Available online at 2019.creative-construction-conference.com/proceedings/

CCC 2019

Proceedings of the Creative Construction Conference (2019) 097

Edited by: Miroslaw J. Skibniewski & Miklos Hajdu

https://doi.org/10.3311/CCC2019-097

Creative Construction Conference 2019, CCC 2019, 29 June - 2 July 2019, Budapest, Hungary

Forecasting heating degree days for energy demand modeling

Merve Kuru[a], Gulben Calis[a]*

[a]*Ege University, Department of Civil Engineering, Izmir,35040, Turkey.*

Abstract

Heating degree day (HDD) is a technical index taking into consideration outdoor temperature and average room temperature to describe the need for the heating energy requirements of buildings. HDD can be used to normalize the energy consumption of buildings with respect to heating since the amount of energy needed to heat a building in a given frequency is directly related to the number of heating degree days in that particular frequency. In order to understand the heating demand of the buildings, it is important to investigate the HDD patterns and to construct forecasting models. This study aims at constructing short-term forecast models by analysing the patterns of the HDD. Within this context, time series analysis was conducted by the monthly HDD data in France between 1974 and 2017. The performance of the models were assessed by the adjusted R^2 value, residual sum of squares, the Akaike Information Criteria (AIC) and the Schwarz Information Criteria (SIC) as well as the analysis of the residuals. As a result, the most suitable model was determined as SARIMA $(2,0,1)(1,0,1)_{12}$. The results of the study show that there is a potential to integrate time series models of HDD for short term load forecasting.

Keywords: Heating degree days; short term forecasting; time series;Box-Jenkins method;SARIMA models

1. Introduction

Today, buildings have become the main consumers of world energy use [1]. It is stated that 50% of the energy is spent for heating and cooling in buildings and industry [2]. Moreover, heating and hot water alone constitute 79% of total final energy use in the EU households [2]. The heating demand is affected by several factors such as the building shell, the type of heating system, outdoor temperature, and occupant behaviour. Among all factors, outdoor temperature is the only one directly affected by the climate change. In order to understand variations in the demand for energy required to heat a building due to climate change, a technical index, called as Heating Degree Days (HDD) based on outdoor temperature and average room temperature has been developed [3]. HDD can be used to normalize the energy consumption of buildings with respect to heating since the amount of energy needed to heat a building in a given frequency is directly related to the number of heating degree days in that particular frequency. It is indicated that HDD is a more reliable measure of climatic impact on energy consumption than temperature alone [4]. Therefore, it is important to investigate the HDD patterns to understand the heating demand of the buildings. Accordingly, HDD was integrated in the building energy demand models to improve the prediction accuracy [5–7]. D'Amico et al. [8] identified the relationship between HDD and heating energy performance. Fan et al. [9] estimated the impacts of climatic factors including HDD and Cooling Degree Day (CDD) on electricity demand in China. Kurekci [10] investigated the optimum insulation thickness for building walls by using HDD and CDD values of Turkey's provincial

*Corresponding author: Gulben Calis email: gulben.calis@ege.edu.tr

centers. Kohler et al. [11] developed a new degree-day method that provides accurate estimates of annual building energy demand for space heating. It should be noted that the performance of the models might be affected by the selection of HDD data. Elizbarashvili et al. [12] estimated daily, monthly and annual HDD and CDD for fourteen different sites of Georgia based on daily mean air temperature data for 30-year period (1961–1990). The results show that there are significant differences for HDD and CDD among the fourteen examined cities of Georgia. OrtizBeviá et al. [13] estimate trends and inter-annual variability in the evolution of HDD and CDD in Spain from observations at 31 stations for an extended period of 1958–2005. The results show that there is a trend which is found to be statistically significant at roughly 2/3 of the Spanish stations used in the study. Although, OrtizBeviá et al. [13] conclude that these trends are similar to those obtained from observations in other parts of Europe, France is stated to be the most temperature sensitive country in Europe [14]. This paper aims at constructing short-term forecast models by analysing the patterns of the HDD. Within this context, time series analysis was conducted by the monthly HDD data in France. The following sections of the paper describe datasets and methodology. Then, findings and conclusions are presented.

2. Dataset and methodology

2.1. Dataset

In this study, the monthly HDD data in France were obtained from the official website of the Statistical Office of the European Union [15]. A total of 528 data covering the period of January 1974 and December 2017 was used to develop the forecasting model. The unit of data is °C * day. Table 1 presents a part of the HDD data used in the analysis.

Table 1. A part of the HDD data from 1974 to 2017.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1974	434.2	385.8	327.9	248.8	128.5	24.71	4.658	2.940	30.13	150.3	297.2	398.4
1975	368.1	345.4	386.7	270.5	186.6	77.73	17.10	11.21	40.27	215.6	331.3	469.7
2015	414.6	402.0	320.5	200.5	117.4	31.28	9.428	8.295	77.75	184.5	222.4	279.9
2016	363.3	341.3	347.9	254.1	135.3	41.40	12.62	8.048	21.08	180.9	289.1	390.3
2017	486.8	321.9	256.0	236.1	109.4	24.20	6.902	10.01	71.63	113.2	300.3	400.1

2.2. Methodology

Box-Jenkins method, also known as the autoregressive integrated moving average (ARIMA) model, was used for time-based analysis and time-based modeling of HDD data. This method applies the autoregressive moving average (ARMA) or ARIMA models to find the best fit of a time series model to the historical values of a time series [16,17]. The seasonal autoregressive integrated moving average model (SARIMA) is an expanded form of ARIMA. SARIMA processes are designed to model trends, seasonal patterns and correlated time series, and have proven to be successful in estimating short-term fluctuations [16,18]. The SARIMA model consists of (1) automatic regression, (2) difference and (3) moving average. SARIMA model is represented as SARIMA (p, d, q) $(P, D, Q)_S$ in which p, d, q represent the degree of non-seasonal linear autoregressive model (AR), non-seasonal difference and degree of non-seasonal moving average model (MA), respectively whereas P, D, Q represent the degree of seasonal AR, seasonal difference and the degree of seasonal MA. In addition, the seasonality length is represented by S. The d and D parameters are taken into account when the series is not stationary.

In this study, to select the most appropriate SARIMA (p, d, q) $(P, D, Q)_S$ model for HDD series, the significance of the coefficients of the models was checked by the Ljung-Box-Pierce Chi-square statistics and t-test. In addition, the corrected R^2 values, the sum of the squares of the residuals, the Akaike Information Criteria (AIC) and the Schwarz Information Criteria (SIC) were taken into account. Furthermore, the residuals analysis including the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the residuals were conducted. Minitab 18.0 and EViews 10.0 packet programs were used to calculate all the statistics and the model selection criteria.

3. Findings

Fig. 1 shows the plot of time series graph to evaluate the overall behaviour of the HDD series over time. Fig. 1 indicates that the HDD series shows similar periodic fluctuations per month which specifies that the series has a seasonal effect, in other words, it has a seasonality characteristic. Therefore, it can be concluded that the series is not stable.

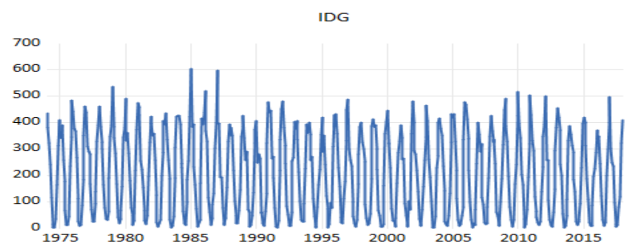
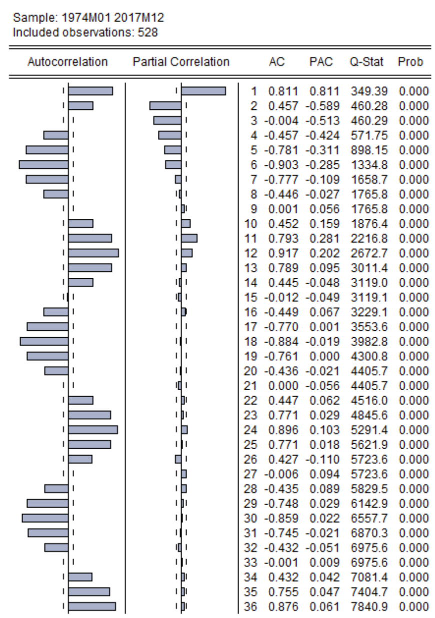


Fig 1. Time series graph of HHD series.

In addition to time series graph, the correlogram of the series was plotted for the k=36 month delay (Fig. 2). This correlogram was created in EViews in order to examine the seasonality.


Sample: 1974M01 2017M12
Included observations: 528

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.811	0.811	349.39	0.000
		2	0.457	-0.589	460.28	0.000
		3	-0.004	-0.513	460.29	0.000
		4	-0.457	-0.424	571.75	0.000
		5	-0.781	-0.311	898.15	0.000
		6	-0.903	-0.285	1334.8	0.000
		7	-0.777	-0.109	1658.7	0.000
		8	-0.446	-0.027	1765.8	0.000
		9	0.001	0.056	1765.8	0.000
		10	0.452	0.159	1876.4	0.000
		11	0.793	0.281	2216.8	0.000
		12	0.917	0.202	2672.7	0.000
		13	0.789	0.095	3011.4	0.000
		14	0.445	-0.048	3119.0	0.000
		15	-0.012	-0.049	3119.1	0.000
		16	-0.449	0.067	3229.1	0.000
		17	-0.770	0.001	3553.6	0.000
		18	-0.884	-0.019	3982.8	0.000
		19	-0.761	0.000	4300.8	0.000
		20	-0.436	-0.021	4405.7	0.000
		21	0.000	-0.056	4405.7	0.000
		22	0.447	0.062	4516.0	0.000
		23	0.771	0.029	4845.6	0.000
		24	0.896	0.103	5291.4	0.000
		25	0.771	0.018	5621.9	0.000
		26	0.427	-0.110	5723.6	0.000
		27	-0.006	0.094	5723.6	0.000
		28	-0.435	0.089	5829.5	0.000
		29	-0.748	0.029	6142.9	0.000
		30	-0.859	0.022	6557.7	0.000
		31	-0.745	-0.021	6870.3	0.000
		32	-0.432	-0.051	6975.6	0.000
		33	-0.001	0.009	6975.6	0.000
		34	0.432	0.042	7081.4	0.000
		35	0.755	0.047	7404.7	0.000
		36	0.876	0.061	7840.9	0.000

Fig. 2. The correlogram of the HDD series for k=36 month delay.

When the correlogram of the series (Fig. 2) is examined, it is seen that the seasonality shows a structure similar to each other in terms of 12 month delays. In other words, there is a strong association between seasonal neighboring observations. For example, ACF (1), ACF (13) and ACF (25) show similar positive autocorrelations which are statistically significant ($p=0.000<0.05$). In addition, for the delays defined by ACF (3), ACF (15) and ACF (27), the autocorrelations are zero. As a result, it can be confirmed from Fig. 2 that these autocorrelation structures continue systematically and regularly in other delays. Therefore, in addition to the time series graph, the correlogram shows that there is a high association between seasonal observations of the series and that the series is not stationary.

In this study, a total of 79 tentative models were generated. Table 2 presents the 44 of these models. It should be noted that the 44 tentative models shown in the table were generated without constant. Tentative models other than those with a star are also tested by including constant.

Table 2. Tentative models generated in this study.

Models			
*SARIMA $(1,0,1)(1,0,1)_{12}$	SARIMA $(1,0,0)(1,1,1)_{12}$	SARIMA $(1,1,1)(1,1,0)_{12}$	*SARIMA $(1,0,2)(0,1,1)_{12}$
SARIMA $(0,0,1)(1,0,1)_{12}$	*SARIMA $(1,0,1)(0,1,1)_{12}$	SARIMA $(0,1,0)(1,1,1)_{12}$	SARIMA $(1,0,1)(0,1,2)_{12}$
SARIMA $(1,0,0)(1,0,1)_{12}$	SARIMA $(1,0,1)(1,1,0)_{12}$	SARIMA $(1,1,0)(1,1,0)_{12}$	*SARIMA $(0,0,2)(0,1,1)_{12}$
SARIMA $(1,0,1)(0,0,1)_{12}$	SARIMA $(1,0,0)(1,1,0)_{12}$	SARIMA $(0,1,1)(1,1,0)_{12}$	SARIMA $(0,0,1)(0,1,2)_{12}$
SARIMA $(1,0,1)(1,0,0)_{12}$	SARIMA $(0,0,1)(1,1,0)_{12}$	*SARIMA $(0,1,1)(0,1,1)_{12}$	SARIMA $(2,0,0)(0,1,1)_{12}$
SARIMA $(1,0,0)(1,0,0)_{12}$	*SARIMA $(0,0,1)(0,1,1)_{12}$	SARIMA $(1,1,0)(0,1,1)_{12}$	SARIMA $(1,0,0)(0,1,2)_{12}$
SARIMA $(0,0,1)(1,0,0)_{12}$	*SARIMA $(1,0,0)(0,1,1)_{12}$	*SARIMA $(2,0,1)(1,0,1)_{12}$	SARIMA $(2,1,1)(0,1,1)_{12}$
SARIMA $(1,0,0)(0,0,1)_{12}$	SARIMA $(1,1,1)(1,1,1)_{12}$	SARIMA $(1,0,2)(1,0,1)_{12}$	SARIMA $(1,1,2)(0,1,1)_{12}$
SARIMA $(0,0,1)(0,0,1)_{12}$	SARIMA $(0,1,1)(1,1,1)_{12}$	SARIMA $(1,0,1)(2,0,1)_{12}$	SARIMA $(1,1,1)(0,1,2)_{12}$
SARIMA $(1,0,1)(1,1,1)_{12}$	SARIMA $(1,1,0)(1,1,1)_{12}$	SARIMA $(1,0,1)(1,0,2)_{12}$	SARIMA $(0,1,2)(0,1,1)_{12}$
SARIMA $(0,0,1)(1,1,1)_{12}$	*SARIMA $(1,1,1)(0,1,1)_{12}$	SARIMA $(2,0,1)(0,1,1)_{12}$	SARIMA $(0,1,1)(0,1,2)_{12}$

As the first step, the models which comply with both of the following conditions were selected: (1) the models with correlations that are statistically significant; (2) the models with the chi-square statistics of the residuals of k=12,24,36 months delay are statistically insignificant, in other words, the residuals of the models are not correlated with each other. As a result, 9 models were selected for further analysis. In the next step, the adjusted R^2 value, the residual sum of squares, the AIC and the SIC were calculated to select the appropriate model. The selected models and the selection criteria values are presented in Table 3.

Table 3. The selected models in the first step and second step selection criteria.

	Model	Adjusted R^2	Residual sum of squares	AIC	SIC
1	SARIMA $(1,0,1)(1,0,1)_{12}$	0.936654	828032.9	10.33716	10.37759
2	SARIMA $(1,0,1)(0,1,1)_{12}$	0.475825	812361.4	10.26529	10.29820
3	SARIMA $(0,0,1)(0,1,1)_{12}$	0.464422	831655.0	10.27557	10.30026
4	SARIMA $(1,0,0)(0,1,1)_{12}$	0.615053	1090103.0	10.57437	10.59910
5	SARIMA $(1,1,1)(0,1,1)_{12}$	0.721348	787551.5	10.27085	10.30382
6	SARIMA $(0,1,1)(0,1,1)_{12}$	0.715896	804533.5	10.28096	10.30569
7	SARIMA $(2,0,1)(1,0,1)_{12}$	0.936911	816848.7	10.32561	10.37412
8	SARIMA $(1,0,2)(0,1,1)_{12}$	0.487184	793205.5	10.26017	10.30132
9	SARIMA $(0,0,2)(0,1,1)_{12}$	0.467373	825461.2	10.27325	10.30617

Regarding the selection of the most appropriate forecasting model among the models selected in the first step, the criteria shown in Table 5 were taken into account. The higher the model's adjusted R^2 value, the lower the residual sum of squares, the lower the AIC and the SIC are, the better the model is. Accordingly, the model that addresses most of the criteria at the same time is selected. Table 3 shows that the adjusted R^2 value of the models vary between 0.464422 and 0.936911 and the highest values are observed in the 1st and the 7th models. Then, the residuals sum of squares, AIC and SIC values of these models were compared. The values of the aforementioned criteria for the 7th model are smaller than those of the 1st model. As a result, the 7th model, namely SARIMA $(2,0,1)(1,0,1)_{12}$, was selected as the suitable and the final forecasting model. However, it should be noted that the graph of the residuals should be examined to check the appropriateness of the final forecasting model. The graphs obtained via Minitab are presented in Fig. 3.

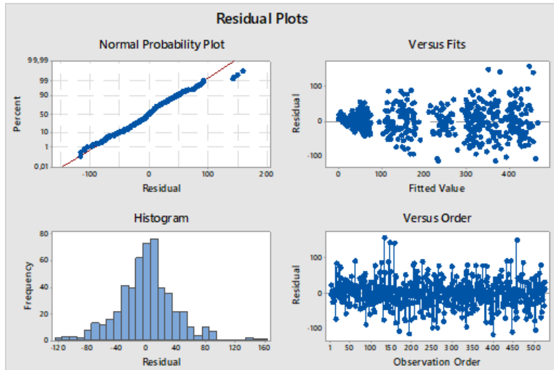


Fig.3. Plots of the HDD series' residuals according to the selected model.

It should be noted that the residuals should be normally distributed with zero mean and constant variance. The normal probability plot and histogram in Fig. 5 shows that residuals except for the last residuals approximately suit to normal distribution. In addition, versus fits graph shows that the variance tends to increase slightly. It should be noted that this situation can be caused by the large variation of the data in the series. In addition to the plots in Figure 3, ACF and PACF plots of the residuals, which are the most examined plots in the model selection, are presented in Fig. 4 and Fig. 5.

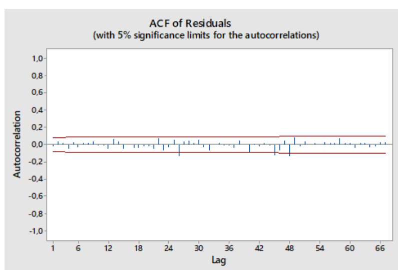


Fig.4. ACF plot of residuals.

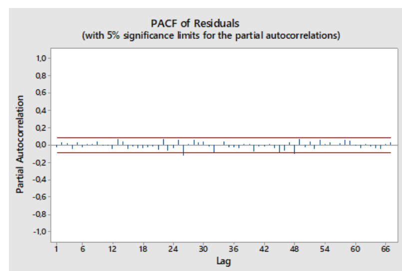


Fig. 5. PACF plot of residuals.

The plots indicate that most of the autocorrelations of the residuals are zero, in other words, they are not correlated. Thus, it can be concluded that the chosen model is appropriate.

The monthly actual HDD values as well as the monthly forecasted HDD values and the confidence intervals obtained via the SARIMA (2,0,1) $(1,0,1)_{12}$ model are shown in Fig. 6. The plot shows that the forecasted HDD values are close to the actual HDD values. In addition, the actual and forecasted HDD values of 2017 are presented in Table 4 which also indicates that the forecasted HDD values are close to the actual HDD values.

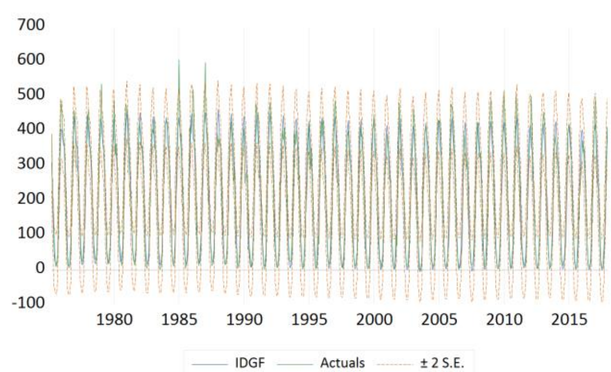


Fig. 6 The actual values, the forecasted values and the confidence intervals of the HDD data according to model SARIMA (2,0,1) $(1,0,1)_{12}$.

Table 4. The actual and forecasted HDD values for 2017

Months	Actual (kWh)	Forecasted (kWh)
2017M01	497.52	417.986
2017M02	303.62	361.840
2017M03	248.77	308.607
2017M04	233.89	217.635
2017M05	95,36	111,711
2017M06	14,73	32,316
2017M07	4,98	4,335
2017M08	10,36	5,148
2017M09	78,48	48,420
2017M010	117,86	151,317
2017M011	321,16	294,122
2017M012	411,22	395,733

4. Conclusion

In this study, time series analysis was conducted by the monthly HDD data in France between 1974 and 2017. Seventy nine SARIMA models were constructed and the models were evaluated according to the adjusted R^2 value, residual sum of squares, AIC and SIC criteria as well as the analysis of the residuals. SARIMA $(2,0,1)(1,0,1)_{12}$ model was selected as the final forecasting model. The results show that the SARIMA models yield fairly acceptable forecasts for supporting short-term forecasting of HDD. Future studies can focus on the integration of these models in the forecasting models of the heating demand in buildings.

References

[1] X. Cao, X. Dai, J. Liu, Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade, Energy Build. 128 (2016) 198–213. https://doi.org/10.1016/j.enbuild.2016.06.089.

[2] EU strategy for heating and cooling, (n.d.). https://ec.europa.eu/energy/en/topics/energy-efficiency/heating-and-cooling.

[3] T.I. Panel, C. Change, U. Nations, E. Programme, T. Ipcc, F.A. Report, T. Report, C. Change, Climate Change: Impacts, Adaptation and Vulnerability, Cambridge University Pres, Canada, 2007. https://doi.org/volume.

[4] M. Mourshed, Relationship between annual mean temperature and degree-days, Energy Build. 54 (2012) 418–425. https://doi.org/10.1016/j.enbuild.2012.07.024.

[5] Q. Meng, M. Mourshed, Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures, Energy Build. 155 (2017) 260–268. https://doi.org/10.1016/j.enbuild.2017.09.034.

[6] A. Özmen, Y. Yılmaz, G.W. Weber, Natural gas consumption forecast with MARS and CMARS models for residential users, Energy Econ. 70 (2018) 357–381. https://doi.org/10.1016/j.eneco.2018.01.022.

[7] D. Trigaux, B. Oosterbosch, F. De Troyer, K. Allacker, A design tool to assess the heating energy demand and the associated financial and environmental impact in neighbourhoods, Energy Build. 152 (2017) 516–523. https://doi.org/10.1016/j.enbuild.2017.07.057.

[8] A. D'Amico, G. Ciulla, D. Panno, S. Ferrari, Building energy demand assessment through heating degree days: The importance of a climatic dataset, Appl. Energy. 242 (2019) 1285–1306. https://doi.org/10.1016/J.APENERGY.2019.03.167.

[9] J.L. Fan, J.W. Hu, X. Zhang, Impacts of climate change on electricity demand in China: An empirical estimation based on panel data, Energy. 170 (2019) 880–888. https://doi.org/10.1016/j.energy.2018.12.044.

[10] N.A. Kurekci, Determination of optimum insulation thickness for building walls by using heating and cooling degree-day values of all Turkey's provincial centers, Energy Build. 118 (2016) 197–213. https://doi.org/10.1016/j.enbuild.2016.03.004.

[11] M. Kohler, N. Blond, A. Clappier, A city scale degree-day method to assess building space heating energy demands in Strasbourg Eurometropolis (France), Appl. Energy. 184 (2016) 40–54. https://doi.org/10.1016/j.apenergy.2016.09.075.

[12] M. Elizbarashvili, G. Chartolani, T. Khardziani, Variations and trends of heating and cooling degree-days in Georgia for 1961–1990 year period, Ann. Agrar. Sci. 16 (2018) 152–159. https://doi.org/10.1016/j.aasci.2018.03.004.

[13] M.J. OrtizBeviá, G. Sánchez-López, F.J. Alvarez-Garcìa, A. RuizdeElvira, Evolution of heating and cooling degree-days in Spain: Trends and interannual variability, Glob. Planet. Change. 92–93 (2012) 236–247. https://doi.org/10.1016/j.gloplacha.2012.05.023.

[14] 2016 Annual Elecetricity Report, France, n.d.

[15] The Statistical Office of the European Union, (n.d.). https://ec.europa.eu/eurostat/web/products-datasets/product?code=nrg_chdd_m.

[16] T. Baldigara, M. Koic, Modelling Occupancy Rates in Croatian Hotel Industry, Int. J. Bus. Adm. 6 (2015) 121–131. https://doi.org/10.5430/ijba.v6n3p121.

[17] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time Series Analysis: Forecasting and Control, 5th Editio, WILEY, 2015.

[18] H.J. Kam, J.O. Sung, R.W. Park, Prediction of daily patient numbers for a regional emergency medical center using time series analysis, Healthc. Inform. Res. 16 (2010) 158–165. https://doi.org/10.4258/hir.2010.16.3.158.