



## Profile analysis of daily electricity workloads in Malaysia

Zunna'aim Zolkepley\* and Maman Abdurachman Djauhari

*Department of Mathematics, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia*

Received 21 October 2011, Revised 15 March 2012, Accepted 15 June 2012, Available online 28 June 2012

### ABSTRACT

We learn from literature review that research on statistical modelling of Malaysia daily electricity workload has received very much attention. However, to our knowledge, profile analysis of daily electricity workload had received far less attention. Profile analysis will provide us with daily time series model which is very important in order to understand the statistical model for annual situation. This paper deal with profile analysis and its advantage.

| Data Analysis| Time Series Model| Linear Regression| Ordinary Least-Squared |

© 2012 Ibnu Sina Institute. All rights reserved.  
<http://dx.doi.org/10.11113/mjfas.v8n4.162>

## 1. INTRODUCTION

In this paper, which is mainly focus only on maximum daily half hourly workload in Malaysia, we proposed a method on profile analysis to visualize comparison across group which actually will help researchers to understand better about the data that they are currently investigating on.

Profile analysis is a broad term that used to describe a data analysis which identifies the patterns and characteristics of groups which is very helpful in determining and identifying group difference when traditional statistical analysis may not be able to detect these differences [1].

In the case of time series analysis, it is always involving the uses of autoregressive model (AR), moving average (MA), autoregressive moving average model (ARMA) and also autoregressive integrated moving average (ARIMA) in the process of forecasting while profile analysis in the daily workload of the electricity in Malaysia will help to determined the characteristic and patterns of the data before start to forecasting them.

This paper is organized as follows. In the next section we shall see about the lag plotting of the daily electricity workloads in Malaysia data with the use of profile analysis and how profile analysis will help in forecasting process.

## 2. DATA PREPARATION

There are many researches that had been done in the Time Series modeling on daily load electricity for example Azadeh et. al 2007, who build the model by integrating the neural network, time series and ANOVA. Bernoud et. al 2006, who uses the neural network. El Telbany and F.El Karmi et. al 2008, use the swarm optimization toward the electricity demand. Mahpol et.al 2004 who uses the SARIMA T model and Mirasgedis et. al and Ramanathan et.al for short run forecasting. For the case of electricity consumption in Malaysia, many different papers are available for example Razak et. al 2006, Razak et.al 2007, Zuhaimy and Rosnalini. These authors also discussing the forecast model of maximum electricity consumption in Malaysia. Razak et.al used the ARIMA model while Zuhaimy and Rosnalini used the exponential smoothing for the whole time series data sets. In our previous observation we found out that time series data of maximum electricity daily consumption in Malaysia are not normally distributed. In the data preparation phase, we have encountered with transformation problems where the data cannot be transformed either by using some of the transformation method such as Box-Cox method or by using Tukey's transformation ladder to normalize the data distribution. That is what had inspired us to check on the characteristics and the patterns of the data first before start forecasting them to determined what are the characteristics of the data that influenced the data transformation process and the analysis which used to analyze the data are called as profile analysis. From the figure 1, we can see how the data of maximum daily electricity workload in Malaysia for the whole 1 year between September 2005 until August 2006 were distributed. The lag plot which were build are between

\*Corresponding author at:

E-mail addresses: [djauhari@utm.my](mailto:djauhari@utm.my) (M.A. Djauhari)

the  $L(t-1)$  data came from the original data of 365 days on the horizontal axis against  $L(t)$  data from the same data sets on the vertical axis. Lag plot are done to test the randomness of the data sets, the autocorrelation between the values of the same set, the best fit model for the data which soon will be used for forecasting process and also by using lag plot, outlier can be easily being detected and be removed later to form the stationary time series model [2].

group of data. Clustering the data with the same characteristics might give the important information about how electricity used each day and thus forming patterns throughout the 364 days. From the lag plot of the maximum daily electricity workload, the characteristic of the data are observed first by doing intellectual exploration towards the distribution of the data.

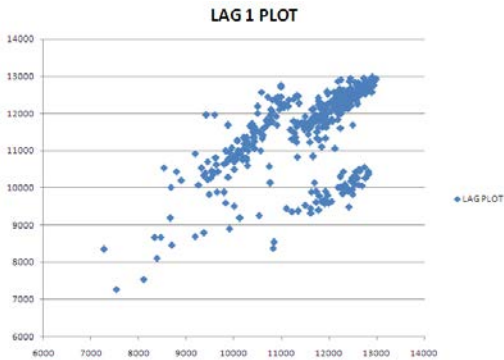


Fig. 1 Lag 1 plot of Maximum Daily Electricity Workload in Malaysia

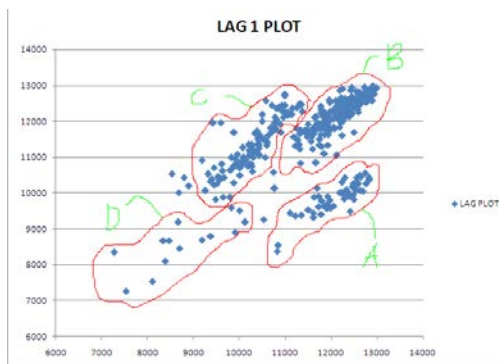


Fig. 2 Clusters produce from Maximum Daily Electricity Workload in Malaysia

During our research we have come up with finding such as the data of maximum daily electricity workload in Malaysia are random and the model for this data has no autocorrelation as the model for the whole 364 daily data after been removed their outlier will produce a cloud of correlation coefficient. Not only that, the lag plotting of the lag 1 data also shows that there are outliers exist in the model proposed for the forecasting method.

### 3. RESULT AND DISCUSSION

#### 3.1 Clustering And Classification

Clustering is one of the techniques used to classify a large of information data sets into manageable meaningful

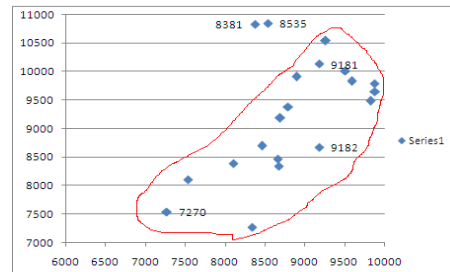


Fig. 3 Cluster D

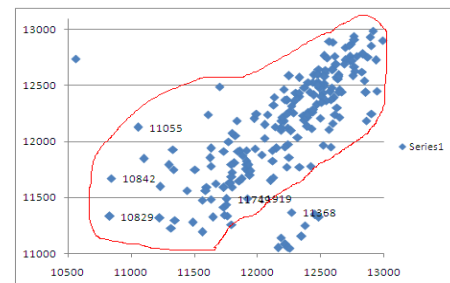


Fig. 3 Cluster B

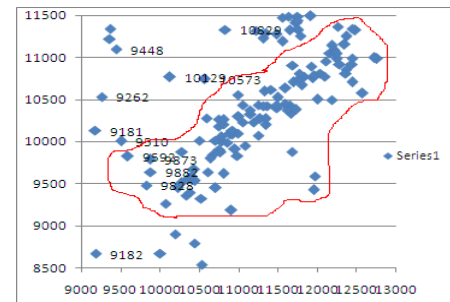


Fig. 3 Cluster C

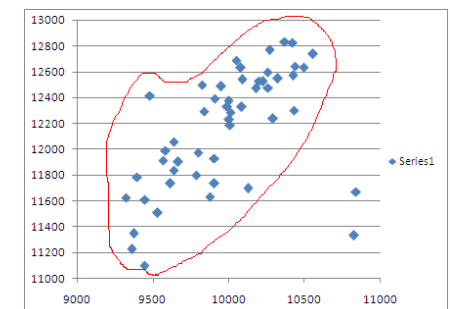


Fig. 3 Cluster A

At first glimpse at the lag plot, we can see that there are almost 4 clusters distributed from the data of maximum daily electricity workload in Malaysia. Then all the clusters

are labelled as cluster A, cluster B, cluster C, and also cluster D as in the figure 2 just by looking at the graph without any implementation of any cluster analysis method yet just to see and investigate what types and characteristics of the data that produce such clusters.

From figure 2, the 4 clusters which can directly be seen from the plotting have each characteristics of what type of day are plotted in the each group of the clusters such as, in cluster A, most of the data distributed in the cluster A are the data that comes from data of Sunday and Monday, cluster C have distributed data mostly from data of Saturday and Friday, while cluster B have distributed data from most of the weekdays data and the last cluster is the cluster D which is the special cluster which contains data from most of the day in the week in that cluster. From our exploration also, we then search for deeper information about what kind of days that selected as the day inside of the cluster of distribution. One of the examples is that data in cluster D which are distributed by the mix data from the entire week type of days but as we can see from the lag plot that the distribution of the data inside the cluster D are distributed far apart from each other like in the figure 3 rather than the other clusters which we can clearly see that the distribution for each cluster are close to each other and the data inside of the clusters have value far more greater than the value of data in cluster D. After certain exploration done to the cluster D, we found out that although the data in cluster D are far apart from each other and have lower maximum electricity workload value but actually the data inside this cluster is consist of only 20 days data but inside of this group there are 5 biggest festival celebrated in Malaysia such as Hari Raya Puasa, Hari Raya Qurban, Chinese New Year, Deepavali and also Christmas. Not only those 5 festivals but most of the data in cluster D is consists of the public holiday and this is what make the cluster D becomes special than others. About the lower value of the maximum daily electricity workload, although this cluster is consist of most the biggest festivals in Malaysia but the an assumption of this event can be made is that during this festival most of the Malaysian will use this opportunity to go back to their village and also it is also public holidays where most of the company and industries sector in Malaysia will stop for certain days for the holidays. This will promotes the less activity which uses the electricity machinery by industries sector and also the less use of air conditioner whether in the workplace or even at home usually in the city area as the air conditioner are the highest selected appliance to be the one which contribute to the highest electric workload in Malaysia. Usually during working days, especially in industries, the electric workload are very high compare to during public holidays as in Malaysia, most of the factories run 24 hour a day by shift. So, during public holidays the factories will be closed down for awhile accordingly to the total amount of holidays set by the Government and the company thus will lowered down the electricity workload on that day.

For cluster B, we can see that most of the data for the whole 364 days data of electricity workload are distributed inside this cluster like in the figure 4 and this is from our previous exploration we know that all the data clustered in cluster B are the data comes from the workdays and during weekdays, electricity usage are higher due the working and the usage of home appliances such as workplace air conditioner and home air conditioner. Not only that, during working days factories, schools, hotels, and supermarkets are all open as usual and even smallest electric equipments such as lamp and fan can also contribute to the high value of electric workload in Malaysia. Air conditioners, lamps and fans themselves in factory consume a lot of electric and there are many factories in Malaysia from small industries for example food industries to the large scale factories such as petroleum and diesel oil plantations.

Not only that, if we take a look at cluster A and cluster C, we can see that this two clusters are not only have cluster of their own clearly by just using our observation but we can also see that cluster C have longer distribution of data rather than in cluster A and this is all because due to the existence of the data from Saturday and Friday in the cluster C are more than the data of Sunday and Monday in cluster A. This is all due to the weekend holiday for certain countries in Malaysia are given not as the same as the major countries for example in the Northern region and Eastern region of Malaysia such as Kedah and Kelantan, these two countries will have weekend holidays on every Friday and Saturday rather than the usual weekend holidays of Saturday and Sunday. Not only that, the cluster A seems to be less distributed with data although the data population in the cluster came from data of Sunday and Monday but actually not every Monday are public holiday like Friday in the countries that have weekend on Friday like in Kedah or Kelantan and this is what makes the data distribution for cluster C is much longer than the distribution for the cluster A like in the figure 5 and figure 6. Not only that, most of Monday in cluster A is actually a public holiday. This is because of in Malaysia whenever there is a public holiday due to event such as Malaysia's Independent Day fall on Sunday then the next Monday will be considered as public holiday as well. Since Mondays is clustered in cluster A due to public holiday, then the value of electricity workload for Monday in cluster A is almost the same as the value of Sunday in the same cluster and that is why those Monday are classified together with Sunday in cluster A. The values of electricity workload for cluster A are much more lower if compare to cluster C is due to public holiday on Sunday and Monday in the cluster A rather than in the cluster C which contains data of Friday and Saturday which only some of the country in Malaysia take those day as public holiday but for other countries Friday is still count as working day and Saturday is a half working day for certain company especially in private sectors and Saturday will be count as public holiday is during the first and the third week.

Table 1 Description for each clusters

Clusters	Description
Cluster A	<ul style="list-style-type: none"> <li>• This cluster is mostly distributed with data of Sunday and Monday</li> <li>• The value of electricity workload in cluster A ranged from 9322 to 12495 kilowatt a day for all 48 days.</li> <li>• The data for Monday in cluster A is a public holiday due to the special public holiday such as Malaysia Independent Day or even Christmas for example that fall on Sunday and the next Monday will be considered as public holiday as well.</li> <li>• This cluster consists of public holidays but just ordinary public holiday and some special public holiday such as Maulidur Rasul and Labour Day.</li> </ul>
Cluster B	<ul style="list-style-type: none"> <li>• This cluster have the largest data distribution among all 4 clusters</li> <li>• Cluster B is consist of all of the day in a week except Sunday and only some of Saturday and Friday data in this cluster due to this cluster is distributed with data comes from working days.</li> <li>• The value of electricity workloads of 201 days are ranged between 10129 kilowatts to 12990 kilowatts a day and most of the data in this cluster have higher value than other clusters and the highest value of electricity workload also lies in this cluster.</li> </ul>
Cluster C	<ul style="list-style-type: none"> <li>• This cluster is majority consist of data from Friday and Saturday but there is a few data from Friday and Monday which can be count as anomalies where the data is clustered to this group is because of near to the public holiday.</li> <li>• The value of electricity workload in this cluster is in ranged between 8535 kilowatts to 12759 kilowatts per day for about 95 days.</li> <li>• In Northern and Eastern part of Malaysia for example Kedah and Kelantan, Friday is consider as public holiday.</li> <li>• For some company especially in private sector Saturday is still a working day but only at first and third Saturday will be considered as public holiday. This what makes the value of electricity workload for Saturday is still high although that Saturday should be considered as a public holiday.</li> </ul>
Cluster D	<ul style="list-style-type: none"> <li>• Cluster D from observation is the smallest group among all cluster since this cluster only have 20 data and in this cluster distributed far apart from each other rather than other cluster which the data are cramp among them.</li> <li>• This cluster have the smallest value of electric workload among other and ranged only between 7270 kilowatts to 9882 kilowatts per day.</li> <li>• This cluster can be consider as special cluster because in this cluster the biggest holidays celebrated in Malaysia are group together for example Hari Raya Puasa, Deepavali, Chinese New Year, and Christmas. There are three major races live in Malaysia which are the Malay, Indian and Chinese with various religions and these will results that there will be many celebration will be celebrated by Malaysian whether it is celebration of religions for example Deepavali which is celebrated by Hindu religion and even celebration due to some special occasion such as celebration of Birthday of Yang Di-Pertua Agong.</li> <li>• Since this cluster consist of the biggest public holiday celebrated by Malaysian so there can be assume that the value of the electricity workload are the lowest among all is due to during this holiday, usually there will be more than one day is given according to each celebration such as Raya Puasa. There usually an addition day of public holiday during Raya Puasa because during this celebration people went home and usually move from town to the countryside for the celebration. During this season also the factories will be closed down for the public holiday for a while and this will promote a lower electricity usage in that certain day.</li> </ul>

The table 1.0 above discuss about the description of each clusters in more details.

#### 4. SELECTION THE NUMBER OF CLUSTERS

##### 4.1 Elbow Criterion

Elbow criterion is one of the methods used to detect the actual maximum number of clusters that can be produce from the whole data set by looking at the percentage of the variance. Before this, exploration done just to see how many cluster should be made from the overall data just by using simple analysis which only by observing the distribution of the data plotting in lag plot. Thus, the initial information of the total number for the cluster from the exploration will be the used as the expected number of maximum cluster which will be detect by this criterion and then by using the cluster analysis method the cluster will be display in the dendrogram form. From the real clusters displayed by the dendrogram we can model each clusters into time series model to be used in forecasting process which will be much better rather than just forecasting the whole data as by using profile analysis first, the outlier for the time series model for each models can be removed and only stationary model will be left for the forecasting process.

#### 5 CONCLUSION

In this paper we show how the profile analysis will help the forecasting process to forecast better rather than just by using the ordinary forecasting steps used by most of the researcher whose research is in the area of Time Series Analysis and forecasting.

#### ACKNOWLEDGEMENT

This research is sponsored by the Ministry of Higher Education of Malaysia under the FRGS vote numbers 4F013 and RUG vote number Q.J130000.7126.02H18. The authors gratefully acknowledge those sponsorships, Government of Malaysia and Universiti Teknologi Malaysia.

#### REFERENCES

- [1] M.Shelton, Basic Concept in Profile Analysis of Means, Texas A&M University,1998
- [2] NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>
- [3] David J. Gerby, Clustering Analysis, Purdue University.
- [4] Zuhaimy, Rosnalini, Short Term Forecasting of Malaysian Electricity Load Demand Based on Day-Type,Universiti Teknologi Malaysia.