Annals of Tourism Research 82 (2020) 102899

Contents lists available at ScienceDirect

Annals of Tourism Research

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

Research article

Group pooling for deep tourism demand forecasting

Yishuo Zhang[a,b], Gang Li[a,b,*], Birgit Muskat[c], Rob Law[d], Yating Yang[b]

[a] School of Information Technology, Deakin University, Geelong, VIC 3216, Australia
[b] Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China
[c] Research School of Management, ANU College of Business & Economics, Australian National University, ACT 2601, Australia
[d] School of Hotel & Tourism Management, The Hong Kong Polytechnic University, 17 Science Museum Rd, TST East, Kowloon, Hong Kong

ARTICLE INFO

Associate editor: Gang Li

Keywords:
Tourism demand forecasting
AI-based methodology
Group-pooling method
Deep-learning model
Tourism demand similarity
Asia Pacific travel patterns

ABSTRACT

Advances in tourism demand forecasting immensely benefit tourism and other sectors, such as economic and resource management studies. However, even for novel AI-based methodologies, the challenge of limited available data causing model overfitting and high complexity in forecasting models remains a major problem. This study proposes a novel group-pooling-based deep-learning model (GP–DLM) to address these problems and improve model accuracy. Specifically, with our group-pooling method, we advance the tourism forecasting literature with the following findings. *First*, GP–DLM provides superior accuracy in comparison with benchmark models. *Second*, we define the novel dynamic time warping (DTW) clustering quantitative approach. *Third*, we reveal cross-region factors that influence travel demands of the studied regions, including "travel blog," "best food," and "Air Asia."

Introduction

Tourism planners, policymakers, hotel managers, retailers, and government agencies rely on accurate forecasting to develop strategies, long-term business plans, and manage infrastructure capacity and visitor numbers in the short term (Pan & Yang, 2017; Song, Qiu, & Park, 2019; Wan & Song, 2018). In 2019, global tourism demand forecasting yielded over USD 2750.7 billion, generating 3.2% of the global GDP (WTTC, 2019). Predicting tourism demand volumes and making reliable assumptions on economic growth and social and environmental impacts require accurate and robust demand forecasting models and methodologies. Concurrently, considerable tourism research and practice rely on accurate demand modeling and forecasting.

Despite many contributions to the literature, research on tourism demand modeling remains "one of the most interesting and important areas of research in tourism studies" (Silva, Hassani, Heravi, & Huang, 2019, p. 134). The high relevance to further advance tourism forecasting methodology lies in addressing the key challenge to improve accuracy and omit overfitting of the developed model (e.g. Cai, Lu, & Zhang, 2009; Li, Wu, Zhou, & Liu, 2019; Wu, Law, & Xu, 2012). Assaf, Li, Song, and Tsionas (2019) further stressed that new tourism demand forecasting research should focus on enhancing the accuracy of data sources and improve processing techniques to generate higher model predictability. Thus, the aim of this study is to fill in these gaps in the existing tourism literature by using innovative pooling-based deep-learning techniques to obtain higher accuracy and enhance forecasting model effectiveness.

Deep-learning methods are advanced artificial neural network models that use network architectures with a high number of interconnected processing layers. For instance, Law, Li, Fong, and Han (2019) used deep learning with an attention mechanism to

* Corresponding author at: School of Information Technology, Deakin University, Geelong, VIC 3216, Australia.
E-mail addresses: yishuo.zhang@tulip.org.au (Y. Zhang), gang.li@deakin.edu.au (G. Li), birgit.muskat@anu.edu.au (B. Muskat), rob.law@polyu.edu.hk (R. Law), yangyt@ms.xjb.ac.cn (Y. Yang).

https://doi.org/10.1016/j.annals.2020.102899
Received 12 August 2019; Received in revised form 21 February 2020; Accepted 28 February 2020
Available online 13 March 2020

construct a large-scale neural network. However, the problem with existing deep-learning methodologies remains that these advanced models often face the overfitting problem (Shi, Xu, & Li, 2017). One key underlying problem that causes an overfitted model is typically the limited data availability (Hawkins, 2004; Tetko, Livingstone, & Luik, 1995), which can be solved by having larger data sample or data diversity (Goodfellow, Bengio, & Courville, 2016).

Group pooling is a method that can integrate time series data with cross-sectional data. The purpose of group pooling is to mitigate limited data availability and also increase data diversity. One advantage of group pooling of data is the ability to explore inter-homogeneity or heterogeneity across different regions (Baltagi & Griffin, 1997). Thus, with this combination of multiple data sources, group-pooling method proves to be an effective technique to address this major problem of limited data availability. In other research areas, such as energy forecasting, Shi et al. (2017) successfully demonstrated that with group pooling, data from different data sources can increase data volume and diversity. However, only limited tourism demand forecasting studies have attempted to employ the group pooling data. One of the few studies is from Long, Liu, and Song (2019), who used group-pooling data from Chinese cities for their tourism demand forecasting. However, the issue with heterogeneity of their data causes their model to perform worse than the benchmark models.

Our study addresses the challenge of generating group-pooling data with added homogeneity. The aim of this study is to improve the accuracy of tourism demand forecasting methods and attempt to alleviate the limitations that cause overfitting in extant studies. Our study is one of the first studies to combine group-pooling method and deep-learning method Methodologically, we advance the tourism literature with three major contributions:

1) We develop a dynamic time warping (DTW) clustering method and a pooling method to identify regions with similar tourism arrival volumes. With this novel group pooling-method, we extend prior pooling forecasting studies and alleviate the issues of limited data availability and model overfitting, and further improve the overall accuracy.
2) For the Asia-Pacific tourism markets, we identify similar cross-country homogenous tourism arrival patterns for the Hong Kong and Macau tourism markets. With our novel group pooling-based deep-learning model (GP–DLM), we complement existing studies that typically use traditional models and offer an innovative forecasting method to determine cultural homogeneity of travel demands.
3) Our GP–DLM applies the pooling method for tourism demand forecasting data for Hong Kong and Macau. Results show improved accuracy for one step ahead forecasting and multiple steps ahead forecasting. We achieve higher forecasting accuracy and confirm that our group-pooling method alleviates the limited data availability issue, adding data diversity, which allows the DLM to mitigate the overfitting in tourism demand forecasting.

Preliminaries and related works

Preliminaries on tourism demand forecasting

Tourism demand forecasting is a time series-forecasting task that utilizes various determinants and indicators to forecast future tourist arrival volumes (Law et al., 2019; Pan, Wu, & Song, 2012). Tourism demand forecasting with search engine data is an accurate and efficient approach. Yang, Pan, Evans, and Lv (2015) for example, integrated the search indicator intensity (SII) data into the tourism demand forecasting.

In this research, we use vector $Y^T = \{y^{(1)}, y^{(2)}, \ldots, y^{(T)}\}$ to represent tourist arrival volume data with T time steps, and $X^T = \left\{\overrightarrow{X}^{(1)}, \overrightarrow{X}^{(2)}, \cdots, \overrightarrow{X}^{(T)}\right\}$ to represent the SII data factors in time series Y^T. For a particular $y^{(i)}$, $\overrightarrow{X}^{(i)}$ is the related SII data factor at time step t in the time series. For $\overrightarrow{X}^{(i)}$, the SII data factors are made by n exogenous features:

$$\overrightarrow{X}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \cdots, x_n^{(t)})^\top \in \mathbb{R}^n \tag{2.1}$$

To perform tourism demand forecasting, we use the SII data factors $\{\overrightarrow{X}^{(i)}\}_{i=1}^{T}$ along with the corresponding past tourist arrival volume $\{y^{(i)}\}_{i=1}{}^{T}$ to form the input data and let $\mathscr{F}$ be the model, which can utilize the input data to forecast future tourist arrival volume $\{y^{i}\}_{i=T+1}{}^{T+\delta}$, where δ is the length of future time steps to be forecasted. This forecasting task is formalized as Eq. (2.2):

$$\{y^{(i)}\}_{i=T+1}^{T+\delta} = \mathscr{F}(\{\overrightarrow{X}^{(i)}\}_{i=1}^{T}, \{y^{i}\}_{i=1}^{T}). \tag{2.2}$$

In tourism demand forecasting, a model $\mathscr{F}$ is expected to capture the nonlinearity and complexity of the relationship between factors and future tourist arrival volumes (Chen, Liang, Hong, & Gu, 2015; Law et al., 2019).

Tourism demand forecasting

Extant tourism demand forecasting literature can be grouped into three categories: time series, econometric, and artificial intelligence (AI) methods (Song, Dwyer, Li, & Cao, 2012). Time series methods, including autoregressive integrated moving average (ARIMA), are the most widely used models for tourism demand forecasting (Li et al., 2018; Pan & Yang, 2017; Yang, Pan, & Song, 2014). ARIMA predicts future tourist arrival volumes on the basis of historical trends and patterns and has been applied to many tourism demand forecasting cases with non-stationary data by using the differencing step during the process (Asteriou & Hall, 2011). Furthermore, seasonal ARIMA is used by defining seasonal parameters during the forecasting process (Song et al., 2019). Apart from

ARIMA, other simple time series methods, such as naïve and exponential smoothing methods have also been widely adopted often used as benchmarks in most research (Long et al., 2019).

Econometric methods analyze the relationship between tourism demand and factors, such as policy on the regional market, income, and any other econometric variables related to origin and destination markets (Gunter, Önder, & Smeral, 2019; Wan & Song, 2018; Li et al., 2005). These factors determine economic growth in tourism demand and provide important insights and recommendations for tourism practitioners and policymakers (Song et al., 2012; Song et al., 2019). Traditional regression methods, such as OLS and error correction model (Kulendran & Wilson, 2000) have been used, including vector autoregression model (Yang et al., 2015), others have used a combination of econometric models to conduct tourism demand forecasting (Song et al., 2019). Many comparisons between time series and econometric methods have been carried out, and studies have shown that forecasting results can vary with context. For example, Witt and Witt (1995) found that time series is relatively accurate than econometric model. However, in general, time series does not outperform all other methods in different situations (Song et al., 2019).

Artificial Intelligence (*AI*) methods have emerged with the advancement in data collection technologies, web traffic, and search engine index crawling have promoted the dynamic increase of data volume and data dimension for tourism demand forecasting (Law et al., 2019). When compared with time series and econometric methods, AI methods have two advantages: finding the nonlinearity within big data and calculation capacity. Support vector machine (Cai et al., 2009) and artificial neural network (Constantino, Fernandes, & Teixeira, 2016) are popular AI methods in literature. Law et al. (2019) proposed a DLM with attention mechanism for tourism demand forecasting. They highlighted the superiority of the method in terms of accuracy and interpretation in the case of Macau tourism demand forecasting, claiming that DLM accurately forecasts tourism arrivals and provides a clear interpretation of factor importance through attention mechanism.

With the overall emerging interests in AI, the need to advance AI methods has increased. For example, decomposition methods offer important advantages, as they help to reduce model complexity, thus leading to improved accuracy in demand forecasting (Li & Law, 2020). The decomposition process includes filtering and decomposing time series data generating decomposed series components. Specifically, this is achieved through singular spectral analysis (SSA), empirical model decomposition (EMD), and seasonal trend decomposition (STL) (Li & Law, 2020).

Hassani, Webster, Silva, and Heravi (2015) used SSA on US tourist arrivals forecasting leading to accurate performance. Moreover, Hassani, Silva, Antonakakis, Filis, and Gupta (2017) experimented with multiple methods, including the SSA and also confirmed higher accuracy. Silva et al. (2019) applied SSA with neural networks and stressed that the model performs with higher accuracy. Finally, Li and Law (2020) used EMD to decompose the Hong Kong arrival data confirming higher accuracy.

Another common decomposition method that has been used in forecasting methods is the STL. STL has the advantage to decompose time series data into only three components: trend, seasonality, and residual (Cleveland, Cleveland, McRae, & Terpenning, 1990). This decomposition method provides additional stationary data in forecasting and has the potential to significantly advance tourism demand forecasting methodology: "decomposing the original search engine data series into distinct components contributes most to the improvement of forecasting accuracy" (Li & Law, 2020, p. 1).

Pooling data model forecasting

Pooling data to alleviate overfitting

First, we introduce group pooling of data to the tourism forecasting literature to address the overfitting problem, which are prevailing in existing studies. Pooling data models are created with certain techniques that combine multiple data sources to model the final pooling data (Otsu, Pesendorfer, & Takahashi, 2016). Pooling data models also integrate time series data with different sectional data and provide additional information regarding forecasting (Long et al., 2019). Tourism demand forecasting can benefit from using pooling data models, which exhibits increased freedom on forecasting. However, pooling data models are rarely adopted in tourism demand forecasting literature (Song & Li, 2008).

Thus far, group-pooling data models are limited in tourism demand forecasting. Instead, group-pooling models have been successfully used in econometric research (Bell & Jones, 2015; Chudik & Pesaran, 2015). Hoogstrate, Palm, and Pfann (2000) stated that pooling can reduce the mean squared error for forecasting the GDP of 18 countries with similar econometric patterns. Baltagi and Griffin (1997) suggested that the gasoline demand forecasting performance on the pooled data is superior to individual time series forecasting.

Further, Shi et al. (2017) pointed out that pooling data with deep learning reduces the overfitting problem within the household load forecasting because of similar power consumption patterns. Their study suggested that pooled data can add regularization on the large-scale deep-learning models and gain accuracy on household load forecasting in the testing stage. However, the limitation of Shi et al.'s work is that their pooling method only randomly selected some customer profiles into the pooling group. Such problem can introduce unnecessary noise into the pooled data and then affect the forecasting performance. Long et al. (2019) used pooling data in tourism demand forecasting indicating that accuracy is diminished when the heterogeneity of the data in use is larger than the data with small heterogeneity. Thus, considering spatial and temporal effects in the model can improve the accuracy performance in tourism demand forecasting.

Pooling data to find similar travel patterns

Another reason to apply group pooling of data is to confirm similar travel patterns of neighboring countries. Existing studies that investigated similarities in cross-country travel patterns, rather retrospective analyses, or qualitative data and offer cultural–behavioral focused contributions—instead of revealing insights into tourism demand forecasting and predictions of different similarities tourist arrival volume data from neighboring countries. Similar visitor travel patterns in neighboring countries have often been explained by cultural homogeneity or cultural convergence (e.g. Correia, Kozak, & Ferradeira, 2011; Muskat, Muskat, & Richardson, 2014) and similar cultural experiences (e.g. Lee & Bai, 2016). For example, prior research used survey data and single data sources to examine cross-cultural differences on the individual tourist level (Correia et al., 2011) or used qualitative data to understand cross-cultural differences in perception of destination image (Lee & Bai, 2016). Muskat et al. (2014) used panel data and applied correspondence analysis to understand demand travel patterns for several European tourism markets.

Methodology

The quality of input data is highly relevant for the forecasting performance (Liu, Tseng, & Tseng, 2018). Traditionally, data on tourism demand forecasting are obtained for one specific region. Thus, the forecasting model was only applicable for this region. However, comparing tourism demand forecasting models with different demands across different regions is complex because models of different regional tourism demand data had to be consolidated and each model used different forecasting factors.

In addition to this complexity, two common issues exist with these traditional tourism demand-forecasting methods. First, limited data accessibility may lead to insufficient data diversity for the model training, which may further result in overfitted forecasting models. For instance, the constructed model performs well on the training stage but performs poorly in the actual forecasting scenario (Wolpert & Macready, 1997). Second, finding the homogeneity factors that determine cross-country travel demand patterns for multiple demand markets requires the construction of multiple models.

As indicated above, model overfitting is a major problem for quantitative forecasting models (Law et al., 2019; Lever, Krzywinsky, & Altman, 2016). For DLM, overfitting in general is caused by insufficient data diversity in use for the large size of neural layers with a large number of parameters. The complexity of DLM allows the model to focus on fitting the available demand data but exhibits weakness in generalizing future tourism demands. For example, in total, the parameter for DLM is over 100,000, which indicates the complexity of DLM. In this work, the pooling of tourism demand data from regions with similar tourism demands can be regarded as increasing the diversity on all given regional tourism demand data, thus, providing a promising strategy to alleviate overfitting.

The pooling of data from regions with similar tourist arrival patterns allows the fusion of similar data into one model without increasing the noises of the overall data and the factors utilized in the forecasting model work on all those regions within the group. Homogeneity factors, which are used for forecasting tourist arrival volumes should be consistent and easily explored over those similar regions, such as weather, transportation in nearby regions, similar public holidays, and similar cultural attractiveness. To further elaborate, according to the studies of (Baltagi, Bresson, & Pirotte, 2008; Glass, 1976; Howell, 2009), the pooling data will have similarities with one another, which can be either similar dependent variables or consistent covariate variables across all samples. Therefore, the pooling of tourism demand data will require the tourism arrival volume data to have similar patterns. In addition, the homogenous factors found from the model should be consistent across all pooled regions and time steps. Therefore, the pooling strategy will have the similarity on tourism arrival volume data and the consistency on homogenous factors. For generating the pooling data. The tourism arrival volume data similarity will be first identified by the group-pooling method below, and then the homogenous factor consistency will be found in the experiment by using the DLM attention mechanism.

The rationale discussed above claims that identifying regions with similar historical tourist arrival patterns is crucial for the tourism demand group-pooling method. In general, tourism demand data are obtained using time series method, and the trend and seasonal patterns reflected by the time series data reveal the tourism demand similarity among different regions. Therefore, DTW distance clustering, a commonly used method for the similarity across time series, is incorporated with the STL decomposition to identify the pooling data for cross-country travel patterns.

In this section, we propose the framework of GP–DLM for tourism demand forecasting with SII factors. Fig. 1 summarizes the following two stages of the GP–DLM method: “group pooling of tourism demand” and “tourism demand forecasting using DLM”.

Group pooling of tourism demand

The GP–DLM method is based on the observation that many tourist regions exhibit similar tourism demand patterns. In this subsection, we initially present the DTW clustering method to identify groups of regions with similar tourism demands. We subsequently illustrate the details on how tourism demand group pooling data are generated.

Identifying pooling groups

To identify pooling groups on the basis of cross-country travel patterns, we calculate the similarity on the given tourism arrival volume data across regions. We use trend and seasonality patterns from tourism demand data, followed by the DTW clustering.

Trend and seasonality patterns are obtained from the STL decomposition (Cleveland et al., 1990). We use weighted regression with a smooth filter to decompose the time series data into the three components, *trend*, *seasonality*, and *residual*. When compared with prior decomposition methods, such as SSA (Silva et al., 2019) and EMD (Li & Law, 2020), the generated STL provides less decomposed components. Fig. 2 shows the STL decomposition for Macau tourism arrival volume data.

STL decomposition results reveal that the trend and seasonality components capture the major pattern for the regional tourist arrival volume data. The combination of these patterns forms the global trend component, which can be used to compare similarities across different regions. Thus, we use the global trend component with DTW clustering to identify pooling groups.

DTW clustering is a hierarchical clustering method that uses DTW as the distance measure within clusters (Izakian, Pedrycz, & Jamal, 2015). DTW cluster builds the solution by assigning similar time series to their own clusters and continuously merges and

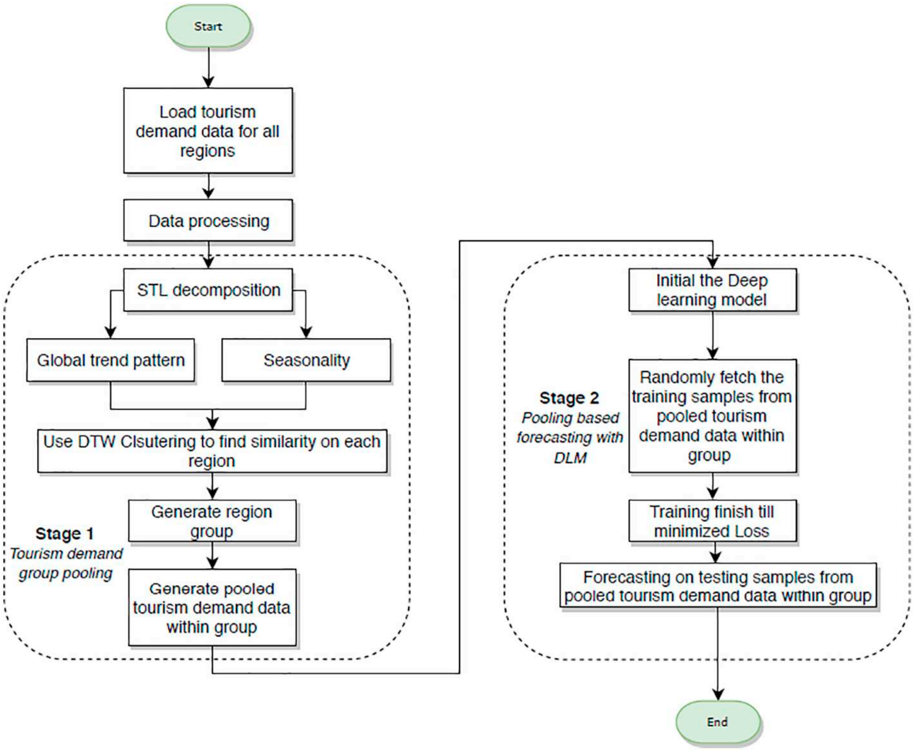

Fig. 1. Group pooling tourism demand forecasting: GP-DLM.

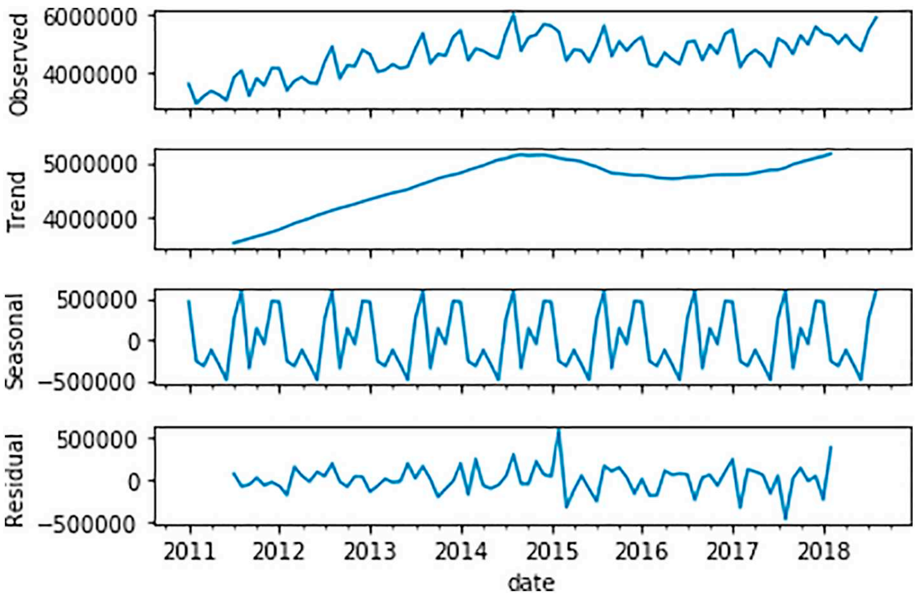

Fig. 2. STL decomposition.

selects the clustering until all-inclusive clusters are formed. The distance measure DTW is a robust similarity measure for time series. DTW can allow the matching of time series with similar shapes, even when they possess different lengths (Berndt & Clifford, 1994; Keogh & Ratanamahatana, 2005).

We assume two tourist arrival volume data from two regions ${Y_1}^m = \{{y_1}^{(1)},{y_1}^{(2)},\ldots,{y_1}^{(m)}\}$ and ${Y_2}^n = \{{y_2}^{(1)},{y_2}^{(2)},\ldots,{y_2}^{(n)}\}$, where m and n are the total time steps for the two time series and $m \neq n$. To align and calculate the distances of these two time series, a matrix with n rows and m columns is constructed. The element (i,j) in the matrix is the alignment distance $D({Y_1}^{(i)}, {Y_2}^{(j)})$ between ${Y_1}^{(i)}$ and ${Y_2}^{(j)}$. In the matrix, each element (i,j) denotes the alignment distance of the two points from the series of Y_1 and Y_2. Figs. 3 and 4

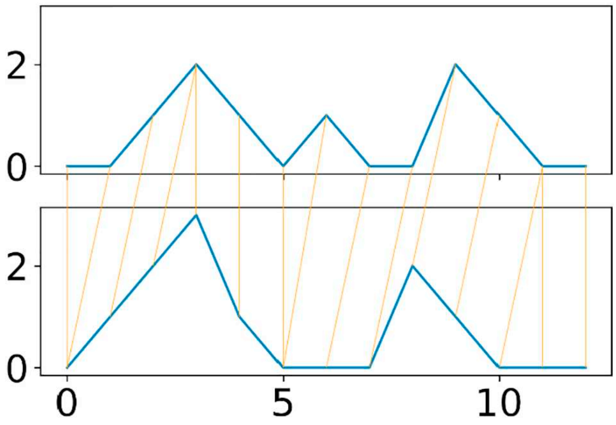

Fig. 3. Alignment in DTW.

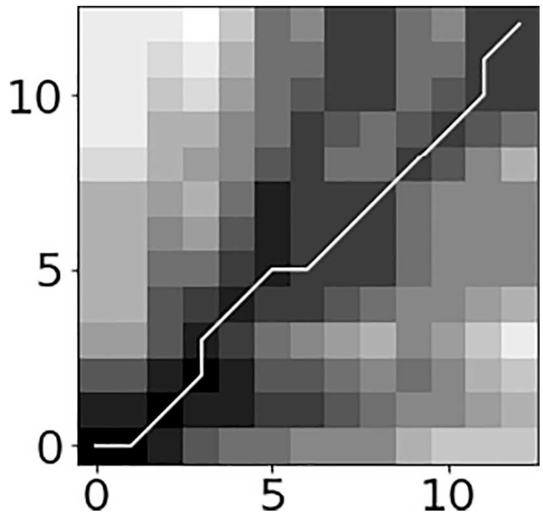

Fig. 4. Best path in DTW.

demonstrate this alignment and the best path.

In Fig. 3, the X-axis represents the time step for two compared time series. The number of time steps is different in many cases; the Y-axis shows the normalized numeric value on each time step for two given time series. The link between time series represents the alignment on time step from one to another compared time series. Fig. 4 shows the best path from DTW for two compared time series. The X axis and Y-axis are the time step from two time series. The path in the Fig. 4 is the best alignment path according to the distance on each time step. The darker shade on the grids signifies closer distance between two time steps from two given time series and vice versa.

By applying DTW, path $P = P_1, P_2, \ldots, P_k$, which is formed by the matrix elements, represents the mapping alignment on the two series. Given that many paths can be generated, the best path for DTW is the path that minimizes the warping distance as defined in Eq. (3.1) below.

$$DTW\left(Y_1, Y_2\right) = min\left(\sqrt{\sum_{k=1}^{k} P_k}\right), \tag{3.1}$$

where k is the total number of elements in the path, including the total number of aligned point pairs from the two time series, which has the limitation on $max\ (m,n) \leq k < m + n - 1$. The distance from DTW can be represented as the following equation:

$$D\left(Y_1^{(i)}, Y_2^{(j)}\right) = \left|\left(Y1^{(i)}, Y_2^{(j)}\right)\right| + min\left(D\left(Y_1^{(i-1)}, Y_2^{(j-1)}\right), D\left(Y_1^{(i-1)}, Y_2^{(j)}\right), D\left(Y_1^{(i)}, Y_2^{(j-1)}\right)\right).$$

During the DTW clustering, the similarities between time series and the global trend patterns from regions are measured between each pair of regions, thereby forming the distance matrix. By providing the desired number of clusters, the method can repeatedly form the cluster by paring the regions or clusters with the shortest distance from the distance matrix. Once a new cluster is formed, the distance matrix is updated by using the maximum/minimum distance between time series within the cluster to calculate the distances between clusters. Here, we use the maximum distance to represent the distance between clusters. DTW clustering results reveal that the most similar time series (tourist arrival patterns) from regions are grouped together, and the pooling groups are identified accordingly.

Tourism demand group pooling

After the DTW clustering on the global trend patterns across regions, the pooling groups are identified accordingly. The initial part of the Methodology section mentions that the group pooling of tourism demand data can solve the overfitting problem and the issue about finding homogeneity in factors. Hence, for tourism demand forecasting, we pool the tourism demand data for the pooling group to support the construction of tourism demand forecasting models. Importantly, our novel group-pooling method identifies the similarities in tourism demand patterns across regions, by increasing data volume and data diversity without adding extra noises.

In this study, the tourist arrival volume data are captured monthly. Therefore, the monthly tourist arrival volume data and SII data for each region in the pooling group are used to generate the pooling data.

We take vectors $Y_1{}^T = \{y_1{}^{(1)}, y_1{}^{(2)}, \ldots, y_1{}^{(T)}\}$ and $Y_2{}^T = \{y_2{}^{(1)}, y_2{}^{(2)}, \ldots, y_2{}^{(T)}\}$ as the tourist arrival volume data with T months for two regions Y_1 and Y_2 in one pooling group, respectively. Their global trend patterns are the most similar based on the clustering results in subsection of identifying pooling groups. Correspondingly, the SII data with T months on regions Y_1 and Y_2 can be expressed as $X_1^T = \left\{\overrightarrow{X_1}^{(1)}, \overrightarrow{X_1}^{(2)}, \cdots, \overrightarrow{X_1}^{(T)}\right\}$ and $X_2^T = \left\{\overrightarrow{X_2}^{(1)}, \overrightarrow{X_2}^{(2)}, \cdots, \overrightarrow{X_2}^{(T)}\right\}$, respectively. Preliminaries on Section 2 indicates the inputs of regions Y_1 and Y_2, which are used for tourism demand forecasting, can be described as $\left(\{\overrightarrow{X_1}^{(i)}\}_{i=1}^T, \{y_1^i\}_{i=1}^T\right)$ and $\left(\{\overrightarrow{X_2}^{(i)}\}_{i=1}^T, \{y_2^i\}_{i=1}^T\right)$, respectively.

The tourism demand pooling data are generated through two steps:

Step 1: Data Splitting: For both regions, the input data are split at each time step. Given that the data used in this work are generated monthly, the input data for regions Y_1 and Y_2 are divided into monthly level.
Step 2: Data Fusion: After splitting the input data for regions Y_1 and Y_2 in the same group, the tourism demand pooling data can be generated by merging the SII data and tourism arrival volume data for both regions at each time step (i.e. monthly) level. Therefore, the newly generated tourism demand pooling data can be expressed as $\left(\left\{\overrightarrow{X_1}^{(i)}, \overrightarrow{X_2}^{(i)}\right\}_{t=1}^T, \{y_1^i, y_2^i\}_{t=1}^T\right)$, which has similar time steps T but with large data volume size.

Group pooling and tourism demand forecasting

In this section, the group pooling of tourism demand data is used with the DLM algorithm by Law et al. (2019) for forecasting. DLM is briefly recapped, followed by the detailed discussion of the GP–DLM method.

DLM

DLM (Law et al., 2019) worked on the historical time series of tourism demand data by utilizing the long short-term memory (LSTM) with the attention mechanism. Fig. 5 describes the LSTM cell structure.

The LSTM used in this study aims to overcome the gradients' vanishing issues from recurrent neural network (RNN), which is the commonly used DLM for time series applications. LSTM provides multiple memory blocks with three gates, namely, input, forget, and

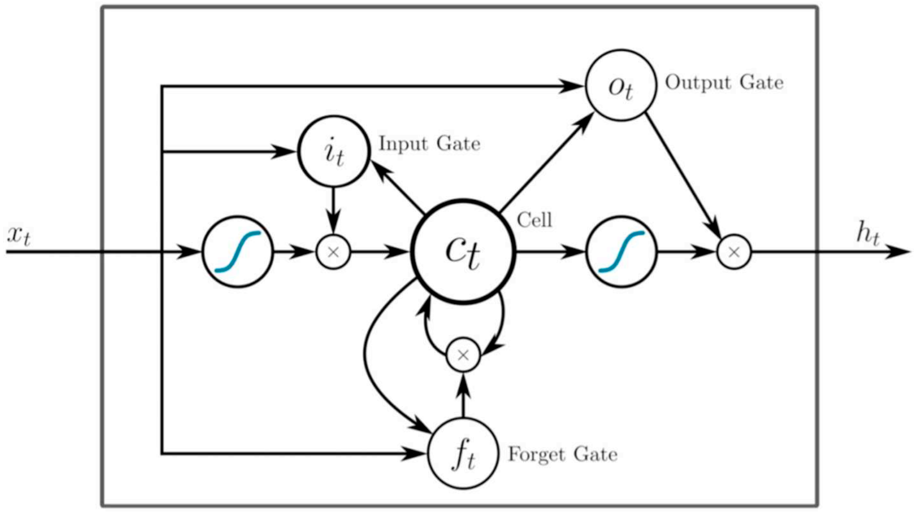

Fig. 5. Long short terms memory (adapted from Law et al., 2019).

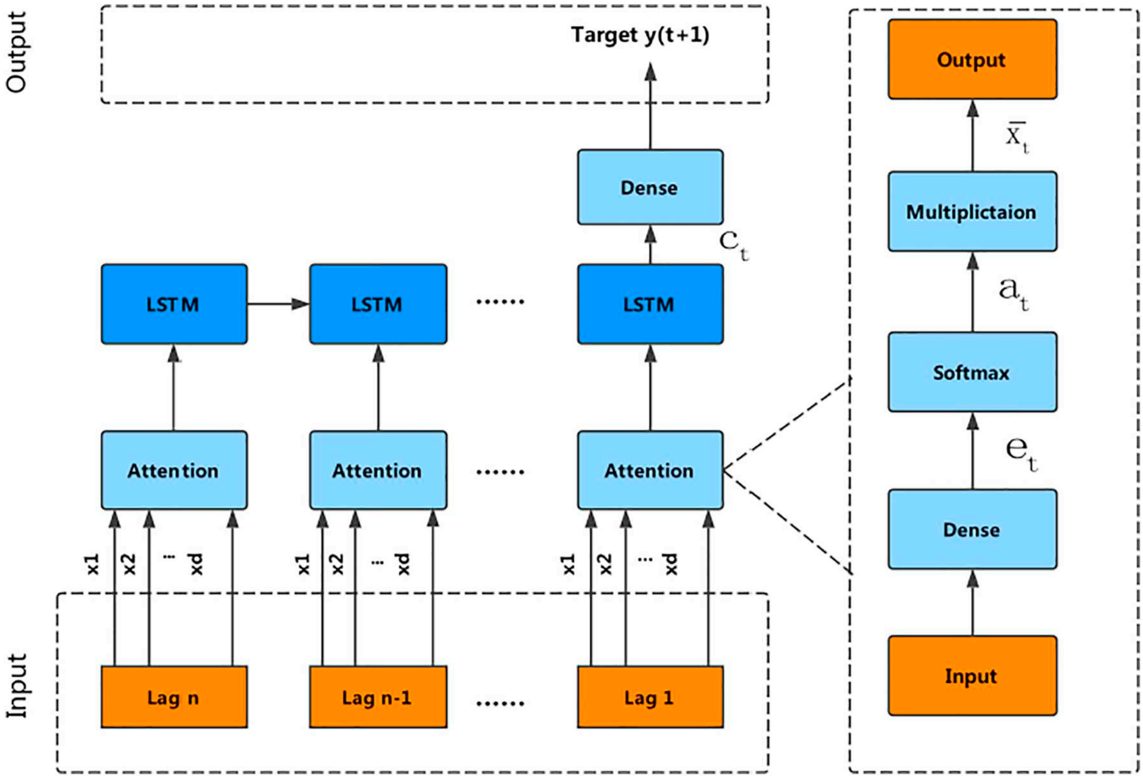

Fig. 6. Deep learning model for tourism demand forecasting (adapted from Law et al., 2019).

output gates. Below are the equations on updating long-range dependencies.

$$
\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1} x_t] + b_f) \\
i_t &= \sigma(W_i \cdot [h_{t-1} x_t] + b_i) \\
\widehat{C}_t &= tanh(W_C \cdot [h_{t-1} x_t] + b_C) \\
C_t &= f_t * C_{t-1} + i_t * \widehat{C}_t \\
o_t &= \sigma(W_o \cdot [h_{t-1} x_t] + b_o) \\
h_t &= o_t * tanh(C_t)
\end{aligned}
$$

From the equations above, at time step t in the time series, cell state C_t, which denotes long-range dependencies, is updated by the output h_{t-1} from the last time step and the current input x_t through the input and forget gates. The output of the current time step t is calculated using the updated cell state C_t and the result o_t from the output gate.

Another important component of DLM is attention mechanism. Input attention has been used in the input data for feature selection, lag order selection, and model interpretation. The attention mechanism of DLM applies the dense layer with the softmax function on the input data and learns the parameter within the deep-learning training process. The dense layer and softmax function here provide the importance weight on the input and form the attention score matrix with similar size as the input data. Therefore, the attention mechanism of DLM can provide two kinds of essential information for model interpretation within tourism demand forecasting—lag order selection (temporal relationship) and feature importance—on the basis of the impacts on tourism demand. Fig. 6 illustrates the DLM architecture.

GP–DLM forecasting

We assume that the pooled tourism demand data are completed and expressed as $\left(\left\{\overrightarrow{X_1}^{(i)}, \overrightarrow{X_2}^{(i)}\right\}_{t=1}^{T}, \{y_1^i, y_2^i\}_{t=1}^{T}\right)$ with a group of two regions. Subsequently, the GP–DLM forecasting must perform two steps.

1. First, DLM is trained with the training data by stochastically taking data for each batch from the group pooled tourism demand data. In each epoch, DLM converges to the minimal training error. Here, we use the loss function described as in Eq. (3.3):

$$
\mathscr{L}(Y_{pred}, Y) = \frac{1}{N}\sum_{i=1}^{N} \left|\frac{Y^i - Y^i_{pred}}{Y^i}\right|, \tag{3.3}
$$

where i is the particular sample on the training data. N represents the total number of training samples from the group pooled tourism demand data.

2. Second, the trained DLM is used in future time steps to obtain tourist arrival forecasting volumes.

For the model training, the parameters of GP–DLM in each batch are updated to reduce the loss in Eq. (3.3). Given that the grouped regions exhibit a similar tourism demand pattern, pooled tourism demand data add robustness to the model through the updating parameter. Thus, GP–DLM can learn sufficient diversity from the input, which can also exhibit a great chance of learning nonlinearity from the tourism demand data.

Experiment and analysis

This section describes the implementation of the proposed GP–DLM, which includes data description, similar group identification, model evaluation, and important homogeneity factors' analysis.

Data description

According to subsection on identifying the pooling group, a group of regions with similar demands must be identified before commencing the group-pooling method. In this study, we use the tourism demand series, such as the tourist arrival volume data from nine Asia-Pacific regions to identify the similar groups and support the group pooling. These nine regions include *Macau*, *Hong Kong*, *China*, *Australia*, *New Zealand*, *Singapore*, *Japan*, *Thailand*, and the *Philippines*. These nine regions are major tourism countries in Asia-Pacific and close to one another. South Korea was not in the selected country given that the monthly tourism arrival data were not available when conducting the collection, and also many potential SII factors were not available due to language issues.

With regard to the tourist arrival volume data, the open data set "Macau2018" from Law et al. (2019) is used to collect *Macau's* tourist arrival volume data. The rest of the tourist arrival volume data are collected from the government-subverted organizations and the statistical bureaus in particular regions. For example, *Hong Kong's* tourist arrival volume data are collected from the *Hong Kong Tourism Board*; those from *China* are collected from the *Statistical Bureau of China*; *Australia* and *New Zealand's* tourist arrival volume data are taken from the *Australian Statistical Bureau* and *New Zealand Statistical Bureau*, respectively; *Singapore's* tourist arrival volume data are collected from the *Singapore Statistical Bureau*. Similarly, *Japan*, *Thailand*, and the *Philippine's* tourist arrival volume data are obtained from their respective statistic bureaus. After collecting tourist arrival volume data, those obtained across these regions are processed into the same time period from January 2012 to December 2018.

As discussed in the methodology in subsection on identifying the pooling group, once similar groups are identified, the SII data of each region in the group are collected to prepare the group's pooled tourism demand data. Therefore, not all nine regions must collect their SII data unless they are belonging to the similar pooling group. Following the method of SII data collection in previous studies (Law et al., 2019; Li, Pan, Law, & Huang, 2017), the seed search keywords from seven categories are initially determined, such as *dining, lodging, transportation, tour, clothing, shopping* and *recreation*. Subsequently, the tourism-related keywords for each seed search keyword are retrieved from different *Google Trend* search regions. For each region in the group, the tourism-related search keywords are either the same or similar to the objects that are searched for. For instance, assuming that regions of Hong Kong and Macau belong to the similar group, the search keyword "Hong Kong weather" means the weather search intensity in Hong Kong SII data, whereas the search keyword "Macau weather" is collected for the weather search intensity in Macau SII data. In the final pooled tourism demand data for Macau and Hong Kong, "weather" represents the SII factor of "Hong Kong weather" and "Macau weather". Particularly, the search intensity representation "weather" is formed by two parts of the SII data accordingly.

Identifying the pooling group

As discussed in the methodology in subsection on identifying the pooling group, nine tourist arrival volume data are placed into the DTW clustering to find the pooling groups. Before using DTW, the tourist arrival volume data for all nine regions are normalized by the minimum–maximum scaler. The normalization employed in this study aims to place the tourist arrival volume difference from different regions in the same range for convenient comparison and clustering. The pooling group will be compared with two major decomposition methods in many recent tourism literatures: STL decomposition and SSA.

Figs. 7 and 8 demonstrate the DTW distances for nine regions by using trend and seasonality from STL decomposition and SSA. Dark color indicates that the distance is large, whereas the light color indicates small distance. Figs. 7 and 8 reveal that Macau and Hong Kong exhibit a shorter distance than that between Macau and other regions and vice versa. The distance between Australia and New Zealand is the shortest among all paired-wise distances. However, the distance between China and other regions is consistently high, and Singapore exhibits the same pattern regarding its distance with other regions. Figs. 7 and 8 indicate that according to the DTW distance, STL decomposition and SSA provide similar patterns with minor differences in the given nine regions.

Using the DTW distance from STL decomposition, DTW clustering is implemented for identifying the pooling group here. Fig. 9 reveals that China and Singapore cannot be grouped with any other regions, whereas the cluster number equals to five. Macau with Hong Kong, and Thailand with the Philippines belong to the same group. Furthermore, Australia, New Zealand, and Japan belong to the same group.

Based on the findings discussed above, Macau–Hong Kong group is successfully identified from the DTW clustering. In this study,

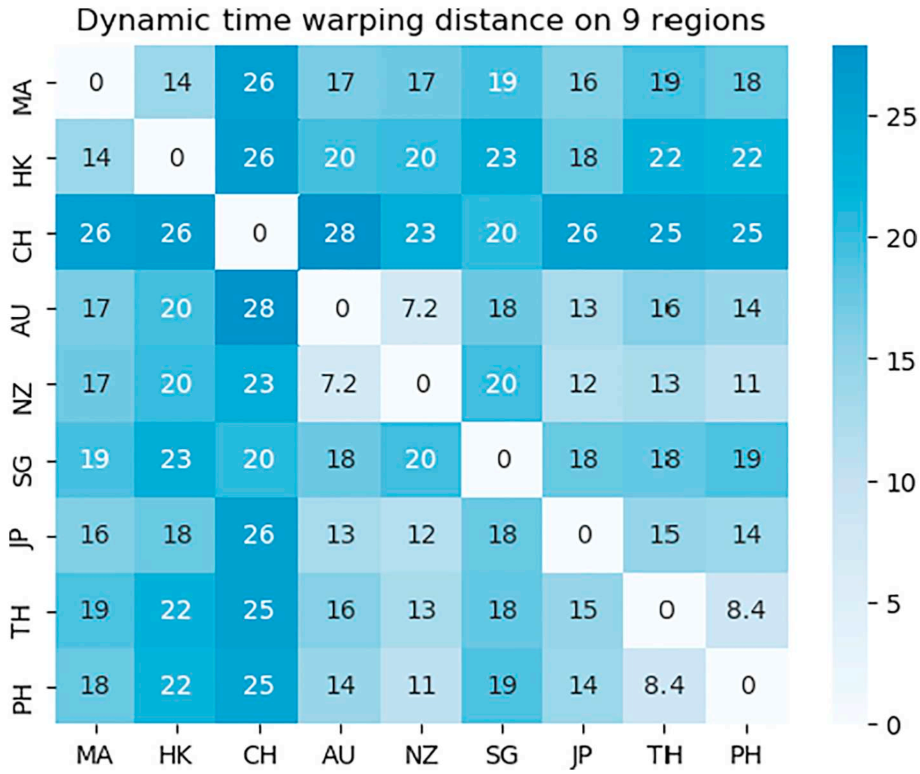

Fig. 7. Dynamic time warping distance comparison of nine regions on STL.

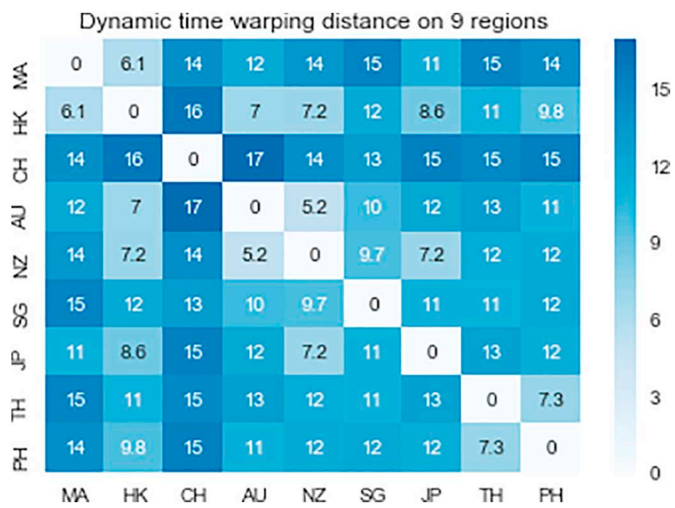

Fig. 8. Dynamic time warping distance comparison of nine regions on SSA.

we generate the pooled tourism demand data for the Macau–Hong Kong group by using the method presented in subsection on tourism demand group pooling.

Macau and Hong Kong tourism demand pooling forecasting

In last section, the Macau–Hong Kong group is identified with similar tourism demands. Subsequently, the pooled tourism demand data for this group can be established on the basis of the discussion in subsection on tourism demand group pooling. To perform the proposed GP–DLM forecasting, this section describes the validation setting for the experiment and the result comparison between the proposed GP–DLM forecasting and certain benchmark methods, including the state-of-the-art deep-learning tourism demand forecast method (Law et al., 2019), the support vector machine regressor (SVR) (Cherkassky & Ma, 2004), the extreme boosting tree

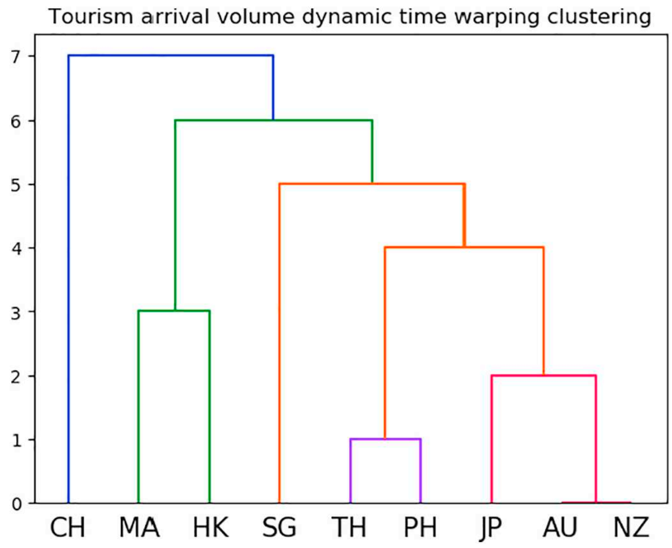

Fig. 9. Dynamic time warping clustering on nine regions.

regressor (XGBTR) (Chen & Guestrin, 2016), the ARIMA, and also the multivariate exponential smoothing, which is called vector local trend seasonal model (VLTS) (Athanasopoulos & de Silva, 2012). The forecasting is conducted through the one-step ahead (one month) and multiple steps ahead (three months).

Validation setup

To find the best parameter setting for the methods under comparison (GP–DLM and benchmark methods), we use the *walk-forward validation* method, following the study of Law et al. (2019). *Walk-forward validation* simulates the real-world forecasting scenario where the test data are only available for the next time steps and are placed back to form the new training data once forecasted. In this study, we initially set the first 24 months (2012 and 2013) from the pooled tourism demand data as the training data for GP–DLM. Similarly, we set the first 24 months (2012 and 2013) as the training data on the Hong Kong and Macau tourism demand data for the benchmark methods. Afterward, all methods are validated on the future time steps in corresponding training data to obtain the best parameters.

Through the *walk-forward validation*, we use the grid search to obtain the parameters for GP–DLM as follows: the LSTM hidden unit $p \in \{128,256,512\}$, the dropout rate $d \in \{0.01,0.3,0.5\}$, and the size of the dense layer $q \in \{32,64,128\}$. Similarly, the benchmark method DLM has the LSTM hidden unit $p \in \{128,256,512\}$, the dropout rate $d \in \{0.01,0.3,0.5\}$, and the size of the dense layer $q \in \{32,64,128\}$ before the output layer. Subsequently, for SVR and XGBTR, the parameters are also selected from the grid search to perform the comparison.

Results comparison

The pooled tourism demand data contain two parts of the tourist arrival volume data from Macau and Hong Kong. Therefore, we use the mean absolute percentage error (MAPE) and root means square error (RMSE) as the evaluation metrics, which are defined below.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{Y^i - Y^i_{pred}}{Y^i}| \times 100\%,$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y^i - Y^i_{pred})^2}$$

where Y^i is the actual tourist arrival volume at time step *i*, and ${Y_{pred}}^i$ is the forecasted tourist arrival volume at time step *i*. MAPE and RMSE are evaluated through the *walk-forward validation* on future tourist arrival volumes from the pooled tourism demand data from Macau and Hong Kong.

All methods are tested for 48 months from 2014 to 2017 via one step ahead forecasting and multiple steps ahead forecasting. Tables 1–2 show the one step ahead forecasting results.

MAPE and RMSE in Tables 1–2 indicate that the proposed GP–DLM outperforms DLM, SVR, XGBTR, multivariate exponential smoothing (MES), and ARIMA on Hong Kong and Macau tourism demand forecasting by one step ahead. Comparing GP–DLM with DLM for Hong Kong tourism demand forecasting, the improvement rate on MAPE is approximately 25% for years 2014, 2016, and 2017 in MAPE. For the data in 2015, the forecasting improvement rate on MAPE is over 49%, despite the decline of tourist arrival volume since 2012. The forecasting result for Hong Kong elaborates that by using the group pooling tourism demand data on training

Table 1
MAPE of Hong Kong and Macau (one step ahead).

Year	Model on Hong Kong						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	1.81	2.41	8.97	9.37	9.18	9.59	24.9%
2015	1.50	2.98	11.59	10.07	8.11	7.37	49.7%
2016	1.64	2.14	6.87	7.47	8.04	8.10	23.4%
2017	2.01	2.71	6.72	6.50	5.83	7.13	25.9%
Year	Model on Macau						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	0.879	0.922	6.701	7.132	7.207	7.772	4.66%
2015	0.930	1.132	4.993	4.552	6.664	5.446	17.84%
2016	2.242	2.425	5.689	5.920	6.474	6.470	7.55%
2017	1.376	1.671	8.571	8.483	6.533	7.822	17.7%

[a] The improvement rate is from the comparison between GP-DLM and DLM.

Table 2
RMSE of Hong Kong and Macau (one step ahead).

Year	Model on Hong Kong						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	123,047	192,416	589,850	620,582	598,745	662,893	36.1%
2015	87,495	186,198	655,053	583,720	481,838	439,800	53.1%
2016	109,247	128,967	399,801	469,803	487,775	489,371	15.3%
2017	140,754	172,938	462,302	441,050	398,969	486,826	18.6%
Year	Model on Macau						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	25,015	27,100	231,583	248,485	250,353	263,518	7.76%
2015	32,779	40,020	179,026	168,944	209,249	186,916	18.1%
2016	78,442	88,802	181,985	192,159	202,699	200,745	11.7%
2017	40,500	50,604	265,493	268,707	238,867	249,339	19.9%

[a] The improvement rate is from the comparison between GP-DLM and DLM.

DLM, the GP–DLM in this study is more robust than DLM alone. This finding further confirms that the group pooling data can alleviate the overfitting problem, which worsens DLM performance on unseen testing data.

For Macau's tourism demand forecasting via one step ahead, although DLM can accurately forecast Macau's tourist arrival volume data, the proposed GP–DLM remains to outperform DLM and all other methods, such as the 2015 and 2017 data results. The improvement rate also suggests that GP–DLM is more robust than DLM. Moreover, results imply that the GP–DLM can alleviate the overfitting problem on Macau and Hong Kong tourism demand forecasting.

For multiple steps ahead forecasting, GP–DLM is compared with DLM and also other benchmarking models. Results for multiple steps ahead forecasting on MAPE and RMSE are listed in Tables 3–4.

Tables 3 and 4 presents that GP–DLM outperforms other models on Hong Kong tourism demand forecasting by three months steps ahead. Particularly, the forecasting accuracy on GP–DLM was improved by 53.9% compared with DLM on MAPE in 2015 for Hong Kong. For RMSE, the accuracy improved by 54.5% compared with DLM on Hong Kong tourism demand forecasting. Similarly, from Tables 3 and 4, the results show that the GP–DLM largely reduced the forecasting error on MAPE and RMSE on Macau tourism demand forecasting in three months steps ahead. In 2015 and 2016, the improvement rate on MAPE and RMSE is over 30% on Macau tourism demand forecasting.

To verify the significance of the improvement, we perform the Diebold–Mariano (DM) test for the one step ahead forecasting and multiple steps ahead forecasting results on Hong Kong and Macau accordingly, and their DM statistics are listed below.

Table 5 reveals that the GP–DLM one step ahead forecasting results from Hong Kong and Macau data are significantly different from DLM and all other methods on Diebold–Mariano test at 5% and 10% level. Based on the test, the statistics for Hong Kong and Macau indicated that GP–DLM can significantly improve the accuracy.

Similarly, Table 6 indicates that the GP–DLM multiple steps ahead forecasting results from Hong Kong and Macau data are significantly different from DLM and all other methods at 5% and 10% level. The null hypothesis that GP–DLM multiple steps ahead

Table 3
MAPE of Hong Kong and Macau (multiple steps ahead – 3 months).

Year	Model on Hong Kong						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	3.51	4.41	12.23	9.96	9.78	11.22	20.4%
2015	2.10	4.56	7.16	7.56	10.13	9.26	53.9%
2016	3.64	4.94	6.88	6.51	11.33	9.80	26.3%
2017	3.31	5.06	8.24	9.45	9.02	8.53	34.6%

Year	Model on Macau						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	2.11	3.31	6.92	8.25	12.21	8.59	36.2%
2015	2.50	3.98	5.04	5.17	9.21	8.77	37.2%
2016	2.92	4.46	6.01	6.34	8.76	9.17	34.5%
2017	3.71	4.57	9.24	9.73	9.89	10.13	18.8%

[a] The improvement rate is from the comparison between GP-DLM and DLM.

Table 4
RMSE of Hong Kong and Macau (multiple steps ahead – 3 months).

Year	Model on Hong Kong						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	224,696	295,088	794,946	686,706	668,903	730,122	18.3%
2015	139,937	337,464	425,305	408,366	591,830	570,012	54.5%
2016	204,747	290,895	407,099	362,704	601,903	510,931	29.6%
2017	162,417	326,719	594,220	657,102	633,098	600,938	50.3%

Year	Model on Macau						IR%[a]
	GP-DLM	DLM	SVR	XGBTR	ARIMA	VLTS	
2014	74,684	101,108	238,163	282,109	401,911	335,609	26.1%
2015	81,431	153,205	179,100	188,760	312,011	271,002	46.8%
2016	91,877	136,218	196,470	210,878	300,210	340,912	32.5%
2017	120,025	165,970	288,225	310,010	331,101	345,102	27.7%

[a] The improvement rate is from the comparison between GP-DLM and DLM.

Table 5
Diebold-Mariano statistics and *P* value on one step ahead forecasting.

Region	DM statistics (MAPE)				
	GP-DLM VS DLM	GP-DLM VS SVR	GP-DLM VS XGBTR	GP-DLM VS ARIMA	GP-DLM VS VLTS
Hong Kong	−3.4047 (0.042)	−3.4968 (0.039)	−4.9123 (0.016)	−5.5582 (0.012)	−6.7213 (0.007)
Macau	−2.8548 (0.063)*	−3.8468 (0.031)	−3.8198 (0.032)	−14.283 (0.001)	−5.8657 (0.009)

Region	DM statistics (RMSE)				
	GP-DLM VS DLM	GP-DLM VS SVR	GP-DLM VS XGBTR	GP-DLM VS ARIMA	GP-DLM VS VLTS
Hong Kong	−3.0906 (0.048)	−4.3549 (0.022)	−5.7489 (0.010)	−5.5682 (0.011)	−4.8598 (0.016)
Macau	−2.1972 (0.089)*	−4.6232 (0.019)	−4.4224 (0.021)	−7.9768 (0.004)	−5.5039 (0.011)

The * indicates the DM statistics is significant at 10% level and all others are significant at 5% level.

forecasting result is similar with all other models can be rejected. From the Hong Kong and Macau forecasting experiments, the accuracy from GP–DLM is better than DLM. The improvement on accuracy is further validated that the group pooling can overcome the overfitting problem that exists in the DLM. As mentioned in the methodology, to solve the overfitting, the common solution is increasing the data sample size and diversity. In GP–DLM, the pooling data has larger data diversity compared with the Hong Kong or Macau tourism demand data. Therefore, the improved accuracy from GP–DLM can indicate the advance on group pooling toward

Table 6
Diebold-Mariano statistics on multiple steps ahead forecasting.

Region	DM statistics (MAPE)				
	GP-DLM VS DLM	GP-DLM VS SVR	GP-DLM VS XGBTR	GP-DLM VS ARIMA	GP-DLM VS VLTS
Hong Kong	−3.5499 (0.038)	−2.4156 (0.084) *	−4.7239 (0.017)	−9.5225 (0.002)	−3.2989 (0.045)
Macau	−7.6813 (0.005)	−3.5040 (0.039)	−3.5169 (0.040)	−4.2044 (0.024)	−8.1931 (0.004)
Region	DM statistics (RMSE)				
	GP-DLM VS DLM	GP-DLM VS SVR	GP-DLM VS XGBTR	GP-DLM VS ARIMA	GP-DLM VS VLTS
Hong Kong	−2.9796 (0.058)*	−2.8640 (0.064)*	−3.0918 (0.047)	−19.658 (0.0002)	−33.002 (0.0001)
Macau	−3.1280 (0.048)	−4.3941 (0.021)	−4.1616 (0.025)	−4.5812 (0.019)	−9.7106 (0.002)

The * indicates the DM statistics is significant at 10% level and all others are significant at 5% level.

overcoming the overfitting problem. As a consequence, from the above experiment results for Hong Kong and Macau forecasting, we conclude that group pooling increases the data diversity and thus, alleviate the overfitting, and this results in improved accuracy when compared with DLM.

Analysis on homogenous factors

An analysis on homogenous factors is conducted to investigate the feature importance on forecasting for Macau and Hong Kong tourist arrival volume data. In GP–DLM, the attention layer inside DLM (Law et al., 2019) provides the attention score of each SII factor used in the DLM as the feature for forecasting. Based on the method used in Law et al. (2019), the pooled tourism demand data are trained through DLM. The future tourist arrival volume data from Macau and Hong Kong are forecasted correspondingly. The weights on the attention layer from DLM are learned from the training and are further transformed into attention scores via the *softmax* function after the attention layer. Hence, the testing data can use those weights to calculate the attention score of interpreting the feature importance (SII factor importance) in the forecasting stage.

To fairly assess the attention scores of Macau and Hong Kong forecasting (as default, we use the one step ahead forecasting to obtain the attention score because of higher accuracy), we use the testing data (2014 to 2017) from the pooled tourism demand data that are generated. The testing data are initially separated as Macau and Hong Kong for their respective inputs and are subsequently used as well-trained inputs for extracting the output of the attention layer that corresponds to Macau and Hong Kong. After extracting the output of the attention layer, we can obtain the attention scores of all SII factors for 12 months accordingly by using the pooled tourism demand testing data. Furthermore, the attention scores of all SII factors are averaged to generate the average attention score. Tables 7 and 8 show the differences in the average attention scores of the top-10 SII factors of Macau and Hong Kong.

Tables 7 and 8 reveal that those important SII factors for Macau and Hong Kong forecasting can be observed. In Table 7, the factor "Macau travel blog" is the most important for Macau forecasting. Similarly, the factor "Hong Kong travel blog" in Table 8 also becomes the most essential factor for Hong Kong forecasting. "Map of Macau" and "Map of Hong Kong" indicate that travelers use these maps to prepare for their future trips. "Macau best food" and "Hong Kong best food" are also highly regarded as important SII factors for Macau and Hong Kong tourism demand forecasting, indicating that travelers search for the best foods during the trip to their destinations. Conversely, "Hong Kong cheap flights" and "Macau cheap flights" largely contribute to the tourism demand forecasting of Hong Kong and Macau, respectively. Overall, the attention score tables show the common SII factors in pooled tourism demand data, which represent Macau and Hong Kong SII factors, such as "travel blog", "map of region", "best food", "shopping", "what to do in region", and "Air Asia". These SII factors are the homogenous factors that indicate the tourism demand for Macau and Hong Kong collectively. The attention score tables also present certain SII factors that are influential for Hong Kong but not for Macau, such as "Hong Kong ferry terminal to Macau" and "Hong Kong Packages". These factors suggest that travelers going to Hong

Table 7
Attention score of Macau from GP-DLM.

SII factor	Average attention score
Macau travel blog	0.5402
Map of Macau	0.3016
Macau best food	0.2297
What to do in Macau	0.1034
Macau shopping	0.0543
AGODA	0.0539
Macau cheap flights	0.0519
Singapore to Macau	0.0512
Skyscanner	0.0482
Air Asia	0.0458

Table 8
Attention score of Hong Kong from GP-DLM.

SII factor	Average attention score
Hong Kong travel blog	0.3667
Map of Hong Kong	0.3033
Hong Kong best food	0.1443
Hong Kong ferry terminal to Macau	0.1025
Hong Kong cheap flights	0.0615
What to do in Hong Kong	0.0536
Air Asia	0.0507
Best food in Hong Kong	0.0498
Hong Kong packages	0.0471
Hong Kong shopping	0.0409

Kong also check ferry rides to Macau, and certain travelers going to Hong Kong search for package tours that include accommodation and transportation while purchasing. Conversely, certain SII factors contribute to the Macau tourism demand forecasting but not Hong Kong, such as “Agoda” and “Singapore to Macau”.

Ideally, those homogenous factors, which are important for Hong Kong and Macau tourism arrival volume forecasting in GP–DLM should also be important in DLM. That is, the homogenous factors are consistent cross all the pooled regions. To validate this consistency mentioned earlier, we use DLM to individually compute the Hong Kong and Macau forecasting, without group pooling and the attention score results are in Tables 9–10. Tables 9–10, depict that the top SII factors for Hong Kong and Macau are highly consistent. This consistency on the factor attention score also indicates that the DLM models perform similarly when selecting the factors via attention mechanism. Thus, we conclude that the homogenous factors are the most relevant factors for both regions in the pooling group, given that they have similar tourism arrival volume data patterns and hence, the group pooling strategy is applicable.

Conclusions

This research presents a new framework of GP–DLM for tourism demand forecasting. Particularly, the multiple region tourism demand forecasting is explored for the first time by generating a group of regions with similar tourist arrival patterns. In addition, the novel GP–DLM is proposed by using the group pooled tourism demand data with deep learning for tourism demand forecasting. This study provides three important contributions advancing the tourism forecasting literature.

First, we use the DTW clustering to identify a group of regions with similar tourist arrival volume patterns. The generated group contains the smallest distance with one another based on their historical tourist arrival volume series. This method can further help tourism practitioners in identifying similar regions in terms of their tourism demands.

Second, the performance of GP–DLM on Macau and Hong Kong tourism demand forecasting one step ahead and multiple steps ahead further confirms that GP–DLM provides greatly accurate tourism demand forecasting than the state-of-the-art DLM (Law et al., 2019), including other benchmark models. The pooled tourism demand data can increase data diversity in the training stage of DLM and further alleviate the overfitting problem.

Third, we apply group pooling of tourism demand data in the training of DLM. Thus, the attention layer of DLM captures specific factors that influence the tourist arrival volume in Macau and Hong Kong. Therefore, our attention scores for countries' tourism demand data represent the common factors (i.e., homogeneity factors) of the analysis. For example, we found that “travel blog”, “best food”, “Air Asia”, and “map” are the common factors, which are largely important for tourism demand forecasting. Knowledge about these generated common factors helps tourism practitioners to analyze the cross-country travel demand forecasting. Importantly, these communalities in cross-country tourism demand offer new research avenues to better understand inter-organizational collaboration for destinations and leverage on this tourism demand knowledge as a strategic capability for tourism marketing, even in times of crisis and disasters (e.g. Hansen, Fyall, Spyriadis, Rogers, & Brander-Brown, 2019).

Table 9
Attention score of Macau from DLM.

SII factor	Average attention score
Map of Macau	0.7411
Macau best food	0.5609
Macau weather	0.3502
Macau hotel	0.2011
Macau cheap flights	0.0902
What to do in Macau	0.0783
Macau ferry terminal to Hong Kong	0.0416
Singapore to Macau	0.0203
AGODA	0.0130
Macau currency	0.0117

Table 10
Attention score of Hong Kong from DLM.

SII factor	Average attention score
Hong Kong best food	0.5215
Hong Kong ferry terminal to Macau	0.4722
What to do in Hong Kong	0.4016
Hong Kong travel blog	0.2980
Hong Kong	0.1928
Hong Kong cheap flights	0.1021
Hong Kong shopping	0.0890
AGODA	0.0633
Hong Kong hotel	0.0412
Map of Hong Kong	0.0209

However, this study is not without limitations. Given the data availability, SII factors are used for the input of tourism demand forecasting, while other determinants, such as GDP and weather condition, are not considered inputs. Moreover, this study only uses the pooled group with two regions due to the length limitation. Therefore, future research can group more regions to determine forecasting accuracy and other homogeneity factors. To facilitate future research along this line, we released the collected data sets "HK-MO2018" at https://github.com/tulip-lab/open-data and also open sourced the implementation of our method "GP-DLM" at https://github.com/tulip-lab/open-code.

Acknowledgement

This project was partly supported by a research grant funded by the Hong Kong Polytechnic University. The work was completed when Gang Li was on ASL in Chinese Academy of Sciences, and we would like to thank Deakin University's ASL2019 fund and Xinjiang Science & Technology Research fund with Chinese Academy of Sciences.

References

Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research, 58*(3), 383–397.
Asteriou, D., & Hall, S. G. (2011). ARIMA models and the Box–Jenkins methodology. *Applied Econometrics, 2*(2), 265–286.
Athanasopoulos, G., & de Silva, A. J. (2012). Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research, 51*(5), 640–652.
Baltagi, B. H., Bresson, G., & Pirotte, A. (2008). To pool or not to pool? *The econometrics of panel data* (pp. 517–546). Berlin, Heidelberg: Springer.
Baltagi, B. H., & Griffin, J. M. (1997). Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics, 77*(2), 303–327.
Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modelling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*(1), 133–153.
Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD workshop. Vol. 10, No. 16. KDD workshop* (pp. 359–370). WA: Seattle.
Cai, Z. J., Lu, S., & Zhang, X. B. (2009). *Tourism demand forecasting by support vector regression and genetic algorithm* (pp. 144–146). 8–11 August 2009. (Beijing, China).
Chen, R., Liang, C. Y., Hong, W. C., & Gu, D. X. (2015). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing, 26*, 435–443.
Chen, T., & Guestrin, C. (2016, August). *Xgboost: A scalable tree boosting system* (pp. 785–794). 13–17 August 2016. (San Francisco, CA).
Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks, 17*(1), 113–126.
Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics, 188*(2), 393–420.
Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics, 6*(1), 3–73.
Constantino, H. A., Fernandes, P. O., & Teixeira, J. P. (2016). Tourism demand modelling and forecasting with artificial neural network models: The Mozambique case study. *Tékhne, 14*(2), 113–124.
Correia, A., Kozak, M., & Ferradeira, J. (2011). Impact of culture on tourist decision-making styles. *International Journal of Tourism Research, 13*(5), 433–446.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine learning basics. *Deep Learning, 1*, 98–164.
Gunter, U., Önder, I., & Smeral, E. (2019). Scientific value of econometric tourism demand studies. *Annals of Tourism Research, 78*, 102738.
Hansen, M., Fyall, A., Spyriadis, T., Rogers, D., & Brander-Brown, J. (2019). Motivating stakeholder collaboration within the aerial adventure industry. *International Journal of Tourism Research, 21*(3), 311–322.
Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research, 63*, 112–127.
Hassani, H., Webster, A., Silva, E. S., & Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis. *Tourism Management, 46*, 322–335.
Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1–12.
Hoogstrate, A. J., Palm, F. C., & Pfann, G. A. (2000). Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business & Economic Statistics, 18*(3), 274–283.
Howell, D. C. (2009). *Statistical methods for psychology.* Cengage Learning.
Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence, 39*, 235–244.
Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems, 7*(3), 358–386.
Kulendran, N., & Wilson, K. (2000). Modelling business travel. *Tourism Economics, 6*(1), 47–59.
Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research, 75*, 410–423.
Lee, S., & Bai, B. (2016). Influence of popular culture on special interest tourists' destination image. *Tourism Management, 52*, 161–169.
Lever, J., Krzywinski, M., & Altman, N. (2016). *Points of significance: model selection and overfitting*.
Li, G., Song, H., & Witt, S. (2005). Recent development in econometric modelling and forecasting. *Journal of Travel Research, 44*(1), 82–99.

Li, G., Wu, D. C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research, 75*, 363–378.
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.
Li, X., & Law, R. (2020). Forecasting tourism demand with decomposed search cycles. *Journal of Travel Research, 59*(1), 52–68.
Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57–66.
Liu, Y. Y., Tseng, F. M., & Tseng, Y. H. (2018). Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model. *Technological Forecasting and Social Change, 130*, 123–134.
Long, W., Liu, C., & Song, H. (2019). Pooling in tourism demand forecasting. *Journal of Travel Research, 58*(7), 1161–1174.
Muskat, B., Muskat, M., & Richardson, A. (2014). How do Europeans travel in Australia? Examining cultural convergence in travel behaviour. *Journal of Vacation Marketing, 20*(1), 55–64.
Otsu, T., Pesendorfer, M., & Takahashi, Y. (2016). Pooling data across markets in dynamic Markov games. *Quantitative Economics, 7*(2), 523–559.
Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research, 56*(7), 957–970.
Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*.
Shi, H., Xu, M., & Li, R. (2017). Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid, 9*(5), 5271–5280.
Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research, 74*, 134–154.
Song, H., Dwyer, L., Li, G., & Cao, Z. (2012). Tourism economics research: A review and assessment. *Annals of Tourism Research, 39*(3), 1653–1682.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management, 29*(2), 203–220.
Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research, 75*, 338–362.
Tetko, I. V., Livingstone, D. J., & Luik, A. I. (1995). Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences, 35*(5), 826–833.
Wan, S. K., & Song, H. (2018). Forecasting turning points in tourism growth. *Annals of Tourism Research, 72*, 156–167.
Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting, 11*(3), 447–475.
Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82.
WTTC (2019). *Travel and tourism: Economic impact 2019.* London: World Travel & Tourism Council (WTTC).
Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications, 39*(5), 4769–4774.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–397.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*(4), 433–447.

Yishuo Zhang: His research interests are tourism and hospitality management, and data science.

Gang Li: His research interests are technology applications to tourism and hospitality, data science, and business intelligence.

Birgit Muskat: Her research interests are tourism management, entrepreneurship, and innovation.

Rob Law: His research interests are technology applications to tourism and information management.

Yating Yang: Her research interests are technology applications in tourism, and data science.