Journal of Revenue and Pricing Management (2020) 19:8–25
https://doi.org/10.1057/s41272-019-00221-6

RESEARCH ARTICLE

Hotel daily demand forecasting for high-frequency and complex seasonality data: a case study in Thailand

Naragain Phumchusri[1] · **Phoom Ungtrakul**[1]

Received: 10 January 2019 / Accepted: 8 November 2019 / Published online: 30 November 2019

Abstract

Accurate hotel daily demand forecasting is an important input for hotel revenue management. This research presents forecasting models, both time series and causal methods, for a case study 4-star hotel in Phuket, Thailand. Holt–Winters, Box–Jenkins, Box–Cox transformation, ARMA errors, trend and multiple seasonal patterns (BATS), trigonometric BATS (TBATS), artificial neural network (ANN), and support vector regression are explored. For causal method, independent variables used as regressor inputs are transformed data observed in the past periods, the number of tourist arrivals from main countries to Phuket, Oil prices, exchange rate, etc. Model accuracy is measured using mean absolute percentage error (MAPE) and mean absolute error. Findings suggested that ANN outperforms other models with the lowest MAPE of 8.96%. It shows that Machine Learning techniques studied in this research outperform the advanced time series methods designed for complex seasonality data like BATS and TBATS. Unlike previous works, this research is a pioneer to introduce data transformation as inputs for machine learning models and to compare time series method and machine learning method for hotel daily demand forecasting. The results obtained can be applied to the case study hotel's future planning about the forecasted number of left-over rooms so that they effectively allocate to their discounted online travel agent more effectively.

Keywords Hotel revenue management · Hotel demand forecasting · Complex seasonality · Hotel industry · Machine learning

Introduction

Revenue Management (RM) is essential to hotel business, leading to profit maximization and customer satisfaction. In RM, forecasting is considered to be a key input for management decisions. It is beneficial for management team to have advance knowledge about hotel room demand because there is time to implement strategies to correct the situation. On the other hand, if the time of such insight is only few days in advance, nothing much but inventory control could be implemented. Revenue management system recalculates the approaches. Typically, hotel revenue management update forecasts on a daily basis for occupancy rates in the near future (1–2 weeks) and update on a weekly basis for dates farther away (2–8 weeks) (Weatherford and Kimes 2003; Kouprioucha et al. 2014).

Most of hotels nowadays collect customer information: e.g., spending at different rate category, etc., for marketing purposes. For example, big chain of hotels like Marriot and Hyatt keeps customer profile according to the amount of money spent on amenity, food, and meeting room usage (Baker and Collier 1999). Hotel management collects arrivals data by rate category and length of stay. It is suggested that if disaggregation on both rate category and length of stay are not possible, the hotel should at least collect data by rate category or the type of room (Weatherford et al. 2001).

Phuket, an island province situated in the southern region of Thailand, in Andaman sea, dubbed 'Pearl of Andaman' has served the country as the finest tourist destination for many years. In 2005, people voted Phuket as one of the world's top five travel destinations. And, in 2017, 10 million foreign visitors traveled to Phuket and generated an estimation of 385 billion baht accounting for nearly 14 percent of the year's total GDP (Office of the National Economic 2019). Figure 1 shows the number of total tourist arrivals in

✉ Naragain Phumchusri
naragain.p@chula.ac.th

[1] Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand

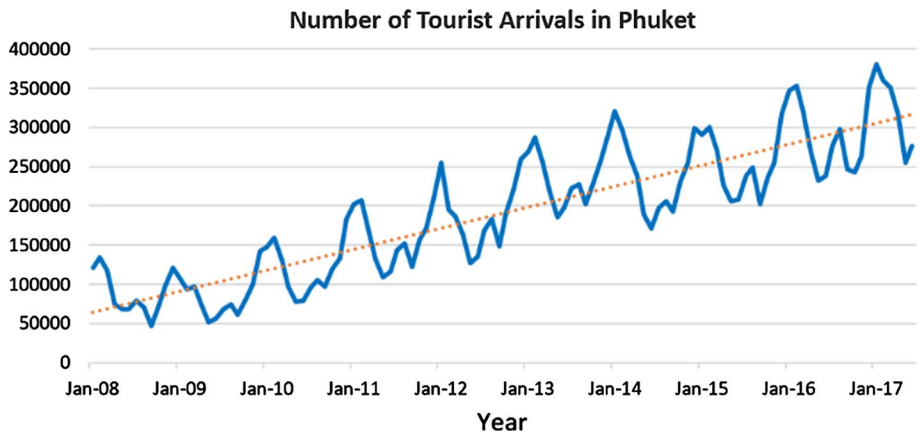

Fig. 1 The number of tourist arrivals in Phuket from 2008 to 2017

Phuket. The graph reveals strong upwards trend and seasonality, expressing long-term growth of the number of tourists. In consequence, there are opportunities for hotel business growth. At the same time, the competition in this industry is also likely to be higher.

The case study hotel is operated in Phuket, Thailand, having 152 rooms in total, categorized into 4 types: Deluxe; Seaview; Pool Access; and Villa. Deluxe room type has 48 rooms, taking into account 32% of the total. Seaview type is the largest room type which has 75 rooms, counting into 49% of the total. Pool Access type has 15 rooms and Villa room type has 14 rooms, counting for 10% and 9%, respectively.

In most case, hotel daily time series comprise high-frequency and complex seasonal properties that period of seasons can also be non-integer. Daily data constitute high frequency due to high number of days in one period (365 days in a year). Besides, daily data could propagate complex seasonal pattern, having weekly seasonal pattern with a period of 7 and an annual seasonal pattern with a period of 365.25. In contrast to monthly, quarterly, and yearly data that had smaller period within single seasonal pattern, such period can only be integer (4 for quarterly data or 12 for monthly data).

Figure 2 shows Time series plot of daily demand data of the case study hotel from year 2015 to 2018 and Fig. 3 presents those data for 2018. At the beginning, graph value ranges between 125 and 150, affirming the existence of High season. The rest of the graph scatters, showing the high frequency characteristic of daily data. Hence, this exhibits high-frequency and seasonal properties in daily demand data of this hotel. So, it is interesting to explore the appropriate forecasting methods, both time series and causal models, to accurately forecast this type of data for the case study hotel. Specifically, this paper aims to introduce data transformation as inputs for machine learning forecasting models and to compare them with time series

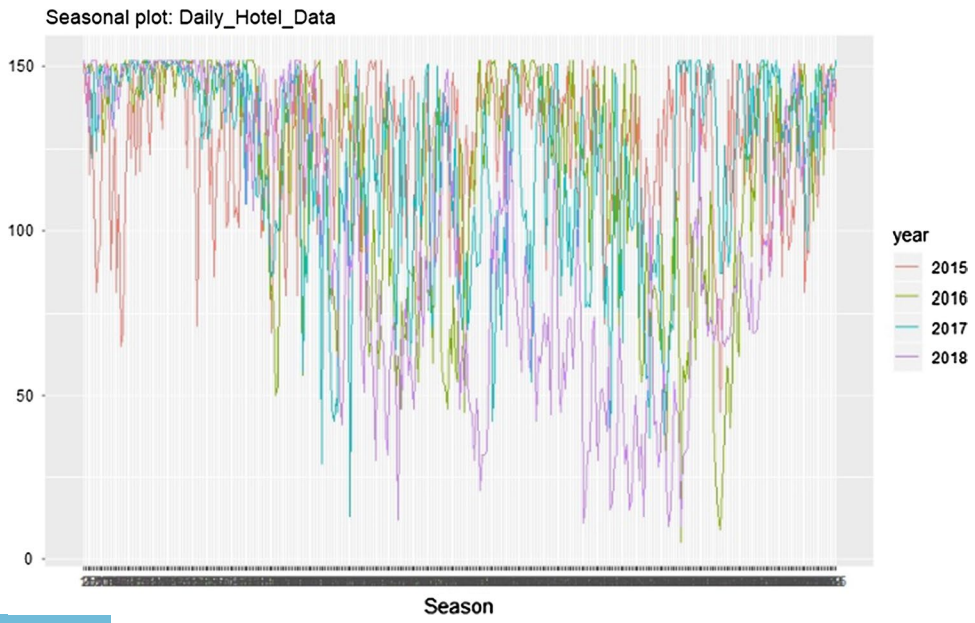

Fig. 2 Time series plot of daily demand data of hotel (2015–2018)

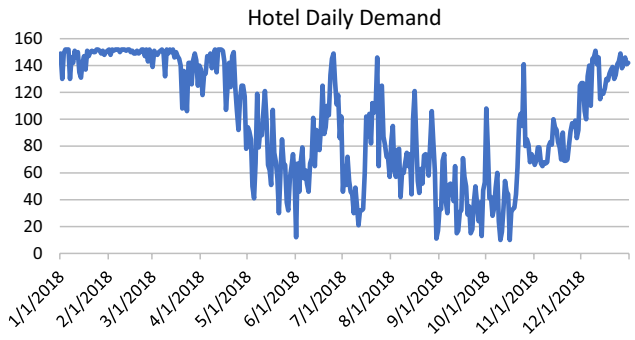

Fig. 3 Time series plot of daily demand data of hotel for year 2018

methods for hotel room occupancy with high-frequency and complex seasonality data, and to find insights on how to obtain the accurate forecast results.

It is more beneficial for the management team to have advance knowledge about hotel room occupancy because there is time to implement strategies to correct the situation. On the other hand, if the time of such insight is only few days in advance, nothing much but inventory control could be implemented (Kouprioucha et al. 2014). The result from this paper can provide the hotel management team with reliable forecasting method on daily hotel room demand both in total and in each category. Also, the insights can assist management team in comprehension of trend, seasonality pattern, and affecting variables that explain the data pattern.

The rest of the article is organized as follows. The next section summarizes literature review on hotel demand forecasting. "Research methodology" section introduces forecasting models used for forecasting daily demand for the case study hotel. "Results and discussion" section presents the results and discussion about model comparison and identify the most suitable model, followed by a discussion of the practical implications of the research outcomes. The final section concludes the results of this paper and presents future research directions.

Literature review

In revenue management, there are many issues to consider in forecasting. Rajopadhye et al. (2001) introduced forecasting technique on unconstrained room demand. This meant that cancelation of the booking is also reflected in this model. Thus, it was imperative to use actual booking activity that includes both completed and incomplete stay nights. Weekly data were collected from both occupancy and arrivals. Weatherford et al. (2001) explored forecast accuracy on four different levels of aggregation with 2-year daily arrival data by rate category and length of stay, reported that disaggregated forecast highly outperforms every aggregated forecast. Therefore, they suggested that hotel should gather arrivals data by rate category and length of stay. But, if disaggregation on both rate category and length of stay are not possible, the hotel should at least collect data by rate category or the type of room. Weatherford and Kimes (2003) discussed these issues in detail and presented the use of daily arrivals data to forecast. Arrivals data from Choice hotels were used to form a suitable forecasting method. Then, Arrivals data from Marriot hotels were used to verify the findings.

Chen and Kachani (2007) proposed optimization model by network flow formulation. Authors used both arrivals and occupancy data on forecasting side, employing, classical and advanced pickup method, simple exponential smoothing, and combined methods (pickup and exponential smoothing). Lim et al. (2009) used hotel–motel monthly occupancy in New Zealand from January 1997 to December 2006 to calculate 1-month ex post forecast for 12 periods. Then, this forecast was compared with occupancy data in year 2007 to quantify forecasting accuracy. Zakhary et al. (2011) introduced Monte Carlo simulation for approximation of arrivals by occupancy data to forecast hotel room demand in Egypt. El Gayar et al. (2011) proposed incorporation of group reservation bookings and integrality constraints into forecasting model using decomposition method to extract components such as reservations, cancelations, duration of stay, no shows, seasonality, trend, arrivals, occupancy rates, etc. Haensel and Koole (2011) studied forecasting model on combination of regions, day of week arrivals, and length of stays. With weekly data, authors used Additive Holt–Winters, and Multivariate Vector Autoregression to forecast accrued booking curve, plus expected number of total reservations in each day.

ARMA errors, trend and multiple seasonal patterns (BATS) and trigonometric BATS (TBATS) were proposed by De Livera (2010) and De Livera et al. (2011), applying these models to forecast weekly number of barrels of motor gasoline product supply in the United States from February 1991 to July 2005. This series had seasonality length of 52.179 with 745 observations in total, 484 observations for training period, and 261 for testing set. Later, Pereira (2016) applied these forecasting methods to forecast high-frequency daily occupancy data from 300-room Portuguese four-star hotel. He mentioned that daily time series are dissimilar to monthly, quarterly, or annual data because daily time series present high-frequency and complex seasonal patterns. In most case, daily time series comprised double seasonal effect that period of seasons can be non-integer. Usually, daily data had weekly seasonal pattern with a period of 7 and an annual seasonal pattern with a period of 365.25. Therefore, conventional time series forecasting methods, such as exponential smoothing or ARIMA/SARIMA, were not appropriate for forecasting high-frequency and complexed time series. Those conventional methods were modeled for smaller period within single seasonal pattern that such

period can only be integer (4 for quarterly data or 12 for monthly data).

For Machine Learning method in hotel forecasting, Urraca et al. (2015) optimized a knowledge discovery in databases (KDD) scheme using genetic algorithms. Booking data were obtained from a hotel located in a small village of La Rioja region in northern Spain and 6 years of data were used to train and validate the models, while a test set was created from the latest year during January and July. Also, 119 attributes of macro-economic indicators, temporal situation, social patterns, meteorological data, and local and regional holidays were selected by experts. Then, these indicators were decreased to 22 by detecting dependencies via scatterplots and matrix correlation. Forecasting accuracies were computed by root mean squared error (RMSE) and generalized degrees of freedom (GDF) was used for the selection of parsimonious models. Results which presented the three best performing models were those that included feature selection (FS).

Later, Martinez-de Pison et al. (2016) compared Grid search (GS) and Genetic Algorithm (GA) on nine regression models which are Model Tree based on Quinlan's M5 algorithm (M5P), instance-based learning (IBL) method, multilayer perceptron (MLP) neural network, support vector regression (SVR), extreme machine learning (ELM) algorithm, locally weighted learning for linear regression (LWLLIN), random forest (RF) algorithm, extreme gradient boosting (XGB) algorithm, and linear ridge regression (LIN) as the reference prediction. Forecasting performance of daily historical booking records of the Spanish hotel for 6 years (2004–2010) was observed. Regressors also included a list of local, regional, and national festivities, and other additional information such as the day of the week, the local weather conditions, some sociological information, several indicators comprising the macro-economic situation of the area, and independent factors affecting the hotel room demand provided by the Spanish National Institute of Statistics. Root mean squared error (RMSE) and mean squared error (MAE) were computed to compare forecasting accuracy, plus computation time for each regression model with Grid search and Genetic Algorithm was shown in comparisons.

Table 1 summarizes research in literature that are related to hotel demand forecasting and points out the main focus of this paper. It can be noticed that there were limited researches in forecasting daily hotel demand using machine learning. The most related research to this paper is Pereira (2016) and Martinez-de Pison (2016). Pereira (2016) focused on time series models, while Martinez-de-Pison (2016) focused on machine learning using GA model. However, it was deemed as heavy burdens for practitioners to forecast using numerous independent variables, since workloads should be impinged on practitioners to select and refine those variables. The existing researches about daily demand forecast focused only on accuracy comparison among time series methods and did not yet compare them with other methods. Therefore, this paper aims to propose hotel daily demand forecasting models using both advanced time series models and machine learning to investigate the use of transformation of time series data.

Research methodology

Research methodology is presented in this section. First, time series demand data are collected from the hotel management software. Second, data are analyzed and used to construct forecasting model. Third, forecasting results from all models are compared. Finally, interpretation and recommendation are provided for management use. The following sub-section provides details about data categories, data division, and proposed models in this research.

Data categories

For time series methods, the interested data to be forecasted are the daily demand, i.e., the number of occupied rooms (from booking history of case study hotel). The forecasting result will be summarized for both aggregate demand (total demand from all room types) and disaggregate demand (separately for each room type)

For casual method using machine learning:

- The dependent variable in this study is the hotel daily demand
- The independent variables are attributes of macro-economic indicators, temporal situation, the number of tourist Arrivals in Phuket, days of the week, month, season, weather conditions, and various forms of transformed data of dependent variable itself. Details of all variable will be explained in "Artificial neural network" section in detail.

Data division

The hotel daily demand data are divided into two groups. Figure 4 shows data separation for model constructing and testing.

Forecasting models

The following sub-section describes the concepts of all models used in this paper.

Table 1 Summary of literature review regarding hotel forecasting

Authors	Variables used	Data frequency	Forecasting model
Rajopadhye et al. (2001)	Unconstrained hotel room demand	Weekly	Holt–Winters method
Weatherford et al. (2001)	Fully aggregated: length of stay with rate category Disaggregated: rate category with length of stay	Daily	Classical-pickup method, moving averages, linear regression, simple exponential smoothing, and random walk
Weatherford and Kimes (2003)	Fully aggregated: by length of stay with rate category Fully disaggregated (by rate category with length of stay)	Weekly	Classical pickup, advanced pickup, multiplicative, and regression
Vu and Turner (2006)	Guest arrivals at accommodation establishments in 9 cities	Monthly	Box–Jenkins SARIMA, and BSM models (decomposition, trend, seasonal, cycle, irregular)
Chen and Kachani (2007)	Demand distribution for each night, for each rate, and for each length of stay	Weekly	Classical and advanced pickup, linear regression, simple exponential smoothing, advanced pickup exponential smoothing, advanced pickup + exponential smoothing
Lim et al. (2009)	Guest night demand	Monthly	Holt–Winters exponential smoothing, and Box–Jenkins ARMA Models
El Gayar et al. (2011)	Arrivals, reservation data, folio history, occupancy rates	Daily	Advanced room demand forecast model and an optimization model
Haensel and Koole (2011)	Combinations of region, arrival day of week, and length of stay	Weekly	Additive Holt–Winters, and the vector autoregressive forecasts
Pereira (2016)	Daily hotel room demand time series	Daily	Holt–Winters, BATS, and TBATS
Martinez-de Pison et al. (2016)	Daily historical booking records of the Spanish hotel and independent variables: list of festivities; day of the week; local weather conditions; sociological information; and local macro-economic indicators	Daily	Grid search (GS) and Genetic Algorithm (GA) on nine regression models
Urraca et al. (2015)	Booking data of La Rioja region in northern Spain and independent variables: macro-economic indicators; temporal situation; social patterns; meteorological data; and regional holidays	Monthly	KDD scheme with GA
This paper	Aggregated and disaggregated daily hotel room demand time series	Daily	Time series method: Naïve, moving average, Holt–Winters, SARIMA, BATS, TBATS Machine learning method: artificial neural network, and support vector regression

Exponential smoothing (Holt–Winters)

Holt (1957) and Winters (1960) incorporated seasonality fraction into Holt's method. The Holt–Winters seasonal method included one forecast equation, and three smoothing equations. These three smoothing equations are level (m_t), trend (b_t), and seasonal (c_t) with α, β, and γ as smoothing parameters. There are two types of Holt–Winters' exponential smoothing: additive and multiplicative. Holt–Winters' multiplicative exponential smoothing is selected for this case study hotel because amplitudes of the seasonal variations are dependent of the data level. Equations 1 to 5 show Holt–Winters' multiplicative exponential smoothing equations.

$$y_t = (m_t + b_t t)c_t, \tag{1}$$

$$\hat{m}_t = \alpha \frac{y_t}{\hat{c}_{t-s}} + (1-\alpha)\left(\hat{m}_{t-1} + \hat{b}_{t-1}\right), \tag{2}$$

$$\hat{b}_t = \beta\left(\hat{m}_t - \hat{m}_{t-1}\right) + (1-\beta)\hat{b}_{t-1}, \tag{3}$$

$$\hat{c}_t = \gamma \frac{y_t}{\hat{m}_t} + (1-\gamma)\hat{c}_{t-s}, \tag{4}$$

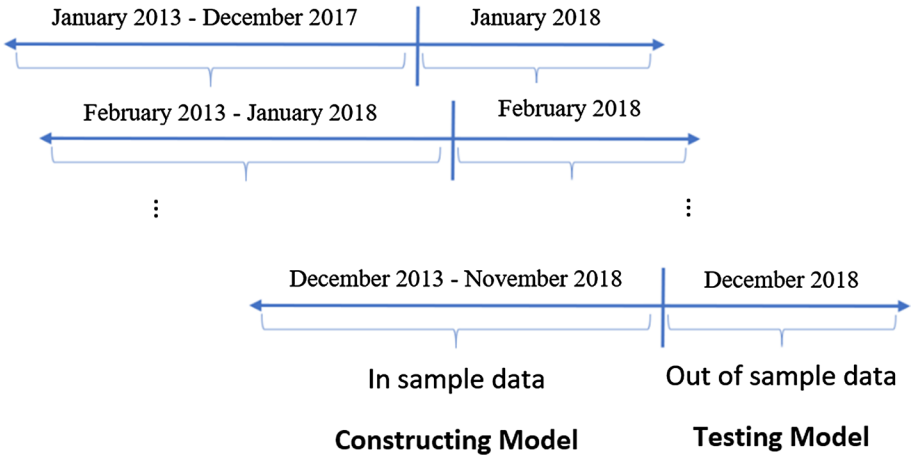

Fig. 4 Data separation for model constructing and testing of forecasting horizon model

$$\hat{y}_{t+r} = \left(\hat{m}_t - \hat{b}_t\tau\right)\hat{c}_{t+\tau-s}, \quad (5)$$

where m_t is mean component that gives the level of the time series at time t, b_t is trend component that indicates direction in which the series is evolving, c_t is the seasonal component that indicates the periodic variations in the level of the series, ϵ_t is random error, α, β, andγ are smoothing constants for the base, trend, and seasonal components respectively, $\hat{y}_{t+r}$ is forecast for future time τ.

SARIMA

Autoregressive Integrated Moving Average (ARIMA) forecasting model was proposed by Box and Jenkins (1970) to predict time series. The method started with identify whether the time series is stationary with autocorrelation function (ACF), and partial autocorrelation function (PACF). If the time series was not stationary by mean, differencing is executed to the time series, but if the time series was not stationary by variance, time series is transformed with square root or natural logarithm. Next, maximum likelihood estimation or non-linear least-squares estimation would be performed on the time series to estimate variables. Then, the predicted model would be tested for stationary univariate process, that is residuals should be constant over time and be independent of each other. This process was checked by Ljung–Box test or plotting autocorrelation and partial autocorrelation of the residuals.

SARIMA (Seasonal Autoregressive Integrated Moving Average) is an extended version of ARIMA, incorporating seasonal effect. A time series $\{X_t | t = 1, 2, \ldots, N\}$ is estimated by SARIMA(p,d,q)(P,D,Q) by the following equation (Box and Jenkins 1970).

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \quad (6)$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the regular autoregressive operator (AR) of order p, $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$ is the seasonal autoregressive operator (AR) of order P, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^q$ is the regular moving average operator (MA) of order q, $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_P B^{Qs}$ is the seasonal moving average operator (MA) of order Q, $\delta = \mu\phi_p(B)\Phi_P(B^s)$ is constant, in which μ is average of stationary time series, N is the number of observations. p, d, q, P, D, and Q are integers, s is the seasonal period length, B is Backward Operator, in which $B^s Y_t = Y_{t-s}$. ε_t is the estimated residual at time t that is identically and independently distributed as a normal random variable with a mean equal to zero and a constant variance.

BATS/TBATS

This original forecasting method was presented by De Livera (2010) and (De Livera et al. (2011). BATS model was an extended version of double-seasonal Holt–Winters, integrated with Box–Cox transformation to handle with non-linear data, and with ARMA model to account for autocorrelation in time series by residuals. De Livera (2010) showed that BATS model improves prediction accuracy compared to simple time series models. Nevertheless, BATS model still did not perform satisfactorily when seasonality is complex and with high frequency. Later, De Livera et al. (2011) proposed TBATS model to include trigonometric functions into the BATS model. TBATS model could reduce model parameters and was flexible for data with seasonality with high frequency. Thus, TBATS could deploy data with non-integer seasonal period, non-nested periods, and high-frequency data.

First, double-seasonal Holt–Winters (DSHW) exponential smoothing equation with additive trend and additive seasonality is shown below. This model (Eqs. 7 to 11), developed by Taylor (2003), was an extension of the Holt–Winters exponential smoothing.

$$l_t = \alpha\left(y_t - s_{t-m_1}^{(1)} - s_{t-m_2}^{(2)}\right) + (1-\alpha)\left(l_{t-1} + b_{t-1}\right), \tag{7}$$

$$b_t = \beta\left(l_t - l_{t-1}\right) + (1-\beta)b_{t-1}, \tag{8}$$

$$s_t^{(1)} = \gamma\left(y_t - l_t - s_{t-m2}^{(2)}\right) + (1-\gamma)s_{t-m_1}^{(1)}, \tag{9}$$

$$s_t^{(2)} = \delta\left(y_t - l_t - s_{t-m_1}^{(1)}\right) + (1-\delta)s_{t-m_2}^{(2)}, \tag{10}$$

$$\begin{aligned}\hat{y}_t(h) = {} & l_t + hb_t + s_{t-m_1+h}^{(1)} + s_{t-m_2+h}^{(2)} \\ & + \phi^h\left[y_t - \left(l_{t-1} + b_{t-1} + s_{t-m_1}^{(1)} + s_{t-m_2}^{(2)}\right)\right],\end{aligned} \tag{11}$$

where l_t and b_t are the smoothed level and trend in period t , respectively, $s_t^{(1)}$ is the seasonal component for the short cycle, $s_t^{(2)}$ t is the seasonal component for the long seasonal cycle, m_1, m_2 are the lengths of the shorter and longer seasonal cycles, respectively, $\hat{y}_t(h)$ is the h step-ahead forecast made from forecast origin t, $\phi^h\left[y_t - \left(l_{t-1} + b_{t-1} + s_{t-m_1}^{(1)} + s_{t-m_2}^{(2)}\right)\right]$ is a simple adjustment for first-order autocorrelation, $\alpha, \beta, \gamma,$ andδ are smoothing parameters.

The following equations show the extension of double-seasonal Holt–Winters (DSHW), called Box–Cox transformation, ARMA errors, trend, and multiple seasonal patterns (BATS). These are expressed by Eqs. 12 to 17 (De Livera 2010).

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}, & \omega \neq 0 \\ \log y_t, & \omega = 0 \end{cases} \tag{12}$$

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + \mathrm{d}_t, \tag{13}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \mathrm{d}_t, \tag{14}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta \mathrm{d}_t, \tag{15}$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i \mathrm{d}_t, \tag{16}$$

$$\mathrm{d}_t = \sum_{i=1}^{p} \varphi d_{t-i} + \sum_{i=1}^{p} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{17}$$

where $m_1, \ldots, m_T$ are the periods of T seasonal patterns, l_t is the local level in period t, b is the long-run trend, b_t is the short-run trend in period t, $s_t^{(i)}$ is the ith seasonal component at time $t (t = 1, \ldots, n; \quad i = 1, \ldots, T)$, d_t is ARMA(p, q) process, ε_t is Gaussian white-noise error term with zero mean and constant variance, ϕ is the damping constant of the trend, $\alpha, \beta,$ andγ_i are smoothing parameters.

The following equations show the extension of BATS model by adapting Eqs. (12) to (17) with the following expressions. This adaptation is called TBATS model (Eqs. 18 to 21) (De Livera et al. 2011).

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + \mathrm{d}_t, \tag{18}$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}, \tag{19}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} \mathrm{d}_t, \tag{20}$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} \mathrm{d}_t, \tag{21}$$

where k_i is the number of harmonics required for the ith seasonal component, $\gamma_1^{(i)}, \gamma_2^{(i)}$ are smoothing parameters, and $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$.

BATS and TBATS forecasting model were estimated using these following steps (De Livera et al. 2011).

- Step 1: Specification of all available model combinations which are to be considered for each series

In the BATS modeling framework, a total of 24 models are available for consideration of each series. This framework consists of 16 model combinations considering each B, A, T, S component and 8 additional models considering a damped trend component. For example, $\omega = 1$ is considered as having no Box–Cox transformation, $\phi = 1$ as having no damping component, p = q = 0 as having no ARMA residual adjustment in the model.

TBATS model formulation is relatively straightforward, the seed states of state space models are treated as random vectors. Since trial values of the unknown parameters, the joint steady-state distributions of stationary states are derived, and then assigned to associated seed states.

- Step 2: Estimation of the models

The initial states x_0, the smoothing parameters, the Box–Cox parameter, the damping parameter, and the coefficients for the ARMA components are estimated using an appropriate

estimation criterion. Three different estimation criteria are considered for non-linear optimization as follows:

(1) Maximize the log likelihood of the estimates (MLE)
(2) Minimize the root mean square error of the original data (RMSE)
(3) Minimize the root mean square error of the transformed data ($RMSE_T$)

- Step 3: Selection of the best of the available models

Akaike information criterion (AIC) is used to compare result among models. The following ARMA fitting approaches are explored.

(1) Setting $\{p = 0, q = 0\}$ assumed that an ARMA residual adjustment is not necessary.
(2) Finding the values for p and q in all possible ARMA combinations up to $p = q = 5$ were considered, and the ARMA(p,q) combination which minimizes the AIC was chosen, or retrieved the values for p and q in a stepwise procedure.

For TBATS model, number of harmonics was selected by constantly adding harmonics, testing the significance of each one using F tests.

- Step 4: Generation of prediction distributions using the best model.

Next, we describe machine learning model considered in this research.

Artificial neural network

The first machine learning model explored in this paper is the artificial neural network (ANN), inspired by the way the human brain works. A human brain can process huge amounts of information using data sent by human senses (especially vision). The processing is done by neurons, working on electrical signals passing through them and applying flip-flop logic, like opening and closing of the gates for signal to transmit through. Those significant variables are then used as inputs for ANN, the non-linear statistical machine learning model to find complex relationships or patterns in data between inputs and outputs. We apply recurrent networks for ANN in this paper (Claveria et al. 2015). The important question is what are factors that should be included as predictor variables for ANN model. Table 2 summarizes all variables selected for backward elimination regression process.

From the literature, Thomason (1999) suggested that forecasting horizon for daily stock price should be long enough to compensate the over-trading resulting in excessive transaction costs, but in forecasting aspect, forecasting horizon should be short because information hidden in financial time series existed in limited duration. Therefore, he recommended that regressors for daily stock price should be converted into four lagged relative difference in percentage of price (RDP), that these transformed data will become more symmetrical and will follow more closely to a normal distribution. To illustrate, RDP-5 is calculated by $\mathrm{RDP}_5 = \frac{A_t - A_{t-5}}{A_{t-5}} * 100$, where A_t is the actual observation at time t. Later, Cao and Tay (2001) applied Support Vector Machine to daily stock price forecasts in comparison with backpropagation (BP) neural network and the regularized radial basis function (RBF) neural network. Standard & Poor 500 stock index futures (CME-SP) were used in their research. Five-day lagged periods (RDP-5, RDP-10, RDP-15, and RDP-20) and one transformed closing price (EMA100) were calculated as regressor variables based on relative difference in percentage of price (RDP) and exponential moving average (EMA).

Thus, in addition to external factors like attributes of macro-economic indicators, temporal situation, the number of tourist arrivals, this paper embedded those idea of data transformation to be some input variables, which has not yet been explored in forecasting hotel daily demand literature.

ANN model comprises input nodes, hidden layers, output nodes, and transfer function. Values of input nodes and output nodes are set with the acquired data. Transfer functions and weights on nodes connection are tested to find the lowest model error. One of the problems for neural network is how to set up hidden layer nodes structure whether Grid Search (GS) is implemented, or to use the better of Genetic Algorithm (GA) (Martinez-de Pison et al. 2016). Figure 5 shows example of neural network structure, containing six input variables, one hidden layer with eight nodes, and four output variables.

Support vector regression

SVR is one of the machine learning models that can be used for both classification and regression problems. SVR can manipulate both linear and non-linear problems by adjusting kernel function and can also operate unsupervised learning approach with unlabeled data by the use of Vapnik's ε -insensitive loss function (Vapnik 1995). Given a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$ $(x_i \in X \subseteq R^n,\ y_i \in Y \subseteq R$, l is the total number of training samples randomly and independently generated from an unknown function, SVR approximates the function using the following form:

$$f(x) = w \cdot \phi(x) + b, \tag{22}$$

Table 2 Independent variables used in casual method

Category	Sub-category	Data	Remarks
Number of tourist arrivals in phuket by country of residence	China	Monthly	Comprises 80% of total number of tourist arrivals
	Russia		
	Australia		
	Korea		
	Malaysia		
	Singapore		
	Sweden		
	United Kingdom		
Customer Price Index of Southern Region of Thailand (CPISouth)		Monthly	Bureau of Trade and Economic Indices
Oil price		Daily	Crude Oil WTI Futures (CLM9) in USD
Exchange rates over thai baht	United Stated (USD)	Daily	
	China (CHY)		
	Russia (RUB)		
	Australia (AUD)		
	Korea (KRW)		
	Malaysia (MYR)		
	Singapore (SGD)		
	Sweden (SEK)		
	United Kingdom (GBP)		
Average room rate (ARR)		Daily	
Days of the week	Monday	Dummy	Wednesday is the baseline
	Tuesday		
	Thursday		
	Friday		
	Saturday		
	Sunday		
Month of the year	January	Dummy	May is the baseline
	February		
	March		
	April		
	June		
	July		
	August		
	September		
	October		
	November		
	December		
Monsoon		Dummy	Monsoon season: March–October
Season	Peak	Dummy	Low season is the baseline
	High		Peak: December 20th to January 20th
			High: November 1st to March 15th
			Low: May 1st to October 31st
RusSanc		Dummy	Sanction on Russia: March 2014–present
Relative difference in percentage of price (RDP)	RDP_7	Daily	RDP transformed data: $RDP_i = \frac{A_t - A_{t-i}}{A_{t-i}} * 100$, where A_t is the actual observation at time t, i = 7, 20, 52, 365
	RDP_{30}		
	RDP_{52}		

Table 2 (continued)

Category	Sub-category	Data	Remarks
Moving average (MA_{360})	RDP_{365}	Daily	MA transformed data: $MA_{360} = y_t - \frac{y_{t-1}+\ldots+y_{t-360+1}}{360}$ Total: 46 variables

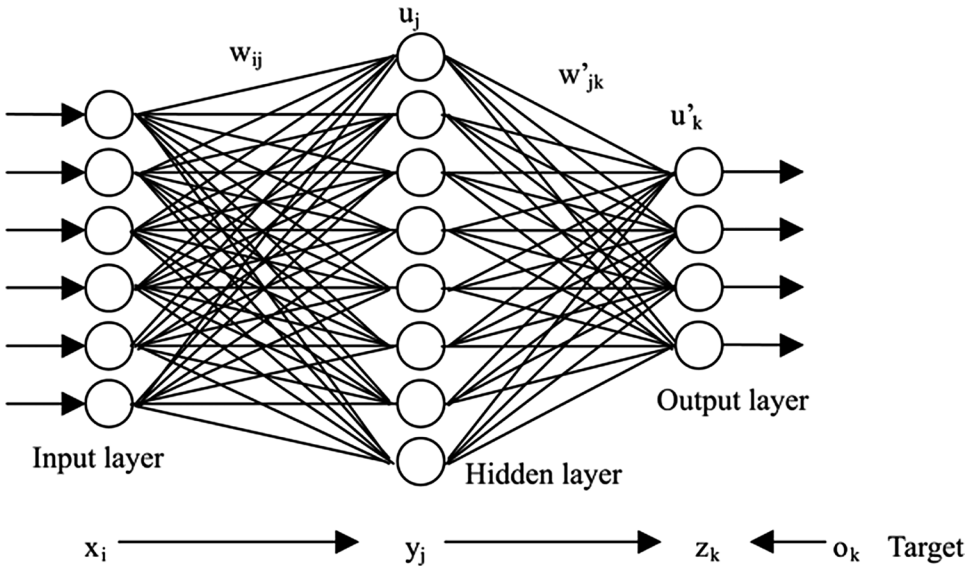

Fig. 5 Example structure of neural network model

where $\phi(x)$ represents the high-dimensional feature spaces which is non-linearly mapped from the input space x. The coefficients w and b are estimated by minimizing the regularized risk function shown in Eq. (23) (Vapnik 1995).

$$\text{Minimize} \quad \frac{1}{2}w^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(y_i, f(x_i)) \tag{23}$$

$$\text{Subjected to} \quad L_\varepsilon\left(y_i, f(x_i)\right) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

The first term (w^2) is called the regularized term. Minimizing w^2 will make a function as flat as possible, thus playing the role of controlling the function capacity. The second term $\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon\left(y_i, f\left(x_i\right)\right)$ in Eq. (24) is the empirical error measured by the ε-insensitive loss function. This loss function provides the advantage of using sparse data points to represent the designed function. C is referred to as the regularization constant, while ε is called the tube size. They are both user-prescribed parameters and determined empirically.

To get the estimations of w and b, Eq. (23) is transformed to the primal objective function shown in (25) by introducing the positive slack variables ξ_i^* ($(*)$ denotes variables with and without $*$) (Vapnik 1995)

$$\text{Minimize} \quad \frac{1}{2}w^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon\left(\xi_i, \xi_i^*\right) \tag{25}$$

$$\text{Subjected to} \quad y_i - w \cdot \phi(x) - b \leq \varepsilon + \xi_i \tag{26}$$

$$w \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, \ldots, l \tag{27}$$

$$\xi_i^{(*)} \geq 0 \tag{28}$$

Finally, by introducing Lagrange multipliers and exploiting the optimality constraints, the decision function (26) has the following explicit form:

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)K(x_i, x) + b. \tag{29}$$

In function (29), $a_i^{(*)}$ are the Lagrange multipliers. They satisfy the equalities $a_i \times a_i^* = 0$, $a_i \geq 0$, and $a_i^* \geq 0$, where $i = 1, \ldots, l$, and they are obtained by maximizing the dual function of (27), which has the following form

$$W\left(a_i^{(*)}\right) = \sum_{i=1}^{l} y_i\left(a_i - a_i^*\right) - \varepsilon \sum_{i=1}^{l}\left(a_i - a_i^*\right) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)K\left(x_i, x_j\right) \quad (30)$$

with the following constraints:

$$\sum_{i=1}^{l}\left(a_i - a_i^*\right) = 0, \quad 0 \le a_i^{(*)} \le C, \quad i = 1, \ldots, l. \quad (31)$$

$K(x_i, x_j)$ is defined as the kernel function. The value of the kernel is equal to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. SVR can manipulate in higher dimensional separation plane with non-linear error function and configure three parameters in order to achieve classification results. These three parameters are kernel function, parameter C, and parameter γ.

First, SVR model transforms data to two-dimensional separation plane. When data separation is completed, the model utilizes kernel function to transform this separation plane back to original dimensions. Available kernel activation functions are linear, polynomial, RBF, and sigmoid. Also, regularization parameter C is adjusted to find the most suitable separation function. If value of C is large, the optimization will choose a smaller margin hyperplane, and if value of C is small, such hyperplane might misclassify more data points. Finally, γ parameter quantifies how the data points are considered in constructing the separation hyperplane. If γ value is high, only data points located close to the separation plane are considered, but if γ value is low, data points far beyond the separation plane up to the boundary of datapoints are considered. Therefore, the most significant characteristic of SVR is the separation hyperplanes which are attuned to construct good margin between groups of data points. That hyperplanes assure that classification of data points is optimized.

Comparison of forecasting model

To identify the most accurate method for hotel daily demand forecasting, mean absolute percentage error (MAPE) and MAE are selected for forecasting accuracy measurement. In addition, to test if different forecasting methods provide significantly different accuracy, randomized complete block design (RCBD) and Tukey test are performed. The p value of 0.05 is used to indicate if there are statistically significant differences between group means. Also, pairs of methods are analyzed using the Tukey test to identify pairs that are significantly different.

Results and discussion

This section reports results from all method discussed in the previous section after applying to the training and testing data. The last part of the section shows model comparison and suggests which model is most suitable for hotel daily demand forecasting.

Holt–Winters method

Figure 6 shows actual values and predicted values of Holt–Winters model for the testing set of data (365 days in 2018). The R Studio software program automatically provides optimized estimates of model parameters (alpha, beta, and gamma) that minimize forecast errors. Holt–Winters model with mean parameter (alpha) = 0.208, trend parameter (beta) = 0, and seasonal parameter (gamma) = 0.498 are selected by the program. Result indicates that Holt–Winters model cannot capture changes in trend component in this hotel daily time series, yielding MAPE 20.69%.

Seasonal ARIMA method

Figure 7 shows actual values and predicted values of SARIMA model for the testing set of data (365 days in 2018). Model adequacy is checked for each forecasting method. Figures 8 and 9 show ACF and PACF plots from SARIMA model. ACF of SARIMA model shows spike at lag 1 though it decays at the end, meaning that trend and seasonality have not been completely removed. Thus, SARIMA model is not appropriate for these data.

Autoregressive, moving average, and integrated factor for both trend and seasonal components (p,d,q,P,D,Q)

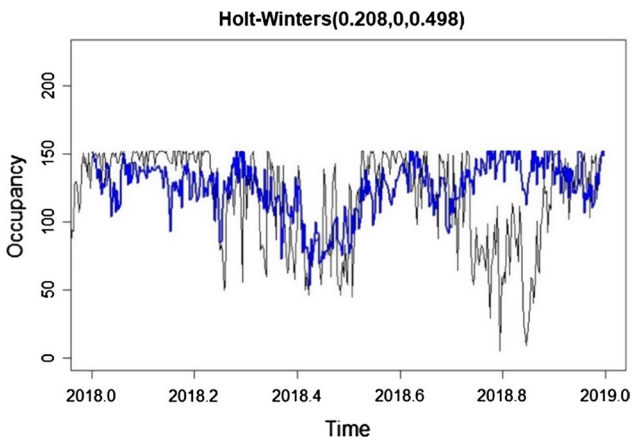

Fig. 6 Holt–Winters forecasting model result (blue line) as compared to actual values (black line). (Color figure online)

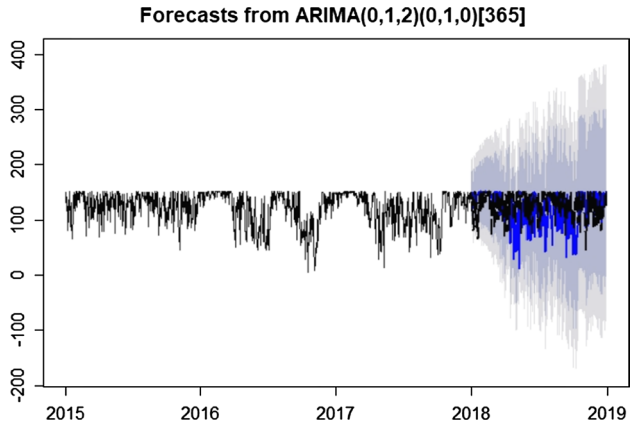

Fig. 7 SARIMA forecasting model result

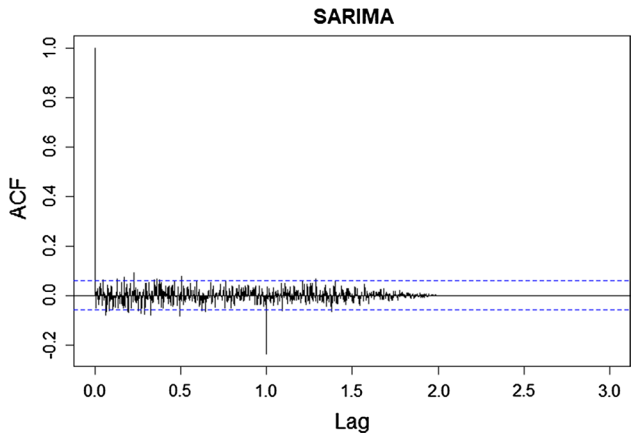

Fig. 8 ACF of residuals from SARIMA model

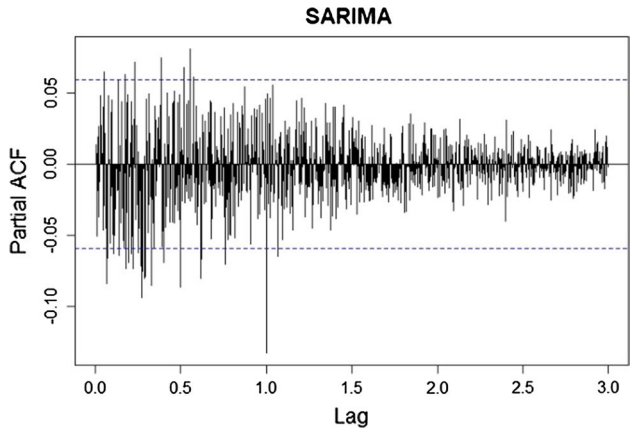

Fig. 9 PACF of residuals from SARIMA model

are adjusted to find lowest AIC and lowest MAPE. Model SARIMA(0,1,2)(0,1,0)365 with AIC = 6957.83, and MAPE = 23.44% is the result of this experiment.

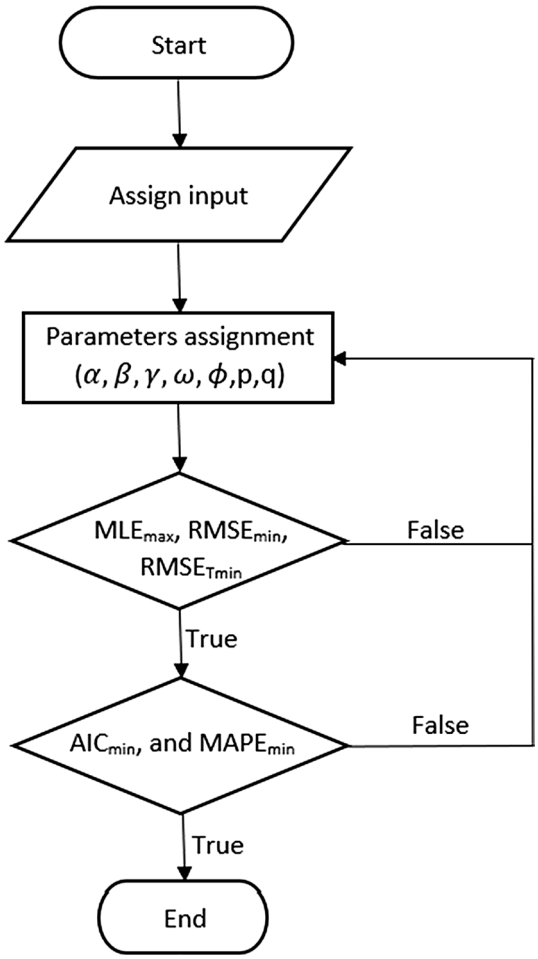

Fig. 10 Flowchart showing BATS/TBATS model selection

BATS and TBATS

Figure 10 shows flowchart of BATS and TBATS model construction. Smoothing parameters (α, β, γ), box–cox transformation parameter (ω), dampening parameter (ϕ), and autoregressive and moving average component (p,q) are adjusted to find maximum log likelihood of the estimates (MLE), minimum Root Mean Square Error of the original data (RMSE) and Root Mean Square Error of the transformed data ($RMSE_T$), minimum Akaike information criterion (AIC), and minimum MAPE). As a result, model BATS $(0.984, \{0,0\}, 0.8, \{365\})$ with MAPE = 19.64% and TBATS $(1, \{2,1\}, -, \{\langle 365, 3\rangle\})$ with MAPE = 17.24% are selected.

Figures 11 and 12 show actual values and predicted values of BATS and TBATS model, respectively. The TBATS model prediction graph shows that seasonality existed in a year that it resembles tourist arrivals in Phuket graph (Fig. 1), peaking at beginning and end of the year and reaching lowest in May and September. This similarity presents

that this model can explain daily hotel data. BATS forecasting model generated BATS (0.984, {0, 0}, 0.8, {365}), representing the use of Box–Cox transformation, no ARMA error function, disposition of damped trend, and seasonal period of 365. TBATS forecasting model generated TBATS (1, {2, 1}, −, {⟨365, 3⟩}), representing no use of Box–Cox transformation, Autoregressive and Moving Average model of 2 and 1, respectively, ARMA(2,1), and seasonal period of 365 with harmonics of 3.

Artificial neural network

After all variables shown in Table 2 were screened using Backward elimination regression at 95% significant level, there are 30 significant variables in total with p value less than 0.05. Those variables that can significantly explain the hotel daily demand are as follows:

- MA-360, RDP-7, RDP-30, RDP-52, RDP-365
- Number of Chinese, Russian, Australian, Korean, Malaysian, Singaporean, Swedish, and British tourists, arrivals to Phuket
- Customer price index of Southern region of Thailand
- Oil price
- Exchange rate of United States Dollar, Malaysian Ringgit, Swedish Krona, and British Pound over Thai Baht
- Average room rate of the hotel
- Month: July, August, September, October, November, December, and February
- Peak and High season dummy variables
- Sanction on Russia dummy variables

There are two types of regressor variables in this study: independent factors and transformed data. Independent data include variables such as Number of Tourist Arrivals in Phuket by Country of Residence, etc. Transformed data are RDP-7, RDP-30, RDP-52, RDP-365 (relative difference in percentage), and MA-360 (occupancy subtracted with moving average). The number of tourists from every country is deemed significant because these are source of hotel occupancy. Customer price index of Southern region of Thailand

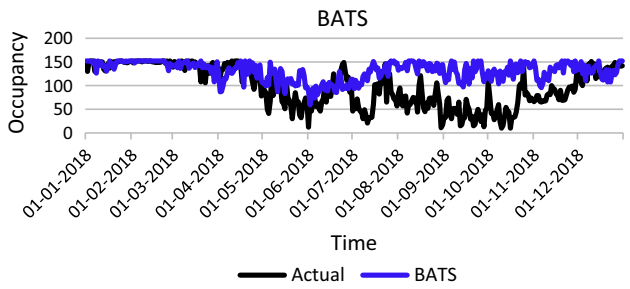

Fig. 11 BATS forecasting model result as compared to actual values

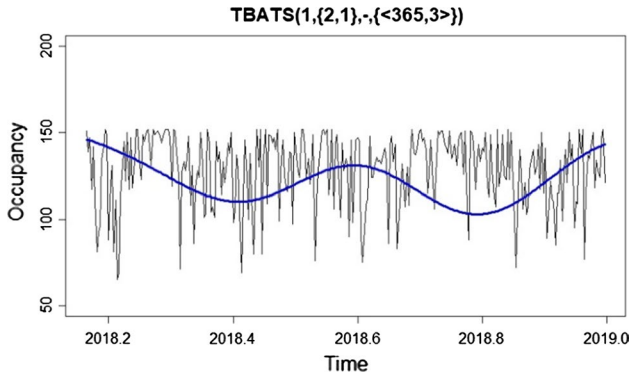

Fig. 12 TBATS forecasting model result (blue line) as compared to actual values (black line)

and oil price is significant because they determine purchasing power of tourist in the region. Some of the exchange rates are considered significant because they have relation to number of tourists in the island. Only months of the year but not days of the week are significant showing seasonal pattern occurs on monthly basis, not on daily basis. Peak season, high season, and sanction on Russia dummy variables are significant as they can explain data pattern in the daily occupancy. Average room rate of the hotel is significant since it is directly disproportionated to the occupancy.

Transformed data of the hotel, i.e., MA-360, RDP-7, RDP-30, RDP-52, RDP-365, are significant showing those data is highly related to daily occupancy. Table 3 shows Adjusted R^2 of Linear Model with three different types of variables. From regression, it can be noticed that models with both independent factors and transformed data provides the highest adjusted R^2.

Table 4 shows parameters selection of neural network model including algorithm and activation function. Sum of squared errors (SSE) is used as error function, while cross-entropy (CE) error function can be used only for binary response. Only 'rprop-' and 'rprop+' algorithm can achieve threshold of 1 under 10^8 stepmax, the lesser threshold and larger stepmax employ longer computation time which is not practical. Logistic function (logistic) and tangent hyperbolicus (tanh) are used as activation function, and both are tested. Results show that neural network model with

Table 3 Adjusted R^2 of three different linear regression models

Variables	Adjusted R^2 (%)	No. of significant variables
Independent factors	44.78	25
Independent factors + Transformed data	99.88	30
Transformed data	97.41	5

Table 4 Error results on different parameters of Neural Network model

Algorithm	Activation function	MAPE	MAE
rprop−	logistic	11.7975	6.983
rprop−	tanh	12.2517	7.238
rprop+	logistic	11.7974	6.983
rprop+	tanh	12.1301	7.175

Table 5 Forecasting results on types of input variables from Neural Network model

Variables	Neural network		
	MAPE	MAE	#Nodes
Independent factors	71.39	37.26	100
Independent factors + transformed data	74.33	37.59	35
Transformed data	8.955	5.603	15

'rprop+' as an algorithm and logistic function as activation function performs best with the lowest error. Thus, 'rprop+' and logistic function are selected to be used in this research.

Different sets of input variables are tested with neural network model to find the forecasting accuracy. Table 5 shows MAPE and MAE results of those input variables from neural network model. It can be observed that transformed data as input variables for neural network model perform better than both independent factors and independent factors with transformed data in two aspects: MAPE is exceptionally better and number of nodes utilized in the neural network structure is lower.

The process of neural network model selection starts with input assignment and error threshold setting at 1. Algorithm and activation function are selected to test for lowest MAPE, then structure of neural network model is tested by adding number of nodes into its structure to find the most accurate model. An experiment has been made to find the most accurate neural network structure, with data up to 1826 points and computation time for neural network is of $O(n^5)$, by assuming that gradient descent runs for n iterations, and that there are n layers each with n neurons. The structure of NN model is limited to 1 hidden layer. First, numbers of node vary from 1 to 25 at lag of 5, then within the range that gives the best model numbers of node are changes by lag of 2 to find all possible models. Table 6 shows error results of experiment on the number of nodes.

To conclude, the applicable models are 1 hidden layer with 12,14,15,16,18, and 20 nodes, yielding MAPE under 10%. Experiments are conducted to find the most suitable error function and activation function that will deliver with the lowest MAPE. These ANN models employ resilient backpropagation with and weight backtracking as an algorithm in calculating models' weights. Figure 13 shows actual values and predicted values of neural network model with backpropagation for the testing set of data in year 2018. Figure 14 shows plot of the most accurate neural network model.

Table 6 Error results at different number of nodes in ANN model

Number of nodes	Errors of ANN	
	MAPE	MAE
1	10.72	6.500
5	10.47	6.313
10	10.50	6.579
15	8.955	5.603
20	9.140	5.826
25	10.86	6.683
…	…	…
12	9.334	5.913
14	9.241	5.850
16	9.118	5.835
18	9.660	6.235

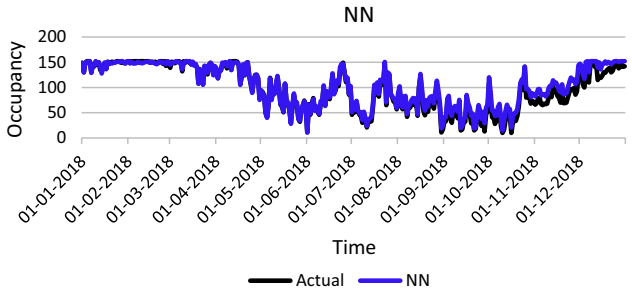

Fig. 13 Neural network forecasting model result as compared to actual values

Support vector regression

SVR experiments are conducted to find the most suitable regression function and kernel function. These SVR models utilize epsilon type regression function, and radial basis function as kernel function. Experiments were made to adjust parameters for SVR (shown in Table 7). Gamma and regularization parameter are fixed at first, varying regression type and Kernel function to find the most accurate settings. Later, Gamma and regularization parameter are adjusted but resulted accuracy did not change significantly. Differences of accuracy between eps and nu-regression type are negligible. However, eps-regression facilitates best performance by controlling amount of errors in the model, in which solution model could be complex, but nu-regression returns fewer support vectors, requirement for small solution.

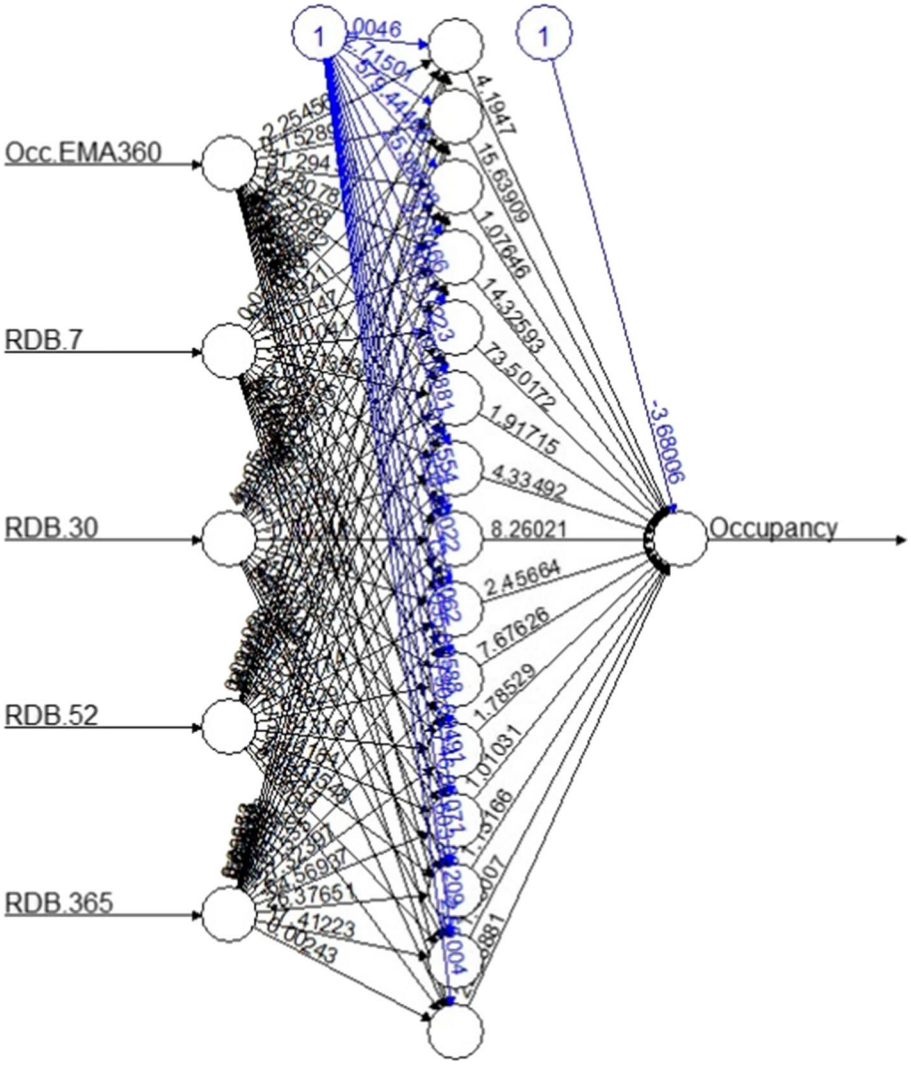

Fig. 14 Structure of the most accurate neural network model

Table 7 Experiment on parameters of support vector regression

Regression type	Kernel function	MAPE	MAE
eps	Polynomial	63.37	46.17
	Radial basis	13.75	7.779
	Sigmoid	1156.42	434.8
nu	Polynomial	64.29	51.62
	Radial basis	13.82	13.42
	Sigmoid	1080.15	469.2

Table 8 Forecasting results on different sets of input variables for SVR

Variables	Support vector regression	
	MAPE	MAE
Independent factors	84.86	41.56
Independent factors + transformed data with p value < 0.05	43.21	22.00
Transformed data	13.75	7.779

Table 8 shows forecasting result on different sets of input variables from SVR. It can be observed that transformed data as input variables for SVR model perform better than both independent factors and independent factors with transformed data.

Kernel function, regularization parameter, and gamma parameter are selected and tested to find the most accurate settings by MAPE. Radial basis as Kernel function with eps-regression type provides the best accuracy, while regularization and gamma parameter are set at default. Then, regularization and gamma parameter are varied to find the best performance. Figure 15 shows actual values and predicted values of SVR model for the testing period (year 2018), giving MAPE as low as 13.75%.

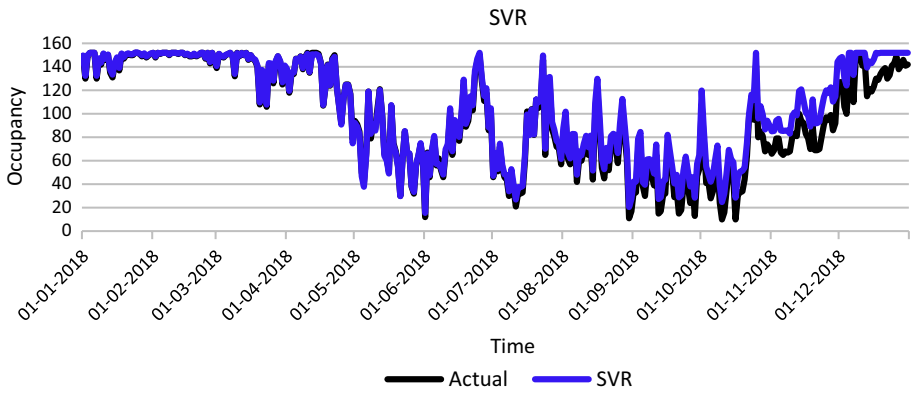

Fig. 15 Support vector regression forecasting model result as compared to actual values

Table 9 Forecasting results from each model

Method	MAPE (%)	MAE
Holt–Winters (0.208,0,0.498)	20.69	16.66
SARIMA (0,1,2)(0,1,0)365	23.44	27.46
BATS (0.984, {0, 0}, 0.8, {365})	19.64	21.60
TBATS (1, {2, 1}, –, {⟨365, 3⟩})	17.24	19.77
Artificial neural network	8.955	6.617
Support vector regression	12.30	8.760

Models comparison and selection

In this section, a comparison among all models is discussed. "Aggregate data" section shows error results for aggregate data (all rooms of the hotel). "Disaggregated data" section presents results for disaggregate data (each room type is forecasted separately). Finally, "Tukey pairwise comparison" section discussed Tukey pairwise comparisons and pointed out the most suitable method for hotel daily demand forecasting.

Aggregate data

Table 9 summarizes error results from each model. The result indicates that ANN model provides the most accurate forecasting value, followed by SVR, TBATS, BATS, Holt–Winters' exponential smoothing, and SARIMA model.

Disaggregated data

Table 10 presents forecasting results of disaggregated data measured by MAPE and MAE, respectively. ANN and TBATS are selected as they are the most accurate model for machine learning and time series method, respectively. Note that MAPE for Pool Access and Villa room type cannot be calculated as there are some zero actual values on some days. It can be seen that ANN also outperforms TBATS for disaggregated data as well.

Table 10 Disaggregated forecasting results according to MAPE and MAE

Type of data	MAPE		MAE	
	TBATS	NN	TBATS	NN
Aggregated data	17.24%	8.955%	20.75	5.603
Disaggregated data				
Deluxe	13.93%	13.94%	4.86	2.742
Seaview	16.76%	14.05%	8.50	2.553
Pool Access	–	–	3.28	0.6819
Villa	–	–	4.70	0.6108

Table 11 Turkey's pairwise comparison results

Factor	N	Mean	Grouping
SDLY	365	34.24	A
SARIMA	365	27.46	B
BATS	365	21.603	C D
TBATS	365	19.768	C D
HW	365	16.662	D
SVM	365	8.760	E
NN	365	6.617	E

Tukey pairwise comparison

Analysis of Variance (ANOVA) at 99% confident interval examines 365-day MAE from each forecasting models. Fisher's LSD Method presents that mean of MAE of the models are grouped differently. Tukey's pairwise comparisons are tested at 95% confidence interval (exhibited in Table 11). SVM and NN are in the same group, giving the lowest MAE (NN) and second lowest MAE (SVM). Same Day Last Year (SDLY) method, whose forecast of the first Monday this year equals the first Monday last year, and so on, is also presented as baseline.

Conclusions

This paper identified the most suitable model to forecast hotel room daily demand occupancy with substantial changes in level and trend. Different models, namely Holt–Winters, Box–Jenkins, Box–Cox transformation, ARMA errors, trend, and multiple seasonal patterns (BATS), TBATS, ANN, and SVR were considered. Both aggregated data and disaggregated data were explored. Unlike previous works, this research compared time series method and machine learning method for hotel daily demand forecasting. This research also introduced the use of data transformation in contrast with the use of traditional independent variables. Results showed that neural network model using independent variables did not provide acceptable forecasting accuracy, while neural network Transformed data, i.e., RDP-7, RDP-30, RDP-52, RDP-365 (relative difference in percentage), and MA-360 (occupancy subtracted with Moving Average), could offer high levels of forecasting accuracy (with MAPE $\leq 10\%$). It was shown that the neural network model could produce high levels of forecasting accuracy and outperformed other time series models and causal models. SVR performed the second most accurate forecast results and did not significantly different compared to ANN. Thus, SVR could be a comparable alternative.

The research is subject to few limitations. The main limitation of this research is availability of daily independent data. We collected 42 independent variables from various sources, yet data such as average room rate of the region are not available. It is expected by expert judgment that average room rate of the region will significantly explain hotel daily occupancy. Secondly, transformed data in regression models cannot be used to explain seasonal or trend in time series data. However, the findings are reliable and indicate that transformed data perform very well in forecasting daily hotel time series.

There are a number of areas that future studies can investigate. Firstly, future research can extend this paper with other machine learning methods to other hotels in order to make comparison and obtain more consistent results. In fact, a further study based on the analysis of different data transformation could be extremely useful to assess forecasting performances. Secondly, future research may apply forecasting results to other attributes regarding hotel revenue management such as pricing or overbooking decisions.

References

Baker, T.K., and D.A. Collier. 1999. A comparative revenue analysis of hotel yield management heuristics. *Decision Sciences* 30: 239–263.

Box, G.E.P., and G. Jenkins. 1970. *Time series analysis, forecasting and control*. San Francisco, CA: Holden-Day.

Cao, L.J., and F.E.H. Tay. 2001. Application of support vector machines in financial time series forecasting. *Omega* 29: 309–317.

Chen, C., and S. Kachani. 2007. Forecasting and optimization for hotel revenue management. *Journal of Revenue and Pricing Management* 6 (3): 163–174.

Claveria, O., E. Monte, and S. Torra. 2015. A new forecasting approach for the hospitality industry. *International Journal of Contemporary Hospitality Management* 27 (7): 1520–1538.

De Livera, A.M. 2010. *Automatic forecasting with a modified exponential smoothing state space framework*. Clayton: Department of Econometrics & Business Statistics, Monash University.

De Livera, A.M., R. Hyndman, and R. Snyder. 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* 106 (496): 1513–1527.

El Gayar, N., M. Saleh, A. Atiya, H. El-Shishiny, A. Zakhary, and H. Habib. 2011. An integrated framework for advanced hotel revenue management. *International Journal of Contemporary Hospitality Management* 23 (1): 84–98.

Haensel, A., and G. Koole. 2011. Booking horizon forecasting with dynamic updating: A case study of hotel reservation data. *International Journal of Forecasting* 27 (3): 942–960.

Holt, C.E. 1957. *Forecasting seasonals and trends by exponentially weighted averages*. O.N.R. memorandum no. 52. Pittsburgh: Carnegie Institute of Technology.

Kouprioushina, L., J. Van der Rest, and Z. Schwartz. 2014. On revenue management and the use of occupancy forecasting error measures. *International Journal of Hospitality Management* 41: 104–114.

Lim, C., C. Chang, and M. McAleer. 2009. Forecasting h(m)otel guest nights in New Zealand. *International Journal of Hospitality Management* 28 (2): 228–235.

Martinez-de Pison, E., J. Fernandez-Ceniceros, A.V. Pernia-Espinoza, F.J. Martinez-de Pison, and A. Sanz-Garcia. 2016. Hotel reservation forecasting using flexible soft computing techniques: A case of study in a Spanish hotel. *International Journal of Information Technology & Decision Making* 15: 1211–1234.

Office of the National Economic and Social Development Council. 2019. https://www.nesdb.go.th/nesdb_en/ewt_dl_link.php?nid=4374&filename=national_account. Accessed 29 Mar 2019.

Pereira, L.N. 2016. An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management* 58: 13–23.

Rajopadhye, M., G.M. Ben, P.P. Wang, T. Baker, and C.V. Eister. 2001. Forecasting uncertain hotel room demand. *Information Sciences* 132: 1–11.

Taylor, J.W. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of Operation Research Society* 54 (8): 799–805.

Thomason, M. 1999. The practitioner methods and tool. *Journal of Computing Intelligence in Finance* 7 (3): 36–45.

Urraca, R., A. Sanz-Garcia, J. Fernandez-Ceniceros, E. Sodupe-Ortega, and F.J. Martinez-de-Pison. 2015. Improving hotel room demand forecasting with a hybrid GA-SVR methodology based on skewed data transformation, feature selection and parsimony tuning. *Hybrid Artificial Intelligent Systems* 9121: 632–643.

Vapnik, V.N. 1995. *The nature of statistical learning theory*. New York: Springer.

Vu, C.J., and L.W. Turner. 2006. Regional data forecasting accuracy: The case of Thailand. *Journal of Travel Research* 45: 186–193.

Weatherford, L.R., and S.E. Kimes. 2003. A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting* 19: 401–415.

Weatherford, L.R., S.E. Kimes, and D.A. Scott. 2001. Forecasting for hotel management: Testing aggregation against disaggregation. *Cornell Hotel & Restaurant Administration Quarterly* 42: 53–64.

Winters, P.R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* 6: 324–342.

Zakhary, A., A.F. Atiya, H. El-Shishiny, and N.E. Gayar. 2011. Forecasting hotel arrivals and occupancy using Monte Carlo simulation. *Journal of Revenue and Pricing Management* 10 (4): 344–366.

Naragain Phumchusri is an assistant professor in Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand. She received master's and doctoral degrees in Industrial Engineering from The H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Georgia, USA in 2010. Her research interests include Operation Research, Revenue Management, Applied Statistics, Stochastic Optimization, and Data Analytics with Machine Learning.