Forecasting Occupancy for Demand Driven HVAC Operations Using Time Series Analysis

Gulben Calis*1, Sevcan Demir Atalay2, Merve Kuru3 and Nalan Mutlu4

1Assistant Professor, Department of Civil Engineering, Ege University, Izmir, Turkey
2Associate Professor, Department of Statistics, Ege University, Izmir, Turkey
3Graduate Student, Department of Civil Engineering, Ege University, Izmir, Turkey
4Graduate Student, Department of Statistics, Ege University, Izmir, Turkey

Abstract
Building heating, ventilation, and air conditioning (HVAC) systems contribute substantially to the energy consumption of buildings. Today, traditional HVAC systems mostly operate according to the maximum occupancy assumption, which in turn increases energy consumption during periods of low occupancy. Although, recently, implementing demand-driven HVAC operations are accepted as an innovate approach for reducing HVAC-related energy consumption, occupancy forecast is important to realize demand-driven HVAC operations in buildings. This study aims at using time series models in order to forecast the daily number of bank customers in a financial center of a bank, which is located in Izmir, Turkey. Data were collected from the computerized tracking system for a period of 60 weeks and two forecasting methods were used: 1) Decomposition Method, 2) Box-Jenkins Method. To determine the final model identified via the Box-Jenkins Method, goodness-of-fit, residual analysis and Akaike information criterion were taken into consideration. The results show that the SARIMA model with a MAPE of 11% yields a good occupancy forecast for supporting demand-driven HVAC operations.

Keywords: demand-driven HVAC operation; occupancy forecast; time series analysis; Box-Jenkins method

1. Introduction
In commercial buildings in the United States nearly 40% of the energy is used by heating, ventilation, and air conditioning (HVAC) systems to maintain comfortable and healthy indoor thermal environments (Yang et al., 2016). Today, the most traditional HVAC systems condition the rooms in a building assuming there is maximum occupancy in the defined zones during operational hours (Yang et al., 2014). Typically, operational settings are dictated according to assumed occupied and unoccupied periods of the day (e.g., 9 am to 6 pm) and do not consider when buildings are partially occupied (Lie et al., 2012). Therefore, occupancy is one of the primary drivers of energy consumption resulting from HVAC systems. In particular, occupancy has a direct impact on various aspects such as heat loads, HVAC system running time, required heating, cooling and distribution of conditioned air, and preferred temperature set points (Li et al., 2012). The main energy consumption components of HVAC systems are (1) cooling or heating and distribution of air by Air Handling Units (AHUs), (2) zone-level flow control device by Variable Air Volume (VAV) boxes, and (3) air humidification or dehumidification (Li et al., 2012; Labeodan et al., 2015). To ensure that the total energy consumed by the HVAC system is kept at the minimum possible and the thermal comfort requirements are met, demand-driven HVAC strategies have been introduced in the building operation phase. Prior research reported that it is possible to save from 10% to 56% of HVAC-related energy consumption by employing demand-driven HVAC controls (Yang et al., 2014). Yang et al. (Yang et al., 2014) demonstrated that 20% of gas and 18% of electricity could be effectively saved if demand-driven HVAC control is implemented. Moreover, it is stated that an HVAC system could waste energy by supplying enough ventilation for 30 people when only 10 actually occupy a room (Erickson et al., 2014). Therefore, especially in dynamic environments, where the setting and occupancy keep changing, knowing the number of occupants and being able to accurately forecast usage patterns may allow significant energy-savings by intelligent control of the HVAC systems compared to strategies assuming fixed occupancy and usage patterns.

*Contact Author: Gulben Calis, Assistant Professor, Ege University, Department of Civil Engineering, Bornova, Izmir, Turkey
Tel: +902323115171 Fax: +902323412957
E-mail: gulben.calis@ege.edu.tr

(Received October 5, 2016; accepted July 12, 2017) DOI http://doi.org/10.3130/jaabe.16.655
Although, occupancy and usage patterns might significantly affect and have the potential to reduce energy consumption in buildings, the large occupancy number and the very high occupancy variation in buildings pose a higher challenge for real-time occupancy number counting (Yang et al., 2016). Obtaining real-time occupancy information requires deploying sensors and/or using cellular/bluetooth data. These occupancy information collection techniques have some drawbacks such as being expensive, yielding low accuracy and raising privacy concerns. Accordingly, many researches have proposed methods to predict occupancy in buildings. Time-series-based models are among the most commonly used methods due to their ability to forecast crowded conditions, especially hour by hour. Schweigler et al. (Schweigler et al., 2009) showed that the models using time series methods generate more accurate short-term forecasts of Emergency Department (ED) bed occupancy than using traditional historical averages models. Kam et al. (Kam et al., 2010) developed and assessed three time series models to predict the daily number of patients visiting the ED of a Korean hospital. The results show that the multivariate SARIMA model exhibits high reliability and forecasting accuracy compared to other models. Andrew et al. (Andrew et al., 1990) empirically investigated the use of two time series models fitted using actual monthly occupancy rates for a major city center hotel. The authors concluded that models’ implementations are relatively easy and the accuracy levels are high. Baldigara and Koic (Baldigara and Koic, 2015) analysed the net-occupancy rate of bed places in Croatian Hotels from January 2005 to August 2014 by using three models of generating time-series forecasts. The results show that the specified models passed all the tests and that they fit the data reasonably well throughout the sample period.

The objective of this study is to develop a model to forecast the number of customers visiting a financial center per day. Two forecasting methods, namely Decomposition and Box Jenkins, are used and the results are evaluated. The following sections introduce the methodology including data collection, forecasting methods and method evaluation. The findings, discussions and conclusions are then presented.

2. Methodology

2.1 Data Collection

In this study, a financial center of a bank, which is located in a business district with many small and medium enterprises, is selected as the case study. The bank is one of the leading ones in Turkey and has a big customer portfolio. The financial center operates 5 days per week unless there is a public holiday. The data set is extracted from the computerized tracking system of the information department. The data utilized for analysis include 60 weeks, from June 2, 2014 until July 24, 2015. The data from June 2, 2014 to July 17, 2015 was used to develop the forecasting models and the rest was used to evaluate the models.

2.2 Forecasting Methods

In this study, two forecasting methods were used: 1) Decomposition Method, 2) Box-Jenkins Method. The decomposition method is based on the analysis of the individual component of the time series. One of the decomposition approaches is the multiplicative decomposition, in which time series can be expressed as the product of the four components of the series and are \( Y = T \cdot S \cdot C \cdot I \), where \( Y \) represents the original data whereas \( T, S, C, \) and \( I \) represent the trend, seasonal, cyclical and irregular component of the series, respectively. The advantage of this method is its capability of removing irregular changes and seasonal factors within the forecast.

The Box-Jenkins method applies autoregressive moving average (ARMA or ARIMA) models to find the best fit of a time-series model to past values of a time series (Adhikari and Agrawal, 2013, Baldigara and Koic, 2015, Box et al., 2015, Moosazadeh et al., 2014, Sarppong, 2013, Lee, 2011). The seasonal autoregressive integrated moving average model (SARIMA) is an expanded form of the ARIMA. SARIMA processes are designed to model time series with trends, seasonal patterns and short time correlation (Baldigara and Koic, 2015). Therefore, it is known to be effective when the components of a time series change rapidly over time, and this model has proven useful in the forecasting of short-term volatility (Kam et al., 2010). The SARIMA model consists of the 1) auto-regression, 2) difference and 3) moving average and is represented as SARIMA \((p, d, q)(P, D, Q)\), in which \( p, d, q \) present non-seasonal AR order, non-seasonal differencing and seasonal MA order respectively, whereas \( Q, D \) and \( S \) present seasonal MA order, seasonal differencing and the length of seasonality, respectively.

The seasonal autoregressive integrated moving average model is

\[
\phi(B) \Phi(B^s)(1 - B)^d(1 - B^s)^D Y_t = \Theta_0 + \Theta(B) \Theta(B^s) \varepsilon_t
\]

where

\[
\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \cdots - \phi_p(B^p)
\]

\(p\) order non-seasonal AR model

\[
\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - \cdots - \theta_q(B^q)
\]

\(q\) order non-seasonal MA model

\[
\Phi(B^s) = 1 - \Phi_1(B^s) - \Phi_2(B^{2s}) - \cdots - \Phi_P(B^{Ps})
\]

\(P\) order seasonal AR model

\[
\Theta(B^s) = 1 - \Theta_1(B^s) - \Theta_2(B^{2s}) - \cdots - \Theta_Q(B^{Qs})
\]

\(Q\) order seasonal MA model

\((1 - B)^d;\) non-seasonal differencing of order \( d \)

\((1 - B^s)^D;\) seasonal differencing of order \( D \)

\(\varepsilon_t;\) error term~IID(0,\( \sigma^2 \))

\(B;\) backward shift

\(S;\) seasonal order
Minitab software is used for the identification of SARIMA models. The flowchart for constructing the models is presented in Fig.1. Further explanation regarding the steps can be found in (Maddala, 2002).

2.3 Method Evaluation

To compare the adequacy and performance of the methods, residual analysis including autocorrelation function (ACF) and partial autocorrelation function (PACF) of residuals was conducted. In time series modelling, the selection of a best model fit to the data is directly related to whether residual analysis is performed well. One of the assumptions of the SARIMA model is that, for a good model, the residuals must follow a white noise process. That is, the residuals have a mean value of zero, constant variance and are uncorrelated. In addition, the ACF of the residuals depicts that the autocorrelation of the residuals are all zero, that is to say they are uncorrelated.

The Akaike Information Criterion (AIC) was used to compare the models which are evaluated by $\text{AIC} = \frac{N \times \ln (\text{RSS/N}) + 2 \times k}{n}$, where n, RSS and k represent the number of observations, sums of square residuals and number of predicted parameters, respectively. Then Ljung–Box–Pierce chi-square statistics and t-test for examination of null hypothesis of the parameters equalization to zero are calculated. Finally three measures were used for determining the accuracy of the fitted values: MAPE (Mean absolute percentage error), MAD (Mean absolute deviation) and MSD (Mean squared deviation) which are formulated as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n} \times 100 \quad (y_t \neq 0)$$

$$\text{MAD} = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n}$$

$$\text{MSD} = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}$$

MAPE represents the relative scale of the prediction error between the forecasted value, which is a series variable, and the observed value; the smaller the error is, the more accurate the prediction is (Kam et al., 2010). Besides, it is valid for MAD and MSD; the smaller MAD and MSD are, the more accurate the prediction is.

3. Findings

Data collected from June 2, 2014 to July 17, 2015 is used to identify the forecasting models. The total number of customers during that period is 77379 with a daily average of 262. Fig.2. shows the time series plot of the observed (original) data which is used to evaluate patterns and behavior in data over time. Fig.2. shows that there is a regularly repeating pattern of highs and lows related to the days of the week, which imply seasonal and trend effects.

3.1 Results of the Decomposition Method

The decomposition method can forecast the customer number with seasonal, trend, cycle and irregular effects. Firstly the moving average of order 5 (representing the working days of the week) on the observed data set is calculated (H), then the observed data set (Y) is divided by H (Y/H). This data preprocessing allows the removal of the irregular components from the observed data set. Secondly, the seasonal component is removed...
from the observed data, the resulting values are called the 'seasonally adjusted' data. For a multiplicative model, the seasonally adjusted values are obtained by using the formula $Y/(a \times b)$, where $a$ and $b$ denote the seasonal index and correction factor, respectively. Following these procedures, the graph of seasonally adjusted data, which is shown in Fig.3., is obtained.

The third step is constructing a trend equation. The parameter of the equation is estimated by the least square estimation method. Then, the trend values ($T$) can be obtained by using this equation. The trend component is removed from the data set by dividing the seasonally adjusted data by trend values $(Y/(a \times b))/T$. Finally, some alpha values are calculated to remove the cycling component. In this study, the alpha values ranging from 0.1 to 0.9 with an interval of 0.1 are taken into consideration and the optimum value of alpha was determined as 0.9. The forecasted values according to the multiplicative method are obtained through this procedure and are given in Fig.4.

3.2 Results of Box-Jenkins Method

By using Box – Jenkins Procedure, different models can be obtained for various combinations of AR and MA individually and collectively. In this study, the following 14 tentative models were generated.

Table 1. The Adequacy and Performance of the SARIMA Models

| Model type | Lag | t    | p   |
|------------|-----|------|-----|
| SARIMA (1,0,1) (1,0,1) | 5   | 28.96| 0.000 |
| AR  | 1   | 3.96 | 0.000 |
| MA  | 1   | 21.73| 0.000 |
| SMA | 5   | 29.46| 0.000 |
| Chi-square: 16.4 | p: 0.037 | AIC: 2160.889 | coef: significant |
| SARIMA (1,1,1) (1,0,1) | 5   | 265.16| 0.000 |
| AR  | 1   | 3.96 | 0.000 |
| MA  | 1   | 21.73| 0.000 |
| SMA | 5   | 30.54| 0.000 |
| Chi-square: 10.1 | p: 0.259 | AIC: 2158.401 | coef: significant |
| SARIMA (1,0,1) (0,1,1) | 5   | 10.52| 0.000 |
| MA  | 1   | 21.53| 0.000 |
| SMA | 5   | 40.12| 0.000 |
| Chi-square: 18.0 | p: 0.035 | AIC: 2153.329 | coef: significant |
| SARIMA (0,1,1) (0,1,1) | 5   | 4.12 | 0.000 |
| AR  | 1   | 4.12 | 0.000 |
| MA  | 1   | 4.12 | 0.000 |
| SMA | 5   | 4.12 | 0.000 |
| Chi-square: 11.8 | p: 0.222 | AIC: 2149.551 | coef: significant |
| SARIMA (0,1,1) (0,1,1) | 5   | 15.65| 0.000 |
| MA  | 1   | 15.65| 0.000 |
| SMA | 5   | 43.71| 0.000 |
| Chi-square: 20.6 | p: 0.024 | AIC: 2157.57 | coef: significant |
| SARIMA (0,1,1) (0,1,1) | 5   | -6.60| 0.000 |
| AR  | 1   | -6.60| 0.000 |
| SMA | 5   | 45.11| 0.000 |
| Chi-square: 38.7 | p: 0.000 | AIC: 2195.427 | coef: significant |

The residual analysis was conducted to the SARIMA (1,1,1) (0,1,1) model. The ACF and PACF of residuals are shown in Fig.5.a and 5.b. It should be noted that the standardized residual reveals that the residuals...
of the model have zero mean and constant variance. Moreover, the ACF of the residuals indicate that the autocorrelation of the residuals are all zero, in other words, they are uncorrelated. Accordingly, the fitness of the model was assured.

Fig. 6. illustrates the observed values and forecasted values according to the SARIMA (1,1,1) (0,1,1)\textsubscript{5} model.

4. Discussions and Conclusions

In this study, two methods, namely Decomposition and Box-Jenkins, are used to forecast the number of customers in a financial center. The data utilized for analysis include 60 weeks and 1-week data which were used for evaluating the forecasting results of the methods against actual data. The results for the next 5-term customer visits are shown in Table 2.

The results indicate that the empirical results are good since the forecasted values, in fact, are close to the actual values. On the other hand it seems that the forecasted values obtained by the SARIMA model are closer to the actual values.

The values of MAPE, MAD and MSD both for the results obtained by the SARIMA model and the decomposition method are shown in Table 3. According to Witt and Witt (Witt and Witt, 1992), if the MAPE value is less than or equal to 10\%, it can be said that the model has a high accuracy degree, if this value is between 10\% and 20\%, the model can be called a 'true prediction model'. Furthermore, according to Lewis (Lewis, 1982), if the MAPE values are below 10\%, it can be said that the model is 'very good', and for the values between 10\% and 20\%, it can be said that the model is 'good'. Therefore, it can be seen from Table 3. that the SARIMA model and the decomposition method are "true prediction models" according to Witt and Witt whereas they are "good models" according to Lewis. Moreover, the results indicate that the SARIMA model provided a slightly better fit than the decomposition method since its MAPE, MAD and MSD values are smaller than the decomposition method's.

Table 3. MAPE, MAD and MSD Values of Two Methods

| SARIMA Model | Decomposition Method |
|--------------|----------------------|
| MAPE | 10.99339 | 11.62327 |
| MAD | 28.96386 | 31.43497 |
| MSD | 1460.925 | 1915.945 |

The results show that the Box-Jenkins Methodology yields fairly acceptable occupancy forecasts for supporting demand-driven HVAC operations. It should be noted that a certain level of error can be tolerated since the HVAC systems do not need to be too sensitive to changes in the occupancy number. Instead

| Table 2. Forecasted Values of the Next 5 Term Customer Number |
|-----------------|-----------------|-----------------|-----------------|
| Period | Actual | Forecast | Lower | Upper | \%95 Limits |
| 296 | 263 | 307.971 | 232.649 | 383.294 | 334.4146 |
| 297 | 238 | 249.838 | 167.941 | 331.735 | 277.1793 |
| 298 | 221 | 229.232 | 145.109 | 313.356 | 256.1273 |
| 299 | 192 | 209.164 | 123.649 | 294.678 | 234.9297 |
| 300 | 232 | 242.737 | 156.051 | 329.423 | 268.903 |
it is indicated that adding or subtracting the number of occupants to a certain level in a room should not cause changes in the HVAC operations unless the room switches from unoccupied to occupied or vice versa (Yang et al., 2012). In particular, integration of prediction models in the HVAC operation can allow the adjustment of the temperature of air supplied by AHUs, the air volume provided to each zone, and the air volume provided to each room. It can be concluded that demand-driven HVAC operations based on the occupancy prediction models could be beneficial to determine temperature set points which will help to save energy consumption and increase the occupant thermal comfort.

References
1) Adhikari, R. and Agrawal, R. (2013) An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.
2) Andrew, W. P., Crange, D. A., and Lee, C. K. (1990) Forecasting hotel occupancy rates with time series models: A empirical analysis. Hospitality Research Journal, 14 (2), pp.173-181.
3) Baldićan, T. and Koic, M. (2015) Modelling Occupancy Rates in Croatian Hotel Industry. International Journal of Business Administration, 2015, 6 (3), 121 Erickson, V.L., Carreira-Perpiñán, M.A. and Cerpa, A.E. (2014) Occupancy modeling and prediction for building energy management. ACM Transactions on Sensor Networks (TOSN), 10 (3), p.42.
4) Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) Time series analysis: forecasting and control. John Wiley & Sons.
5) Erickson, V.L., Lin, Y., Kamthi, A., Brahma, R., Surano, A., Cerpa, A.E., Sohn, M.D. and Narayanan S. (2009) Energy efficient building environment control strategies using real-time occupancy measurements. Berkeley, California: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings.
6) Kam, H.J., Sung, J.O. and Park, R.W. (2010) Prediction of daily patient numbers for a regional emergency medical center using time series analysis. Healthcare informatics research, 16 (3), pp.158-165.
7) Labeodan, T., Zeiler, W., Boxem, G. and Zhao, Y. (2015) Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. Energy and Buildings, 93, pp.303-314.
8) Lee, M.H. (2011) Forecasting of tourist arrivals using subset, multiplicative or additive seasonal Arima Model. Matematika, 27, 169-182 Lewis, C.D., (1982) Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting Li, N., Calis, G. and Becerik-Gerber, B., (2012) Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Automation in Construction, 24, pp.89-99.
9) Maddala, G.S. (2002) Introduction to econometrics. New York, John Wiley & Sons Ltd., Third Edition.
10) Moosazadeh, M., Nasehi, M., Bahrampour, A., Khanjani, N., Sharafi, S. and Ahmadi, S. (2014) Forecasting tuberculosis incidence in Iran using box-Jenkins models. Iranian Red Crescent Medical Journal, 16 (5).
11) Sarpong, S.A. (2013) Modeling and forecasting maternal mortality; an application of ARIMA models. International Journal of Applied, 3 (1).
12) Schweigler, L.M., Desmond, J.S., McCarthy, M.L., Bukowski, K.J., Ionides, E.L. and Younger, J.G. (2009) Forecasting models of emergency department crowding. Academic Emergency Medicine, 16(4), pp.301-308.
13) Witt, S.F. and Witt, C.A., (1992) Modeling and forecasting demand in tourism. Academic Press Ltd. Butterworth-Heinemann.
14) Yang Z., Ghuframani, A. and Becerik-Gerber, B. (2016) Building occupancy diversity and HVAC (heating, ventilation, and air conditioning) system energy efficiency. Energy, 109, pp.641-649.
15) Yang, Z., Li, N., Becerik-Gerber, B. and Orosz, M. (2013) A systematic approach to occupancy modeling in ambient sensor-rich buildings. Simulation 3 (July), 90 (8): 960-77.
16) Yang, J., Santamouris, M. and Lee, S.E. (2016) Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy and Buildings, 121, pp.344-349.
17) Yang, Z., Li, N., Becerik-Gerber, B. and Orosz, M. (2012) A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations. San Diego, CA., USA, SimAUD ’12 Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design, 2.