



RESEARCH ARTICLE

Predictivity of tourism demand data

Yishuo Zhang^{a,b}, Gang Li^{a,*}, Birgit Muskat^c, Huy Quan Vu^d, Rob Law^e^a School of Information Technology, Deakin University, VIC 3125, 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 Department of Information Systems and Business Analytics, Deakin University, VIC 3125, Australia^e School of Hotel & Tourism Management, The Hong Kong Polytechnic University, 17 Science Museum Rd, TST East, Kowloon, Hong Kong

ARTICLE INFO

Article history:

Received 13 December 2020

Received in revised form 2 May 2021

Accepted 8 May 2021

Available online xxxx

Associate editor: Gang Li

Keywords:

Data characteristics

Entropy

Predictivity

Tourism demand forecasting

ABSTRACT

As tourism researchers continue to search for solutions to determine the best possible forecasting performance, it is important to understand the maximum predictivity achieved by models, as well as how various data characteristics influence the maximum predictivity. Drawing on information theory, the predictivity of tourism demand data is quantitatively evaluated and beneficial for improving the performance of tourism demand forecasting. Empirical results from Hong Kong tourism demand data show that 1) the predictivity could largely help the researchers estimate the best possible forecasting performance and understand the influence of various data characteristics on the forecasting performance.; 2) the predictivity can be used to assess the short effect of external shock – such as SARS over tourism demand forecasting.

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Tourism demand forecasting models aim to predict the number of tourists arriving at a destination in a certain period. The relevance of these models is immense as they are essential for policy makers and practitioners in making long-term strategic and short-term operational plans (H. Song et al., 2019). Existing research on tourism forecasting offers a variety of approaches with various predictive models, which are based on either time series, econometric, or artificial intelligence (AI) methods (Wen et al., 2021; Y. Zhang, Li, Muskat, Law, 2020). Recently, significant progress has been made in tourism forecasting research, especially with AI-based models, which utilize decomposition and deep learning techniques to improve forecasting model performance and accuracy (Y. Zhang, Li, Muskat, Law, 2020).

Nevertheless, accurate forecasting of tourism demand remains a challenging task. The two key factors researchers must consider, when constructing forecasting models, are 1) the data quality and 2) the characteristics of data used in the modelling process. In this context, data quality refers to the predictive power of the available data, which plays an important role in determining the performance of the forecasting models (Goh & Law, 2002). To date, only a few tourism studies have investigated how data quality influences performance of forecasting models. The studies conducted by S.F. Witt et al. (2003) and Song et al. (2010) are examples of the related works on data quality, but they focused on model selections based on performance measures, such as mean absolute percentage error (MAPE) or root mean squared percentage error (RMSPE) (S.F. Witt & Witt, 1992).

* Corresponding author at: School of Information Technology, Deakin University, VIC 3125, Australia.

E-mail addresses: Zhangyis@deakin.edu.au, (Y. Zhang), gang.li@deakin.edu.au, (G. Li), birgit.muskat@anu.edu.au, (B. Muskat), q.vu@deakin.edu.au, (H.Q. Vu), rob.law@polyu.edu.hk, (R. Law).

The second factor is referring to data characteristics, such as the number of observations (*data length*), data structure (*data decomposition*), and time interval (*data granularity*) (Ghods et al., 2018), which have been identified as highly relevant to the performance of tourism demand forecasting models. Ghods et al. (2018) used different decomposition formats to test the best forecasting performance on given time series, though the best forecasting performance was associated with the *Singular Spectrum Analysis* method rather than with any possible method. Other researchers pre-processed the data based on certain characteristics prior to the modelling stage (Law et al., 2019). However, very limited knowledge exists on how these data characteristics influence the performance of the forecasting models. The lack of understanding about the influence of data quality and characteristics on forecasting model performance presents a major challenge for tourism researchers because it remains unclear how to obtain an optimal forecasting model from the available data with the highest possible forecasting accuracy.

C. Song, Qu, et al. (2010) described *maximum predictivity* as the theoretical maximum forecasting accuracy for a given tourism demand data, which is also referred to as the predictive power of the data. Identifying the maximum predictivity could largely help researchers to evaluate the forecasting performance of the selected models. In addition, the maximum predictivity sheds light on whether the available data are sufficient for optimal model construction; if not, the practitioners could have some clues on how to improve the quality of available data. However, limited research in tourism literature has been devoted to exploring the maximum predictivity of the available tourism demand data.

To consolidate these gaps in the tourism demand forecasting, this study aims to address the following two research questions:

RQ1. *What is the maximum predictivity that can be achieved by models constructed from the available tourism demand data?*

RQ2. *How do the data characteristics influence the maximum predictivity in tourism demand forecasting?*

Thus far, a wide range of practical techniques are available to improve tourism demand forecasting performance by exploiting various characteristics of the available tourism demand data. These techniques include training data with different lengths (L. Cai & Zhu, 2015), adjusting the granularity of the data (Ott et al., 2013), and decomposing the data (X. Li & Law, 2020; Silva et al., 2019). However, these techniques are criticized to be either predetermined or directly utilized in the forecasting practice without an explicit understanding of their respective impact on the outcomes (C. Song, Qu, et al., 2010). Subsequently, how these data characteristics might influence the maximum predictivity of the tourism demand data remains an unsolved puzzle.

In this study, we introduce an innovative approach to evaluating the quality (predictivity) of tourism demand data, to reveal the relationship between data characteristics (i.e., *data length*; *data decomposition*, and *data granularity*) and the performance of forecasting models. To evaluate the predictability we draw on *entropy*, a measure of data complexity that has proven to be effective for assessing the predictive power of the data used to construct forecasting models (Molgedey & Ebeling, 2000). We capture two types of entropies: 1) *sample entropy* that refers to the degree of predictivity for a given period of tourism demand data, and 2) *multi-scale entropy* that characterizes predictivity over the heterogeneity of a given period of tourism demand data at various granularities, such as *daily*, *monthly*, and *quarterly*.

To evaluate the maximum predictivity of a given tourism demand data set, Fano's Inequality is used (C. Song, Qu, et al., 2010), which is computed from sample entropy and multi-scale entropy measures (T. Xu et al., 2017). Data used for tourism demand forecasting are captured in the form of time series including volumes of tourist arrival to specific locations or destinations. The nature of tourism demand data is unique in terms of seasonality and fluctuations, when compared with time series data used in other domains (e.g. financial data in economic and finance). This paper is devoted specifically to evaluating the quality of tourism demand data.

This research uses Hong Kong tourism demand data to initially demonstrate the evaluation process to achieve maximum predictivity of available data-and to subsequently reveal the influence of data characteristics on the maximum predictivity. Our approach and findings regarding the maximum predictivity evaluation are beneficial to tourism scholars and practitioners in designing and improving the performance of tourism demand forecasting, such as determining the length of the training data to obtain the best forecasting performance and clear understanding of the effectiveness in the use of different granularities and decompositions during tourism demand forecasting.

This paper is structured as follows: Section 2 reviews related works on tourism demand forecasting and predictability of data in forecasting models; Section 3 develops our approach to measure data quality and characteristics; Section 4 presents a case study with Hong Kong tourism demand data, together with a discussion on theoretical and practical implications. Section 5 concludes this research and offers future research directions.

Literature review

Tourism demand forecasting models

Decision-making from tourism practitioners heavily relies on accurate forecasting of tourism demand, which also influences the economic growth of the related industries, such as hospitality, events, attractions, transportation, and retail (Xie et al., 2020; Y. Zhang, Li, Muskat, Law, 2020; Y. Zhang, Li, Muskat, Law, 2020). The data used to forecast tourism demand can be collected from multiple sources (H. Li et al., 2020) and at various scales, such as on a daily, weekly, monthly, quarterly, and yearly basis (Gunter & Önder, 2015).

Existing approaches on tourism demand forecasting can be grouped into three major categories: time-series models, econometric models, and artificial intelligence (AI) models, whose details are discussed below. *Time series models* leverage on historical

data to predict future tourism demands. Forecasting models based on *autoregressive integrated moving average* (ARIMA) and its variants have been widely adopted (G.P. Zhang, 2003). Goh and Law (2002) proposed the SARIMA method, which captures the seasonality inside the univariate time series. With multivariate time series, Lim et al. (2009) discussed the ARIMAX model, which adopts the auto-regression to handle multiple variate time series. The *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) model, which is an extension of autoregressive models, has been found successful in tourism demand forecasting (Chan et al., 2005). Additionally, in early tourism demand forecasting research, the *exponential smoothing* (ETS) (C. Song, Qu, et al., 2010) and Naive methods were proposed (Goh & Law, 2002).

Econometric models utilize the relationships between the tourism demands and explanatory variables for tourism demand forecasting (Law et al., 2019). Traditional econometric models are based on regression methods, such as *Ordinary Least Squares* (OLS), *Autoregressive Distributed Lag Model* (ADLM), *Vector Autoregressive Model* (VAR), and error correction model (Kamel et al., 2008), have been commonly used. Other econometric models include the Bayesian vector autoregressive (Y. Xu et al., 2012) and novel Bayesian FAVAR model (Wong et al., 2006). Recently, the mixed data sampling model was proposed by using Google search data to improve the performance of tourist demand forecasting (X. Li & Law, 2020). However, econometric models have limitations on their feature selection process. Particularly, how to efficiently model with a large number of variables remains a problem in tourism demand forecasting.

AI-based models: AI-based models have also been successfully adopted for tourism demand forecasting. Z.-J. Cai et al. (2009) used the generic SVR algorithm and found that G-SVR required fewer parameters but with better performance than ARIMAs. Wong et al. (2006) proposed a Bayesian network model to forecast Hong Kong tourism demand with satisfactory results. Cankurt (2016) used the regression tree for Turkish tourism demand forecasting. As a non-linear forecasting model, artificial neural networks (ANNs) have also been experimented on tourism forecasting (Teixeira and Fernandes, 2014). Recent works (Law et al., 2019; Y. Zhang, Li, Muskat, Law, 2020) introduced the deep learning models to tourism demand forecasting and found deep learning models provide improved generalization capability and high accuracy.

Meta-methods for tourism demand forecasting

As the performance on each forecasting model typically varies, meta-methods have been used to boost the performance with the advantage of being less biased. In the machine learning literature, 'meta-methods' are defined as mostly as ensemble and boosting methods, which use multiple models rather than single model (Aggarwal, 2013; Wang et al., 2009). In the tourism demand forecasting literature, meta-methods include those which are not limited to any particular type of data. Examples of meta-methods that have been used widely in tourism demand forecasting studies are decomposition and Bagging (X. Li & Law, 2020). Meta-methods can be categorized into two types: 1) *model-oriented meta-methods* and 2) *data-oriented meta-methods*.

Model-oriented meta-methods

These methods use multiple models to perform forecasting rather than a single model alone (H. Song et al., 2019) and are known as ensemble methods. *Bagging* is one of such methods, which provides the flexibility to generalize training data more than a single model would and therefore yields a better forecasting performance. Ensemble methods are independent of specific base models and thus robust in forecasting performance (Zhao et al., 2019). Cankurt (2016) employed regression tree-based ensemble method for tourism demand forecasting in Turkey. Y. Zhang, Li, Muskat, Law (2020) made the pooling data ensemble with deep learning models on both Macau and Hong Kong tourism demand data. Xie et al. (2020) also found the decomposition with ensemble could produce accurate forecasting. The theoretical analysis of ensemble methods has been well understood in machine learning community.

Data-oriented meta-methods

This type of meta-models uses data characteristics, such as the length of the data, the granularity of the data, and the decomposition components. These data characteristics are based on the nature of the tourism demand data and can be processed into different forms before modelling (Y. Zhang, Li, Muskat, Law, 2020). Data characteristics are freely used with any tourism demand forecasting model, and they have been explored extensively for tourism demand forecasting in the last decade (Peng et al., 2014). According to Peng et al. (2014) and L. Cai and Zhu (2015), the accuracy of tourism demand forecasting is sensitive to data length, namely, the time period of historical data (Bangwayo-Skeete & Skeete, 2015; Ott et al., 2013). The data granularity has been known to improve the outcome of tourism demand forecasting in certain situations. Although many studies have discussed the impacts of those data characteristics, no clear understanding exists on how those data characteristics influence the best possible forecasting outcomes, regardless of the particular models in use. Here, the best possible forecasting performance can be regarded as a theoretical concept to reflect predictivity of time series data, according to information theory (Rényi, 1961).

In addition, those data characteristics are often used as pre-determined processing before constructing a forecasting model. Theoretical analysis is needed to justify if such processing is suitable in obtaining the optimum forecasting outcome. Apart from raw data characteristics, data decomposition can also help improve the forecasting accuracy (Chan et al., 2005; H. Hassani et al., 2015; X. Li & Law, 2020; Hassani et al., 2017). X. Li and Law (2020) claimed that the empirical model decomposition could be used to generate stationary series and accurate forecasting performance. Along the same line of research, H. Hassani et al. (2015) used *Singular Spectrum Analysis* (SSA) decomposition in tourism development analysis with US tourism demand forecasting. Y. Zhang, Li, Muskat, Law (2020) concluded that the *Seasonal Trend Loses* (STL) decomposition with deep learning could improve the tourism demand forecasting. Yet, existing studies are limited to particular forecasting models such as SSA or STL. To

extend existing forecasting studies, this work theoretically evaluates the impact of data-oriented meta methods on predictivity, which is not relying on any particular forecasting method.

Predictivity of data

Since C. Song, Qu, et al. (2010) introduced the concept of *maximum predictivity*, several predictivity evaluation methods have been proposed. Predictivity evaluation methods based on Information Theory was the most popular, which utilizes Entropy to estimate the predictivity of a given data set (C. Song, Qu, et al., 2010). Chen et al. (2016) utilized entropy to explore the predictivity of demand data based on online check-ins. Salisu et al. (2019) used the predictivity estimated from financial time series to forecast future patterns. Prior works have found that predictivity depends on the characteristic of the given tourism demand data. For instance, changing the data length and the data granularity can significantly influence predictivity (T. Xu et al., 2017). However, understanding has not been obtained about the influence of other data characteristics (e.g., decomposition) on predictivity and predictivity evaluation.

In the tourism literature, studies have so far concentrated to improve model performance (S.F. Witt et al., 2003), but very limited understanding exists on enhancing the predictivity of data. Some research explored related problems regarding to the best possible performance on giving tourism demand data from various models' selection (S.F. Witt et al., 2003). Their study compared different models, using the same tourism demand data and selected the model with the best forecasting performance with the lowest forecasting error. Owing to the available range of models in consideration, their method of model comparison towards the best forecasting performance is limited, and the actual best possible performance on a given tourism demand data remains an open issue. Researchers also developed techniques to efficiently utilize their available data (Zhang, Li, Muskat, Law, Yang, 2020), yet the influence of the varied data characteristics on the forecasting performance is still not clearly understood.

We aim to develop a method which can directly evaluate the quality (predictivity) of tourism demand data. Our study also provides a theoretical proof behind many data characteristics that influence the forecasting outcomes of tourism demand. The introduced approach and results are highly important for tourism researchers in selecting appropriate process to tourism demand data before the modelling work to achieve the optimum or the best possible forecasting performance.

Methodology

This section presents the methodology for evaluating the predictivity of tourism demand data. We incorporate two well-established information theory concepts, namely, *Entropy* and *Fano's Inequality*, into our evaluation process. This process of predictivity evaluation is carried out in some major steps, as outlined in Fig. 1. Briefly, tourism demand data with different characteristics (e.g., varied lengths, varied granularities, and varied decompositions) are first extracted from the raw data. Then, *sample entropy* and *multi-scale entropy* are computed to measure the complexity of the processed tourism demand data. *Sample entropy* values are applied to the same tourism demand data with *varied data lengths and varied decompositions*. *Multi-scale entropy* is applied to tourism demand data with *varied granularities*. Next, *Fano's Inequality* is used from the calculated entropies of the processed tourism demand data with varied characteristics. Fano's Inequality provides a lower bound on its error probability in terms of the mutual entropy and links the entropy values with the predictivity of the datasets (Richman & Moorman, 2000). Hence, Fano's Inequality can be used to evaluate the predictivity of the tourism demand data in the final step.

Next, we formalize the task of tourism demand forecasting, with details on the computation of entropies and Fano's Inequality. The approach to evaluate data predictivity using Fano's Inequality is then presented.

Formalization of tourism demand forecasting

We define tourism demand forecasting task in this work as the prediction of future tourist arrivals to a destination based on the historical tourism demand. Forecasting is based solely on the historical data in the form of univariate time series (Law et al., 2019; Song and Li, 2008; Yang, Pan & Song, 2014).

Let vector $\mathbf{Y}^T = \{y^{(1)}, y^{(2)}, \dots, y^{(T)}\}$ be the time series data with T time steps, where $\{y^{(i)}\}_{i=1}^k$ is the input, $\{y^{(i)}\}_{i=k+1}^{k+\delta}$ is the output, and δ denotes the forecasting steps. The tourism demand forecasting model \mathcal{F} utilizes $\{y^{(i)}\}_{i=1}^k$ to predict the future tourism arrival volume $\{y^{(i)}\}_{i=k+1}^{k+\delta}$ with δ steps ahead. The performance of tourism demand forecasting model is usually evaluated by its *Root Mean Square Error* (RMSE) and *Mean Absolute Error* (MAE). Lower values indicate better performance.

Entropy computation

Entropy measures the complexity of data and associates probabilities of all possible values in the given data (Dugdale, 2018; Jost, 2006; Richman & Moorman, 2000). Yet, the data complexity varies at different time steps due the nature of time series data. Directly measuring the entropy value for the tourism demand data in the form of time series is difficult (Costa et al., 2002). Therefore, we use the variants of *Entropy*, namely, the *sample entropy* and *multi-scale entropy*, to accommodate this task.

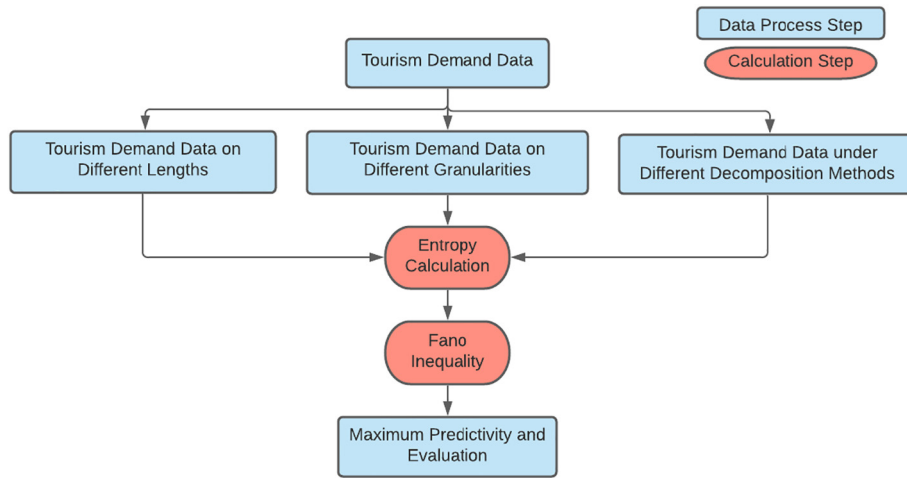


Fig. 1. Process for predictivity evaluation.

Given the time series data Y^T with length T , *sample entropy* is the negative natural logarithm of the probability that if the distance between two sub-series data of length m from Y^T is less than r , then, the Chebyshev distance between two simultaneous sub-series data of length $m + 1$ is also less than r (Richman & Moorman, 2000):

$$\text{SampEn} = -\ln \left[\frac{d^{m+1}(r)}{d^m(r)} \right], \tag{1}$$

$$d^m = \|Y_i^m - Y_j^m\|$$

where $d^m(r)$ is the count of two sub-series data for which the lengths are m and their distance is less than r , and $d^{m+1}(r)$ is the count of two sub-series data for which the lengths are $m + 1$ and their distance is also less than r . r is the acceptance threshold in $[\sigma * 0.2, \sigma * 0.5]$, where σ is the standard deviation. During the implementation, the value of r is selected from the mean absolute error of the forecasting on given tourism demand data. The value of r should not be too small as this can cause the problem that the predictivity becomes too small to be measurable. The increase in sample entropy values also increases the complexity in the time series data, which then increases the difficulty in achieving high forecasting performance.

Costa et al. (2002) argued that sample entropy is suitable for the application to time series data of small length but is unreliable for the data spanning a long period. As such, multi-scale entropy was developed (Costa et al., 2002) to effectively measure the complexity of the time series data with multiple scaling levels. Typical scale levels, also known as the granularity of time series data in tourism demand forecasting, are *weekly*, *fortnightly*, or *monthly*. Multi-scale entropy can be computed in the same way as the sample entropy but with different scale factors τ .

$$\text{MulSampEn} = -\ln \left[\frac{d^{m+1}(r)_{Z^\tau}}{d^m(r)_{Z^\tau}} \right] \tag{2}$$

where $d^m(r)_{Z^\tau}$ is the count of two coarse-grained sub-series data both with the length m , and their distance is less than r ; and $d^{m+1}(r)_{Z^\tau}$ is the number of coarse-grained sub-series data pairs both with the length $m + 1$, and their distance is less than r . $\{Z^\tau\}$ is coarse-grained sub-series, which can be computed from one-dimensional time series tourism demand data Y^T as

$$Z_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} y^{(i)}, 1 < j < \frac{T}{\tau} \tag{3}$$

Similar to sample entropy, multi-scale entropy captures the complexity of the time series data but with varied granularities. Forecasting with data that produce high multi-scale entropy is more complex than with data that produce low multi-scale entropy.

Predictivity evaluation

Predictivity on the time series Y^T with length T is defined as the time averaged probability for an algorithm to correctly predict the future time step based on the given sub-series time series.

$$\begin{aligned}
 p &= P(y^{(i)} = \hat{y}^{(i)} | Y^{i-1}) \\
 \Psi_i &= \sum_{Y^{i-1} \in Y^T} pP(Y^{i-1}) \\
 \Psi &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_i^T \Psi_i
 \end{aligned} \tag{4}$$

where p is a random value from the prediction distribution made by all predictive values regarding the expected value $y^{(i)}$. Ψ_i is the predictivity at time step i based on the historical time series Y^{i-1} , where $P(Y^{i-1})$ is the probability to observe the historical sub-series time series Y^{i-1} with length of $i - 1$, and the sum is taking all possible historical sub-series time series Y^{i-1} . The overall predictivity Ψ of all-time series is the average predictivity over all future time steps. *Maximum predictivity*, denoted as Ψ_i^{max} , is a special case of Ψ when the theoretical best prediction on the future time steps occurs (C. Song, Qu, et al., 2010).

Let $S(Y)$ be the entropy (sample entropy or multi-scale) computed from the given time series data and $S(\Psi)$ be the entropy of the predictivity value Ψ , which reflect whether the time series could be successfully forecasted based on the probability of correct forecast. $S(\Psi)$ can be computed as

$$S(\Psi) = -\Psi \log_2 \Psi - (1-\Psi) \log_2 (1-\Psi) \tag{5}$$

According to *Fano's Inequality*, the complexity of given time series $S(Y)$ is less than or equal to the sum of complexity of correct prediction and the complexity of possible incorrect prediction on the future time steps (Verdú, 1994), which is presented as

$$S(Y) \leq S(\Psi) + (1-\Psi) \log_2 (T-1) \tag{6}$$

where T is the length of the time series Y^T . The equation holds only when Ψ is the maximum predictivity Ψ_i^{max} on the given time series data.

$$S(Y) = -\Psi^{max} \log_2 \Psi^{max} - (1-\Psi^{max}) \log_2 (1-\Psi^{max}) + (1-\Psi^{max}) \log_2 (T-1) \tag{7}$$

Therefore, we can use the $S(Y)$ to estimate the upper bound of predictivity or the maximum predictivity Ψ_i^{max} on the given time series Y^T . This procedure is applied to evaluate the maximum predictivity of time series data with varied lengths, varied granularities, and varied decompositions.

Case study

This section presents a case study to demonstrate the effectiveness of the presented approach for evaluating data predictivity in tourism demand forecasting. Description on the case study and data collection is first presented. The predictability of the collected data (with varied characteristics) is then evaluated following the method outlined in Fig. 1. In addition, we carry out a number of experiments with the popularly used forecasting model SARIMA (e.g. S.F. Witt et al., 2003) to validate the reliability of the calculated maximum predictivity following our proposed method. The predictivity implementation and all experiments with conducted with Python programming platform.

Data description

Our predictivity case study is based on tourism demand data of Hong Kong, a popular tourism destination in Asia. Monthly tourism demand data of Hong Kong are available from the website of the Hong Kong Tourism Board (<https://partnet.hk.com/en/home/index.html>). We used Hong Kong tourism demand data for our case study, for two major reasons. First, the main reason is that numerous studies have used Hong Kong data in their forecasting case studies (X. Li & Law, 2020; Y. Zhang, Li, Muskat, Law, 2020), and thus comparisons with the extant literature can be drawn for our data set. Second, this further allows to validate the effectiveness of our approach. The data are in the form of time series capturing monthly tourist arrival from January 1996 to April 2019 (see Fig. 2). Apparently, tourist arrival increases steadily from 1996 to 2009, except for the SARS pandemic period in 2003, and then slowly increases until 2015. The increase in tourist arrival slowed down for the period between 2015 and 2019, probably due to various events that happened in Hong Kong.

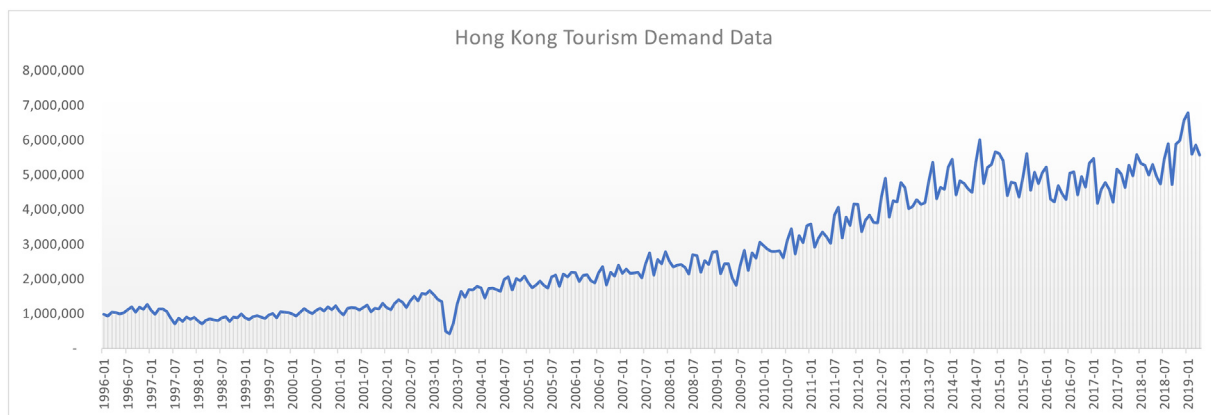


Fig. 2. Monthly Hong Kong tourism arrivals in 1996 – 2019.

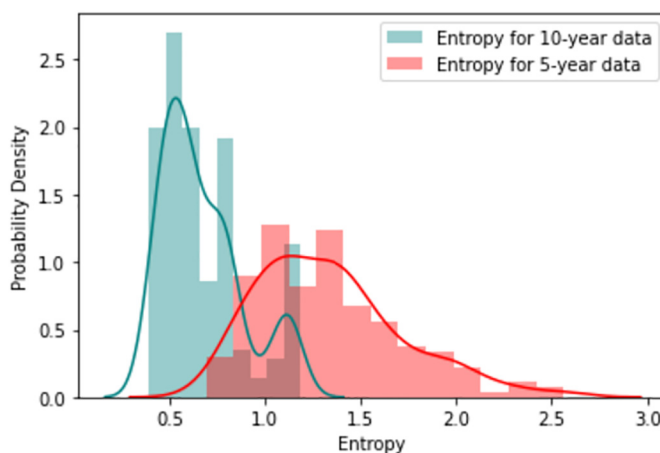


Fig. 3. Entropy values of 5-year and 10-year sub-series data

Predictivity evaluation of Hong Kong tourism demand data

Varied data lengths

This section evaluates the relationship between the data length and the maximum predictivity of Hong Kong tourism demand data. Given that the time series data span for a long period of 280 months, which correspond to the number of months from Jan-1996 to Apr-2019, we evaluate their predictivity based on the samples of multiple sub-series extracted from the original data. Let the original monthly Hong Kong tourist arrivals be denoted as $Y^T = \{y^{(1)}, y^{(2)}, \dots, y^{(T)}\}$, where $T = 280$. The sub-series samples can be defined as $\{y^d\}_{i=d-m}^d$, $m + 1 \leq d \leq 280$, where m corresponds to the number of months in each sub-series sample. In this experiment, we consider the sub-series samples of varied lengths such as $m = 60$ (equivalent to five years) and $m = 120$ (equivalent to 10 years). A total of 120 sub-series samples of five years and 80 sub-series samples of 10 years are included.

The sample entropies are then calculated for the extracted sub-series samples, based on Eq. (1). We set the distance (tolerance) $r = 0.55 * \sigma$ and the dimension of $m = 3$. The choice of $m = 3$ on sample entropy calculation is based on the work of (Udhayakumar et al., 2016) that the smaller m will ensure the accurate entropy values to be obtained, as large m value may result in null value for the entropy calculation. Then, $r = 0.55 * \sigma$ is used as the distance, whose values could be realistic on the absolute forecasting errors according to prior studies on Hong Kong tourism demand forecasting data (Law et al., 2019; Y. Zhang, Li, Muskat, Law, 2020). Also, the choice of r will result in different entropy values but the overall pattern on the higher or the lower maximum predictivity is not impacted (some of the small distance values will make entropy values extremely large and the predictivity will not be available across all conditions). The computed entropy values are shown as the probability density distributions in Fig. 3. We see that the sample entropy values on 10-year sub-series data are generally lower than the sample entropy values on 5-year sub-series data. In other words, 10-year sub-series data have less complexity than the 5-year sub-series data.

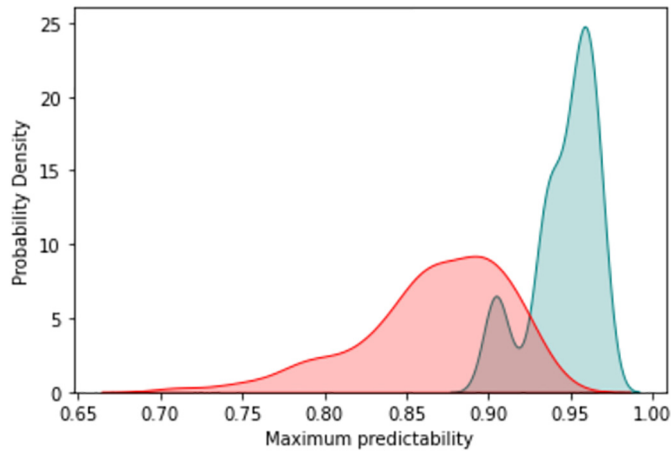


Fig. 4. Maximum predictivity of 5-year and 10-year sub-series data

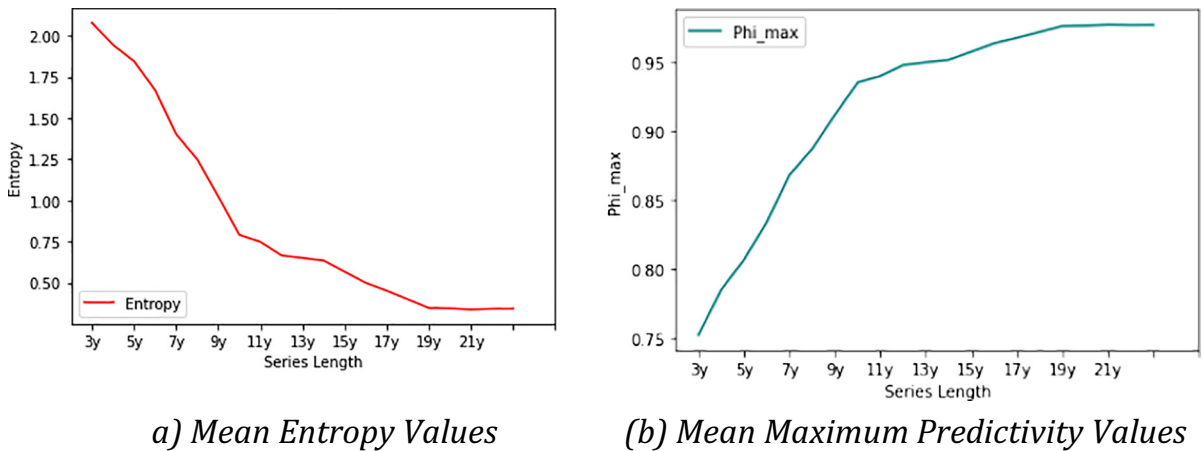


Fig. 5. Predictivity evaluation with varied data lengths.

In addition, we compute the maximum predictivity Ψ_i^{\max} values on the bases of Eq. (6) for the sample sub-series and present them as probability density distributions in Fig. 4. The maximum predictivity values on the 10-year sub-series data are generally larger than on 5-year sub-series data.

In general, the visualization on both sample entropy and the maximum predictivity shows that Hong Kong tourism demand data of longer length provide larger maximum predictivity or larger theoretical maximum forecasting performance. Here, a question arises as follows: does the maximum predictivity keep increasing if we continuously increase the data length, or does an upper bound exist on the maximum predictivity? We explore these issues in the next experiment.

We repeated the similar computation of entropy and maximum predictivity for the sub-series samples of varied lengths, three years ($m = 36$) to 22 years ($m = 264$) in 19 steps. For ease of interpretation, Fig. 5 shows the mean values of the sample entropies and maximum predictivity rather than as probability distributions in the previous case. Apparently, the mean entropy values are decreasing as the data length increases (Fig. 5a), and the maximum predictivity is monotonically increasing with the data length (Fig. 5b). A significant increase happens in the maximum predictivity from the data lengths of three to 10 years. Then, the increase in maximum predictivity slows down. More importantly, when the data length is beyond 19 years, the maximum predictivity tends to stabilize.

The result indicates that the complexity of the data decreases when the data of greater length are considered, which improve the maximum predictivity. However, an upper bound exists, which means that the maximum predictivity will not improve further if the data length is sufficient. In other words, once sufficient data have been collected, further increasing the data length would not help improve the tourism demand forecasting performance. These findings are also consistent with machine learning theory, that the forecasting performance will be influenced by the model as well as the sample complexity (Valiant, 2013). An optimal size exists on the data length to achieve a desired forecasting performance. However, data length beyond the optimal size will not be helpful in further improving the performance.

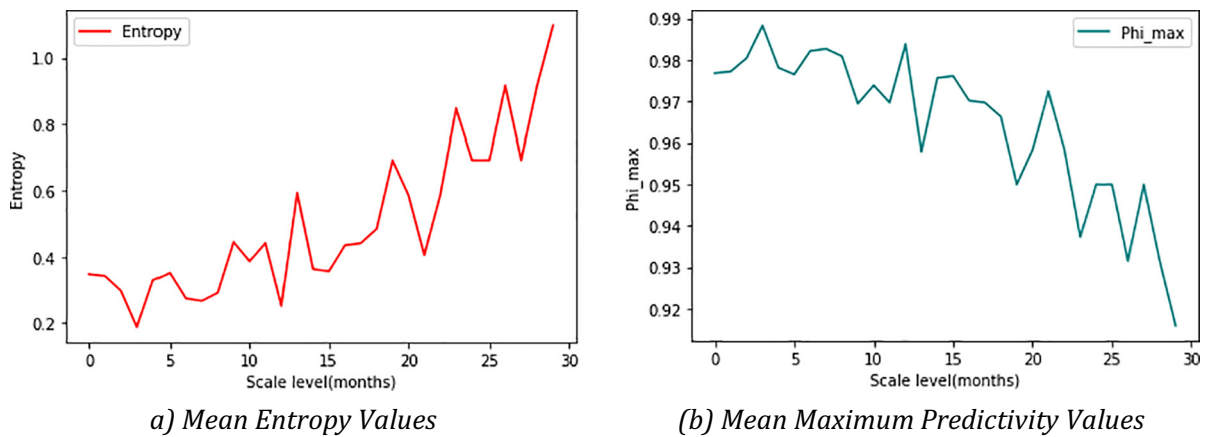


Fig. 6. Predictivity Evaluation with varied data granularities.

Varied data granularity

Data granularity is an important issue in forecasting. For example, daily data have higher data granularity than weekly data, and monthly data have higher granularity than quarterly data. Tourism practitioners could make a choice on the granularity of the data to improve forecasting accuracy. This section explores the maximum predictivity with respect to varied granularities by scaling the data. To clearly address the term of scaling data, we use lower scale of data to describe the data granularity become smaller such as yearly to monthly; and the higher scale of data means the data granularity become larger such as monthly to yearly. First, we sample the data with varied scales, one month to 30 months for each data point. The scale of one month per data point is equivalent to the original data with monthly tourist arrivals. The scale of 30 months per data point transforms the original data into a new time series data with only eight data points. We then calculate the multi-scale entropy of the original Hong Kong tourism demand data with varied granularities. The multi-scale entropy is calculated with the parameter setting as distance $r = 0.55 * \sigma$ and the dimension of $m = 3$. Then, the maximum predictivity values are estimated on the bases of the computed entropy values.

The entropy rises (Fig. 6a), and the maximum predictivity drops (Fig. 6b) when the scale level increases probably because additional new patterns could come from the data with higher scales than the original time series. Moreover, we found that forecasting based on the monthly data (scale = 1) is more accurate than on the quarterly data (scale = 3). This finding is because the length of quarterly data is shorter than the monthly data for the same period, which is also consistent with the previous finding on data length that shorter data length decreases the maximum predictivity. In summary, the lower scale data such as yearly to monthly could have better maximum predictivity as the data length is larger than higher scale data, such as weekly to monthly. However, in real forecasting scenario, the data granularity is also depending on the total data length and any other requirements.

Notably, the distance parameter r has an influence on determining the forecasting results. Forecasting is considered correct if the forecasting error is within the distance parameter r from true tourism demand data. Here, r is usually expressed as the percentage of standard deviation on time series. Thus, we carried out another experiment to examine the impact of r on predictivity. Following the same parameter on dimension $m = 3$, we vary the distance parameter r from $0.1 * \sigma$ to $0.7 * \sigma$ with a step of $0.05 * \sigma$, which produces a total of 16 different tolerance levels. We perform the analysis with respect to 6 different scale levels as shown in Fig. 7.

In general, the entropy values for all scale levels decrease as the distance r (or the tolerance level) increases (Fig. 7a). Lower distance values produce higher entropy values for data with different scale levels. Similarly, lower distance values result in smaller maximum predictivity values (Fig. 7b). These findings suggest that increasing distance r will result in less complexity of the data and the reduction of the data complexity is the reason for the increasing maximum predictivity value, which allows for better possible forecasting performance.

Varied data decomposition

In this part, we use the entropy and maximum predictivity to validate the effectiveness of data decomposition on tourism demand forecasting. Intuitively, when comparing the original series, decomposed tourism demand data should contain less noise and are supposed to improve forecasting accuracy. Hence, our experiment aims to confirm if the maximum predictivity of decomposed time series is larger than the original time series. Two well established decomposition techniques, namely, STL (Cleveland et al., 1990) and SSA (Silva et al., 2019), are evaluated in this study.

Following the work of Cleveland et al. (1990), we first decompose the Hong Kong tourism demand data using the STL technique to extract the global trend series, which is the summation of trend and seasonality components. Then, the sample entropy value of de-noised global trend series and the original time series are calculated with the same setting of distance $r = 0.55 * \sigma$ and the dimension $m = 3$. The maximum predictivities of both the de-noised global trend series and the original time series are then estimated on the basis of the computed sample entropy values and Fano's Inequality. The experiments were also carried

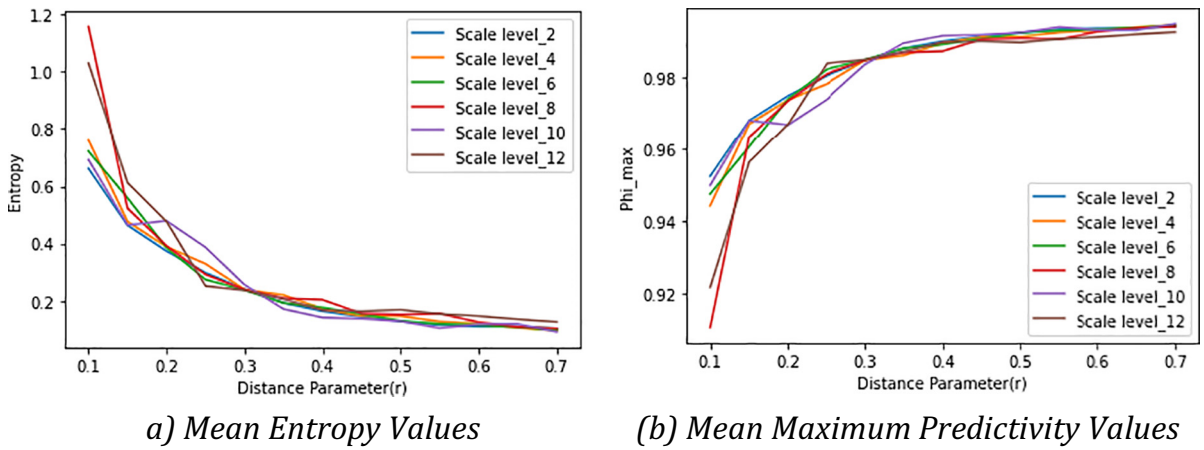


Fig. 7. Predictivity evaluation with varied distance parameters.

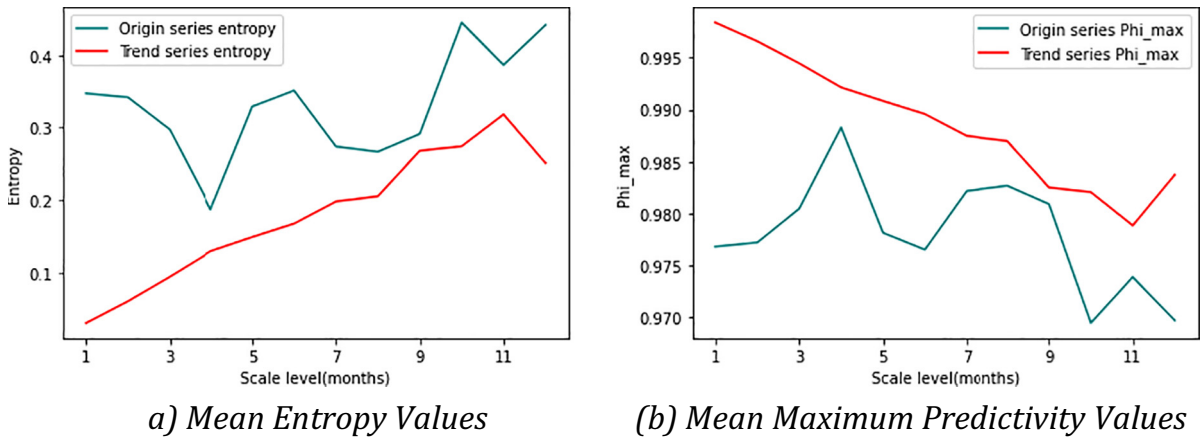


Fig. 8. Predictivity Evaluation for the original and global time series based on STL decomposition.

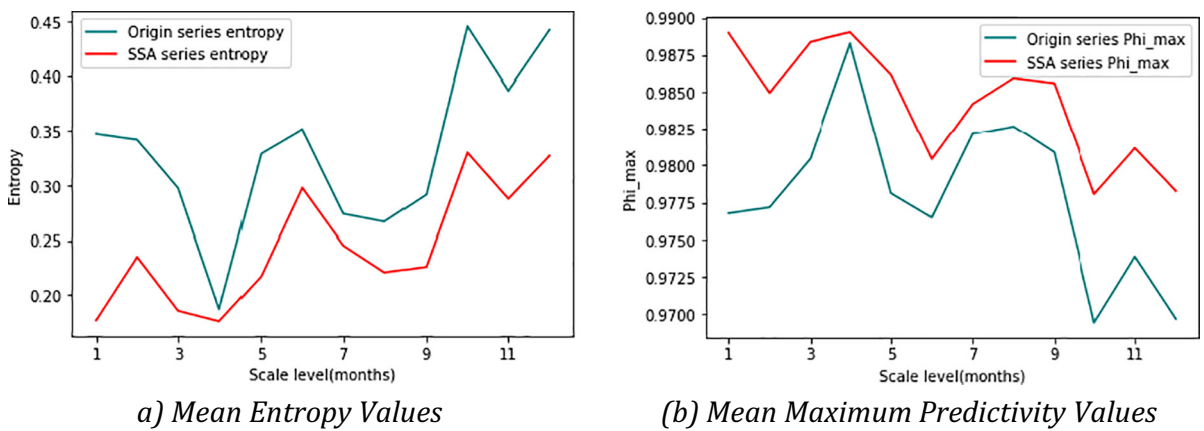


Fig. 9. Predictivity Evaluation for the original and decomposed time series based on SSA.

out with varied scale levels. The STL period parameter is 12 for the month seasonality, and the number of components is 24 with the window size = 12 for SSA.

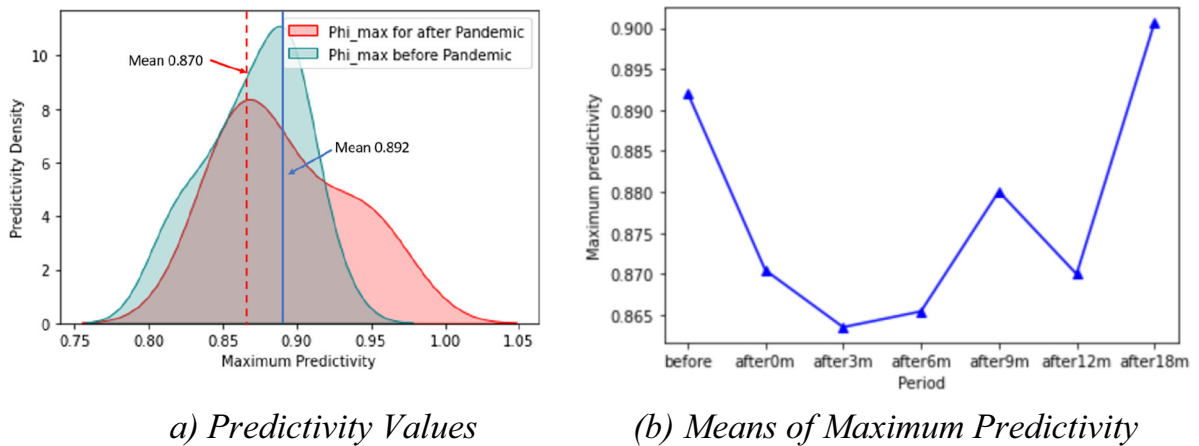


Fig. 10. Predictivity Evaluation of time series data before and after the SARS pandemic.

Fig. 8a shows that the sample entropy values of global trend series are generally lower than the original series. Fig. 8b shows that the maximum predictivity on global trend is generally larger than that of the original series. This result aligns with the common understanding that the complexity of the global trend should be less than the original series. Thus, with proper decomposition strategy, researchers could improve the accuracy of the tourism demand forecasting outcomes.

A similar analysis was carried out to compare the predictivity of the original time series data and reconstructed data using the SSA decomposition technique (Silva et al., 2019). The entropy value of reconstruction data is generally lower than the original series (Fig. 9a), and the maximum predictivity of the reconstruction data is generally larger than the original series (Fig. 9b). These results suggest that both STL decomposition and SSA could filter the noise in the time series, which reduce the complexity level of the time series data. In summary, the entropy value decreases, and the maximum predictivity increases, which are effective in improving forecasting accuracy.

Predictivity evaluation over external shock

External shocks refer to the global-scale events that could have a severe impact on international travel behavior and tourism demand (Khalid et al., 2020). Examples of these events are the global pandemics such as SARS (in 2003) and COVID-19 (in 2020), natural disasters (e.g., bushfires, flooding), and political conflicts. Predicting such disasters and their consequences is often difficult, and how they impact the forecasting for future tourism demand is unclear. This section is devoted to examining this issue by evaluating the effects that external shock would bring to tourism demand forecasting performance.

As the Hong Kong tourism demand data cover the period from 1996 to 2019, we use the SARS pandemic as the external shock period and break the data into two series: before and after the pandemic. The SARS pandemic in Hong Kong happened from Jan 2003 to June 2003. Therefore, the Hong Kong tourism demand data from Jan 1998 to Jan 2003 are used as before the pandemic series, and July 2003 to July 2008 data are used as after the pandemic series (Hung, 2003). Both series have the same length (5 years) and the same data granularity (monthly). The distance $r = 0.55 * \sigma$ and the dimension of $m = 3$ are used for both before and after the pandemic series. To fairly compare the maximum predictivity, the two series are repeatedly sampled with data length of 12 months, and the maximum predictivity distribution will be used for comparison.

Fig. 10a shows the distribution density of maximum predictivity for the two sampled series. Before the SARS pandemic, the mean maximum predictivity on data length of 12 months is 0.892. However, after the pandemic, the mean maximum predictivity declines to 0.870. The comparison shows clearly that the SARS pandemic had created the irregular data pattern on Hong Kong tourism demand data than before the pandemic period. The complexity of the Hong Kong tourism demand data has increased, and therefore the maximum predictivity has decreased. The findings show that the external shock caused by the SARS pandemic has led to a noticeable difference in tourism demand forecasting through the data and dramatically change the predictivities.

We further assess the impact of the SARS pandemic on future predictivity performance. To achieve this, we compute the means of maximum predictivity for various periods after the pandemic, namely, Oct 2003 to Oct 2008 (three months after the pandemic), Jan 2004 to Jan 2009 (six months after the pandemic), Apr 2004 to Apr 2009 (nine months after the pandemic), and then July 2004 to July 2010 (12 months after the pandemic). Fig. 10b shows that immediately after the pandemic, the maximum predictivity declined and then bounced back after six months when the pandemic stopped. Two exceptions are notable: at 12 months (possibly due to another external shock of protests in July 2004) and at 18 months after the pandemic, the maximum predictivity shows higher values than that prior to the pandemic. In short, the influence of the SARS pandemic to predictivity could be said to be in a short period and that the shock to the tourism demand forecasting performance is therefore short term.

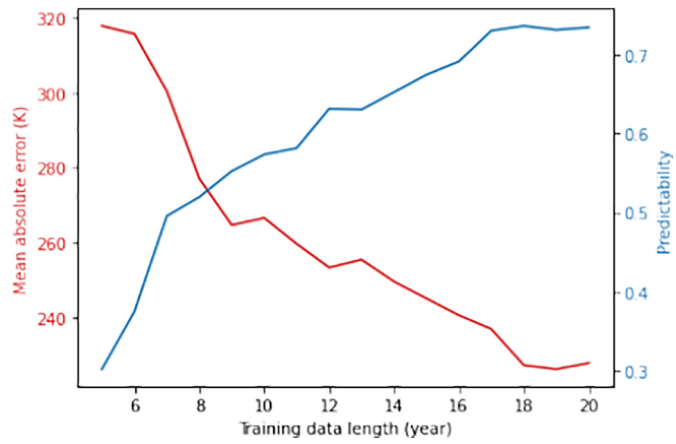


Fig. 11. Forecasting error (MAE) and Forecast maximum predictivity

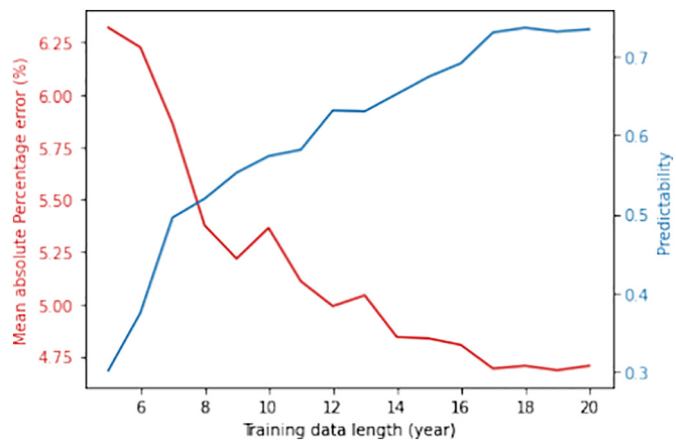


Fig. 12. Forecasting error (MAPE) and Forecast Maximum predictivity

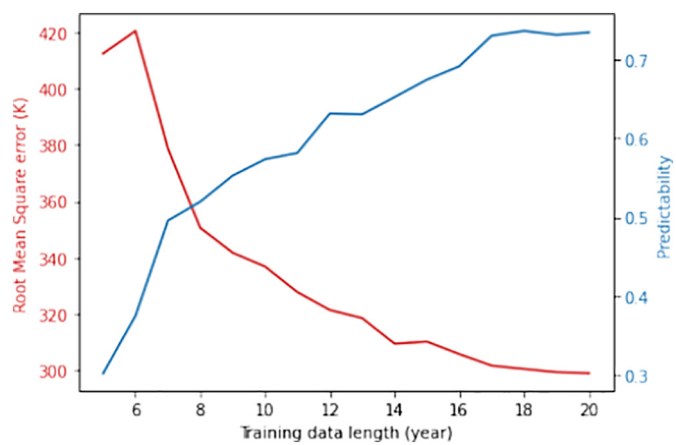


Fig. 13. Forecasting error (RMSE) and Forecast Maximum predictivity

Predictivity validation with forecasting model

Prior sections have demonstrated the influence of various data characteristics on the theoretical predictivity of Hong Kong tourism demand data. This section validates such experimental findings in the actual forecasting tasks. The SARIMA model (Asrin et al., 2015) is employed to make forecasting on the time series data with varied lengths, granularities, and decompositions.

Predictivity validation with varied data lengths

To validate the relationship between the maximum predictivity and the data length, we must first sample historical tourism demand data with varied lengths for model construction. The constructed models are then used to forecast tourism demand in the future time steps. For a fair comparison, the models should be evaluated on the same future time steps, as such, the period from May 2018 to Apr 2019 (12 months) is used for evaluation purpose. The prior time series data with varied lengths, from five to 20 years, are used for model training purpose.

SARIMA has four main parameters: p as the order for auto-regression; q as the order for moving average; d as the degree for difference which is used for obtaining the stationary time series; and m for the period cycle for seasonality pattern. We set $m = 12$, which means that the seasonality is at 12 months. For parameters, we run the grid search with the range of $p = [0,1,2,3]$, $q = [0,1,2,3]$, and $d = [0,1,2,3]$ to identify the best parameter setting, which is then tested with the lowest Akaike Information Criterion (AIC) (Sakamoto et al., 1986). SARIMA is first trained on the sampled time series data and then made prediction on one future time step. The prediction of the next time step is based on the historical data in the prior steps.

After the SARIMA forecasting, we calculate the maximum predictivity on the basis of forecasting results with the same parameters in the prior sections. Namely, $r = 0.55 * \sigma$ is used to compute the sample entropy (data length). If the *mean absolute forecasting error* (MAE) is below $0.55 * \sigma$, then, one successful forecast is counted. With the period of 12 months from May 2018 to April 2019, the success forecasting rate is regarded as the empirical maximum predictivity.

Fig. 11 shows that as the data length increases, the MAE gradually decreases. When the data length reaches a certain level (e.g., 15 years), the improvement on MAE slows down. MAE does not improve further when the data length reaches 17 years. A similar pattern is observed with the empirical maximum predictivity. The experimental results on empirical maximum predictivity are consistent with the analysis on maximum predictivity with varied data lengths in the previous section. Notably, the empirical maximum predictivity is obtained by the experiment results on the actual forecasting models, which are different from the theoretical maximum predictivity. Same patterns on the forecast maximum predictivity could be observed from MAPE and RMSE in Figs. 12 and 13. In general, the forecasting error decreases when the forecast maximum predictivity increases by adding the training data length.

Furthermore, the forecasting on the future period from May 2018 to April 2019 notably reaches the best performance, when training data length is of 17 years. We also find that the current empirical maximum predictivity calculated on the basis of SARIMA is (approximately 0.74 in Fig. 13) largely lower than the theoretical maximum predictivity (approximately 0.96 in Fig. 5b), which implies that further improvement is possible by using other sophisticated forecasting models.

Predictivity validation with varied data granularities

This section presents the experiment to validate the data predictivity with varied granularities based on empirical maximum predictivity. To prepare the data sample for different data granularities, 20-year data from Jan 1999 to Dec 2018 are used and then scaled into different granularities. The testing data are set as the latest 12 months from the 20-year data. SARIMA is trained and then used to forecast future time steps, whose results are shown in Table 1. Apparently, increasing the data scales causes a decrease in the forecasting maximum predictivity. This finding is consistent with the findings on previous analysis with theoretical maximum predictivity on data granularity.

Predictivity validation with varied decompositions

We carried out the experiment with varied decompositions using the same 20-year data sample, and the forecasting is made on the latest 12 months. STL and the SSA decomposition techniques are utilized, which produce decomposed global trend series and reconstruction, respectively. The empirical maximum predictivity based on the forecasting results was calculated based on the same setting of distance $r = 0.55 * \sigma$. The results in Table 2 show that the empirical maximum predictivity is higher on decomposed series than the original time series (last column). In general, the results from the validation experiment are consistent with the findings on the theoretical maximum predictivity, that is, the decomposed series will lead to higher forecasting performance than the original series. In addition, we tested for significant differences in forecast accuracy using the Diebold Mariano (DM) test (Harvey et al., 1997). The DM test result suggests that SSA and STL decomposition improvement is significant comparing to original data on the forecasting results as well as the predictivity results. Interestingly, SSA decomposition results in 0.94 empirical maximum predictivity, which is close to the results on theoretical maximum predictivity (Fig. 8b). Therefore, SSA decomposition could make the empirical tourism demand forecasting performance close to the theoretical best performance.

Discussion and implications

The result in the presented case study confirms that time series data of greater length produce better maximum predictivity and benefit the forecasting performance. However, an upper bound of the maximum predictivity occurs when the data of certain length are considered. In addition, the data with high granularities (small scales) produce high maximum predictivity. The

Table 1
Validation forecast on data granularity

Granularity level	MAE	Empirical maximum predictivity
1-Month level	265,000	0.870
2-Month level	287,100	0.856
3-Month level	286,010	0.862
4-Month level	320,020	0.801
5-Month level	356,190	0.789
6-Month level	410,931	0.732

Table 2
Validation forecast on data decomposition and DM test

Measurements	STL	SSA	Original
MAE	225,032	203,827	250,112
Empirical Maximum Predictivity	0.929	0.940	0.882
DM tests	STL vs Original	SSA vs Original	SSA vs STL
MAE	-2.978	-3.652	-2.003*
Empirical Maximum Predictivity	-2.573	-2.890	-2.363

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

decomposition on time series data also offers better maximum predictivity than the original data. The experiment with STL and SSA decompositions suggests that when the data could become less noisy, the better maximum predictivity can be achieved. This result directly explains why the decomposition method could efficiently improve the forecasting accuracy in many studies (Silva et al., 2019).

The practical implications of this work are manifold. *First*, according to the results, the predictivity and its relationships with data characteristics could have significant implications for the predictivity study on tourism demand forecasting. Second, the research question to obtain the best possible forecasting performance remains open in the tourism demand forecasting literature. This work is the first to formally define the measurement for finding the maximum predictivity of available data. In addition, maximum predictivity could provide a performance evaluation on forecasting models for tourism practitioners. The proposed work offers practical guidance to explore the influence of data characteristics on maximum predictivity. Practical questions such as “*how long is the ideal data length in tourism demand forecasting?*” and “*which decomposition methods are most effective?*” could be thoroughly answered by our proposed approach.

On the one hand, the study of predictivity over data length could largely help define the proper training data length, which has always been a challenge for tourism practitioners in performing tourism demand forecasting. Prior predictivity studies on data length indicate that sufficient data length used in tourism demand forecasting should be guaranteed, but they did not define an upper bound or what should be an optimal data length for the best possible performance. This work goes beyond this issue and provides guidance on whether more data are needed or whether the model requires further improvement.

On the other hand, our approach can evaluate the effectiveness of using different data granularity and decomposition methods. In traditional tourism demand forecasting, the granularity pattern on tourism demand data is essential for forecasting performance (Goh & Law, 2002), such as weekly and monthly patterns. In prior studies, these granularity patterns were directly adopted in many models, such as *SARIMA*, *Prophet*, but without providing evaluation or proper justification (Yang et al., 2013). Our work addresses this issue by evaluating the contribution of those patterns regarding forecasting performance. As such, the most effective data characteristics could be designed and brought into the forecasting models to achieve maximum predictivity. Similarly, the decomposition on tourism demand forecasting is highly regarded in existing works (X. Li & Law, 2020; Silva et al., 2019; Y. Zhang, Li, Muskat, Law, 2020). Within many decomposition methods (X. Li & Law, 2020; Silva et al., 2019; Y. Zhang, Li, Muskat, Law, 2020), tourism practitioners are required to choose the optimal one for the best possible performances. The advantage of our approach is its capability to evaluate the effectiveness of various decomposition methods by comparing their resulting maximum predictivity. Therefore, the introduced approach has extraordinary potential to support practitioners in tourism demand forecasting.

Conclusions

This paper formally defines the predictivity of tourism demand data and to develop the knowledge on how different data characteristics influence the best possible performance of tourism demand forecasting models. In this work, we define predictivity as the possible forecasting performance and proposed a theoretical method for evaluating predictivity for tourism demand data. Sample and multi-scale entropies are exploited with *Fano's Inequity* to calculate the maximum predictivity, which represents the expected forecasting performance based on tourism arrival volume data. The proposed method is evaluated in the experiment

with Hong Kong tourism demand data. We found that 95% maximum predictivity could be achieved by using a 10-year training data on $r = 0.55 * \sigma$ distance value. Further analysis revealed that the monthly tourism demand data are more appropriate than the quarterly and yearly data to achieve the best possible forecasting performance. The decomposed tourism demand data could provide better possible forecasting performance than original tourism demand data.

In summary, this research contributes to the tourism demand forecasting literature by defining the measurement on the theoretically best possible forecasting performance—the maximum predictivity. Moreover, the introduced approach enables the tourism practitioners to answer many practical questions regarding the design of the optimal data characteristics for forecasting models. This paper complements existing studies with an approach on how to evaluate efficiency factors of three data characteristics: data length, data granularity, and data decomposition.

Our work is not without limitation. The maximum predictivity we define in this work is only related to the univariate time series data. However, many recent tourism demand forecasting methods are based on the multivariate time series, such as search intense indicators and econometric determinants. To facilitate the further research along this line, the package “PREDICTIVITY” for calculating the predictivity and the Hong Kong tourism demand data have been released for public access at <https://github.com/tulip-lab/open-code>. Future works could be conducted on how to measure the multi-variate tourism demand data which are widely used in many existing tourism demand forecasting methods.

Declaration of competing interest

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

Acknowledgement

The work was completed when Gang Li was on ASL in Chinese Academy of Sciences in 2019, and we would like to thank Deakin University's ASL2019 fund and Xinjiang Science & Technology Research fund with Chinese Academy of Sciences.

References

- Aggarwal, C. (2013). Outlier ensembles: Position paper. *ACM SIGKDD Explorations Newsletter*, 14(2), 49–58.
- Asrin, K., Pouya, F., & Khalid, A. R. (2015). Modeling and forecasting of international tourism demand in ASEAN countries. *American Journal of Applied Sciences*, 12(7), 479–486.
- Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454–464.
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14(2), 1–10.
- Cai, Z.-J., Lu, S., & Zhang, X.-B. (2009). Tourism demand forecasting by support vector regression and genetic algorithm. In 2009 2nd IEEE International Conference on Computer Science and Information Technology, pp. 144–146, Beijing, China.
- Cankurt, S. (2016). Tourism demand forecasting using ensembles of regression trees. *2016 IEEE 8th International Conference on Intelligent Systems (IS)* (pp. 702–708) (Sofia, Bulgaria).
- Chan, F., Lim, C., & McAleer, M. (2005). Modelling multivariate international tourism demand and volatility. *Tourism Management*, 26(3), 459–471.
- Chen, W., Gao, W., & Xiong, H. (2016). Temporal predictability of online behavior in foursquare. *Entropy*, 18(8) (Article 296).
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics*, 6(1), 3–33.
- Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, 89(6), Article 068102.
- Dugdale, J. S. (2018). *Entropy and its physical meaning* (2nd ed.). London: Taylor & Francis.
- Ghodsii, M., Hassani, H., Rahmani, D., & Silva, E. S. (2018). Vector and recurrent singular spectrum analysis: Which is better at forecasting? *Journal of Applied Statistics*, 45(10), 1872–1899.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23(5), 499–510.
- Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, 46, 123–135.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.
- 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.
- Hung, L. S. (2003). The SARS epidemic in Hong Kong: What lessons have we learned? *Journal of the Royal Society of Medicine*, 96(8), 374–378.
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375.
- Kamel, N., Atiya, A. F., El-Gayar, N., & El-Shishiny, H. (2008). Tourism demand forecasting using machine learning methods. *ICGST International Journal on Artificial Intelligence and Machine Learning*, 8, 1–7.
- Khalid, U., Okafor, L. E., & Shafiullah, M. (2020). The effects of economic and financial crises on international tourist flows: A cross-country analysis. *Journal of Travel Research*, 59(2), 315–334.
- 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.
- Li, H., Hu, M., & Li, G. (2020). Forecasting tourism demand with multisource big data. *Annals of Tourism Research*, 83, Article 102912.
- Li, X., & Law, R. (2020). Forecasting tourism demand with decomposed search cycles. *Journal of Travel Research*, 59(1), 52–68.
- Lim, C., McAleer, M., & Min, J. C. H. (2009). ARMAX Modelling of international tourism demand. *Mathematics and Computers in Simulation*, 79(9), 2879–2888.
- Molgedey, L., & Ebeling, W. (2000). Local order, entropy and predictability of financial time series. *European Physical Journal B-Condensed Matter and Complex Systems*, 15(4), 733–737.
- Ott, H. C., Heilmayer, S., & Yee, C. S. S. (2013). Granularity dependency of forecast accuracy in semiconductor industry. *Research in Logistics & Production*, 3(1), 49–58.
- Peng, B., Song, H., & Crouch, G. I. (2014). A Meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45, 181–193.
- Rényi, A. (1961). On measures of entropy and information. *Berkeley Symposium on Mathematical Statistics and Probability*, 4(1), 547–561.
- Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6), H2039–H2049.

- Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 81(10.5555), 26853.
- Salisu, A. A., Isah, K., & Akanni, L. O. (2019). Improving the predictability of stock returns with Bitcoin prices. *The North American Journal of Economics and Finance*, 48, 857–867.
- 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, C., Qu, Z., Blumm, N., & Barabási, A. -L. (2010). Limits of predictability in human mobility. *Science*, 327(5968), 1018–1021.
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, 29(2), 203–220.
- Song, H., Li, G., Witt, S. F., & Fei, B. (2010). Tourism demand Modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1), 63–81.
- Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research*, 75, 338–362.
- Teixeira, J. P., & Fernandes, P. O. (2014). Tourism time series forecast with artificial neural networks. *Tékhné*, 12(1–2), 26–36.
- Udhayakumar, R. K., Karmakar, C., Li, P., & Palaniswami, M. (2016). Effect of embedding dimension on complexity measures in identifying Arrhythmia. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6230–6233) (Lake Buena Vista (Orlando), Florida USA).
- Valiant, L. (2013). *Probably approximately correct: Nature's algorithms for learning and prospering in a complex world*.
- Verdú, S. (1994). Generalizing the Fano inequality. *IEEE Transactions on Information Theory*, 40(4), 1247–1251.
- Wang, X., Smith-Miles, K., & Hyndman, R. (2009). Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. *Neurocomputing*, 72(10–12), 2581–2594.
- Wen, L., Liu, C., Song, H., & Liu, H. (2021). Forecasting tourism demand with an improved mixed data sampling model. *Journal of Travel Research*, 60(2), 336–353.
- Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection. *Journal of Travel Research*, 42(2), 151–158.
- Witt, S. F., & Witt, C. A. (1992). *Modeling and forecasting demand in tourism*. London, England: Academic Press.
- Wong, K., Song, H., & Chon, K. (2006). Bayesian models for tourism demand forecasting. *Tourism Management*, 27(5), 773–780.
- Xie, G., Qian, Y., & Wang, S. (2020). A decomposition-ensemble approach for tourism forecasting. *Annals of Tourism Research*, 81, Article 102891.
- Xu, T., Xu, X., Hu, Y., & Li, X. (2017). An entropy-based approach for evaluating travel time predictability based on vehicle trajectory data. *Entropy*, 19(4), 165.
- Xu, Y., Niu, W., Tang, H., Li, G., Zhao, Z., & Ci, S. (2012). A policy-based web service redundancy detection in wireless sensor network. *Journal of Network and Systems Management*, 21(3), 384–407.
- Yang, Y., Pan, B., & Song, H. (2013). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433–447.
- 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.
- Zhang, G. P. (2003). Time series forecasting using a hybrid Arima and neural network model. *Neurocomputing*, 50, 159–175.
- Zhang, Y., Li, G., Muskat, B., & Law, R. (2020). Tourism demand forecasting: A decomposed deep learning approach. *Journal of Travel Research*. <https://doi.org/10.1177/0047287520919522>.
- Zhang, Y., Li, G., Muskat, B., Law, R., & Yang, Y. (2020). Group pooling for deep tourism demand forecasting. *Annals of Tourism Research*, 82, Article 102899.
- Zhao, X., Liang, C., Zhang, N., & Shang, P. (2019). Quantifying the multiscale predictability of financial time series by an information-theoretic approach. *Entropy*, 21(7), 684.

Yishuo Zhang, zhangyis@deakin.edu.au, research student. His research interests are tourism and hospitality management, and data science.

Gang Li, gang.li@deakin.edu.au, Ph.D. His research interests are technology applications to tourism and hospitality, data science, and artificial intelligence.

Birgit Muskat, birgit.muskat@anu.edu.au, Ph.D. Her research interests are tourism management, entrepreneurship, and innovation.

Huy Quan Vu, q.vu@deakin.edu.au, Ph.D. His research interests are data mining applications in tourism marketing and management.

Rob Law, rob.law@polyu.edu.hk, Ph.D. His research interests are technology applications to tourism and information management.