TIME SERIES DECOMPOSITION OF NATURAL GAS CONSUMPTION

[1]MUSTAFA AKPINAR, [2]NEJAT YUMUSAK

[1,2]University Of Sakarya, Department of Computer Engineering
E-mail: [1]akpinar@sakarya.edu.tr, [2]nyumusak@sakarya.edu.tr

Abstract— Natural gas takes place one of the most important topics of energy sector in Turkey. This consumption is divided into subcategories related to the fields. In this research, residential natural gas consumption is studied among these areas. It is known that natural gas usage is affected from seasons and/or cycles. In this research, curve fitting and seasonal affect included decomposition methods are used for demand forecasting. Study fundamentally based on accomplishing daily predictions of the year 2014, on two different datasets between years 2011-2014 and 2011-2013. Unlike few data predictions done in the literature, this study covers excessive data predictions for 365 days. For the first round, data set between 2011-2014 years are used for daily prediction of 2014. Here, additive trend seasonal decomposition method gave the lowest MAPE of 16%. For the second dataset of 2011-2013 years, 26% MAPE is obtained by seasonal additive decomposition method. Trend and seasonal affect included additive decomposition method gave 29% MAPE for this dataset. Additionally, both multiplicative with trend and trend seasonal decomposition methods gave ratio of 27% MAPE. Besides exponential growth trend, linear and quadratic trends tend to fall.

Index Terms—Trend models, time series, decomposition, forecasting, natural gas, consumption, cycling, short term.

I. INTRODUCTION

Time series used in daily life for foreseeing the future are data consisting of numbers and values. They are basically used for predicting conditions of future using the past data or researching the reasons about unusual events occurred in the past. Natural gas consuming could be represented with time series as well. With this approach, long term predictions such as monthly or short term forecasting such as daily are both possible. This will help the reserves to keep under control. Besides natural gas usage is increasing in time, it is also affected from the seasonal changes. Decomposing time series are the most frequently used and the most basic prediction method used for forecasting.

In literature there are variety of studies done about time series decomposition. These works not only contain demand forecasting but also many other related topics. First in the literature, Persons introduced seasonal and cycle concepts in time series [1]. After this research time series get a new direction by Cleveland and Tiao. The researchers developed a model on Cencus X-11 program for decomposing time series [2]. They pointed an additive model with stochastic trend, seasonal and noise component. They showed the results on airway and telephone line data. Theodosiou discussed seasonal-trend decomposition based on loess smooting (STL) approach in his study. STL method is an additive technique and it is relative to other decomposition procedures. The study covers irregular variable within the method as well. This yields a net increase of forecasting accuracy [3]. Li et al. studied time series decomposition on avenue traffic and its benefits [4]. They fundamentally used simple average detrending and principal component analysis based detrending. Researchers explained short-term trends and proved that using multi sensors helps to make more accurate predictions. Liu and Kim studied on detecting real-time stealthy DDoS attacks by time series decomposition method [5]. They focused on this specific technique the fact that detection inefficiency of previously introduced methods. They firstly separated time series on trend and random components, then applied double autocorrelation and improved cumulative sum techniques respectively. By this approach they substantially lower false positives and negatives. Sensation delay is decreased with this approach as well. He et al. researched on year-ahead and hourly prediction of prices written on agreements done at load markets [6]. They also mentioned that obtained results could be used for reliability analysis, day-head scheduling and hour-ahead scheduling. They analyzed continues 6 years hourly gathered data. Prior to analysis they normalized the data. Later, they demonstrated autocorrelations, residuals and observed seasonality and cycles on data.

In this study, forecasting natural gas consumption by using curve fitting with trend models and seasonal based model with historical data are applied. For this purpose, used trend equations and decomposition methods are used.

II. FORECASTING MODELS FOR DEMAND ESTIMATIONS AND ERRORS

In this research, 2 different methodologies are used. Curve fitting with trend modeling is used as a first technique. Secondly, decomposition of time series is studied. Decomposition by multiplicative and additive of time series are computed as analytically.

A. Trend Models

In this section, among trend models 3 methods are used. Trend models traditionally are used to show the tendency in time. They indicate the forthcoming term predictions. They could be either linear, quadratic or exponential. Each trend model has its own characteristic superiorities. Below are the trend models used in this study.

Linear Trend Model: $Y_t = \alpha x_t + \beta + \varepsilon$ (1)

Quadratic Trend Model: $Y_t = \alpha x_t^2 + \beta x_t + \sigma + \varepsilon$ (2)

Exponential Growth Trend Model: $Y_t = Y_0 (1 + \rho)^t$ (3)

In the formulas α, $\beta$and σ represents the trend coefficients while ρ states for growth ratio. x_t shows time steps and ε is the error value between equation prediction and occurrence.

B. Time Series Decomposition

There are 4 modules in time series. These are seasonal component (*S*), trend component (*T*), cycle component (*C*) and irregular component (*E*) which are demonstrated in Eq. 4. [7][8]

$$Y = f\left(S, T, C, E\right) \tag{4}$$

Decomposition of time series have 2 types named as additive and multiplicative model. Multiplicative decomposition mainly multiplies the components one by another (Eq.5). Obviously in the additive decomposition, component predictor variables are added to one another (Eq.6) [7]-[10].

$$Y = S \times T \times C \times E \tag{5}$$

$$Y = S + T + C + E \tag{6}$$

Here trend component indicates long term tendencies. Cycle component indicates longer periodic seasonal movements. Seasonal component represents short periodic oscillations and irregular component represents not expected or couldn't predicted values (Fig. 1).

C. Measuring Prediction Results

Measuring forecasting accuracy is the most important next step after analysis and result gathering are done. Measuring the errors in estimated periods, different statistical diagnostics could be used. Here, the model is evaluated by relative absolute error (RAE), mean absolute percentage error (MAPE), root mean square (RMSE) and standard percent error (PE) [8]-[12].

The mean absolute percentage error (MAPE) is implemented to locate residuals and forecasting errors is showed below.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_{Fi} - y_i}{y_i} 100\% \right| \tag{7}$$

In the equation, *n* stands for the number of observations, $\hat{y}_{Fi}$ stands for the predicted result in the *i*.th time step, while y_i is the actual value in *i*.th time step. Here, absolute value is an important point such that negative and positive percent errors should not eliminate each other.

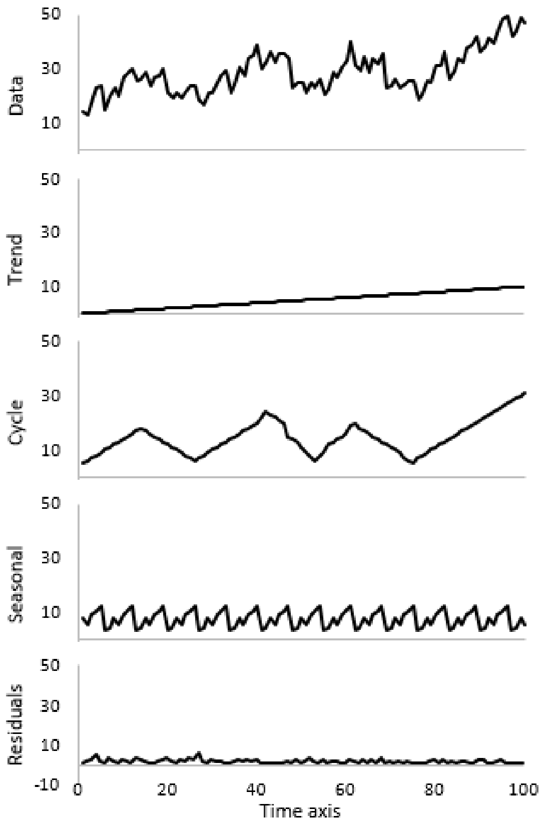

Fig. 1 Time series decompositions [13].

Another coefficient used in the study is adjusted coefficient of determination ($\check{R}^2$). This parameter implies the amount of estimated value corresponds the actual values. Commonly it is showed as follows:

$$\overline{R}^2 = 1 - \left[1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2} \right] \frac{n-1}{n-k-1}, \quad 0 \le \overline{R}^2 \le 1 \tag{7}$$

In the formula *i* represents the *i*.th time, $\bar{y}$ represents the average of actual values, y_i represents accruing value and $\hat{y}_i$ indicates the estimation. Here, $\check{R}^2$ is adjusted by the *(n-1)/(n-k-1)*, where n is data amount and *k* is variable amount. Eventually, as R^2 parameter increases with addition of new values, problem is solved by $\check{R}^2$.

III. MODELING DETAILS

The data contain daily household natural gas consumption and are collected daily in 4-year period. In this study, data are prepared for 2 separate conditions. Plain household natural gas consumption data is obtained by deducting industry users and high consumption users from total daily consumption data. 4-year consumption data varies between 27.765m^3 and 974.960m^3 (Fig. 2). Each year consumption data changes independently. Among the consumptions, the lowest consumption value is 27.765m^3 while the highest is 974.960m^3. In 2011, $5x10^4$ m^3 and under consumptions take 25% of the total consumption. However, this consumption ratio goes down in 2014 and takes only 1% of total consumption. Between $69x10^4$ m^3 and $85x10^4$ m^3, consumption is 1% of 2011, while for the same range consumption increases up to 8% of 2014. Another vital situation is that between $5x10^4$ m^3 and $21x10^4$ m^3 consumption range, the consumption takes 20% of 2011 while in 2014 this ratio sharply rises to 49%. These three incidents prove that consumption has tendency to increase.

Natural gas consumption is affected from environmental situations. For instance, in household consumption, lower temperatures lead consuming more gas, while higher temperatures results decreasing in the consumption rates. This study aims only univariate consumption of 1-year data prediction. 2 independent scenarios are prepared for consumption forecasting. First scenario uses 2011-2014 data to estimate consumption values of 2014. Other, uses 2011-2013 data in order to predict 2014 consumptions. Scenarios are abbreviated as “S1” and “S2”, respectively.

Another point needs to be addressed in the study is consumption tendency is computed with trend models.

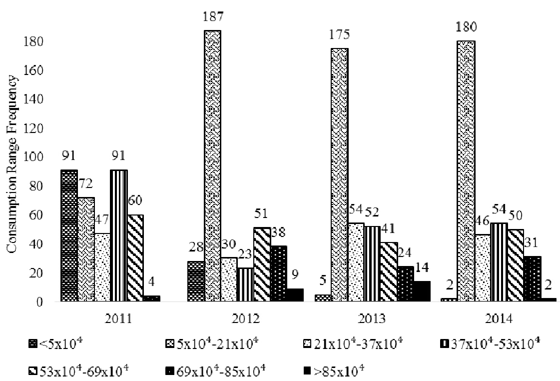

Fig. 2. Histogram graph of consumptions range by year

Trend is calculated by using 3 separate models such as linear, quadratic and exponential growth models. There abbreviates as “TL”, “TQ” and “TE”, respectively. Within this approach, long-term tendencies are targeted.

Decomposition of time series is used for finding estimations on both 2 scenarios. Decomposition of time series is archived by applying “seasonal” and “trend” models. Thus, prediction results could be analyzed without any trend affect. Another fact is having additive either multiplicative component for a used model. For both scenarios and models, additive and multiplicative computations are used together. Multiplicative seasonal, additive seasonal, multiplicative trend-seasonal and additive trend-seasonal are abbreviated as follows: “MDS”, “ADS”, “MDTS”, “ADTS”. The essential goal of this study is determining accuracy of household natural gas demand forecasting by using the naïve time series prediction technique.

Eventually, for 2 different computations, 2 different models and 2 independent scenarios; 8 forecasting methods are created.

Table. I. Decomposition models

Model Name	Seasonal		Trend + Seasonal	
Multiplicative	S1 – MDS	S2 – MDS	S1 – MDTS	S2 – MDTS
Additive	S1 – ADS	S2 – ADS	S1 – ADTS	S2 – ADTS

IV. RESULTS

In the table, trend and decomposition model results are presented separately. Clearly, for both scenarios, linear trend and exponential growth models are inclined to negative (Table 2).

Table. II. Trend Model Results

Scenario	Trend Type	Equation
S1	TL	$Y_t = 299{,}201 – 11.3362\ t$
	TQ	$Y_t = 338{,}094 – 170.8\ t + 0.1091\ t^2$
	TE	$Y_t = 177{,}523\ (1.00001)^t$
S2	TL	$Y_t = 308{,}065 – 37.9703\ t$
	TQ	$Y_t = 374{,}589 – 401.5\ t + 0.3314\ t^2$
	TE	$Y_t = 190{,}206\ (0.999824)^t$

Another outcome is that having bigger coefficient in S2 then S1 in linear trend models, resulted a decrease at consumption in 2014.

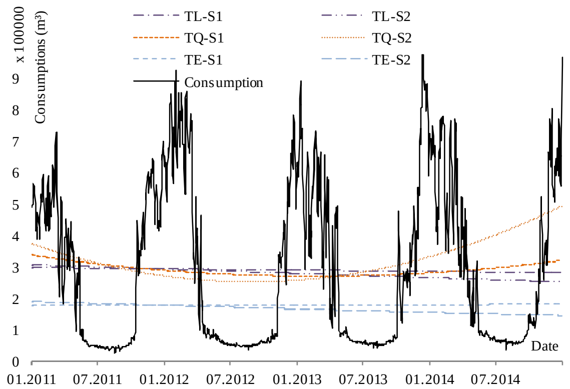

Fig. 3. Consumption trend models

Decomposition method, another step of this study, contains cycle affect in natural gas consumption. Cycle affect is expressed as seasonally within the time series data. In this stage 8 independent estimation are processed. For S1, done on whole data, determination of coefficient is only 0.8333. MAPE is found as 23% for this stage. Used decomposition methods in the prediction have seasonal and trend affects. Since there is a tendency to negative in trend models, models having trend give more accurate results.

Prediction that aims estimating actual data on daily continues 1 year by using actualized data (S1), gave best outcome with 16% MAPE, and 0.868 $Ř^2$ on ADTS method. When data to be estimated is used (S2), 26% MAPE and 0.802 $Ř^2$ values are obtained. However, since annual data

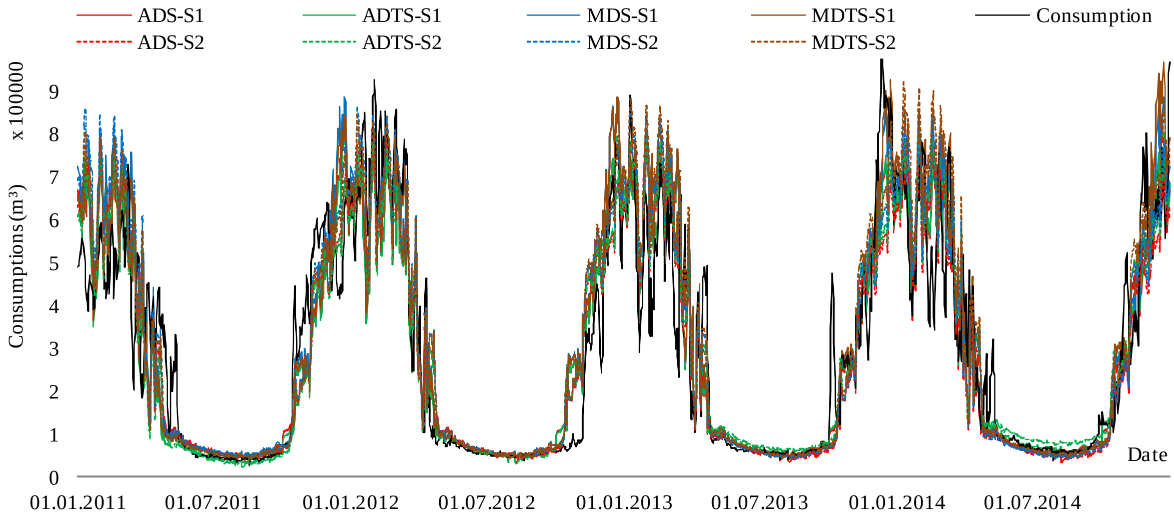

Fig. 4. Consumption decomposition result

to be estimated was not used here, decomposition methods having trend made estimations with higher error rates. Only seasonal included estimation model gave better outcomes

As seen on series, there is not any big variation on data. Here, the lowest MAPE and Ř² are 23% and 0.833 for S1, respectively. MAPE rate did not change for S2 and stayed 26%. Although, Ř² value is increased for whole series up to 0.821.

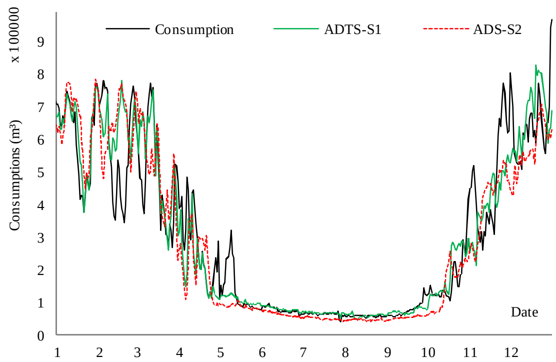

Fig. 5. Best prediction of scenarios

There is an inverse correlation between *Ř²* and MAPE. While MAPE is 198%, *Ř²* is observed as 0.000. On contrary, when MAPE has the lowest value with 16%, *Ř²* is computed as 0.868.

CONCLUSION

This paper mainly focuses on estimation natural gas consumption by implementing trend models and decomposition of time series. When only consumption data presents, decomposition of time series is the simplest way of estimation. Method's ability to apply is proved by achieving 26% consumption rate estimation. Another conclusion is that consumption behavior could be highly determined only by using its own data.

As a future work, other univariate methods such as Simple Smoothing, Exponential Smoothing, Holt's Two Parameter Trend Model. Winters' Three-Parameter Exponential Smoothing, Autoregressive Integrated Moving Averages (ARIMA) will be used on estimation. Besides statistical techniques, learning methods such as Neural Networks will be used in forecasting as well [14][15]. In addition, daily data will be prepared in monthly format and used on other predictions as a following study.

Table. III. MAPE and adjusted coefficient of determination

Date Part	Modelling	Statistics	TL	TQ	TE	ADS	ADTS	MDS	MDTS
For Year 2014	S1	MAPE	150%	160%	95%	19%	16%	20%	21%
		Ř²	0.032	0.009	0.032	0.867	0.868	0.854	0.855
	S2	MAPE	135%	229%	80%	26%	29%	27%	27%
		Ř²	0.032	0.013	0.035	0.802	0.805	0.799	0.804
All Series	S1	MAPE	198%	196%	121%	25%	23%	27%	26%
		Ř²	0.000	0.004	0.000	0.831	0.833	0.831	0.833
	S2	MAPE	191%	211%	114%	26%	27%	27%	27%
		Ř²	0.000	0.001	0.000	0.821	0.824	0.822	0.827

REFERENCES

[1] The Index of General Business Conditions, “The Index of General Business Conditions,” The Review of Economics and Statistics, vol. 1, pp. 3-10, 1919.

[2] W. P. Cleveland and G. C. Tiao, “Decomposition of Seasonal Time Series: A Model for the Census X-11 Program,” Journal of the American Statistical Association, vol. 71(355), pp. 581-587, 1976.

[3] M. Theodosiou, “Forecasting monthly and quarterly time series using STL decomposition,” International Journal of Forecasting, vol. 27, pp. 1178-1195, 2011.

[4] L. Li, X. Su, Y. Zhang, J. Hu, Z. Li, “Traffic Prediction, Data Compression, Abnormal Data Detection and Missing Data Imputation: An Integrated Study Based on the Decomposition of Traffic Time Series,” IEEE 17th International Conference on Intelligent Transportation Systems, pp. 282-289, October 2014.

[5] H. Liu and M. S. Kim, “Real-Time Detection of Stealthy DDoS Attacks Using Time-Series Decomposition,” IEEE International Conference on Communications, pp. 1-6, May 2010.

[6] Y. He, M. Hildmann, G. Andersson, “Modeling the wind power in-feed in Germany by data decomposition and time series analysis,” IEEE Power and Energy Society General Meeting, pp. 1-8, July 2012.

[7] R. Yaffee, M. McGee, “Introduction to Time Series Analysis and Forecasting: With Applications in SAS and SPSS,” Academic Press, pp. 45-63, 2000.

[8] S. A. DeLurgio, “Forecasting Principles and Applications,” McGraw-Hill, pp. 175-190, 1998.

[9] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, “Forecasting Methods and Applications,” John Wiley & Sons, pp. 106-112, 1998.

[10] R. Burtiev, F. Greenwell, V. Kolivenko, “Time Series Analysis of Wind Speed and Temperature in Tiraspol, Moldova,” Environmental Engineering and Management Journal, vol. 12(1), pp. 23-33, January 2013.

[11] R. Jain, A. S. Siddiqui, M. Jamil, C. P. Gupta & Preeti, “Maximising Profits of Electrical Entities Using Load and Price Forecasting,” International Journal of Electrical and Electronics Engineering Research, vol 3(5), pp. 49-58, December 2013.

[12] J.S. Armstrong, F. Collopy, “Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons,” International Journal of Forecasting, vol. 8(1), June 1992, 69-80.

[13] M. Akpinar, N. Yumusak, “Forecasting Household Natural Gas Consumption with ARIMA Model: A Case Study of Removing Cycle,” IEEE 7th International Conference on Application of Information and Communication Technologies, pp. 1-6, October 2013.

[14] R. Deshmukh ,A. Mahor, “Artificial Neural Network Based Approach for short load forecasting,” International Journal of Advanced Computer Research, vol. 1(2), pp.104-108, December 2011.

[15] R. Deshmukh ,A. Mahor, “Comparisons of Short Term Load Forecasting using Artificial Neural Network and Regression Method ,” pp.96-100, December 2011.

★★★