Research Paper

Merve Kuru and Gulben Calis*

Application of time series models for heating degree day forecasting

Abstract: This study aims at constructing short-term forecast models by analyzing the patterns of the heating degree day (HDD). In this context, two different time series analyses, namely the decomposition and Box–Jenkins methods, were conducted. The monthly HDD data in France between 1974 and 2017 were used for analyses. The multiplicative model and 79 SARIMA models were constructed by the decomposition and Box–Jenkins method, respectively. The performance of the SARIMA models was assessed by the adjusted R^2 value, residual sum of squares, the Akaike Information Criteria, the Schwarz Information Criteria, and the analysis of the residuals. Moreover, the mean absolute percentage error, mean absolute deviation, and mean squared deviation values were calculated to evaluate the performance of both methods. The results show that the decomposition method yields more acceptable forecasts than the Box–Jenkins method for supporting short-term forecasting of the HDD.

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 (Cao et al. 2016). It is stated that 50% of the energy is spent on heating and cooling in buildings and industry (*EU Strategy for Heating and Cooling* 2019). Moreover, heating and hot water alone constitute 79% of total final energy use in EU households (*EU Strategy for Heating and Cooling* 2019). The heating demand is affected by several factors such as the building shell, the

type of heating system, outdoor temperature, and occupant behavior. Among all factors, the outdoor temperature is the only one directly affected by climate change (Wu et al. 1992; Neto and Fiorelli 2008; Amber et al. 2018; Li 2018). To understand the variations in the demand for energy required to heat a building due to climate change, a technical index, called heating degree days (HDD) has been developed (IPCC, 2007). HDD is defined as the intensity of coldness (in other words, the need for heating) in a given period of time, considering the outdoor temperature and the average room temperature (Statistical Office of the European Union 2019). The calculation of HDD is based on the base temperature, which is defined as the lowest daily average air temperature that does not result in indoor heating. The value of the base temperature depends, in principle, on various factors related to the building and the environment. Using a general climatic approach, the base temperature is set to a constant 15°C in the HDD calculation. Moreover, it is assumed that the building does not need any heating for any temperature above 18°C. The HDD is calculated over a period of time (typically 1 year) by adding the differences between the daily average temperature and 18°C (weatheronline. co.uk 2019).

If $T_m \leq 15°C$ [HDD = Σi (18°C - T_{im})]; else [HDD = 0]

in which T_{im} is the average air temperature of day *i*. As a result, it can be said that if there are 10% more degree days on any day/week/month/year, 10% more heating energy consumption is expected in that day/week/month/year when all other variables remain the same, and, thus, the amount of energy needed to heat a building in a given frequency is directly related to the number of HDD in that particular frequency. Therefore, HDD can be used to normalize the energy consumption of buildings with respect to heating. In addition, it is indicated that HDD is a more reliable measure of climatic impact on energy consumption than temperature alone (Mourshed 2012). Moreover, models based on HDD can be beneficial to evaluate the impact of design parameters (i.e., area of windows) with less data input (Durmayaz et al. 2000; Durmayaz and

*Corresponding author: Gulben Calis, Ege University, Department of Civil Engineering, Bornova, Izmir, 35040, Turkey.
E-mail: gulben.calis@ege.edu.tr

Kadioglu 2003; Sarak and Satman 2003; Dombayci 2007; Bolattürk 2008; Yu et al. 2009; Idchabani et al. 2015). Accordingly, many researchers integrated HDD in the building energy demand prediction models to improve accuracy (Meng and Mourshed 2017; Trigaux et al. 2017; Özmen et al. 2018). D'Amico et al. (2019) identified the relationship between HDD and heating energy performance. Fan et al. (2019) estimated the impacts of climatic factors including HDD and cooling degree day (CDD) on electricity demand in China. Kurekci (2016) investigated the optimum insulation thickness for building walls by using HDD and CDD values of Turkey's provincial centers. Kohler et al. (2016) 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.

Moreover, several studies focused on investigating the HDD patterns to understand and predict the heating demand of the buildings. Elizbarashvili et al. (2018) estimated daily, monthly, and annual HDD and CDD for 14 different sites of Georgia based on daily mean air temperature data for the 30-year period (1961–1990). The results show that there are significant differences for HDD and CDD among the 14 examined cities of Georgia. OrtizBeviá et al. (2012) estimated trends and interannual 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 two-thirds of the Spanish stations used in the study. Although OrtizBeviá et al. (2012) concluded that these trends are similar to those obtained from observations in other parts of Europe, and France is said to be the most temperature-sensitive country in Europe (*Annual Electricity Report* 2016). In a recent study, SARIMA models were constructed based on HDD data of France between 1974 and 2017. The results show that the SARIMA models yield fairly acceptable forecasts for supporting short-term forecasting of HDD (Kuru and Calis 2019).

This article is based on and is an extended version of a paper presented at the CCC2019 (Kuru and Calis 2019). This study aimed to construct short-term forecast models by analyzing the patterns of the HDD. In this context, two-time series analyses, namely the decomposition and Box–Jenkins methods, were conducted, and the performance of the models was assessed. The following sections describe datasets and the methodology. Next, the findings are presented. Finally, discussion and conclusions are provided.

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 (2019). A total of 528 data covering the period of January 1974 and December 2017 were 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.

2.2 Methodology

A time series consists of four different components, namely trend, seasonal, cyclical, and random components. The individual component analysis of the time series is the basis for the decomposition method. There are two decomposition approaches, which are additive and multiplicative decomposition. In the analysis of consecutive data, the multiplicative decomposition approach is preferred. Therefore, in this study, this approach was used, in which time series can be expressed as the product of the four components of the series and are Y = T.S.C.I, where Y represents the original data, whereas T, S, C, and I represent the trend, seasonal, cyclical, and random

Tab. 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

components of the series, respectively (Calis et al. 2017). In this method, first, the effect of the random change is reduced by applying an exponential correction to the data. Then, the seasonal and trend components of the data are eliminated by calculating the seasonal index and trend equation, respectively. Finally, the model was created assuming that the seasonal and trend components are in multiplication.

The 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 (Baldigara and Koic 2015; Box et al. 2015). 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 (Kam et al. 2010; Baldigara and Koic 2015). 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 nonseasonal linear autoregressive model (AR), nonseasonal difference, and degree of nonseasonal moving average (MA) model, 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 considered 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-squared 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 considered. 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.

To evaluate the performances of the models developed by the decomposition and Box–Jenkins methods, the accuracy of the fitted values was determined by calculating mean absolute percentage error (MAPE), mean absolute deviation (MAD), and mean squared deviation (MSD), which are formulated as follows:

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

$$MAD = \frac{\sum_{t=1}^{n} \left| \left(y_t - \hat{y}_t\right) \right|}{n}$$

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

MAPE corresponds to the relative scale of the prediction error between the forecasted value and the actual value. It should be noted that the smaller the error is in MAPE, the more accurate the prediction is (Kam et al. 2010). Similarly, the smaller the MAD and MSD are, the more accurate the prediction is.

3 Findings

Figure 1 shows the plot of a time series graph to evaluate the overall behavior of the HDD series over time. Figure 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.
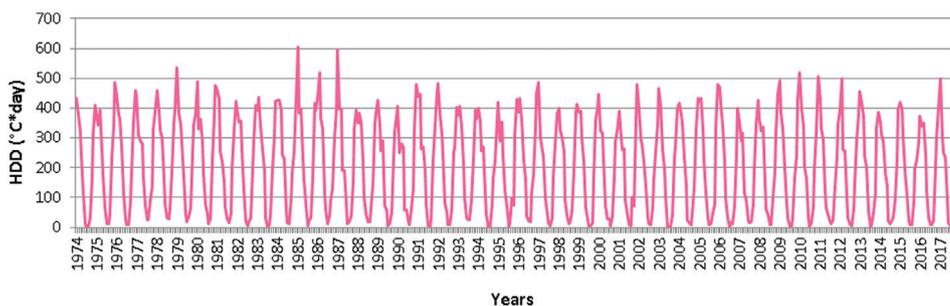

Fig. 1: Time series graph of the HHD series.

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

When the correlogram of the series (Figure 2) is examined, it is observed 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 Figure 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.

3.1 Findings of the decomposition method

In this study, the ratio-to-moving-average method was used for calculating the seasonal index. In this method, first, the moving averages were calculated and the
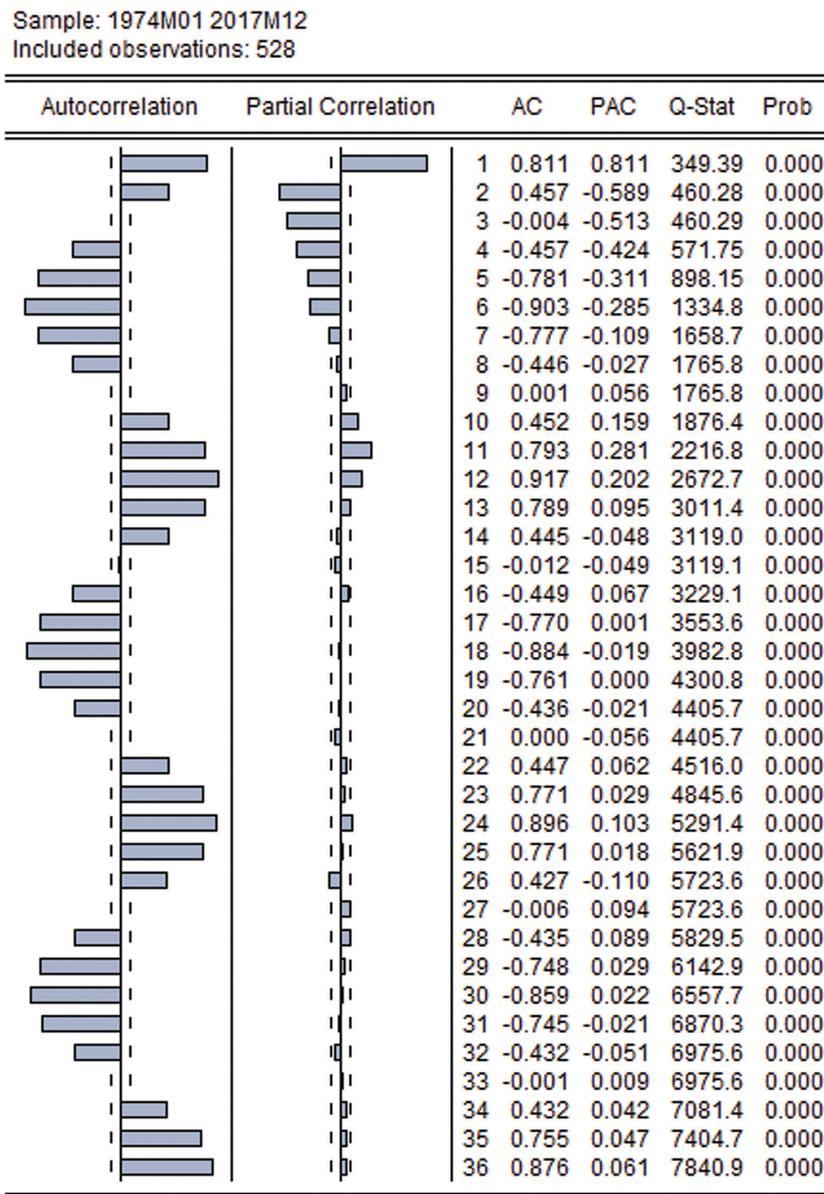
Sample: 1974M01 2017M12
Included observations: 528

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 delays.

effects of seasonal and random changes were eliminated. Since the period numbers were equal to the season type numbers and the months of the year represented the seasons in the HDD data set, 12-months moving averages were calculated. Then, using the calculated moving averages, the central moving averages and the ratios of the actual data to these averages were calculated. Most of the random changes were eliminated by calculating the arithmetic mean of these ratios.

It should be noted that the sum of the calculated averages of each month should be 12.00. However, the fact that the sum of the averages was obtained as 11.99, the correction factor (1.0008) was calculated and applied to the calculated averages of the months. The corrected averages, in other words, seasonal indices, were obtained by multiplying the average ratio value of each month by this correction factor (Table 2). It can be said that the number of HDD in January is 109% higher than the typical index which is 100%, whereas in July and August, it is 94% less.

Next, the actual HDD data were divided into the seasonal indices to obtain the seasonally adjusted data. Actual data and the seasonally adjusted data are shown in Figure 3.

After the seasonal effect was eliminated from the series, the trend effect of the data was determined and eliminated to develop the forecasting model. For this purpose, the trend equation was determined for the seasonally adjusted data. It should be noted that the time

data are numbered as 1, 2, 3..., 528 for the trend equation. Accordingly, the trend line and trend equation were obtained, as shown in Figure 4.

The trend values were calculated by using the trend equation. The trend component was removed from the data set by dividing the seasonally adjusted data by the trend values. The graph of trend corrected data is presented in Figure 5.

After eliminating the trend effect, the random component effect in each time series was removed by the exponential correction method. In this method, the corrected time series data at time t was calculated by substituting the numbers from 0.1 to 0.9 with intervals of 0.1 instead of α in the equation below.

$$Y_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots + (1-\alpha)^{t-1} y_1$$

$$(0 \leq \alpha \leq 1; t = 1,2,3,\ldots n)$$

In addition, the sum of the squares of the differences between the actual data and the corrected data was calculated for each α value. The least squares sum was obtained for $\alpha = 0.9$. Therefore, this α value was chosen for forecasting future observations. As a result of the first-order exponential correction by taking $\alpha = 0.9$, the random component is removed and the graph of the obtained cyclic indices is given in Figure 6.

As a result, since the random component is eliminated, the forecast equation is obtained as follows:

Tab. 2: Monthly rate averages and seasonal indices.

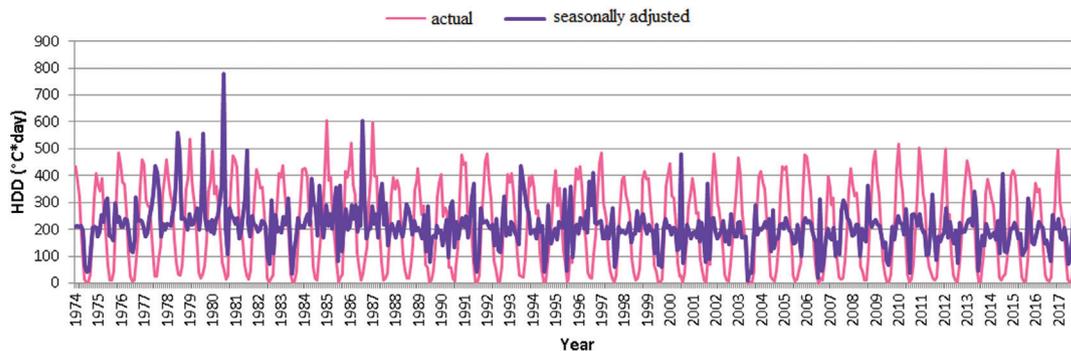|	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Arithmetic mean of ratios	2.09	1.78	1.54	1.13	0.59	0.21	0.06	0.06	0.27	0.79	1.50	1.97	11.99
Seasonal index	2.09	1.78	1.54	1.13	0,59	0.21	0.06	0.06	0.27	0.79	1.50	1.97	12.00
Seasonal index(%)	209	178	154	113	59	21	6	6	27	79	150	197	1200

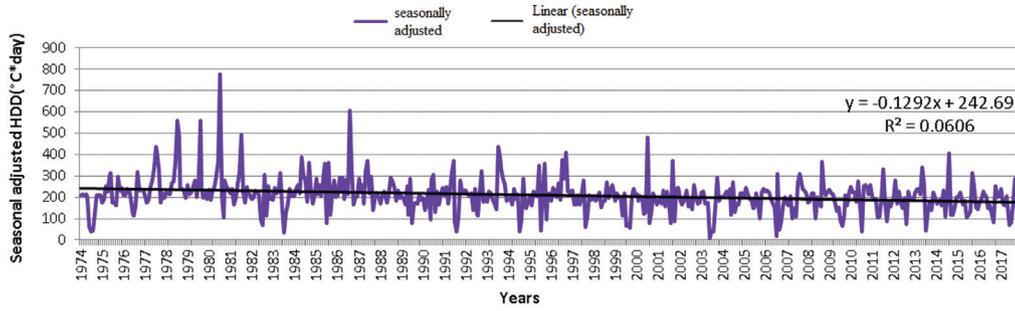**Fig. 3:** The plot of actual and seasonal adjusted HDD data sets.

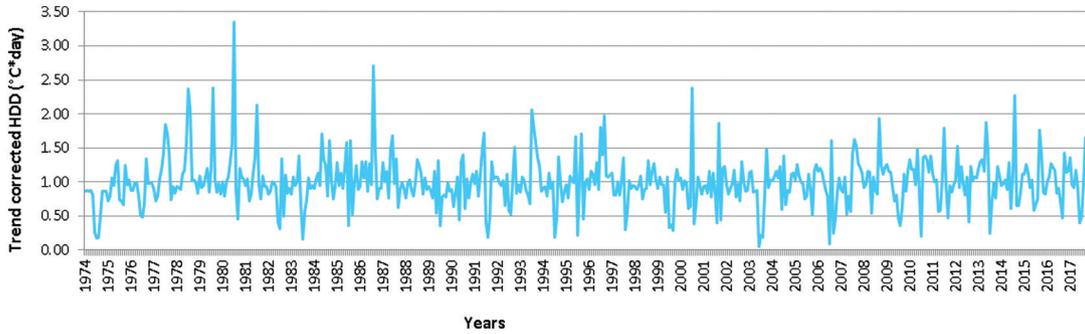**Fig. 4:** The plot of seasonally adjusted data and trend line.

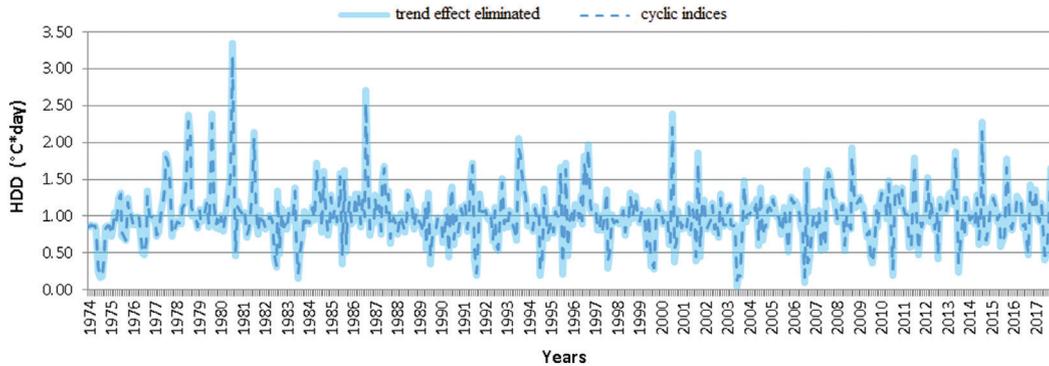**Fig. 5:** The graph of trend corrected data.

Fig. 6: Distribution of cyclic indices.

$$Y_c = T_c * S * T$$

where Y_c represents the forecasted HDD values; T_c represents the trend and the seasonality effect eliminated HDD values; S is the seasonal index and T is the trend index. The actual HDD values and the forecasted values calculated by using the equation are presented in Figure 7.

3.2 Findings of the Box–Jenkins method

In this study, 79 tentative models were generated. Table 3 presents 44 of these models. It should be noted that 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.

As a 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-squared 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, nine 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 4.
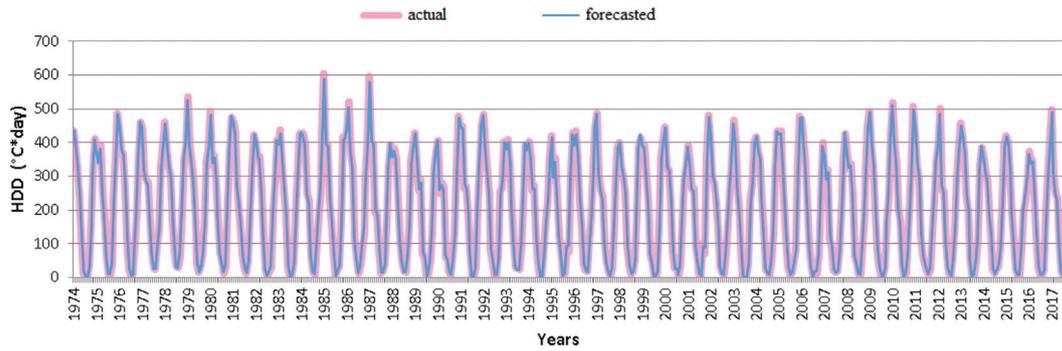
Fig. 7: Actual and forecasted values obtained by the decomposition method.

Tab. 3: 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)$_{12}$

Tab. 4: 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 considered. 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 varies between 0.464422 and 0.936911 and the highest values are observed in the first and the seventh models. Then, the residuals sum of squares and AIC and SIC values of these models

were compared. The values of the aforementioned criteria for the seventh model are smaller than those of the 1st model. As a result, the seventh 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 through Minitab are presented in Figure 8.

It should be noted that the residuals should be normally distributed with zero mean and constant variance. The normal probability plot and histogram in Figure 5 show

Tab. 5: The actual and forecasted HDD values for 2017.

Months	Actual (kWh)	Forecasted (°C*day)	Forecasted (°C*day)
2017M01	497.52	490.2	417.986
2017M02	303.62	315.16	361.840
2017M03	248.77	251.1	308.607
2017M04	233.89	228.96	217.635
2017M05	95.36	97.75	111.711
2017M06	14.73	16.71	32.316
2017M07	4.98	4.95	4.335
2017M08	10.36	9.84	5.148
2017M09	78.48	75	48.420
2017M010	117.86	127.85	151.317
2017M011	321.16	313.33	294.122
2017M012	411.22	411.29	395.733
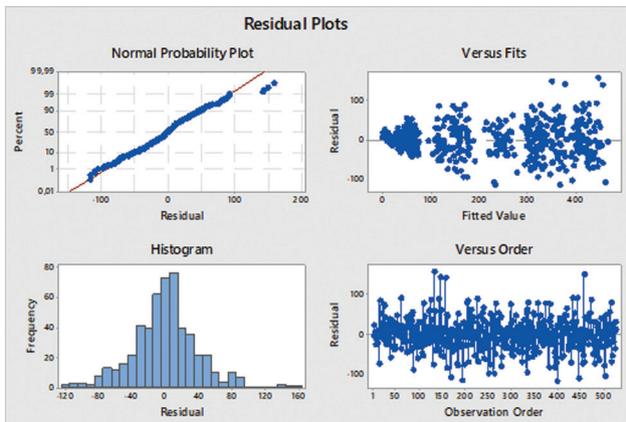
Fig. 8: Plots of the HDD series' residuals according to the selected model.

Fig. 9: ACF plot of residuals.

Fig. 10: PACF plot of residuals.

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 Figures 9 and 10.

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 through the SARIMA $(2,0,1)$ $(1,0,1)12$ model, are shown in Figure 11. The plot shows that the forecasted HDD values are close to the actual HDD values. In addition, the actual and forecasted HDD
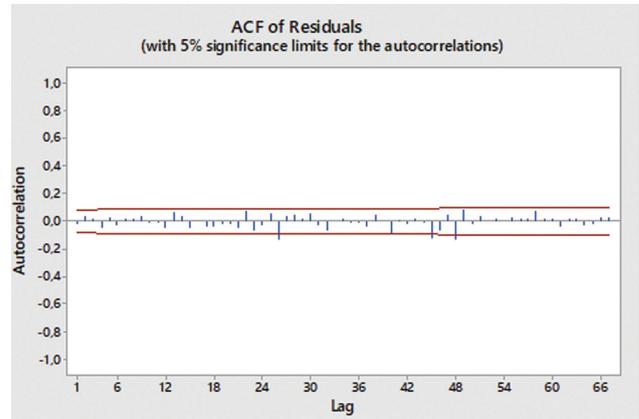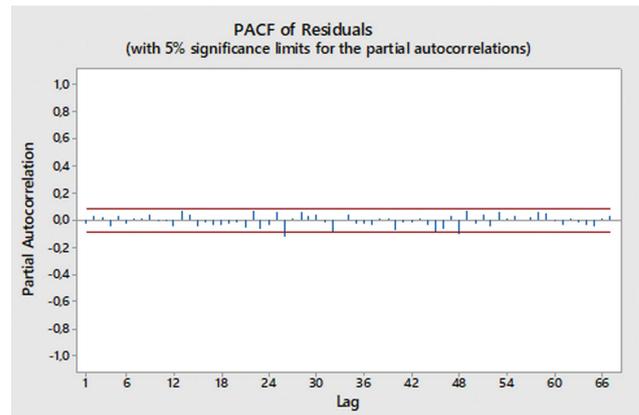
values of 2017 are presented in Table 4, which also indicates that the forecasted HDD values are close to the actual HDD values.

4 Discussion

In this study, two methods, namely the decomposition and Box–Jenkins, are used to forecast the number of HDD of France. The results of both methods show that forecasted values are close to the actual values. The results for the next 12 months are shown in Table 5.

In addition to the comparison of actual and forecasted values, MAPE, MAD, and MSD values have to be evaluated to determine the performance of the model. MAPE, MAD, and MSD values for both the decomposition and the Box–Jenkins methods are shown in Table 6. According to 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." Furthermore, according to Witt and Witt (1992), if the MAPE
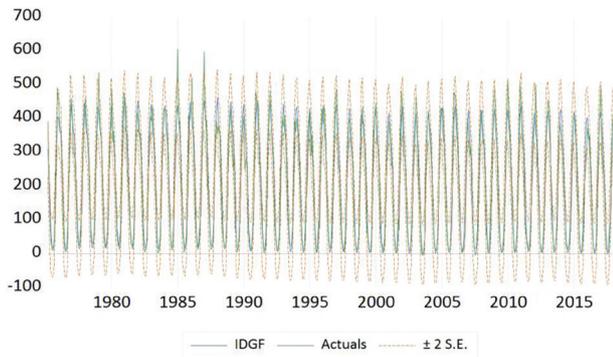
Fig. 11: The actual values, the forecasted values, and the confidence intervals of the HDD data according to model SARIMA (2,0,1) (1,0,1) $_{12}$.

Tab. 6: MAPE, MAD, and MSD values of two methods.

	Multiplicative model	**SARIMA (2,0,1)(1,0,1)$_{12}$ model**
MAPE	3.771	18.7363
MAD	4.3675	19.9735
MSD	33.3013	969.8330

value is ≤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." As can be seen in the Table, the multiplicative model constructed by the decomposition method is "very good" and has a high accuracy degree, whereas the SARIMA model constructed by the Box–Jenkins method is "good" and a "true prediction model."

The results show that the decomposition method yields an acceptable number of HDD forecasts for supporting accurate energy prediction models. It can be concluded that the decomposition method can be preferred to forecast the number of HDD based on time. In addition to its performance, the advantage of this method is its capability of removing irregular changes and seasonal factors within the forecast.

It can be concluded that the decomposition method can be selected to forecast the number of HDD. In addition to its performance, the advantage of this method is its capability of removing irregular changes and seasonal factors within the forecast. For seasonal (monthly, weekly, or quarterly) data, the decomposition methods are often as accurate as of the ARIMA methods and they provide additional information about the trend and cycle, which may not be available in ARIMA methods (NCSS data analysis 2019).

5 Conclusion

In this study, two different time series analysis methods, namely the decomposition and Box–Jenkins, were conducted by the monthly HDD data in France between 1974 and 2017. The multiplicative model and 79 SARIMA models were constructed by the decomposition and Box–Jenkins method, respectively. The SARIMA 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. Finally, MAPE, MAD, and MSD values were calculated to evaluate both multiplicative and SARIMA models. As a result, the SARIMA (2,0,1)(1,0,1)$_{12}$ model was selected as the final forecasting model. In addition, the multiplicative model constructed by the decomposition method is "very good" and has a high accuracy degree, whereas the SARIMA model constructed by the Box–Jenkins method is "good" and a "true prediction model." Therefore, the results show that the decomposition method yields 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

Amber, K. P., Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., & Khan, M. S. (2018). Intelligent techniques for forecasting electricity consumption of buildings. *Energy, 157*, pp. 886-893. doi: 10.1016/j.energy.2018.05.155.

Annual Electricity Report (2016). France.

Baldigara, T., & Koic, M. (2015). Modelling occupancy rates in croatian hotel industry. *International Journal of Business Administration, 6*(3), pp. 121-131. doi: 10.5430/ijba.v6n3p121.

Bolattürk, A. (2008). Optimum insulation thicknesses for building walls with respect to cooling and heating degree-hours in the warmest zone of Turkey. *Building and Environment, 43*(6), pp. 1055-1064. doi: 10.1016/j.buildenv.2007.02.014.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*, 5th edn. Wiley.

Calis, G., Atalay, S. D., Kuru, M., & Mutlu, N. (2017). Forecasting occupancy for demand driven HVAC operations using time series analysis. *Journal of Asian Architecture and Building Engineering, 16*(3), pp. 655-660. doi: 10.3130/jaabe.16.655.

Cao, X., Dai, X., & Liu, J. (2016). Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and Buildings, 128*, pp. 198-213. doi: 10.1016/j.enbuild.2016.06.089.

D'Amico, A., Panno, D., Giuseppina, C., & Ferrari, S. (2019). Building energy demand assessment through heating degree days: The importance of a climatic dataset. *Applied Energy, 242*(December 2018), pp. 1285-1306. doi: 10.1016/J.APENERGY.2019.03.167.

Dombayci, Ö. A. (2007). The environmental impact of optimum insulation thickness for external walls of buildings. *Building and Environment, 42*(11), pp. 3855-3859. doi: 10.1016/j.buildenv.2006.10.054.

Durmayaz, A., & Kadioglu, M. (2003). Heating energy requirements and fuel consumptions in the biggest city centers of Turkey. *Energy Conversion and Management, 44*(7), pp. 1177-1192. doi: 10.1016/S0196-8904(02)00097-3.

Durmayaz, A., Kadoğlu, M., & En, Z. (2000). An application of the degree-hours method to estimate the residential heating energy requirement and fuel consumption in Istanbul. *Energy, 25*(12), pp. 1245-1256. doi: 10.1016/S0360-5442(00)00040-2.

Elizbarashvili, M., Chartolani, G., & Khardziani, T. (2018). Variations and trends of heating and cooling degree-days in Georgia for 1961–1990 year period. *Annals of Agrarian Science, 16*(2), pp. 152-159. doi: 10.1016/j.aasci.2018.03.004.

EU Strategy for Heating and Cooling. (2019). *Journal of Chemical Information and Modeling.* Available at: https://ec.europa.eu/energy/en/topics/energy-efficiency/heating-and-cooling.

Fan, J. L., Hu, J. W., & Zhang, X. (2019). Impacts of climate change on electricity demand in China: An empirical estimation based on panel data. *Energy, 170*, pp. 880-888. doi: 10.1016/j.energy.2018.12.044.

Idchabani, R., Garoum, M., & Khaldoun, A. (2015). Analysis and mapping of the heating and cooling degree-days for Morocco at variable base temperatures. *International Journal of Ambient Energy, 36*(4), pp. 190-198. doi: 10.1080/01430750.2013.842497.

IPCC (2007). In Climate Change 2007: Synthesis Report. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Core Writing Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden P.J. & Hanson C.E. (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kam, H. J., Sung, J. O., & 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. doi: 10.4258/hir.2010.16.3.158.

Kohler, M., Blond, N., & Clappier, A. (2016). A city scale degree-day method to assess building space heating energy demands in Strasbourg Eurometropolis (France). *Applied Energy, 184*, pp. 40-54. doi: 10.1016/j.apenergy.2016.09.075.

Kurekci, N. A. (2016). Determination of optimum insulation thickness for building walls by using heating and cooling degree-day values of all Turkey's provincial centers. *Energy and Buildings, 118*(825), pp. 197-213. doi: 10.1016/j.enbuild.2016.03.004.

Kuru, M., & Calis, G. (2019). Forecasting heating degree days for energy demand modeling. pp. 8-13.

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting.* Butterworth Scientific, London, Boston.

Li, X. X. (2018). Linking residential electricity consumption and outdoor climate in a tropical city. *Energy, 157*, pp. 734-743. doi: 10.1016/j.energy.2018.05.192.

Meng, Q., & Mourshed, M. (2017). Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures. *Energy and Buildings, 155*, pp. 260-268. doi: 10.1016/j.enbuild.2017.09.034.

Mourshed, M. (2012). Relationship between annual mean temperature and degree-days. *Energy and Buildings, 54*, pp. 418-425. doi: 10.1016/j.enbuild.2012.07.024.

NCSSdataanalysis. (2019). *Time Series and Forecasting Methods in NCSS.* Available at https://www.ncss.com/software/ncss/time-series-and-forecasting-in-ncss/.

Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings, 40*(12), pp. 2169-2176. doi: 10.1016/j.enbuild.2008.06.013.

OrtizBeviá, M. J., Sánchez-López, G., Alvarez-Garcìa, F. J., & Ruizdeelvira, A. (2012). Evolution of heating and cooling degree-days in Spain: Trends and interannual variability. *Global and Planetary Change, 92-93*, pp. 236-247. doi: 10.1016/j.gloplacha.2012.05.023.

Özmen, A., Yılmaz, Y., & Weber, G. W. (2018). Natural gas consumption forecast with MARS and CMARS models for residential users. *Energy Economics, 70*, pp. 357-381. doi: 10.1016/j.eneco.2018.01.022.

Sarak, H., & Satman, A. (2003). The degree-day method to estimate the residential heating natural gas consumption in Turkey: A case study. *Energy, 28*(9), pp. 929-939. doi: 10.1016/S0360-5442(03)00035-5.

StatisticalOfficeoftheEuropeanUnion. (2019). *Energy Statistics - Cooling and Heating Degree Days.* Available at ec.europa.eu/eurostat/cache/metadata/fr/nrg_chdd_esms.htm.

The Statistical Office of the European Union. (2019). Available at https://ec.europa.eu/eurostat/web/products-datasets/product?code=nrg_chdd_m.

Trigaux, D., Oosterbosch, B., De Troyer, F., Allacker, K. (2017). A design tool to assess the heating energy demand and the associated financial and environmental impact in neighbourhoods. *Energy and Buildings, 152*, pp. 516-523. doi: 10.1016/j.enbuild.2017.07.057.

Weatheronline.co.uk. (2019). *Heating Degree Days and Cooling Degree Days Indices.* Available at www.weatheronline.co.uk/faq/hdd_cdd.html.

Witt, S. F., & Witt, C. A. (1992). *Modeling and forecasting demand in tourism.* Academic Press Ltd. Butterworth-Heinemann, London.

Wu, J., Reddy, T. A., & Claridge, D. (1992). Statistical Modeling of Daily Energy Consumption in Commercial Buildings Using Multiple Regression and Principal Component Analysis. In: *Proceedings of the Eighth Symposium on Improving Building Systems in Hot and Humid Climates.* Dalla, Texas, pp. 155-164. doi: 10.20595/jjbf.19.0_3.

Yu, J., Yang, J., Tian, L., & Liao, D. (2009). A study on optimum insulation thicknesses of external walls in hot summer and cold winter zone of China. *Applied Energy, 86*(11), pp. 2520-2529. doi: 10.1016/j.apenergy.2009.03.010.