] N. Güler Dincer, Ö. Akkuş, A new fuzzy time series model based on robust clustering for forecasting of air pollution, Ecol. Inform. 43 (2018) 157–164, http://dx.doi.org/10.1016/j.ecoinf.2017.12.001.
[36] Y.L. Huang, S.J. Horng, M. He, P. Fan, T.W. Kao, M.K. Khan, J.L. Lai, I.H. Kuo, A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization, Expert Syst. Appl. 38 (2011) 8014–8023, http://dx.doi.org/10.1016/j.eswa.2010.12.127.
[37] P. Jiang, H. Yang, J. Heng, A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting, Appl. Energy. 235 (2019) 786–801, http://dx.doi.org/10.1016/j.apenergy.2018.11.012.
[38] W. Lu, X. Chen, W. Pedrycz, X. Liu, J. Yang, Using interval information granules to improve forecasting in fuzzy time series, Internat. J. Approx. Reason. 57 (2015) 1–18.
[39] S.M. Chen, B.D.H. Phuong, Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors, Knowl.-Based Syst. 118 (2017) 204–216, http://dx.doi.org/10.1016/j.knosys.2016.11.019.
[40] S.M. Chen, X.Y. Zou, G.C. Gunawan, Fuzzy time series forecasting based on proportions of intervals and particle swarm optimization techniques, Inf. Sci. (Ny). 500 (2019) 127–139, http://dx.doi.org/10.1016/j.ins.2019.05.047.
[41] C.H. Cheng, J.H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing 302 (2018) 33–45, http://dx.doi.org/10.1016/j.neucom.2018.04.014.
[42] A. Rubio, J.D. Bermúdez, E. Vercher, Improving stock index forecasts by using a new weighted fuzzy-trend time series method, Expert Syst. Appl. 76 (2017) 12–20, http://dx.doi.org/10.1016/j.eswa.2017.01.049.
[43] H.J. Sadaei, F.G. Guimarães, C.J. Da Silva, M.H. Lee, T. Eslami, Short-term load forecasting method based on fuzzy time series, seasonality and long memory process, Internat. J. Approx. Reason. (2017).
[44] J. Jackson, Developing regional tourism in China: The potential for activating business clusters in a socialist market economy, Tour. Manag. 27 (2006) 695–706.
[45] J.C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c -means clustering algorithm, Comput. Geosci. 10 (1984) 191–203.
[46] O. Kesemen, Ö. Tezel, E. Özkul, Fuzzy c-means clustering algorithm for directional data (FCM4DD), Expert Syst. Appl. 58 (2016) 76–82, http://dx.doi.org/10.1016/j.eswa.2016.03.034.
[47] Y. Ding, X. Fu, Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm, Neurocomputing 188 (2016) 233–238, http://dx.doi.org/10.1016/j.neucom.2015.01.106.
[48] D.B.T.-I.I.C. on D.L. Mei, A new fuzzy information optimization processing technique for monitoring the transformer, 2000, pp. 273–275.
[49] C. Huang, Y. Shi, Towards Efficient Fuzzy Information Processing, Physica-Verlag HD, 2002.
[50] L.A. Zadeh, Fuzzy sets *, Inf. Control. 8 (1965) 338–353.
[51] J. Liu, L. Shuai, W. Jin, X. Liu, J. Zhang, Research of influence of sample size on normal information diffusion based on the Monte Carlo method: risk assessment for natural disasters, Environ. Earth Sci. 77 (2018) 480.
[52] W. Zhao, L. Wang, Z. Zhang, A novel atom search optimization for dispersion coefficient estimation in groundwater, Future Gener. Comput. Syst. (2018) http://dx.doi.org/10.1016/j.future.2018.05.037.
[53] W. Zhao, L. Wang, Z. Zhang, Atom search optimization and its application to solve a hydrogeologic parameter estimation problem, Knowl.-Based Syst. 163 (2018) 283–304, http://dx.doi.org/10.1016/j.knosys.2018.08.030.
[54] J. Whitney, Testing for differences with the nonparametric Mann–Whitney U test, J. Wound Ostomy Cont. 24 (1997) 12.
[55] G.D. Ruxton, The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test, Behav. Ecol. 17 (2010) 688–690.
[56] S. Zhang, J. Wang, Z. Guo, Research on combined model based on multi-objective optimization and application in time series forecast, Soft Comput. (2018) 0123456789.