Tourism forecasting research: a bibliometric visualization review (1999–2022)

XiaoXi Wu, Jinlian Shi and Haitao Xiong

Abstract

Purpose – This paper aims to analyze the research highlights, evolutionary process and future research directions in the field of tourism forecasting.

Design/methodology/approach – This study used CiteSpace to conduct a bibliometric analysis of 1,213 tourism forecasting articles.

Findings – The results show that tourism forecasting research has experienced three stages. The institutional collaboration includes transnational collaboration and domestic institutional collaboration. Collaboration between countries still needs to be strengthened. The authors’ collaboration is mainly based on on-campus collaboration. Articles with high co-citation are primarily published in core tourism journals and other relevant publications. The research content mainly pertains to tourism demand, revenue management, hotel demand and tourist volumes. Ex ante forecasting during the COVID-19 pandemic has broadened existing tourism forecasting research. The future forecasting research focuses on the rational use of big data, improving the accuracy of models and enhancing the credibility of forecasting results.

Originality/value – This paper uses CiteSpace to analyze tourism forecasting articles to obtain future research trends, which supplements existing research and provides directions for future research.

Keywords Tourism forecasting, Bibliometric analysis, CiteSpace, Model optimization, Future research trends, Big data

Paper type Research paper

Investigación sobre previsión turística: una revisión de visualización bibliométrica (1999–2022)

Resumen

Objetivo: Este artículo pretende analizar los aspectos más destacados de la investigación, el proceso evolutivo y las futuras orientaciones de la investigación en el campo de la previsión turística.

Diseño/metodología/enfoque: Este estudio utilizó CiteSpace para realizar un análisis bibliométrico de 1,213 artículos sobre previsión turística.

Received 16 March 2023
Revised 11 May 2023
Accepted 17 June 2023

This work was supported by the Humanities and Social Science Project of Ministry of Education of China (grant no. 21YJJCZH186) and the National Natural Science Foundation of China (grant no. 72171004).
Resultados: Los resultados muestran que la investigación sobre previsión turística ha experimentado tres etapas. La colaboración institucional incluye la colaboración transnacional y la colaboración institucional nacional. La colaboración entre países aún debe reforzarse. La colaboración entre autores se basa principalmente en la colaboración dentro del campus. Los artículos con una alta cocitación se publican principalmente en las principales revistas de turismo y en otras publicaciones relevantes. El contenido de la investigación se refiere principalmente a la demanda turística, el revenue management, la demanda hotelera y los volúmenes turísticos. La previsión previa y durante la pandemia de la COVID-19 ha ampliado la investigación existente sobre previsión turística. La futura investigación sobre previsiones se centrará en el uso racional de los big data, la mejora de la precisión de los modelos y el aumento de la credibilidad de los resultados de las previsiones.

Originalidad/valor: Este artículo utiliza CiteSpace para analizar artículos de previsión turística con el fin de obtener futuras tendencias de investigación, lo que complementa la investigación existente y proporciona orientaciones para futuras investigaciones.

Palabras clave Previsión turística, Análisis bibliométrico, CiteSpace, Optimización de modelos, Tendencias futuras de la investigación, Big data

Tipo de papel Trabajo de investigación

1. Introduction

Compared with other industries, tourism is more vulnerable to seasonal variations, so accurate tourism forecasting can solve the problem of crowded destinations and wasted resources, providing scientific guidance for the operation and sale of hotels, restaurants, attractions and other industries (Chen et al., 2015; Gunter, 2021; Li et al., 2022a; Yu and Chen, 2022). Furthermore, accurate tourism forecasts can inform governments or managers to make decisions and formulate appropriate development policies, promoting sustainable tourism development (Yao and Cao, 2020; Guizzardi et al., 2021; Hu, 2022; Song et al., 2022).

Tourism forecasting research began in the 1970s, and currently, its methods can be divided into two main categories: quantitative and qualitative. The Delphi method is the most prominent among the latter methods, and its application includes two central aspects, tourism forecasting and adjusting quantitative method forecasts. For example, Lee et al. (2008) used the Delphi method to forecast the number of visitors to an Expo, while Lin et al. (2014) used the opinions of the expert group to make judgmental adjustments to the econometric forecast of tourism demand, with results indicating that this approach improved forecasting accuracy. On the other hand, quantitative methods can be divided into the following four categories:

1. time series models;
2. econometric models;
3. artificial intelligence models; and
4. combined forecasting models.

1. Time series models are frequently used in tourism forecasting research because of their ability to capture historical patterns reasonably (Song et al., 2019). Among these models, the autoregressive integrated moving average (ARIMA) models and their extensions are commonly used. For example, Lim and McAleer (2002), Gounopoulos et al. (2012) and Ismail (2020) used the ARIMA model to forecast the number of tourists. In addition, Liang (2014), Alvarez-Diaz and Gupta (2016) and Basnayake and Chandrasekara (2022) extended the ARIMA model with the seasonal autoregressive integrated moving average model, using seasonal tourist behavior to forecast the tourist volumes. Finally, Zhao et al. (2022) proposed an approach based on time-series trajectory similarity and achieved high accuracy in forecasting daily tourist arrivals.

2. Econometric models: scholars use econometric models, including the vector autoregressive (VAR) and time-varying parameter (TVP) models, to find the causal relationship between
tourism demand and various economic factors. For example, Song and Witt (2006) used a VAR model to predict tourist flows to Macao. Song et al. (2011) constructed an structural time series model and time-varying parameter model to forecast the quarterly number of Hong Kong tourists in four major source markets. In addition, Assaf et al. (2019) constructed a Bayesian global vector autoregression model to predict international tourism flows.

3. Artificial intelligence models: frequently used artificial intelligence models include support vector regression (SVR), BP neural networks, kernel extreme learning machines and deep learning. Regarding their application in research, Li et al. (2018) and Lin et al. (2018) proposed the principal component analysis-adaptive differential evolution algorithm-back propagation neural network and empirical mode decomposition-back propagation neural network forecasting models to predict tourist volumes. Chen et al. (2015) and Wu and Cao (2016) constructed the fruit fly optimization algorithm-support vector regression model to forecast the tourist flow. Sun et al. (2019) found that the kernel extreme learning machine model can significantly improve the accuracy and robustness of analysis in forecasting performance. Finally, Tsang and Benoit (2020) indicated that Gaussian processes could be used for point and interval forecasting.

4. Combined forecasting models combine many single models by averaging, weighting and regression. Shen et al. (2008) used the simple average combination method, the variance–covariance combination method and the discounted mean square forecast error method to combine the forecasting models. Chen (2011) combined the linear and nonlinear statistical models. Most studies show that combined forecasting models have better accuracy than single models (Shen et al., 2011). In summary, many models are available for tourism forecasting, but no single model applies to every case (Song and Li, 2008).

Witt and Witt first studied the literature review of tourism forecasting and discussed the main techniques used in the field before 1995. The research methods include descriptive analysis and bibliometric analysis. Song, Law, Goh, Li and other scholars published review articles on tourism forecasting, mostly adopting descriptive analysis techniques. By sorting, summarizing and analyzing the existing tourism forecasting literature, the authors obtained the research course, research priorities and research trends. These tourism forecasting review articles have a variety of focuses, such as tourism demand forecasting literature reviews, hotel demand forecasting literature reviews and airline demand forecasting literature reviews. Some authors also conduct literature reviews for articles that use specific forecasting methods or predictor variables. Descriptive analysis methods are commonly used, but the discipline’s traditional descriptive analysis is highly subjective, casting doubt on the research results attained from its use.

In contrast, bibliometric analysis is a comprehensive, systematic and computationally driven approach. The bibliometric analysis can yield the current state of collaboration, literature co-citation, research priorities and future research trends. Thus, some scholars have started to use bibliometric methods to analyze tourism forecasting articles. For example, Liu et al. (2019) and Zhang et al. (2020) used CiteSpace to analyze the literature on tourism forecasting, finding that integrating Web-based data will be the focus and direction of future research. The literature analyzed in these articles covers all aspects of the field of tourism forecasting. At present, compared with descriptive analysis, few articles use bibliometric analysis to examine the tourism forecasting literature, and even fewer articles include the period of COVID-19. Since the outbreak of COVID-19, the global tourism industry has taken a huge hit. Forecasting the trend and recovery of tourism is particularly important for the future development of tourism. So, this study uses bibliometric methods to analyze the development lineage of tourism forecasting, research priorities and future research directions, providing guidelines for where future research may venture. This study is not only a supplement to the bibliometric analysis of tourism forecasting articles by Liu and Zhang, but also a new exploration of tourism forecasting research during the COVID-19 epidemic (Table 1).
Table 1 Main studies of tourism forecasting review

<table>
<thead>
<tr>
<th>Authors</th>
<th>Research contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Witt and Witt (1995)</td>
<td>They discussed the main techniques used in tourism forecasting before 1995</td>
</tr>
<tr>
<td>Song and Li (2008)</td>
<td>They reviewed 121 tourism demand forecasting articles since 2000; they found that no single model can be applied to every case of tourism forecasting</td>
</tr>
<tr>
<td>Lin and Song (2015)</td>
<td>They reviewed published research using the Delphi prediction of the hotel and tourism industry in the past 40 years</td>
</tr>
<tr>
<td>Li et al. (2021)</td>
<td>They reviewed the articles on internet data used for tourism forecasting between 2012 and 2019</td>
</tr>
<tr>
<td>Wang and Gao (2021)</td>
<td>They reviewed 87 air travel demand articles from 2010 to 2020</td>
</tr>
<tr>
<td><strong>Bibliometric analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Liu et al. (2019)</td>
<td>They retrieved articles on tourism forecasting from the Web of Science (WoS) Core Collection database from 1973 to April 2018. They found that a new trend is to use search engine data</td>
</tr>
<tr>
<td>Zhang et al. (2020)</td>
<td>They retrieved articles on tourism forecasting from the WoS Core Collection database from 1999 to 2018. They found that econometric, time series and intelligence models were the primary methods; likewise, integrating Web-based data is the focus and direction of future research</td>
</tr>
</tbody>
</table>

Source: Produced by the authors

2. Methodology

2.1 Data collection

In this paper, we searched from the Web of Science (WoS) Core Collection for tourism forecasting literature. The WoS Core Collection database is the most reputable and authoritative bibliographic database; it contains numerous high-impact journal and conference proceedings. Therefore, it enables us to find high-quality articles and ensure the authenticity and reliability of the data sources (Li et al., 2017; Fang et al., 2018). We searched the literature using the keywords: “tourism,” “forecasting,” “tourism” and “prediction.” We searched 2,704 records while excluding articles that only researched forecasting or tourism by reading these articles’ titles, abstracts and keywords. As a result, 1,213 records (from 1999 to 2022) were obtained.

2.2 Data analysis

CiteSpace is a popular bibliometric analysis software developed by Professor Chen Chaomei at Drexel University. In recent years, some scholars have used CiteSpace to analyze tourism crises, inbound tourism, cultural tourism or other topics in the field. However, scholars rarely use CiteSpace to analyze the tourism forecasting literature systematically and comprehensively.

The primary function of CiteSpace is to generate knowledge maps of different categories. The first tool it uses for map generation is the co-occurrence network, where, if different keywords, categories or terms appear in the same literature, a connection exists between them. CiteSpace can thus form co-occurrence network diagrams based on connections or links between two keywords or categories that appear in the same article. The collaboration network is the second tool for creating knowledge maps, mainly used to analyze the authors of different countries and institutions in the same article. If there is a collaboration between different countries, institutions and authors, here are links in the collaboration network diagram. The third one is the co-citation network. Traditionally as long as two documents,
authors or journals are cited within the third article, those documents, authors or journals are considered co-citations. Therefore, whenever there is a link between two authors, journals or documents, they are cited by a third article simultaneously. The fourth one is document bibliographic coupling, which means if a document is cited by two other articles simultaneously, a literature coupling can be said to exist between these two articles.

In addition to the four types of maps, burst detection and timeline are also noteworthy features of CiteSpace. Here, a “burst” is a surge in the frequency of particular events (Kleinberg, 2003). CiteSpace supports those different types of burst detection by displaying, for example, the frequencies of keywords; the number of publications by institutions, countries or authors; and the number of citation counts of cited references. Citation bursts contain burst intensity and duration (Li et al., 2022b). In a timeline map, each keyword is placed on a different timeline depending on when it appears, reflecting the evolution of the research focus over time and helping researchers identify emerging trends (Chen, 2006). CiteSpace also calculates structure-related indicators, such as silhouette score, betweenness centrality and modularity. A cluster’s silhouette score measures the cluster’s homogeneity; clusters with high silhouette scores are considered more meaningful and worthy of scholars’ attention. Betweenness centrality measures how likely an arbitrary shortest path in the network will go through the node. For example, a node in the middle of a sub-network or between two large communities will have a high betweenness centrality (Brandes, 2001; Qiao et al., 2022). Finally, modularity can reflect the internal structure of the network. The higher the modularity, the closer the coefficient is to 1 (Newman, 2006).

2.3 Analysis framework

Figure 1 shows the analytical framework for this study. First, the articles on tourism forecasting from the WoS database were searched. This study conducted descriptive statistics and bibliometric analysis of the searched articles on tourism forecasting. Second, this study uses descriptive statistics to analyze the evolution over time of the number of articles. Third, for bibliometric analysis, this study uses CiteSpace to obtain the maps of collaboration networks, co-citation networks, keywords co-occurrence and clusters. The collaborative networks analyze which institutions, countries and authors study tourism forecasts and which institutions, countries and authors collaborate. The co-citation networks mainly analyze the co-citation and burst detection of authors and journal articles.

**Figure 1** Analysis framework

![Analysis framework diagram](image-url)
Theme analysis mainly includes keyword co-occurrence analysis, timeline analysis and burst detection. Fourth, cluster analysis helps summarize the research themes of tourism forecasting literature. Finally, based on descriptive statistics and bibliometric analysis, this study summarizes the models or methods for tourism forecasting research and gets the current hotspots and development trends in the tourism forecasting field to assist subsequent research.

3. Results and discussion

3.1 Descriptive statistical analysis

Figure 2 shows the number of tourism forecasting articles published per year between 1999 and 2022; the number of articles continues to grow. From Figure 2, by sorting the number of articles published into categories of less than 20, 20–60 and more than 60, we can divide the study of tourism forecasting into three stages. The first stage (1999–2007) was a low and steady development stage. The number of articles was low at this stage, with no more than 20 published yearly. However, the number of articles grew over time. The second stage (2008–2016) had steadily rising development. The number of articles was between 20 and 60, illustrating how tourism forecasting research was gaining more scholars’ attention. Finally, the third stage (2017–2022) is one of rapid development. At this stage, the number of articles published shows a linear upward trend, evidencing that tourism forecasting has become a research priority for scholars.

3.2 Collaboration network analysis

3.2.1 Collaborations between institutions. According to the data in Table 2 and Figure 3, Hong Kong Polytechnic University is the best institution for research on tourism forecasting, followed by the University of Chinese Academy of Sciences and the University of Surrey. The centrality of the Hong Kong Polytechnic University is greater than 0.1. Thus, this institution is central to the institutional collaboration network, connecting different institutions. The centrality of the University of Surrey, the University of Chinese Academy of Sciences and the University of Macau is greater than 0.05; thus, these institutions are at the sub-center of the institutional collaboration network. While the institutional collaboration network is loose, many institutions collaborate with other institutions on the periphery of the

![Figure 2: Annual number of publications from 1999 to 2022](image-url)

Source: Produced by the authors
collaboration network, all of which have a centrality of 0. The institutional collaboration includes transnational collaboration and domestic institutional collaboration.  

3.2.2 Collaborations between countries/regions. From Table 3 and Figure 4, China and the USA are the first countries to research tourism forecasting; England, Spain and Australia followed closely. The centrality of China is 0.5, showing that this country is at the center of the international collaboration network, connecting different countries. The centrality of England is 0.36, and is in second place, which means England is at the sub-center of the collaboration network. Several countries, such as Greece and Portugal, are on the network’s edge, as they only collaborated with China, the United Kingdom, and the USA; thus, the centrality of these countries is 0. China, the USA and England have the earliest collaboration with other countries.  

3.2.3 Collaborations between authors. Table 4 and Figure 5 show that the authorial collaboration network structure is loose, with multiple small, independent collaboration networks. Haiyan Song is at the top of the cooperative author network, followed by Gang Li

<table>
<thead>
<tr>
<th>Count</th>
<th>Centrality</th>
<th>Year</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>0.18</td>
<td>2000</td>
<td>Hong Kong Polytechnic University</td>
</tr>
<tr>
<td>50</td>
<td>0.06</td>
<td>2013</td>
<td>University of Chinese Academy of Sciences</td>
</tr>
<tr>
<td>24</td>
<td>0.07</td>
<td>2003</td>
<td>University of Surrey</td>
</tr>
<tr>
<td>14</td>
<td>0.02</td>
<td>2021</td>
<td>Xi’an Jiao Tong University</td>
</tr>
<tr>
<td>13</td>
<td>0.02</td>
<td>2013</td>
<td>Sun Yat-sen University</td>
</tr>
<tr>
<td>10</td>
<td>0.02</td>
<td>2011</td>
<td>Bournemouth University</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>2011</td>
<td>Monash University</td>
</tr>
<tr>
<td>9</td>
<td>0.05</td>
<td>2021</td>
<td>University Macau</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>2019</td>
<td>University of Nottingham Ningbo China</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>2017</td>
<td>Nanjing University of Aeronautics and Astronautics</td>
</tr>
</tbody>
</table>

Source: Produced by the authors
and Rob Law. The former is an expert in conducting tourism forecasting research and has cooperative relationships with many scholars. Law’s centrality is 0.07, and Song’s is 0.06, indicating they are central to the author collaboration network. Finally, the centrality of Gang Li and Shouyang Wang is each greater than 0.01. These authors often collaborate and
serve as the center of their respective networks, connecting other authors. Although the existing author collaboration is mainly based on on-campus collaboration, it is necessary to enhance the diversity of author collaboration.

3.3 Co-citation analysis

Each article cites other literature, and this process can be seen as a flow of different research themes toward current research. Therefore, the evolution of tourism forecasting and the flow of knowledge can be obtained by analyzing the references co-cited in tourism forecasting articles (Liu et al., 2019). The co-citation analysis includes cited author analysis and cited journal analysis.

3.3.1 Cited author analysis. From Table 5 and Figure 6, Haiyan Song’s articles published in 2001 and 2006 have the highest number of citations. Rob Law, Christine Lim and Stephen Witt followed Haiyan Song. The highest centrality from FongLin Chu, Haiyan Song, Rob Law, Christine Lim and C. Goh is greater than 0.05, meaning those authors’ articles are substantially connected. In addition, the articles of these highly cited authors were published earlier. The 20 authors with citation bursts are listed in Figure 7. The end time of the 20 authors with citation bursts was before 2014, consistent with the time listed in Table 5. Thus, the articles of highly cited authors are mainly published in the first two stages, laying the foundation for the rapid development of the third stage.

3.3.2 Cited journal analysis. From Table 6, Tourism Management is one of the most highly cited journals. In addition to Tourism Management, the Annals of Tourism Research, the International Journal of Forecasting and the Journal of Travel Research rank among the most cited journals. The articles are published in journals specializing in tourism and core journals in other specialist majors, such as applied mathematics, computer science and statistical probability. The Annals of Tourism Research’s centrality is 0.11, placing it at the
Tourism Management, the Journal of Travel Research, Expert Systems with Applications and the Journal of Forecasting have a relevance greater than 0.05, indicating that the tourism forecasting research articles published in these journals were widely cited. By analyzing the journals with citation bursts, the duration of the citation burst across journals was long, and the journals with citation bursts are the top journals in every field (Figure 8).

### 3.4 Themes analysis

Keywords are words that summarize and indicate the content of an article. By analyzing the co-occurrence of keywords, we can get a clear overview of the knowledge structure of the subject. From Figure 9, according to the size of the nodes, it can be obtained that demand, model, tourism demand, forecasting, time series, neural network and arrival are the keywords with the highest number of co-occurrences. These keywords represent the

---

**Table 5** Table of cited authors

<table>
<thead>
<tr>
<th>Count</th>
<th>Centrality</th>
<th>Year</th>
<th>Cited authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>355</td>
<td>0.06</td>
<td>2001</td>
<td>Haiyan Song</td>
</tr>
<tr>
<td>301</td>
<td>0.09</td>
<td>2006</td>
<td>Haiyan Song</td>
</tr>
<tr>
<td>231</td>
<td>0.09</td>
<td>2000</td>
<td>Rob Law</td>
</tr>
<tr>
<td>228</td>
<td>0.06</td>
<td>2000</td>
<td>Christine Lim</td>
</tr>
<tr>
<td>204</td>
<td>0.04</td>
<td>1999</td>
<td>Stephen Witt</td>
</tr>
<tr>
<td>194</td>
<td>0.03</td>
<td>2006</td>
<td>Gang Li</td>
</tr>
<tr>
<td>161</td>
<td>0.05</td>
<td>2005</td>
<td>C.Goh</td>
</tr>
<tr>
<td>137</td>
<td>0.11</td>
<td>2002</td>
<td>FongLin Chu</td>
</tr>
<tr>
<td>137</td>
<td>0.03</td>
<td>2001</td>
<td>N. Kulendran</td>
</tr>
<tr>
<td>136</td>
<td>0.02</td>
<td>2009</td>
<td>George Athanasopoulos</td>
</tr>
</tbody>
</table>

*Source: Produced by the authors*
research focus on tourism demand forecasting and forecasting models. Accuracy, big data, machine learning and genetic algorithms are frequently used, highlighting how academic research often focuses more on improving forecasting models’ accuracy. Two methods to improve forecasting accuracy exist: adding exogenous variables and optimizing existing models.

From Table 7 and Figure 10, we can get the time evolution of tourism forecasting research. The keywords can be classified into three types during the low and steady development stages. First, the centrality of tourism demand and accuracy is higher than 0.05, meaning that tourism demand and accuracy are the research focus in this stage. Second, the tourism forecasting model mainly includes time series, neural networks and econometric forecasts. Third, the research subjects mainly include Hong Kong and the USA, which
aligns with the findings of the country collaboration network analysis. Finally, scholars must add exogenous variables such as income and price into the forecasting models (Law and Au, 1999; du Preez and Witt, 2003; Song et al., 2003). Compared with the time series and econometric models, the forecasting accuracy of artificial intelligence models is higher (Law, 2000; Burger et al., 2001; Chen and Wang, 2007).
The keywords can be roughly classified into five types during the steadily rising stage. Including seasonality, arrival, revenue management and combination performance, each has a centrality higher than 0.05, indicating their prevalence during this stage. In addition, forecast combination, algorithm, support vector regression, optimization, machine learning and fuzzy time series forecasting model each have a centrality higher than 0.05, which means that combination forecasting models, AI forecasting models, and forecasting model

<table>
<thead>
<tr>
<th>Development stage</th>
<th>Category of keywords</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low and steady</td>
<td>Tourism forecasting model</td>
<td>Model (0.16); Time Series (0.05); Neural Network (0.09); Genetic Algorithm (0.04); Econometric Forecast (0.00)</td>
</tr>
<tr>
<td>Research content</td>
<td></td>
<td>Tourism Forecasting (0.04); Demand (0.05); Accuracy (0.11); Forecasting (0.1); Tourism Demand (0.04); Tourism (0.13)</td>
</tr>
<tr>
<td>Research subjects</td>
<td></td>
<td>Hong Kong (0.00); USA (0.01)</td>
</tr>
<tr>
<td>Rapid development stage</td>
<td>Tourism forecasting model</td>
<td>Artificial Neural Network (0.04); Econometric Model (0.01); Sarima (0.01); Forecast Combination (0.05); BP Neural Network (0.00); Arima (0.03); Algorithm (0.05); Support Vector Regression (0.07); Optimization (0.05); Machine Learning (0.07); Fuzzy Time Series Forecasting Model (0.06)</td>
</tr>
<tr>
<td>Research subjects</td>
<td></td>
<td>Australia (0.01); Spain (0.01); South Korea (0.00); UK (0.00)</td>
</tr>
<tr>
<td>Research purpose</td>
<td></td>
<td>Management (0.03); Decision Making (0.00)</td>
</tr>
<tr>
<td>Data sources</td>
<td></td>
<td>Google (0.00); Big Data (0.04)</td>
</tr>
<tr>
<td>Research purpose</td>
<td></td>
<td>Design (0.00); Sustainability (0.01)</td>
</tr>
<tr>
<td>Tourism forecasting model</td>
<td></td>
<td>Deep Learning (0.02); Time-Varying Parameter (0.00); LSTM (0.02); Artificial Intelligence (0.01); Hybrid Approach (0.01); Fuzzy Time Series (0.00); Decision Tree (0.00); Grey Prediction (0.00); Particle Swarm Optimization (0.00); CNN (0.00); ANN (0.00); Grey Model (0.00); Machine Learning (0.00); SARIMAX (0.00); Random Forecast (0.00); Box-Jenkins Methodology (0.00); Multi-view Learning (0.00)</td>
</tr>
<tr>
<td>Research content</td>
<td></td>
<td>Destination (0.04); Growth (0.03); Big Data Analytics (0.00); Hotel (0.03); Tourism Industry (0.00); Hotel Tourist Arrival (0.00); Traffic Growth (0.00); Volume (0.02); Passenger (0.00); Traffic Prediction (0.00); Tourist Flow (0.00); Occupancy (0.02); Climate Change (0.01); Covid-19 (0.01); Risk (0.00); Uncertainty (0.00); Sentiment Analysis (0.00); Motivation (0.00)</td>
</tr>
<tr>
<td>Data sources</td>
<td></td>
<td>Google Trend (0.00); Baidu Index (0.00); Composite Search Index (0.00); Search Query Data (0.00); Dynamic Panel Data (0.00); Search Engine (0.01); Online Review (0.00); Social Media (0.00); Index (0.01)</td>
</tr>
<tr>
<td>Research subjects</td>
<td></td>
<td>China (0.02); USA (0.00)</td>
</tr>
</tbody>
</table>

Note: * and ** represent the centrality of each keyword
Source: Produced by the authors
optimization have become common models for tourism forecasting (Cai et al., 2009; Chen et al., 2015; Cao and Wu, 2016; Sakhuja et al., 2016). In addition to panel data, big data has also begun to become the data for tourism forecasting. Big data uses include examining Web traffic volume (Yang et al., 2014), big weather data (Lee et al., 2015) and search engine data (Artola et al., 2015; Bangwayo-Skeete and Skeete, 2015; Yang et al., 2015). Finally, Australia, Spain and the UK have significantly progressed in tourism forecast research. Many scholars believe that accurate tourism forecasting is beneficial to tourism management and tourism managers to make correct decisions.

During the rapid development stage, keywords can be divided into five main categories; destination, growth, hotel, volume, occupancy, climate change and COVID-19’s centrality are all higher than 0.01, evidencing how the research content gained richness over time. Scholars believe accurate forecasting is conducive to sustainable tourism development and related policy design (Li et al., 2019; Ramos et al., 2021; He et al., 2022). The research core is to find suitable forecasting models to improve accuracy. The centrality of deep learning, long short-term memory, artificial intelligence and a hybrid approach is higher than 0.01, which means that artificial intelligence (AI) forecasting models and combined forecasting models were the common models. AI and combined forecasting models have become the main models, while econometric and time series models were primarily used as benchmarks (Li et al., 2018; Sun et al., 2019; Wang, 2021). With the in-depth development of Internet technology, scholars have begun to use deep learning to forecast hotel demand and tourist volumes (Zhang et al., 2019; Huang and Zheng, 2021; Peng et al., 2021; He et al., 2022). AI forecasting models can improve forecasting progress, but they cannot be explained theoretically, and the results are questionable. Introducing big data into tourism forecasting can significantly improve forecasting models. However, big data is mainly based on search engine data (Law et al., 2019; Sun et al., 2019; Wen et al., 2019). In addition, multiple types of big data sources, such as social media (Topal et al., 2018; Ampountolas and Legg, 2021; Kovacs et al., 2021; Tian et al., 2021), news (Park et al., 2021) and online review (Li et al., 2020; Hu et al., 2022) are comprehensively used. Still, the current research difficulty pertains to dealing with big data and combining big data and traditional statistical data into forecasting models.

Particularly in the wake of the COVID-19 outbreak, the significance of tourism forecasting is more pronounced. In the case of the COVID-19 pandemic, the forecasting accuracy of
traditional forecasting models has significantly decreased, because they cannot handle exceptional situations. To solve this problem, some scholars use data on confirmed cases of COVID-19 disease, vaccinations and policy responses to construct a COVID-19 indicator and add it to the original tourism forecasting models (Prilistya et al., 2021; Turtureanu et al., 2022; Zhang and Lu, 2022). Most of the existing tourism forecasting articles are post-forecasting, and the innovation lies in improving the forecasting accuracy of the model. Tourism recovery has become the focus of attention after the COVID-19 epidemic. Some scholars use probabilistic forecasting models or scenario-based judgmental forecasting methods to predict tourism recovery. These forecasting models are mostly a combination of quantitative and qualitative methods and are mostly ex ante forecasts (Arbulu et al., 2021; Liu et al., 2021; Zhang et al., 2021). The outbreak of the COVID-19 epidemic has greatly changed tourists’ psychology, so tourism sentiment forecasting is also a research focus after the COVID-19 epidemic (Karl et al., 2021). The addition of relevant variables of COVID-19 and ex ante forecasting broadens the existing tourism forecasting studies.

Citation bursts analysis of the keywords yielded Figure 11, showing that keywords such as tourism forecasting, time series, tourism demand, support vector regression and artificial neural networks have been cited for a number of years. Therefore, these keywords have long been the research focus of scholars.

3.5 Literature clusters analysis

There are three clustering analysis algorithms in CiteSpace: LSI, LLR and MI. In this study, based on comparing three algorithms and drawing on previous research by Chen et al. (2012), the LLR algorithm was chosen for clustering analysis. From Table 8, 11 main clusters were obtained by analyzing the literature cluster. On the one hand, the modularity (Q) of the tourism

![Figure 11 Keywords with the strongest citation bursts](image-url)
forecasting network is 0.736, which proves that the network structure fits well. On the other hand, the mean silhouette of the overall network is 0.4461, less than 0.5, which proves the overall clustering is poor. The silhouette core of each cluster is greater than 0.7, which evidences that each clustering result is credible:

1. The 11 clusters obtained by cluster analysis can be divided into three theme groups:
2. tourism forecasting models;
3. tourism forecasting research content; and
4. big data.

The theme group of the tourism forecasting models includes Cluster #1, LSTM; Cluster #4, spatial correlation; Cluster #5, particle swarm optimization; Cluster #6, fuzzy time series forecasting model; and Cluster #12, integrated approach. This group includes the AI forecasting model, time series model, combined forecasting model and optimization algorithm. The silhouette score of Clusters #6 and #12 is greater than 0.8, meaning these clusters have high credibility. Clusters #1 and #12 appear later, consistent with the above keywords’ co-occurrence analysis, which means that as time passes and technology advances, deep learning and integrated analysis will be the common research methods for tourism forecasting.

The research content theme group includes Cluster #3, seasonality; Cluster #7, sentiment analysis; Cluster #9, COVID-19; and Cluster #11, revenue management. The research in this theme group was conducted after 2015, and the silhouette scores of each cluster were high. The tourism industry has seasonal characteristics, and considering seasonality can effectively change the forecasting results (Shen et al., 2009; Song et al., 2011; Vergori, 2012; Vergori, 2017; Elamin and Fukushige, 2018; Mishra et al., 2018). The main tourism forecasting contents are sentiment analysis, revenue management and COVID-19. The research content has become richer and more detailed, especially after the outbreak of COVID-19; forecasting the recovery of the tourism industry has become scholars’ research focus.

The theme group of big data includes Cluster #2, search query data, and Cluster #8, Google. Clusters #2 and #8 appear in the steadily rising stage, consistent with the above keywords’ co-occurrence analysis. Since then, search engine data has continued to be added to tourism forecasting models because it has high stability and can improve forecasting accuracy. Thus, search engine data will be one of the leading big data sources in future studies (Figure 12).

In summary, the tourism forecasting research content mainly pertains to tourism demand, revenue management, hotel demand and tourist volumes. The research method is mainly

<table>
<thead>
<tr>
<th>Table 8</th>
<th>Summary of the clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster ID</strong></td>
<td><strong>Size</strong></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

Source: Produced by the authors
quantitative forecasting models. Compared with the time series and econometric models, artificial intelligence forecasting models have higher accuracy but cannot be explained theoretically. After the COVID-19 epidemic, combining quantitative and qualitative forecasting of tourism recovery and tourism sentiment has become the focus of scholars' research. Future research focuses on more rational use of big data, improving tourism forecasting accuracy and increasing the interpretability of artificial intelligence models.

4. Conclusion and future research

This paper uses the bibliometric method to analyze 1,213 tourism forecasting articles extracted from the WoS core database. The results show that tourism forecasting research has experienced three stages. In addition, Hong Kong Polytechnic University is the most-cited institution researching tourism forecasting, and institutional collaboration includes transnational and domestic institutional collaboration. China and the USA are the first countries to research tourism forecasting, while Haiyan Song is the most cited author researching tourism forecasting, followed by Gang Li and Rob Law. Furthermore, the co-citation analysis yielded that the articles in the first two stages were widely cited by the authors in the third stage, laying the foundation for its rapid development. Finally, *Tourism Management*, *Annals of Tourism Research* and the *International Journal of Forecasting* rank among the most cited journals.

The theme and cluster analyses revealed that some keywords such as model, demand, neural network, time series and tourism represent the research focus on tourism demand and tourism forecasting models. In addition, the analyses demonstrated that the task of improving the accuracy of forecasting models is a research hotspot. Adding exogenous variables and optimizing existing models are two methods that can improve forecasting accuracy. According to the analysis of the temporal evolution of keywords, the research content is becoming increasingly rich, and AI forecasting models and combined forecasting models have higher forecasting accuracy, but they cannot be explained theoretically. Applying big data from sources such as social media, search engines and online reviews in tourism forecasting can improve predictive accuracy and broaden the research of tourism forecasting. After the COVID-19 pandemic, traditional forecasting models have significant errors. Therefore, using probability or forecasting models to
forecast tourism recovery has become a research focus. Future research may focus on more rational use of big data, improving tourism forecasting accuracy and increasing the interpretability of artificial intelligence models.

However, to ensure data quality, only the WoS Core Collection database was chosen as the sample data source for this study, and articles on Scopus or Google Scholar on tourism forecasting were not included in the analysis. Unfortunately, this choice has resulted in incomplete data and less comprehensive findings obtained. Moreover, this study used only articles in English for analysis, so the findings may have language bias. Therefore, in the follow-up research, literature from different databases should be introduced into the research as much as possible. Finally, to compensate for CiteSpace’s shortcomings, we shall consider combining qualitative and quantitative analyses in future research.

References


Topal, I., Ucar, M.K. and Ieee (2018), International Conference on Artificial Intelligence and Data Processing (IDAP), Ieee.


**About the authors**

XiaoXi Wu is a doctoral candidate in the School of International Economics and Management, Beijing Technology and Business University. Her research interests include smart tourism, tourism forecasting and tourism big data.

Jinlian Shi was born in Hohhot, Inner Mongolia, China, in 1972. She joined Beijing Technology and Business University as a Full Professor at the School of International Economics and Management in 2020. Her current research focuses on sustainable tourism, ecotourism in protected area and outdoor recreation management in national park.

Haitao Xiong was born in Jiujiang, Jiangxi, China, in 1983. Since 2021, he has been a Professor at the School of International Economics and Management, Beijing Technology and Business University, Beijing, China. His current research interests include tourism forecasting, machine learning and business intelligence. Haitao Xiong is the corresponding author and can be contacted at: xionghaitao@btbu.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com