Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Predicting residential electricity consumption using neural networks: *A case study*

IOP Conf. Series: Journal of Physics: Conf. Series **1072** (2018) 012005 doi:10.1088/1742-6596/1072/1/012005

Predicting residential electricity consumption using neural networks: *A case study*

Sang Guun Yoo[1,2,3,*] and Myriam Hernández-Álvarez[1]

[1] Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Ladrón de Guevara, E11-253, Quito, Ecuador
[2] Smart Lab, Escuela Politécnica Nacional, Ladrón de Guevara, E11-253, Quito, Ecuador
[3] Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Av. General Rumiñahui s/n, Sangolquí, Ecuador

Email: *sang.yoo@epn.edu.ec

Abstract. Electricity demand prediction plays an essential role in short-term load allocation and long-term planning for new generation and transmission infrastructures. An accurate prediction also allows to take better decisions in terms of cost and energy efficiency. In this aspect, this paper proposes a model for predicting the electricity consumption of residential area in Seoul using neural network. This work has analyzed several particular characteristics of the aforementioned city to extract variables that could have direct influence in the electricity consumption pattern. Using the extracted variables, this paper could forecast the residential electricity consumption with an average error rate of 2.0375% and it could demonstrate how the elderly population is the parameter that influences with major weight at the moment of forecasting residential electricity consumption. Additionally, the presented work illustrates how executing the supervised learning process using a data set organized by months of the year can reduce considerably the error rate. Furthermore, the analysis or results delivers interesting findings related to the energy consumption in Seoul.

1. Introduction

Electricity demand prediction plays an essential role in short-term load allocation and long-term planning for new generation and transmission infrastructures. An accurate prediction also allows to take better decisions in terms of cost and energy efficiency. In this aspect, responding to the significance of forecasting electricity demand, several works have dedicated their content to this field [1-7]. Some of those works have predicted the energy consumption based on weather variables (e.g. temperature and relative humidity, or season of the year) [3, 4] and others forecasted based on socio-economic and demographic variables [6, 7].

Even though it is important to predict the electricity consumption based on general variables, we believe that it is possible to have a better result if we understand the particular characteristics of the population and zone where the electricity is consumed. To confirm this assumption, this paper proposes the prediction of residential electricity consumption of a specific city i.e. Seoul, the capital city of South Korea, based on real data and unique characteristics of the aforementioned place.

Seoul (officially the Seoul Special City) is the capital and largest city of Republic of Korea. Seoul is considered one of the largest cities in the world with approximately 10 million people living in the city, 12.7 million people living in the urban area, and 25.5 million people living in the metropolitan area.

Additionally, Seoul is considered as a leading and growing global city, resulting from an economic and financial explosion called the Miracle on the Han River which converted it into the fourth biggest metropolitan economy of the world with a GDP of US$845.9 billion in 2014 just behind Tokyo, New York City and Los Angeles.

Even though Seoul has many similarities with most of megacities in the world, it has several particular characteristics (see section 2 for more details). The authors of this work believes that those particular characteristics have a direct effect in the residential electricity consumption in Seoul. They also believe that such particularities can explain the trend change of residential electricity consumption occurred in the last few years (consumption decrease after decades of constant growing) (see Figure 1).

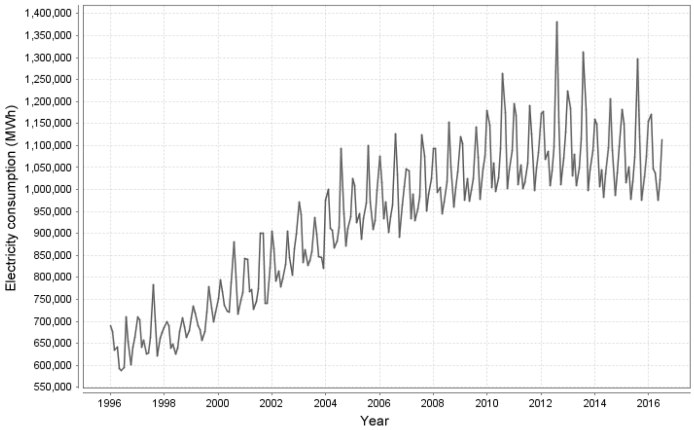

Figure 1. Residential electricity consumption in Seoul, Republic of Korea

With this antecedent, this work proposes the prediction of the residential electricity consumption using variables extracted from aforementioned characteristics. This paper contributes in the field of electricity consumption forecasting in several ways. (1) First, this paper delivers a neural network prediction model for Seoul based on real data. (2) It also analyses the performance of the model when using different data sets (training data grouped by months of the year or single block). (3) Finally, this work makes an analysis of different variables that affect the prediction and determines which of them have the highest level of influence.

The rest of the paper is organized as follows. Section 2 and 3 deliver the hypothesis to be proved in this work and the way how they are going to be proved (methods and metrics). Then, Section 4 analyses the results obtained after executing the modelling process. Finally, Section 5 concludes this paper.

2. Research Scheme and Hypothesis

The research scheme to be used is presented in figure 2. As mentioned before, this work analyses the particular characteristics of Seoul and uses them as variables to predict its residential electricity consumption. Again, those particular characteristics are as follows.

(1) First, Koreans (including people living in Seoul) are working 1.2 times longer on average than their peers in other developed countries. According to the Organization for Economic Cooperation and Development (OECD), Koreans worked on average 2113 hours in 2015; 347 hours more than the OECD average, 323 hours longer than people from United States, and 742 hours longer than Germans [8].

(2) According to the Ministry of Health, Welfare and Family Affairs of Korea, Korean young students study an average of three hours more per day than teenagers in other OECD countries [9]. The Korean adolescents also spent on private tutoring on average one hour and 59 minutes per week which is leaded by a large margin the 22 minutes in Japan, 19 minutes in the United States, 16 minutes in the United Kingdom, and three minutes in Finland [9].

(3) Since several years ago, South Korea is experiencing extremely low fertility and a rapidly increasing number of elderly people. The inequitable social and economic situation in Korean have encouraged people to delay marriage and childbearing [10]. Statistics also indicates that single-person households are increasing rapidly [11].

(4) Seoul is a city with four distinct seasons, and therefore, its weather varies depending on the period of the year. In summer, there is a hot and humid weather with temperatures elevated as high as 35 °C because of the influence of the North Pacific high-pressure system. In winter, the temperature and humidity (as low as -20 °C) drops influenced by Siberian high-pressure and west winds.

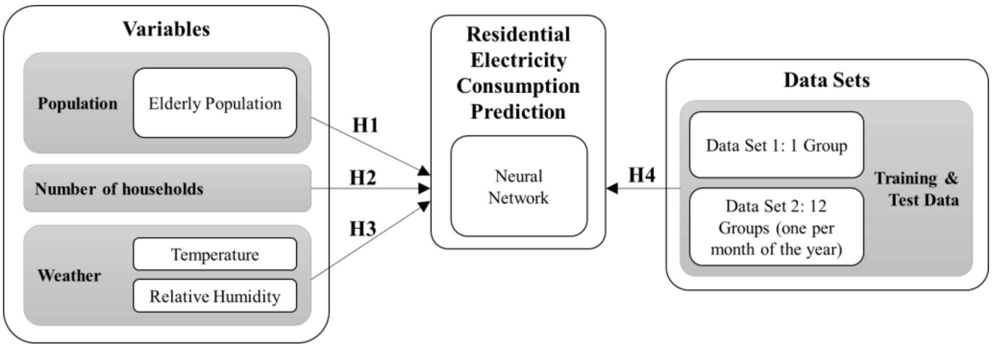

Figure 2. The Proposed Research Scheme

Based on the aforementioned data, it is possible to deduce that students (young population) spent most of their time in schools and private tutoring, while the economically active adults spend most of their time at work. Consequently, it is possible to reason that that elder people are the segment of the population that spends more time at their home. Having this background, this paper hypothesizes the following: H1: Elderly population has an important influence when predicting the electricity consumptions in residential areas of Seoul. Additionally, since the single-person households are increasing, it is possible to hypothesize the following: H2: Number of households influences when predicting the electricity consumptions in residential areas in Seoul. Finally, since Seoul has four different seasons with high variation of temperature and humidity, electricity consumption pattern will depend on the period of the year. Therefore, this work also hypothesizes: H3: Temperature and relative humidity influence when predicting the electricity consumptions in residential areas in Seoul, H4: Electricity prediction based on months of the year will be more precise than the yearly based prediction. In summary, the intention of this work is to confirm the influence of several factors i.e. total population, elderly population, weather (temperature and humidity) and number of households in predicting the electrical consumption in residential areas of Seoul, and to understand which of those factors are the most influential. Additionally, this work pretends to show that prediction based on months of the year is a better option than a prediction based on a yearly organized data set.

3. Research Method

3.1. Data

For forecasting the residential electricity consumption in Seoul, we have collected real data of approximately 20 years i.e. from January 1996 to July 2016. Seoul´s demographic and residential electricity consumption data were delivered by the Seoul Open Data Plaza [12], while the weather data were provided by the Korea Meteorological Administration [13]. All the accumulated data was normalized to use only the following attributes (to be used as variables): date, total population (called as population in the rest of the paper), elderly population (defined by OECD as people aged 65 and over [14]), average temperature (called as temperature in the rest of the paper), average relative humidity (called as humidity in the rest of the paper), and number of households (called as households in the rest of the paper). Once normalized the collected data, it was used to create two different data sets. (1) The first data set was created by dividing the filtered data into training data (data from January 1996 to December 2012) and test data (data from January 2013 to July 2016) (see Figure 3). (2) On the other hand, the second data set was created by dividing the normalized data per month of the year (see Figure 4) and separating them into two groups: one for training purposes (data from January 1996 to December 2012) and another for performance testing purposes (data from January 2013 to July 2016) (see figure 4).

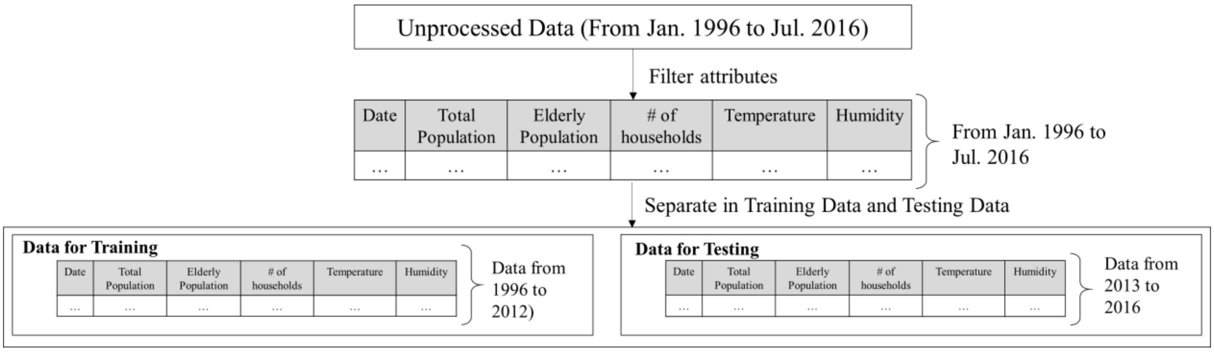

Figure 3. Data Set 1

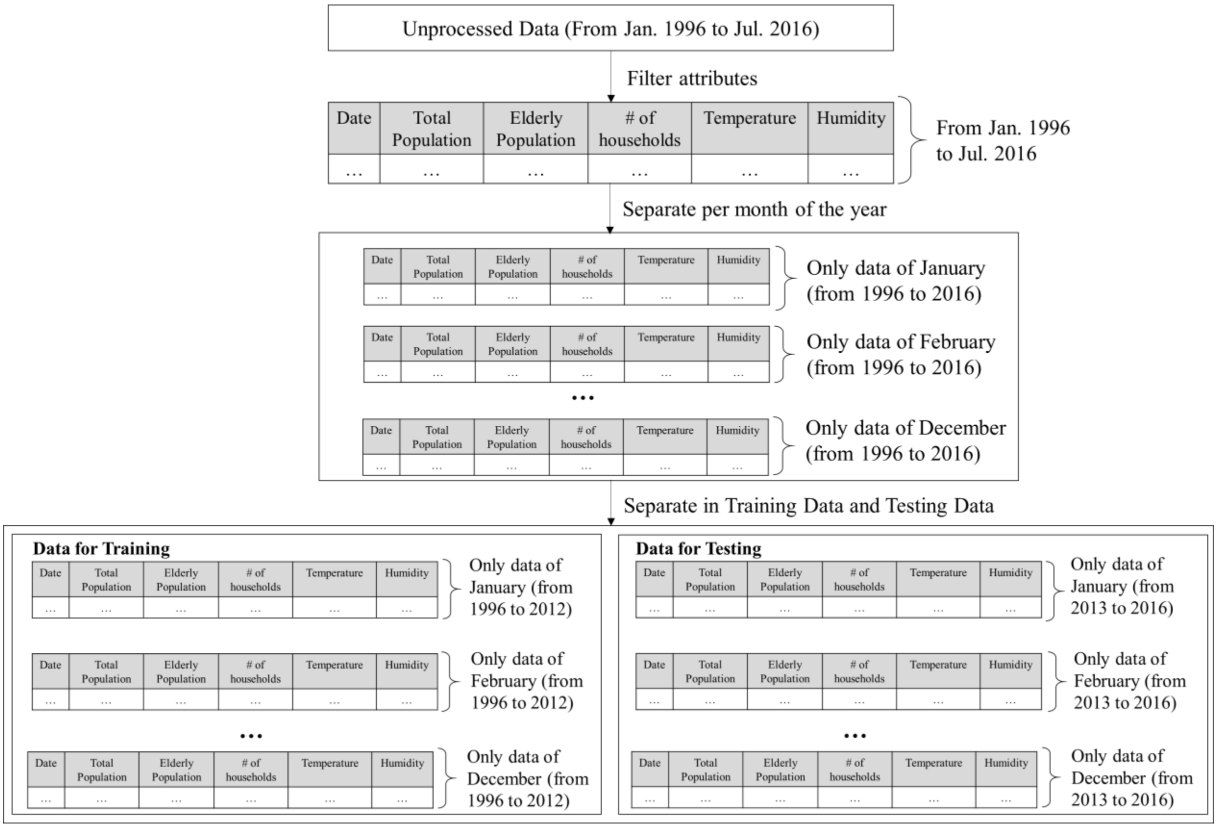

Figure 4. Data Set 2: Data separated per months of the year

3.2. Machine Learning Model and Variables

For this work, we have applied the neural network, since it was the best supervised machine learning model in predicting the electrical energy consumption in Seoul. We executed the prediction process using other models, such as linear regression, polynomial regression, and support vector machine, but the obtained results were less exact than the results gotten using neural networks. Furthermore, the

reasons behind selecting the neural network are its capability to predict future values of more than one variable at the same time and to model the nonlinear relation in the data structure, and because their widespread usage for predicting values [15-17].
Once selected the machine learning mechanism, we defined the variables to use. As mentioned before, this work has selected five variables i.e. temperature, humidity, households, population, and elderly population. To understand, which of those variables has more influence in forecasting the electricity consumption, we have executed the prediction model using all possible combinations i.e. (1) using only elderly population, (2) using only population, (3) using population and elderly population, (4) using only households, (5) using households and elderly population, (6) using population and households, …, (31) using temperature, humidity, households, population and elderly population.

3.3. Performance Metrics

This work has used the average of three metrics i.e. coefficient of variance, mean bias error, and mean absolute percentage of error to measure the performance of the prediction model. This work has decided to use the average of the aforementioned metrics to reduce the limitations that each one of those metrics could have. The details of such metrics are described below.

1) Coefficient of Variance (CV): It defines how much the prediction error varies with respect to the target's mean [18]. Lower value means lower error rate. It is calculated as follows:

$$CV = \frac{\left(\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right)^{1/2}}{\bar{y}} \times 100 \quad (1)$$

where $\hat{y}_i$ is the predicted residential electricity consumption, y_i is the real residential electricity consumption, and is the average residential electricity consumption.

2) Mean Bias Error (MBE): It indicates how a particular model is to over-estimate or under-estimate the actual electricity consumption. Lower value means lower error rate. MBE is calculated as follows:

$$MBE = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)}{\bar{y}} \times 100 \quad (2)$$

3) Mean Absolute Percentage of Error (MAPE): It indicates the percentage of error per prediction [3, 19]. MAPE is calculated as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i}) \times 100 \quad (3)$$

4. Data Analysis and Results

4.1. Analysis of Data Set 1

Table 1 illustrates the prediction results of the neural network of all possible variable combinations using data set 1. Such table shows the combinations of variables and the resulted performance metrics (CV, MBE, MAPE, and average). It also indicates which variable combinations had the lowest rate of error i.e. "order". Since not all the variable combinations resulted with the best error rate, we have selected the three best variable combinations (10% of the set) for analysis (see Table 2). This analysis shows how the lowest average error rate (4.26%) is reached applying the variable combination {temperature, population, elderly population}. The comparison of real consumption against the prediction value is detailed in figure 5.

4.2. Analysis of Data Set 2

Table 3 illustrates the neural network's prediction results of all possible variable combinations using data set 2. Such table shows the combinations of variables and the resulted performance metrics (CV, MBE, MAPE, and average). It also indicates which variable combinations had the lowest rate of error i.e. "order". Since not all the combination resulted with the best error rate, we have highlighted the three best variable combinations (10% of the set) of each month of the year for analysis (see Table 3). The analysis of Table 3 allowed the construction of Table 4 with the variables combinations having the three lowest average error rates for each month of the year. The comparison of real electricity consumption value against the prediction value of the best variable combination of each month of the year is detailed in figure 6.

4.3. Comparison of Prediction Executed using Data Set 1 and Data Set 2

Comparing Table 2 and Table 4, and comparing Figure 5 and Figure 6, it is possible to deduce which data set is better for prediction. The results indicate that the error rates of Table 4 are much lower than error rates of Table 2, which indicates that usage of data separated by months of the year has superior performance. The superiority of prediction using Data Set 2 is also visualized in Figure 6, which shows how the prediction line is much more close to the real values than in Figure 5. This analysis corroborates the hypothesis H4 which indicates "Electricity prediction based on months of the year will be more precise than the yearly based prediction".

4.4. Analysis of Variable Combinations

After verifying the superiority of prediction using data set 2 (per months of the year), we have analyzed which variables were the most influential in forecasting the electricity consumption in residential areas. For this, we have selected the 20% of combinations (6 combinations) that delivered the lowest average error rate and verified which variables were more frequently used in electricity consumption predictions. As shown in Table 5, this analysis indicates that the most influential variables for January are elderly population and population (5 repetitions each), for February are population (5 repetitions) and elderly population/temperature/households (3 repetitions each), for March are humidity/temperature (4 repetitions each) and elderly population (3 repetitions), for April are humidity/temperature (4 repetitions each) and elderly population/humidity (3 repetitions each), for May are elderly population and temperature (4 repetitions each), for June are elderly population (4 repetitions) and humidity (3 repetitions), for July are temperature (6 repetitions) and elderly population/households (4 repetitions each), for August are temperature and elderly population (3 repetitions each), for September are elderly population (5 repetitions) and populations/humidity/temperature (4 repetitions each), for October are number of households (5 repetitions) and elderly population/population (3 repetitions each), for November are population (5 repetitions) and elderly population (4 repetitions), and for December are temperature (4 repetitions), humidity (3 repetitions) and elderly populations (2 repetitions).

This data indicates that depending on the month of the year all the variables have a direct influence in predicting the energy consumption in residential areas of Seoul. This situation corroborates the hypothesis H2: Number of households has influence predicting the electricity consumptions in residential areas in Seoul and H3: Temperature and relative humidity has influence predicting the electricity consumptions in residential areas in Seoul.

Table 5 also shows how the most influential variables for the whole year are the elderly population (43 repetitions), temperature (40 repetitions), and population (30 repetitions). This show how the elderly population is the variable that has the biggest influence in electricity consumption in residential areas of Seoul. Therefore, it is important to take account this parameter for the electricity consumption planning in the metropolitan area of Seoul. This corroborates the correctness of hypothesis H1 which says "Elderly population has important influence in predicting the electricity consumptions in residential areas of Seoul".

IOP Conf. Series: Journal of Physics: Conf. Series **1072** (2018) 012005 doi:10.1088/1742-6596/1072/1/012005

Figure 7 corroborates the hypothesis H1. The total population decrease does not affect directly to the residential electricity consumption since population decrease is neutralized by the elderly population increase.

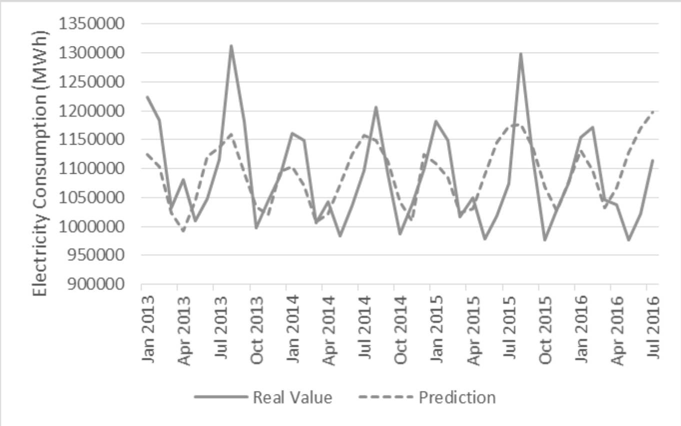

Figure 5. Real Consumption vs Prediction in Best Variables Combination using Data Set 1

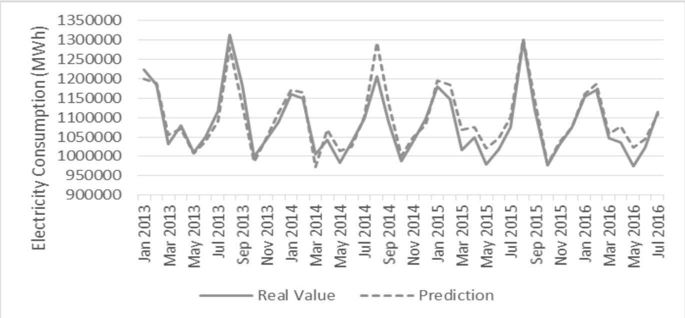

Figure 6. Real Consumption vs Prediction in Best Variables Combination using Data Set 2

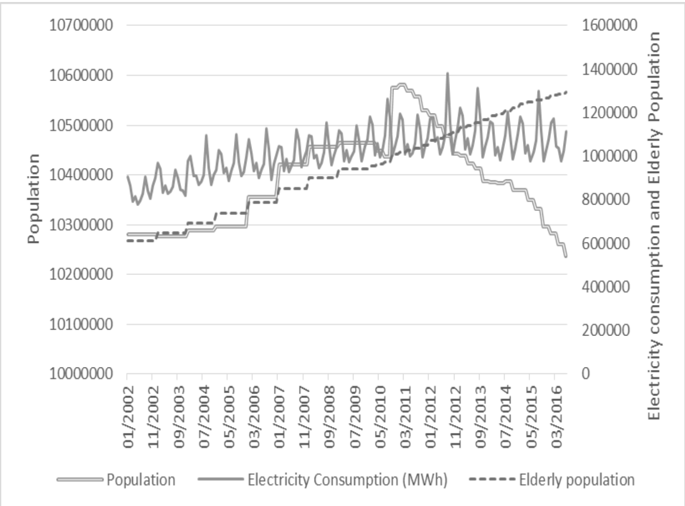

Figure 7. Population vs. Elderly Population in Residential Electricity Consumption

5. Conclusions

This paper has proposed a model for predicting the electricity consumption of residential area in Seoul. This work has analysed several variables that could have direct influence in electricity consumption based on the particularities of the mentioned city. The presented work has demonstrate how the elderly population is the variable that has the biggest influence at the moment of forecasting residential electricity consumption. Additionally, it has shown how the analysis of data separated by months of the year can reduce considerably the error rate. Furthermore, this work has delivered interesting findings related to the energy consumption forecast in Seoul.

IOP Conf. Series: Journal of Physics: Conf. Series **1072** (2018) 012005 doi:10.1088/1742-6596/1072/1/012005

6. References

[1] Hamzacebi C 2007 Forecastirng of Turkey's net electricity energy consumption on sectoral bases *Energy Policy* vol 35 pp 2009-2016

[2] Limanond T. et al. 2011 Projection of future transport energy demand of Thailand *Energy Policy* vol 39 pp 2754-2763.

[3] Dong B et al. 2005 Applying support vector machine to predict building energy consumption in tropical region *Energy and Buildings* vol.37 pp 545-553

[4] Tso G and Yau K 2007 Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks *Energy* vol 20 pp 1761-1768

[5] Ekonomou L 2010 Greek long-term energy consumption prediction using artificial neural networks *Energy* vol 35 pp 512-517

[6] Kankal M et al. 2011 Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables *Applied Energy* vol 88 pp 1927-1939

[7] Sözen A and Arcaklioglu E 2007 Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey *Energy Policy* vol 25 pp 4981-4992

[8] OECD 2016 OECD Data [Online] Available: https://data.oecd.org/emp/hours-worked.htm [Accessed: November 23, 2016]

[9] Chosunilbo 2009 Korean Youth Study Longest Hours in OECD. [Online] Available: http://english.chosun.com/site/data/html_dir/2009/08/10/2009081000200.html [Accessed: November 23, 2016]

[10] Hervery E 2012 Bracing for Low Fertility and a Large Elderly Population in South Korea *Korea Economic Institute Academic Paper Series* April 18.

[11] The Korea Times 2016 Korea faces rapidly aging population [Online] Available: http://www.koreatimes.co.kr/www/news/biz/2016/03/123_201016.html [Accessed: November 23, 2016]

[12] Seoul Metropolitan Government 2016 Seoul Open Data Plaza [Online] Available: http://data.seoul.go.kr [Accessed: November 23, 2016]

[13] Korea Metrological Administration 2016 [Online] Available: http://www.kma.go.kr [Accessed: November 23, 2016]

[14] OECD 2014 Elderly population [Online] Available: https://data.oecd.org/pop/elderly-population.htm [Accessed: November 23, 2016]

[15] Geem Z 2011 Transport energy demand modeling of South Korea using artificial neural network *Energy Policy* vol 39 pp 4644-4650

[16] Lisboa P and Taktak A 2006 The use of artificial neural networks in decision support in cancer: a systematic review *Neural Networks* vol 19 pp 408–415

[17] Murat Y and Ceylan H 2006 Use of artificial neural networks for transport energy demand modeling *Energy Policy* vol 34(17) pp 3165-3172.

[18] Kreider J and Haberl J 1994 Predicting hourly building energy use: the great energy predictor shootout – overview and discussion of results *ASHRAE Transactions* vol 100 (2) pp 1104–1118

[19] Gonzalez P and Zamarreno J 2005 Prediction of hourly energy consumption in buildings based on a feedback artifitial neural network *Energy and Buildings* vol 37 (6) pp 595-601

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Escuela Politécnica Nacional, for the development of the project PIJ-17-08 - "Diseño e implementación de un sistema de parqueadero inteligente".

Table 1. Combinations of Variables and Corresponding Metrics Results using Data Sample 1 (NE=Not Executed)

Variables		Variables combinations															
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
	Humidity	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
	Temperature	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
	# of households	no	no	no	no	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes
	Population	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
	Elderly pop.	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Metric (In %)	**CV**	NE	17.95	10.16	7.74	9.75	8.69	7.80	13.06	6.95	7.82	7.81	6.85	7.85	8.18	8.79	14.23
	MBE	NE	15.61	6.69	0.64	5.83	2.78	0.33	8.68	4.21	0.34	0.50	0.46	0.15	2.71	1.07	11.68
	MAPE	NE	16.51	8.99	6.28	8.59	7.35	6.24	11.41	5.94	6.14	6.28	5.46	6.19	6.72	7.34	12.59
	Average	NE	16.69	8.62	4.88	8.06	6.27	4.79	11.05	5.70	4.77	4.86	4.26	4.73	5.87	5.74	12.83
	Order	NE	**27**	**21**	**13**	**20**	**19**	**8**	**24**	**16**	**7**	**11**	**1**	**5**	**18**	**17**	**25**

Variables		Variables combinations															
		17	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
	Humidity	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
	Temperature	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
	# of households	no	no	no	no	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes
	Population	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
	Elderly pop.	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Metric (In %)	**CV**	12.00	7.78	7.82	8.14	8.00	12.18	15.56	25.80	7.70	6.31	7.31	26.17	7.71	20.14	19.77	7.87
	MBE	8.51	0.50	0.49	1.19	1.65	6.51	10.03	23.26	0.03	1.87	2.05	22.01	0.86	18.60	17.91	0.12
	MAPE	10.45	6.26	6.28	6.25	6.68	10.27	13.12	23.58	6.04	4.85	5.18	22.17	5.49	18.95	18.29	6.23
	Average	10.32	4.85	4.87	5.19	5.44	9.65	12.90	24.21	4.59	4.34	4.85	23.45	4.69	19.23	18.66	4.74
	Order	**23**	**10**	**12**	**14**	**15**	**22**	**26**	**31**	**3**	**2**	**9**	**30**	**4**	**29**	**28**	**6**

Table 2. Three Combinations of Variables with Lowest Error Rates and Corresponding Metrics Results using Data Sample 1

Variables		Variables combinations		
		12	**26**	**25**
	Humidity	no	yes	yes
	Temperature	yes	yes	yes
	# of households	no	no	no
	Population	yes	no	no
	Elderly pop.	yes	yes	no
Metric (In %)	**CV**	6.85	6.31	7.70
	MBE	0.46	1.87	0.03
	MAPE	5.46	4.85	6.04
	Average	4.26	4.34	4.59
	Order	1	2	3

Table 3. Combinations of Variables and Corresponding Metrics Results using Data Sample 2

Variables		Variables combinations															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Humidity	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
	Temperature	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
	# of households	no	no	no	no	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes
	Population	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
	Elderly pop.	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Jan.	**CV**	NE	6.10	4.17	2.53	6.25	7.58	2.08	6.40	4.84	3.23	3.21	2.54	3.43	4.32	5.41	4.77
	MBE	NE	5.33	4.41	2.11	5.44	7.16	0.37	7.47	4.59	0.99	3.14	1.24	1.67	3.08	5.96	5.11
	MAPE	NE	4.85	3.28	1.72	4.96	5.78	1.32	5.59	3.69	2.62	2.33	2.01	2.89	3.67	4.47	3.80
	Average	NE	5.43	3.95	2.12	5.55	6.84	1.26	6.49	4.37	2.28	2.90	1.93	2.67	3.69	5.28	4.56
	Order	NE	23	19	5	24	26	3	25	20	6	13	4	11	18	22	21
Feb.	**CV**	NE	6.31	3.49	9.53	7.93	11.13	8.94	2.90	3.59	4.43	6.50	7.90	5.00	7.06	3.52	11.82
	MBE	NE	6.69	3.46	10.48	8.54	12.16	9.82	2.73	3.70	4.85	7.19	8.14	5.39	7.20	1.78	13.26
	MAPE	NE	5.04	2.61	7.88	6.43	9.15	7.39	2.03	2.79	3.65	5.41	6.14	4.06	5.43	2.62	9.97
	Average	NE	6.02	3.19	9.29	7.64	10.81	8.72	2.55	3.36	4.31	6.37	7.39	4.82	6.57	2.64	11.68
	Order	NE	14	5	28	21	29	26	2	6	8	15	19	11	16	3	31
Mar.	**CV**	NE	6.15	12.08	12.44	6.45	10.71	12.80	10.06	8.10	5.42	9.91	7.48	9.31	8.24	6.59	14.52
	MBE	NE	6.81	13.66	14.03	7.18	11.96	14.41	11.33	9.00	6.03	11.12	8.40	10.42	9.27	7.34	16.25
	MAPE	NE	5.13	10.26	10.54	5.40	8.98	10.82	8.51	6.77	4.54	8.35	6.31	7.83	6.97	5.52	12.20
	Average	NE	6.03	12.00	12.34	6.34	10.55	12.68	9.97	7.96	5.33	9.79	7.40	9.19	8.16	6.48	14.32
	Order	NE	5	24	25	8	22	26	21	15	3	20	12	19	16	9	28
Apr.	**CV**	NE	5.89	10.45	11.34	6.69	9.30	12.01	6.30	3.02	6.19	12.82	7.51	7.96	6.96	9.02	12.34
	MBE	NE	5.97	11.31	12.50	6.88	9.88	13.11	6.70	2.59	6.66	13.85	8.21	8.60	7.55	9.64	13.50
	MAPE	NE	4.52	8.53	9.42	5.20	7.46	9.89	5.06	2.41	5.03	10.45	6.19	6.48	5.70	7.28	10.18
	Average	NE	5.46	10.10	11.09	6.26	8.88	11.67	6.02	2.67	5.96	12.37	7.30	7.68	6.74	8.65	12.01
	Order	NE	5	24	27	9	21	29	7	1	6	31	15	17	11	18	30
May	**CV**	NE	7.17	9.90	5.66	8.09	5.03	6.47	4.87	7.39	4.34	5.65	4.50	5.63	5.05	4.63	8.11
	MBE	NE	7.71	10.87	5.85	8.72	4.97	6.66	4.87	7.98	4.26	5.86	4.40	5.71	5.09	4.50	8.77
	MAPE	NE	5.82	8.19	4.42	6.58	3.76	5.03	3.71	6.01	3.25	4.43	3.37	4.31	3.85	3.46	6.61
	Average	NE	6.90	9.65	5.31	7.80	4.58	6.05	4.49	7.13	3.95	5.31	4.09	5.22	4.66	4.20	7.83
	Order	NE	23	30	14	26	7	19	5	24	2	15	3	11	8	4	27
Jun.	**CV**	NE	3.69	9.59	7.97	5.38	5.93	11.44	9.19	2.93	3.54	11.82	3.56	5.54	3.63	5.02	3.72
	MBE	NE	3.38	10.37	8.81	5.37	6.09	12.56	10.01	2.96	3.68	13.03	3.64	5.86	3.66	5.23	3.84
	MAPE	NE	2.71	7.81	6.63	4.06	4.59	9.45	7.54	2.23	2.77	9.81	2.75	4.42	2.76	3.95	2.90
	Average	NE	3.26	9.26	7.80	4.94	5.54	11.15	8.92	2.71	3.33	11.55	3.32	5.27	3.35	4.73	3.49
	Order	NE	5	27	24	16	21	29	26	3	7	30	6	18	8	14	9
Jul.	**CV**	NE	4.02	9.76	12.38	6.01	6.01	13.91	12.45	7.51	3.46	8.32	5.68	4.08	1.97	3.08	2.46
	MBE	NE	3.77	10.52	13.61	6.14	6.14	15.13	13.78	7.72	2.85	8.72	5.99	3.57	0.06	2.50	1.38
	MAPE	NE	2.88	7.92	10.24	4.64	4.64	11.38	10.37	5.82	2.78	6.57	4.52	3.23	1.44	2.29	1.89
	Average	NE	3.56	9.40	12.08	5.60	5.60	13.47	12.20	7.02	3.03	7.87	5.39	3.63	1.16	2.62	1.91
	Order	NE	10	27	29	18	19	31	30	22	6	26	17	12	1	5	3
Aug.	**CV**	NE	5.48	22.84	36.45	5.20	11.68	35.02	26.73	5.92	8.16	20.53	21.60	13.58	18.11	6.58	37.64
	MBE	NE	3.28	26.83	43.95	2.46	12.89	41.89	32.93	2.98	0.88	25.09	26.04	14.97	22.00	7.81	44.05
	MAPE	NE	3.51	18.07	29.51	3.38	8.75	28.13	22.15	3.77	5.84	16.79	17.44	9.83	14.74	5.17	29.62
	Average	NE	4.09	22.58	36.64	3.68	11.11	35.01	27.27	4.22	4.96	20.80	21.69	12.79	18.28	6.52	37.10
	Order	NE	2	20	28	1	10	24	22	3	4	18	19	11	16	6	29
Sep.	**CV**	NE	7.45	6.97	6.66	7.47	7.88	7.36	11.72	6.88	7.93	9.35	4.52	7.86	5.36	7.20	19.63
	MBE	NE	6.80	6.22	6.04	6.82	7.61	7.19	13.03	5.92	7.67	9.61	1.02	7.50	3.17	6.34	22.58
	MAPE	NE	5.38	5.06	4.67	5.41	5.48	4.92	8.85	5.12	5.51	6.55	3.54	5.52	4.27	5.32	15.25
	Average	NE	6.55	6.09	5.79	6.57	6.99	6.49	11.20	5.97	7.04	8.51	3.03	6.96	4.27	6.29	19.15
	Order	NE	20	15	13	21	23	19	30	14	24	27	1	22	2	16	31
Oct.	**CV**	NE	5.41	6.28	3.25	5.46	2.51	2.45	1.19	3.88	2.12	2.95	3.91	2.77	5.05	2.91	5.53
	MBE	NE	6.27	7.42	3.62	6.31	2.31	2.61	0.80	4.36	1.99	3.14	4.50	2.84	5.87	3.12	6.43
	MAPE	NE	4.19	4.96	2.43	4.22	1.64	1.75	0.93	2.92	1.36	2.10	3.01	1.91	3.92	2.09	4.30
	Average	NE	5.29	6.22	3.10	5.33	2.15	2.27	0.97	3.72	1.82	2.73	3.81	2.51	4.95	2.70	5.42
	Order	NE	22	28	10	23	4	5	2	13	3	8	14	6	21	7	24
Nov.	**CV**	NE	6.34	6.96	0.88	6.66	5.10	1.30	7.36	4.77	4.71	4.49	4.11	5.01	5.34	4.67	6.03
	MBE	NE	7.55	8.38	1.02	7.91	5.96	1.47	8.88	5.67	5.52	5.33	4.91	5.86	6.32	5.48	7.20
	MAPE	NE	5.04	5.59	0.68	5.28	3.98	0.98	5.93	3.79	3.69	3.56	3.28	3.92	4.22	3.66	4.81
	Average	NE	6.31	6.98	0.86	6.62	5.02	1.25	7.39	4.74	4.64	4.46	4.10	4.93	5.29	4.61	6.01
	Order	NE	23	27	1	25	13	2	29	11	10	8	4	12	16	9	20
Dec.	**CV**	NE	4.02	7.49	2.66	4.02	8.49	4.21	15.04	2.09	3.05	3.29	6.49	2.58	3.66	4.48	4.44
	MBE	NE	4.56	8.80	3.15	4.55	10.11	4.86	18.22	0.56	3.10	2.56	7.79	2.85	2.50	4.04	5.22
	MAPE	NE	3.05	5.88	2.10	3.04	6.75	3.25	12.16	1.44	2.05	2.39	5.19	1.90	1.98	2.68	3.48
	Average	NE	3.87	7.39	2.64	3.87	8.45	4.10	15.14	1.36	2.73	2.75	6.49	2.44	2.71	3.73	4.38
	Order	NE	14	24	5	13	25	16	30	2	8	9	20	4	7	11	17

Variables		Variables combinations															
		17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Humidity		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Temperature		no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
# of households		no	no	no	no	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes
Population		no	no	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
Elderly pop.		no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Jan.	CV	7.67	8.94	3.98	1.44	8.61	7.58	3.25	1.62	3.62	3.64	8.95	2.93	3.31	3.13	3.08	4.13
	MBE	7.68	9.25	3.92	0.21	8.80	7.81	3.60	0.26	2.75	2.46	9.49	2.38	2.15	1.55	1.92	2.07
	MAPE	5.82	7.01	3.05	1.12	6.67	5.92	2.69	1.12	2.85	2.40	7.19	2.08	2.03	2.33	2.20	2.80
	Average	7.06	8.40	3.65	0.92	8.03	7.10	3.18	1.00	3.07	2.83	8.54	2.46	2.50	2.34	2.40	3.00
	Order	27	30	17	1	29	28	16	2	15	12	31	9	10	7	8	14
Feb.	CV	4.75	6.16	6.84	3.06	7.88	8.37	4.36	11.60	4.62	6.12	9.05	8.57	7.71	8.74	8.15	2.13
	MBE	4.93	6.51	7.35	2.95	8.45	9.02	4.12	12.90	5.08	6.55	9.75	9.58	8.20	9.64	9.11	2.11
	MAPE	3.72	4.91	5.53	2.23	6.36	6.79	3.12	9.70	3.82	4.93	7.35	7.20	6.17	7.24	6.85	1.59
	Average	4.47	5.86	6.57	2.74	7.56	8.06	3.87	11.40	4.50	5.87	8.72	8.45	7.36	8.54	8.04	1.94
	Order	9	12	17	4	20	23	7	30	10	13	27	24	18	25	22	1
Mar.	CV	5.69	6.32	11.34	8.68	8.42	7.66	6.96	15.53	3.89	7.18	6.26	19.17	7.86	15.68	14.57	3.60
	MBE	6.41	6.99	12.74	9.69	9.39	8.51	7.63	17.45	1.76	8.01	6.91	21.61	8.80	17.54	16.02	3.31
	MAPE	4.82	5.26	9.56	7.28	7.05	6.40	5.74	13.09	3.03	6.02	5.19	16.21	6.61	13.17	12.03	2.50
	Average	5.64	6.19	11.21	8.55	8.29	7.52	6.78	15.36	2.89	7.07	6.12	19.00	7.76	15.46	14.20	3.13
	Order	4	7	23	18	17	13	10	29	1	11	6	31	14	30	27	2
Apr.	CV	4.13	7.61	10.63	7.55	7.89	7.10	9.84	11.14	4.56	5.42	9.00	9.17	6.40	7.71	12.03	9.30
	MBE	4.02	8.00	11.34	8.12	8.18	7.31	10.42	12.07	4.74	5.84	9.67	9.75	6.86	7.87	12.97	10.01
	MAPE	3.09	6.05	8.55	6.13	6.18	5.53	7.87	9.10	3.58	4.40	7.29	7.36	5.18	5.95	9.78	7.55
	Average	3.75	7.22	10.17	7.27	7.41	6.65	9.38	10.77	4.29	5.22	8.65	8.76	6.15	7.17	11.59	8.96
	Order	2	13	25	14	16	10	23	26	3	4	19	20	8	12	28	22
May	CV	6.87	4.07	5.83	5.69	5.33	5.59	6.55	9.80	4.90	5.15	12.87	7.74	5.64	7.25	9.60	6.08
	MBE	7.39	3.91	6.10	5.98	5.29	5.77	6.83	10.68	4.99	5.24	14.22	8.44	5.78	7.64	10.40	6.31
	MAPE	5.57	3.09	4.61	4.51	4.00	4.36	5.15	8.05	3.77	3.96	10.71	6.36	4.36	5.77	7.84	4.76
	Average	6.61	3.69	5.51	5.39	4.87	5.24	6.18	9.51	4.56	4.78	12.60	7.51	5.26	6.89	9.28	5.72
	Order	21	1	17	16	10	12	20	29	6	9	31	25	13	22	28	18
Jun.	CV	2.17	3.36	13.99	5.05	5.73	2.57	6.04	8.88	4.50	4.81	11.07	5.63	5.68	5.32	4.62	4.59
	MBE	1.04	2.91	15.53	5.39	5.78	1.76	6.26	9.45	4.95	5.21	12.00	6.14	6.24	5.78	4.89	4.92
	MAPE	1.71	2.20	11.68	4.06	4.36	1.92	4.72	7.12	3.73	3.92	9.04	4.62	4.70	4.35	3.69	3.71
	Average	1.64	2.82	13.73	4.83	5.29	2.09	5.68	8.48	4.39	4.65	10.71	5.46	5.54	5.15	4.40	4.41
	Order	1	4	31	15	19	2	23	25	10	13	28	20	22	17	11	12
Jul.	CV	8.34	3.97	4.14	6.78	5.65	4.22	8.31	10.13	4.09	3.18	4.82	7.65	2.82	6.72	4.57	4.97
	MBE	7.30	2.58	3.39	7.05	4.95	2.34	8.41	10.43	2.45	1.44	4.38	8.15	0.79	6.40	4.22	3.75
	MAPE	6.51	3.14	3.20	5.32	4.34	3.43	6.35	7.86	3.32	2.27	3.63	6.14	1.91	5.21	3.49	3.99
	Average	7.38	3.23	3.58	6.38	4.98	3.33	7.69	9.47	3.29	2.30	4.28	7.31	1.84	6.11	4.10	4.24
	Order	24	7	11	21	16	9	25	28	8	4	15	23	2	20	13	14
Aug.	CV	7.17	6.48	37.29	35.94	14.46	23.07	35.44	38.30	9.84	9.24	29.13	20.16	16.94	16.04	37.55	14.36
	MBE	7.50	5.43	44.79	43.17	16.28	27.10	42.13	45.80	5.67	11.18	33.56	24.40	20.59	18.52	42.97	17.11
	MAPE	5.12	4.06	30.04	28.98	11.03	18.22	28.24	30.77	5.28	7.51	22.49	16.37	13.79	12.20	28.54	11.36
	Average	6.60	5.32	37.38	36.03	13.92	22.80	35.27	38.29	6.93	9.31	28.39	20.31	17.11	15.58	36.35	14.28
	Order	7	5	30	26	12	21	25	31	8	9	23	17	15	14	27	13
Sep.	CV	10.69	8.52	8.03	6.69	9.29	6.64	7.48	8.08	7.92	6.89	7.55	7.04	7.39	10.18	10.42	7.63
	MBE	8.58	4.91	5.24	2.38	6.32	4.72	2.15	4.26	3.84	2.28	3.30	1.56	3.42	8.51	8.58	2.35
	MAPE	6.76	5.57	5.63	4.87	6.13	4.91	4.85	4.73	4.98	4.77	5.24	4.87	5.27	6.78	6.65	4.97
	Average	8.68	6.33	6.30	4.65	7.25	5.42	4.82	5.69	5.58	4.65	5.36	4.49	5.36	8.49	8.55	4.98
	Order	29	18	17	4	25	10	6	12	11	5	9	3	8	26	28	7
Oct.	CV	5.85	3.82	6.32	3.28	4.24	4.89	1.34	10.42	4.90	5.05	6.40	9.62	4.83	4.95	6.46	3.63
	MBE	6.66	4.14	7.37	3.24	4.73	5.58	0.12	12.37	5.44	5.68	7.35	11.30	5.47	5.47	7.42	4.03
	MAPE	4.46	2.77	4.93	2.17	3.16	3.73	0.95	8.26	3.64	3.80	4.92	7.55	3.66	3.66	4.96	2.70
	Average	5.66	3.58	6.21	2.89	4.04	4.74	0.80	10.35	4.66	4.84	6.22	9.49	4.65	4.69	6.28	3.45
	Order	25	12	26	9	15	19	1	31	17	20	27	30	16	18	29	11
Nov.	CV	10.01	4.55	4.45	6.58	5.57	5.77	5.71	4.51	5.39	7.15	7.11	5.58	6.69	6.56	8.15	4.08
	MBE	10.40	5.28	4.71	7.07	6.09	6.69	6.35	5.30	6.19	7.99	8.34	6.29	7.44	7.24	9.58	4.22
	MAPE	6.96	3.52	3.15	4.73	4.07	4.47	4.24	3.54	4.14	5.34	5.57	4.20	4.97	4.84	6.40	2.82
	Average	9.13	4.45	4.10	6.13	5.24	5.64	5.43	4.45	5.24	6.83	7.01	5.36	6.37	6.21	8.04	3.71
	Order	31	6	5	21	15	19	18	7	14	26	28	17	24	22	30	3
Dec.	CV	4.05	7.33	1.89	12.20	6.33	10.29	12.99	15.54	2.34	3.75	9.41	6.70	3.99	3.89	5.51	6.57
	MBE	4.62	8.70	0.38	14.87	7.41	12.41	15.74	18.64	1.73	2.35	11.45	8.20	2.55	4.59	6.66	8.02
	MAPE	3.09	5.81	1.25	9.92	4.95	8.28	10.50	12.45	1.37	1.94	7.64	5.47	2.23	3.05	4.44	5.35
	Average	3.92	7.28	1.17	12.33	6.23	10.33	13.08	15.54	1.81	2.68	9.50	6.79	2.92	3.84	5.53	6.64
	Order	15	23	1	28	19	27	29	31	3	6	26	22	10	12	18	21

Table 4. Best Combination of Variables and Corresponding Metric Results using Data Sample 2

Month of the Year	Best Average Error Rate	Best Variable Combination (number / combination)
January	0.92	20 / humidity, population, elderly population
	1.00	24 / humidity, households, population, elderly population
	1.26	7 / households, total population
February	1.94	32 / humidity, temperature, households, population, elderly population
	2.55	8 / households, population, elderly population
	2.64	15 / temperature, households, population, elderly population
March	2.89	25 / humidity, temperature
	3.13	32 / humidity, temperature, households, population, elderly population
	5.33	10 / temperature, elderly population
April	2.67	9 / temperature
	3.75	17 / humidity
	4.29	25 / humidity, temperature
May	3.69	18 / humidity, elderly population
	3.95	10 / temperature, elderly population
	4.09	12 / temperature, population, elderly population
June	1.64	17 / humidity
	2.09	22 / humidity, households, elderly population
	2.71	9 / temperature
July	1.16	14 / temperature, households, elderly population
	1.84	29 / humidity, temperature, households
	1.91	16 / humidity
August	3.68	5 / households
	4.09	2 / elderly population
	4.22	9 / temperature
September	3.03	12 / temperature, population, elderly population
	4.27	14 / temperature, households, elderly population
	4.49	28 / humidity, temperature, population, elderly population
October	0.80	23 / humidity, households, population
	0.97	8 / households, population, elderly population
	1.82	10 / temperature, elderly population
November	0.86	4 / population, elderly population
	1.25	7 / households, total population
	3.71	32 / humidity, temperature, households, population, elderly population
December	1.17	19 / humidity, population
	1.36	9 / temperature
	1.81	25 / humidity, temperature

IOP Conf. Series: Journal of Physics: Conf. Series **1072** (2018) 012005 doi:10.1088/1742-6596/1072/1/012005

Table 5. Frequency of Variables in Combinations with Lowest Error Rate

Month of the year	Variable	First 6 Variable Combinations with Lowest Error Rate						Frequency of variable
		1	2	3	4	5	6	
January	Humidity	yes	yes	no	no	no	no	2
	Temperature	no	no	no	yes	no	yes	2
	# of households	no	yes	yes	no	no	no	2
	Population	yes	yes	yes	yes	yes	no	5
	Elderly population	yes	yes	no	yes	yes	yes	5
February	Humidity	yes	no	no	yes	no	no	2
	Temperature	yes	no	yes	no	no	yes	3
	# of households	yes	yes	yes	no	no	no	3
	Population	yes	yes	yes	yes	yes	no	5
	Elderly population	yes	yes	no	yes	no	no	3
March	Humidity	yes	yes	no	yes	no	yes	4
	Temperature	yes	yes	yes	no	no	yes	4
	# of households	no	yes	no	no	no	no	1
	Population	no	yes	no	no	no	yes	2
	Elderly population	no	yes	yes	no	yes	no	3
April	Humidity	no	yes	yes	yes	no	no	3
	Temperature	yes	no	yes	yes	no	yes	4
	# of households	no	no	no	no	no	no	0
	Population	no	no	no	no	no	no	0
	Elderly population	no	no	no	yes	yes	yes	3
May	Humidity	yes	no	no	no	no	yes	2
	Temperature	no	yes	yes	yes	no	yes	4
	# of households	no	no	no	yes	yes	no	2
	Population	no	no	yes	yes	yes	no	3
	Elderly population	yes	yes	yes	no	yes	no	4
June	Humidity	yes	yes	no	yes	no	no	3
	Temperature	no	no	yes	no	no	yes	2
	# of households	no	yes	no	no	no	no	1
	Population	no	no	no	no	no	yes	1
	Elderly population	no	yes	no	yes	yes	yes	4
July	Humidity	no	yes	no	yes	no	no	2
	Temperature	yes	yes	yes	yes	yes	yes	6
	# of households	yes	yes	yes	no	yes	no	4
	Population	no	no	yes	no	yes	no	2
	Elderly population	yes	no	yes	yes	no	yes	4
August	Humidity	no	no	no	no	yes	no	1
	Temperature	no	no	yes	yes	no	yes	3
	# of households	yes	no	no	no	no	yes	2
	Population	no	no	no	no	no	yes	1
	Elderly population	no	yes	no	yes	yes	no	3
September	Humidity	no	no	yes	yes	yes	yes	4
	Temperature	yes	yes	yes	no	yes	no	4
	# of households	no	yes	no	no	no	yes	2
	Population	yes	no	yes	yes	no	yes	4
	Elderly population	yes	yes	yes	yes	yes	no	5
October	Humidity	yes	no	no	no	no	no	1
	Temperature	no	no	yes	no	no	yes	2
	# of households	yes	yes	no	yes	yes	yes	5
	Population	yes	yes	no	no	yes	no	3
	Elderly population	no	yes	yes	yes	no	no	3
November	Humidity	no	no	yes	no	yes	yes	3
	Temperature	no	no	yes	yes	no	no	2
	# of households	no	yes	yes	no	no	no	2
	Population	yes	yes	yes	yes	yes	no	5
	Elderly population	yes	no	yes	yes	no	yes	4
December	Humidity	yes	no	yes	no	no	yes	3
	Temperature	no	yes	yes	yes	no	yes	4
	# of households	no	no	no	yes	no	no	1
	Population	yes	no	no	no	yes	no	2
	Elderly population	no	no	no	no	yes	yes	2
All months	Humidity	yes	no	yes	no	no	yes	30
	Temperature	no	yes	yes	yes	no	yes	40
	# of households	no	no	no	yes	no	no	25
	Population	yes	no	no	no	yes	no	33
	Elderly population	no	no	no	no	yes	yes	43