The Establishment and Prediction of Regression Models of Energy Sales and System Peak Loading By Considering AMI Data for High Voltage Customers

Cho[*], Ming-Yuan Li[*], Chien-Hsing

Abstract

This paper uses the complex regression analysis method to establish the customer's load regression models, which consider economic indicators, temperature and rainfall. Furthermore, the proposed models are used to study the forecasting feasibility of the future energy sales and summer peak load demand. At first, this paper used least-squares techniques to derive regression models for considering economic indicators and temperature of 34 customer energy sales and total energy sales. Besides, the AMI high voltage customer demand data and system generating capacity for 24 hours were adopted to forecast summer peak load. The above-mentioned data analysis tool is used by EViews software to achieve, in order to verify the feasibility of the research framework. The study found that although its forecasting model accuracy is low only when mixed with temperature and high voltage demands. So, when mixed with high voltage demand data and system generating capacity for 24 hours to forecast peak load, the average error is ± 0.87% and in the majority of its energy sales forecasting model of average error is ±3%. This result can provide power company as future reference.

*National Kaohsiung University of Science and Technology

Key Words : Complex Regression, Least-Squares Techniques, Advanced Metering Infrastructure (AMI)

1 - Introduction

Taiwan Power Company(TPC) accomplished installation of Advanced Metering Infrastructure (AMI) of high voltage customers in 2014; Provision of relatively complete data started in 2015. Demand information of AMI equipped HV users can be used to deduce load curve via proper processing, thereby assisting TPC in performing system load forecast, planning and cost analysis of the power price. Reference [1] provided an effective algorithm to reduce dimensions and time of the massive raw data, and demonstrated practicality and effectiveness of the method. Reference [2] and [3] provided considerations on efficiency factors in the process of load curves and assessed data processing performance.

Current technologies are still unable to store massive electricity effectively, therefore TPC needs to depend on load demand to determine its amount of power generation and to forecast the load by estimating power consumption models of industries. Through load forecast results, TPC will be enabled to implement demand bidding based on the knowledge of customer demands, thereby maximizing benefits of load forecasts.

 doi:10.20944/preprints202104.0622.v1

Referencing domestic and overseas load forecasts and temperature sensitivity analyses, such as that of Reference [4], a study on mid-term load forecast model in TPC's power distribution system, where load forecast models are established using least-squares technique and fussy theory, respectively, employing TPC's district offices as test areas of the mid-term load forecasts. Reference [5] deduced estimation mechanism for electricity sales with data of temperature economics etc. Considered for the relationship analysis and for finding the main factors affecting load changes; finally regression relationships between parameters were derived for studying load characteristics.

Reference [6], using the complex regression analysis method, established customer load regression models considering economic indicators, temperature and rainfall. Models proposed can be used for the feasibility studies of future electricity sales and summer peak load demand. By considering economic indicators, temperature and rainfall, a regression model of 34 customer energy sales and total energy sales was developed. Reference [7], based on collected historical load data of the area in combination with theory and error calibration method, regression analysis method was used to forecast local load development in the next 5 years to come; a new method was finally proposed for elevating the accuracy of load forecast in the future. Reference [8] provided a regression based method to forecast daily peak values; also providing a conversion technique. In order to forecast an accurate annual load, considerations shall be made to seasonal load changes, annual load increases and the latest daily load variations.

Combining the viewpoints of the abovementioned documents, this thesis proposes to estimate trade-based electricity sales and total electricity sales using complex regression of high voltage AMI customers' demand and to select more suitable models to forecast electricity sales. Since the average error of the electricity sales forecast falls within an acceptable range, the same method can thereby be used for estimating summer peak load demand as well as using high voltage AMI customers' demand and 24 hour system generation data; and for studying the applicability of the high voltage AMI customers demand data.

2 - Trade-based complex regression modeling of electricity sales

This paper performs complex regression with 1 independent variable and 2 or more dependent variables. The independent variable selected for model establishment is trade-based power sales of 34 TPC customers [6]; the dependent variables are 9 major economic indices and temperature. These 34 trade-based customers are referred to as TPC bulletins; table-1 below shows the independent variable. The above data are derived from TPC, directorate-general of budget, accounting and statistics of executive yuan, and central weather bureau.

 doi:10.20944/preprints202104.0622.v1

Table-1 Selecting Independent Variables for Modeling by trade

Code	Description	Unit
X_1	Industrial production index	Index: 2011 = 100
X_2	Export po index	Index: 2011 = 100
X_3	Composite leading index	Point
X_4	Monitoring indicator (points)	Point
X_5	Producer inventory index	Index: 2011 = 100
X_6	Producer shipment index	Index: 2011 = 100
X_7	Unemployment rate	%
X_8	Overtime hours	Hour
X_9	Number of employees	Million
X10	Temperature	°C

Table-2 shows the best trade-based regression models derived according to the procedures of this paper. Examples are taken from machinery manufacture and repair industry where the highest fit exists. Actual regression modeling results demonstrated insignificant differences between forecast values and actual values; residual values were within reasonable range, stable and without abrupt changes; maximum fit of the regression model was good, approaching 92%; suitable for being used as a reference model for trade-based power sales, as shown in Fig.1.

Table-2 Combination of representable independent variables by trade

Code	Description of variable (electricity sales)	Fit	Combination of representable independent variables
Y_{14}	Rubber product producer	80%	$X_1X_4X_7X_{10}$
Y_{15}	Plastic product producer	84%	X_7 X_{10}
Y_{17}	Non-metallic mineral product producer	80%	$X_3X_6X_7X_{10}$
Y_{20}	Metal product producer	87%	X_1 X_7 X_{10}
Y_{21}	**Machinery manufacture and repair industry**	**92%**	**$X_1X_4X_7X_{10}$**
Y_{22}	Electronic manufacture	88%	$X_1X_9X_{10}X_4$
Y_{24}	Transportation tools manufacture and repair industry	81%	$X_1X_3X_7X_{10}$
Y_{25}	Other industrial product manufacturer	81%	$X_4X_6X_7X_{10}$
Y_{28}	Wholesale, retail and food industry	81%	$X_2X_3X_7X_{10}$
Y_{30}	Finance, insurance and real estate industries	85%	$X_{10}X_2X_3X_9$
Y_{31}	Industrial and commercial services	81%	$X_{10}X_9X_3X_5$
Y_{32}	Social and individual services	85%	X_{10} X_6 X_9
Y_{33}	Public administration	81%	$X_1X_3X_5X_{10}$

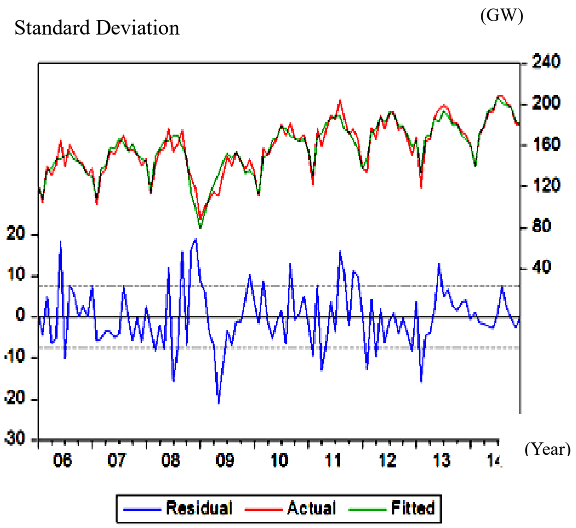

Fig.1 Machinery manufacture and repair industry regression model fit curve

Fit of electronic manufacture, metal products manufacture, financial and insurance and real estate, social and individual services reaches 85% or more. Taking electronic manufacture for an example, complex regression with in-sample forecast performed on trade-based power sales shows fair deviation between forecast values and actual values; residuals within a reasonable range, good maximum fit of regression model approaching 88%; results can be used as reference model of power sales of that trade. See Fig.2.

Combined maximum fit of trade-based power sales in other industrial products, commercial services, public administration services reached over 80%. Taking other industrial products as an example, complex regression with in-sample forecast performed on trade-based power sales shows fair deviation between forecast values and actual values; residuals within reasonable range, good maximum fit of regression model approaching 81%; results can be used as reference model of power sales of that trade. See fig.3.

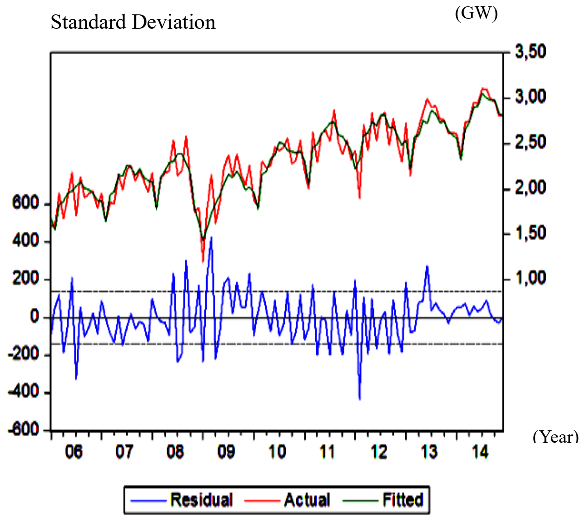

Fig.2 Electronic manufacture industry regression model fit curve

3 - Complex regression modeling of total power sales

In this section we took total power sales (TWh) of TPC as Independent Variable; according to selection criteria of regression modeling, we found that export PO index X_2, employment X_9, and temperature X_{10} were positively affecting total power sales; Table-3 shows deducted regression model, in which forecast value of independent variable can be substituted to obtain total power sales in the future for the reference of TPC. Sources of above data include TPC, directorate-general of budget, accounting and statistics of executive yuan, and central weather bureau.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2021 doi:10.20944/preprints202104.0622.v1

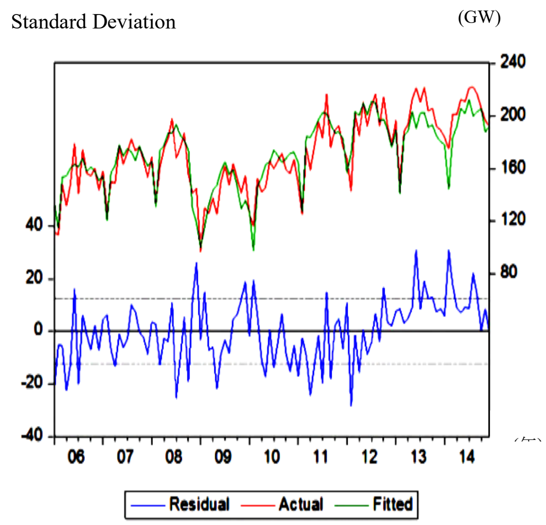

Fig.3 Regression model fit curve of other industrial products industry

Table-3 Regression model established form total power sales

General type power sales	Independent variable	Regression model
Total power sales	X_{10} X_2 X_9	$Y = 1295.72670641 + 277.162013338 * X_{10} + 49.4461804541 * X_2 + 0.000525404829755 * X_9$

Actual results demonstrate that insignificant difference existed between forecast values and actual values; residuals were within reasonable range, there were steady and without abrupt changes; maximum fit of regression model was good and approaching 80%; these can be used as reference model of trade-based power sales in that trade as shown in Fig.4.

4 - Forecasting trade-based power sales and total power sales

I. Forecast results of trade-based power sales

Table-4 shows example of the best fit modeling of the machinery manufacture industry selected by this thesis, with forecast of annual power sales in 2015. average error calculated by dividing the summation of error by the number of total entries was ±3%, this indicated a deviation between actual values and forecast values was insignificant, demonstrating feasibility of the regression modeling for the reference of TPC.

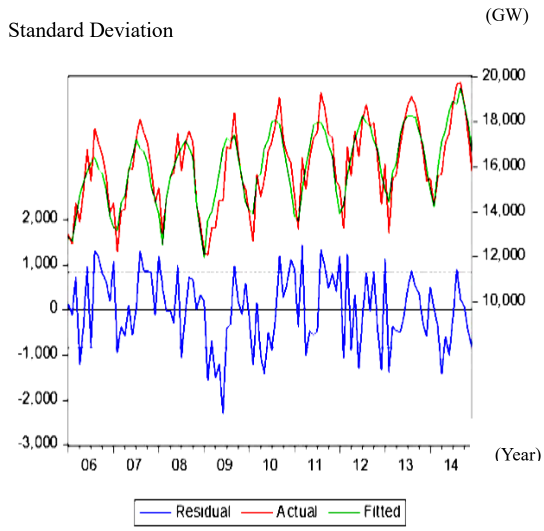

Fig.4 Regression model fit curve of total power sales

II. Forecast results of total power sales

Table-5 shows the forecast of total power sales of 2015 based on the total power sales regression model. Average error calculated by dividing the summation of error by the number of total entries was ±3%, this indicated deviation between actual values and forecast values was insignificant; total power sales can be estimated according to 9 economic indices and temperature, demonstrating feasibility of the regression modeling for the reference of TPC.

Table-4 Comparison between forecast power sales and actual power sales in 2015

Machinery manufacture and repair industry	Actual power sales of tpc (GW)	Forecast value (GW)	Deviation %
Jan.	180.3055	174.4736	3.234445
Feb.	133.3753	145.8986	-9.38957
Mar.	178.8797	186.9345	-4.5029
April	176.7138	186.959	-5.79765
May	191.0349	192.827	-0.93814
June	194.7261	200.4591	-2.94411
July	211.7418	204.642	3.353068
Aug.	193.7257	192.8662	0.44366
Sep.	186.6605	188.8271	-1.16075
Oct.	184.9863	189.8225	-2.61438
Nov.	172.5507	180.3146	-4.49949
Dec.	172.807	176.9324	-2.38731
Average deviation (%)	-2.268		

Forecast results show that the forecast of the total power sales and that of trade-based power sales are both good; probably because of the close co-relations between total power sales changes and economic indices and temperature, regression modeling can thereby be established to obtain fairly good results. Furthermore, a rolling forecast of subsequent years can be performed in the future, predicting trade-based power sales and total power sales of the coming 10 years for the reference of TPC.

Table-5 Comparison between forecast power sales and actual power sales in 2015 (GW)

Period	Actual total power sales of TPC	Forecast of total power Sales	Deviation %
Jan.	15930.386	15746.031	1.157
Feb.	13833.854	14622.614	-5.70
Mar.	15982.186	16468.255	-3.041
April	16035.56	17265.974	-7.673
May	17259.582	17849.142	-3.42
June	17596.702	18684.989	-6.185
July	19728.41	18928.660	4.054
Aug.	19531.714	18523.398	5.162
Sep.	18767.494	19290.039	-2.784
Oct.	18300.347	19026.665	-3.969
Nov.	17300.188	18321.158	-5.902
Dec.	16224.842	16840.474	-3.794
Average Deviation (%)	-2.67		

5 – Estimate peak load via multiple data sources

Power generation is equal to load demand when not considering the line loss. Relationship between high voltage AMI customers' demand and TPC's 24H power generation data was studied using temperature and rainfall data, followed by establishing more explanatory models using complex regression analysis and finally a forecast of TPC's summer peak load was estimated.

(1). Definition of independent variable for estimating peak load

This paper performs complex regression

analysis using 24H system power generation as dependent variable, together with multiple independent variables. Table-6 shows independent variables selected for establishing the model.

Table-6 Independent variables used for estimation of peak load analysis

y_{10}~y_{16}	TPC System 10~16H Power Generation	(MWh)
Hy_{10}~ Hy_{16}	Forecast of 10~16H high voltage demand	(MWh)
Ly_{10}~ Ly_{16}	Forecast of 10~16H low voltage demand	(MWh)
$NEWY_{10}$~ $NEWY_{16}$	Forecast of 10~16H TPC system power generation	(MWh)
Y_{10}~ Y_{16}	Forecast of 10~16H TPC system power generation	(MWh)
T_9~ T_{16}	9~16H Average temperature in entire taiwan	℃
R_9~ R_{15}	9~15H Average rainfall in entire taiwan	mm

(2) Estimate summer peak load using AMI power consumption data and 24h system generation

Estimate peak load using temperature and rainfall data, AMI power consumption data, and TPC's 24H system generation data. Complex regression analysis is used for establishing more explanatory models to estimate the forecast value of the TPC annual peak load. Fig.5 shows the planning procedure.

Study Procedures:

A. Sort out power generation data within the period of 2013-2014 by compiling TPC's 24H power generation data and AMI power consumption data.

B. Simulate to import AMI HV users power consumption data and system 24H power generation data minus low voltage power consumption data of AMI HV users for a relationship analysis of rainfall and temperature, use complex regression analysis to establish a more explanatory model.

C. Obtain forecast value of the summer peak load using the regression model described in this thesis.

D. After comparing with the actual values of the current period, the model is deemed to comply with the forecast requirements if actual values are within the deviation range of forecast values, thereby to select the model with a better fit.

Low voltage consumption forecast data obtained by deducting power consumption data of HV AMI users from 24H system power generation is used for carrying out relationship analyses of temperature and rainfall, and establishing low voltage regression model; results are shown in Table-7. Table-8 shows a summary of the low voltage users regression model.

(3). Peak forecast results

Fig.6 shows the forecast of peak load via high voltage demand data; the average error is ±48%. when forecasting entire system summer peak consumption at the system end using power consumption data of HV AMI users and 24H system generation, the average error is ±0.87% as shown in Fig.7.

Forecast peak load of above possible conditions, summarize both methods by dividing summation of hourly errors with the number of data entries to obtain average error as shown in Fig.9. minimum average error appears in peak load forecast results of the [AMI HV Forecast + LV Forecast] approach, however this approach still requires to combine 24H system generation data for estimating peak load forecasts

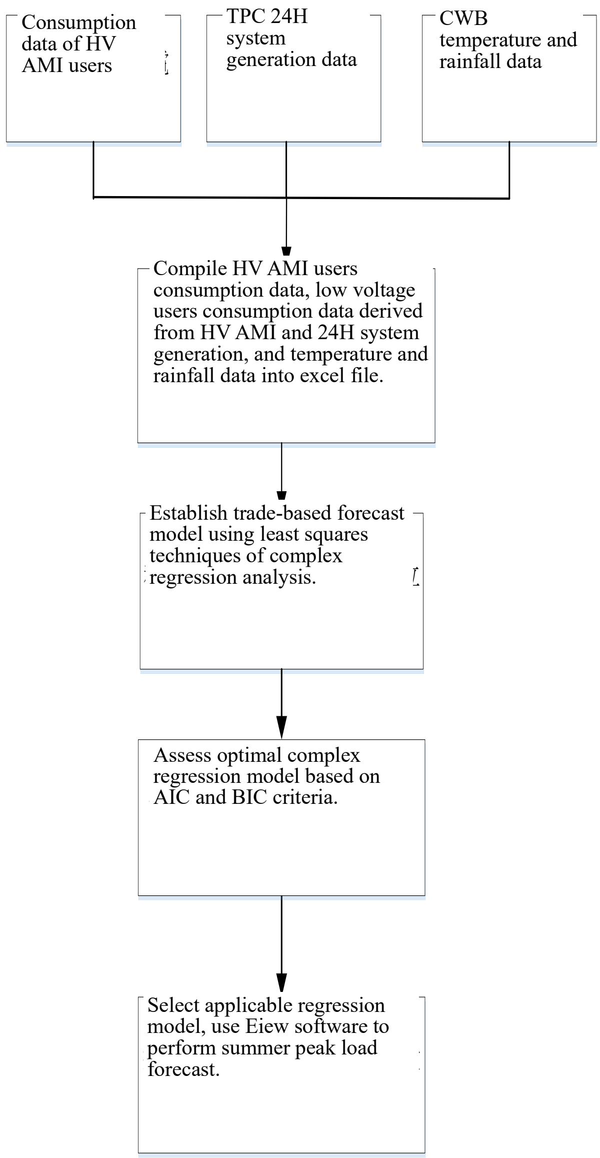

Fig.5 Flowchart of summer peak load estimation using AMI consumption data and 24H system generation data

Table-7 Regression model results of low voltage users consumption data

Forecast time	R-squared (Coefficient of determination)	AIC	BIC	F Test value	P Value
10	0.627385	15.694	15.775	81.942	0
11	0.632131	15.849	15.929	83.627	0
12	0.656416	16.06	16.14	92.977	0
13	0.684327	16.125	16.225	78.584	0
14	0.656356	16.15	16.25	69.237	0
15	0.597276	16.144	16.224	72.177	0
16	0.545503	16.016	16.096	58.411	0

Table-8 Summary of low voltage user consumption data regression model

Forecast time	Combination of independent variables	Regression model
10	T8 T9 R9	$Ly10 = -3578.18482282 + 363.284939283*T8 + 228.936488112*T9 + 165.380400797*R9$
11	T9 T10 R9	$Ly11 = -3711.39300491 + 378.658542458*T9 + 222.632000616*T10 + 208.675587054*R9$
12	T10 T11 R8	$Ly12 = -6127.6814471 + 436.499592047*T10 + 220.128241008*T11 + 166.265155251*R8$
13	T11 T12 R11 R10	$Ly13 = -8210.71273388 + 402.049367072*T11 + 337.773045125*T12 + 379.055280973*R11 + 158.106335896*R10$
14	T12 T13 R13 R12	$Ly14 = -6623.50748347 + 433.130523506*T12 + 257.390321028*T13 + 284.011686842*R13 + 156.144048386*R12$
15	T14 T13 R13	$Ly15 = -3799.92063224 + 220.129334494*T14 + 380.250668206*T13 + 287.136809428*R13$
16	T14 T15 R15	$Ly16 = 233.532472858 + 247.042764118*T14 + 231.81165204*T15 + 87.8134024315*R15$

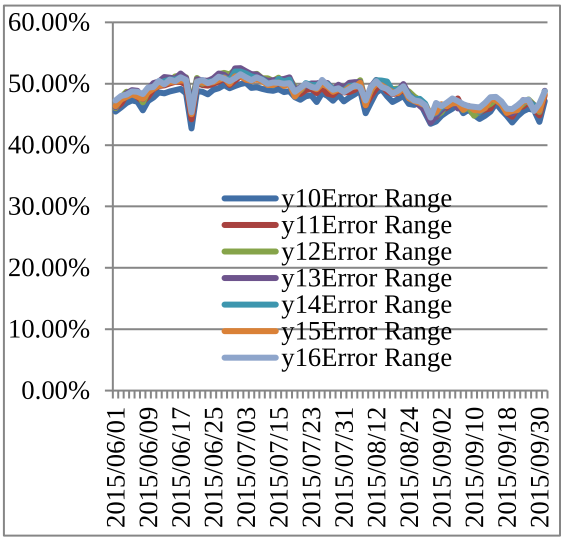

Fig.6 Curve of error range of 2015 summer peak load forecast

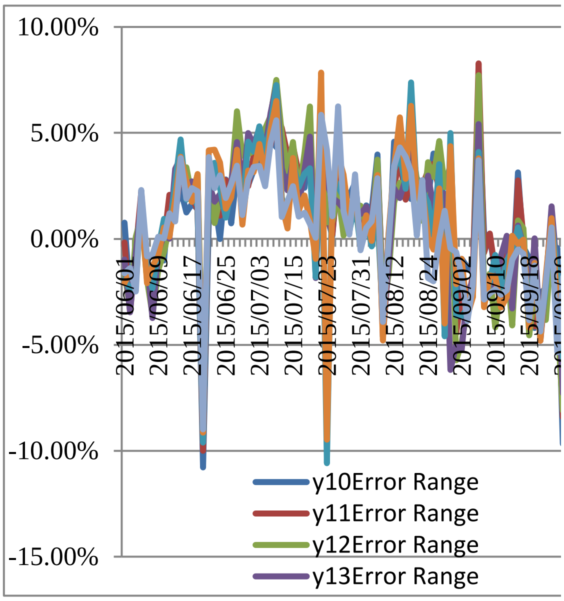

Fig.7 Curve of error range of 2015 summer peak load forecast

Table-9 Estimated peak load average error (%)

Data	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
HV Forecast + LV Forecast	0.86	0.87	0.82	0.70	0.64	0.64	0.58
AMI HV Forecast	47.2	48.2	48.8	48.8	48.7	48.4	47.2

6 – Conclusions and future studies

This paper uses complex regression analysis modeling and least-squares method to forecast power sales and summer peak load, and verifies the accuracy of the least-squares forecast model via forecasting power sales; subsequently, HV AMI users power consumption data and 24H system generation are used to forecast the entire system summer peak load at the system end. Study results show an average error of ± 0.87 % when HV AMI users consumption and 24h system generation are used together to forecast summer peak load; if forecasting peak load by HV AMI users consumption alone, the average error is as high as 48%. Therefore, we propose that in order to estimate the peak load, it is necessary for the regression model to use HV users demand data and 24h system generation data; and that temperature and rainfall are necessary for forecasting the summer peak load demand in the future for the reference of TPC.

7 - References

[1]Mohsen Kojury-Naftchali, Alireza Fereidunian, Hamid Lesani, AMI Data Analytics; "An Investigation of the Self-Organizing Maps Capabilities in Customers Characterization and Big Data Management," 2017 Smart Grid Conference,

[2]A. Albert and R. Rajagopal Smart Meter Driven Segmentation: What Your Consumption Says About You, IEEE Transactions on power systems,Vol.28,Nov,2013.

[3] J. Kwac, J. Flora, and R. Rajagopal, Household Energy Consumption Segmentation using hourly data, IEEE Transactions on Smart Grid, Vol.5, jan,2014

[4] R. Torkzadeh, A. Mirzaei, M. M. Mirjalili, A. S. Anaraki, M. R. Sehhati and F. Behdad, "Medium term load forecasting in distribution systems based on multi linear regression & principal component analysis: A novel approach," 2014 19th Conference on Electrical Power Distribution Networks (EPDC), Tehran, Iran, 2014, pp. 66-70, doi: 10.1109/EPDC.2014.6867500.2008。

[5] C. S. Chen, M. Y. Huang and C. C. Chen, "Expansion planning of substation capacity by considering service reliability," 2004 International Conference on Power System Technology, 2004. PowerCon 2004., Singapore, 2004, pp. 58-62 Vol.1, doi : 10.1109/ICPST.2004.1459966。

[6] Jeng-Feng Hsu ; Jyh-Ming Chang ; Ming-Yuan Cho ; Yi-Hang Wu ; Wen- Yao Chang Chin-Tun Wang ," Development of Regression Models for Prediction of Electricity by Considering Prosperity and Climate"，IEEE 2016 3rd International Conference on Green Technology and Sustainable Development (GTSD) 26 December 2016.

[7] "Effects of Climate Change in Electric Power Infrastructures," by Daniel Burillo, Submitted: June 18th 2018Reviewed: October 20th 2018Published: December 10th 2018 DOI: 10.5772/intechopen.82146

[8] Chen Chen，Jie Na Zhou，"Application of Regression Analysis in Power System Load Forecasting" Advanced Materials Research, vol. 960–961, Trans Tech Publications, Ltd., June 2014, pp. 1516–1522. Cross ref, doi:10.4028/www.scientific.net/amr.960-961.1516. Scientific.Net

[9] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," in IEEE Transactions on Power Systems, vol. 9, no. 4, pp. 1788-1794, Nov. 1994, doi: 10.1109/59.331433.

[10] Nelson Fumo and Mantosh Biswas, "Regression analysis for prediction of residential energy consumption", Renewable and Sustainable Energy Reviews 47:332-343 · March 2015 . Reads DOI: 10.1016/j.rser.2015.03.035

[11] Yunsun Kim, Heung-guSon, SahmKim," Short term electricity load forecasting for institutional buildings", Energy Reports Volume 5, November 2019, Pages 1270-1280.

[12] Hourly electric load by multiple linear regression with interactions, IEEE PES General Meeting,QuantaTechnology,LLC,Raleigh,2010.

[13] Alexander，WolF-Peter schill，"On the representation of demand-side management in power system model"，Energy Volume 84，1 May 2015，Pages 840-845

[14] Geoffrey K.F. Tso，K. K.W. Yau，" Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks"， Energy Volume 32, Issue 9, September 2007, Pages 1761-176

[15] M.R.Braun, H.Altan, S.B.M.Beck, "Using regression analysis to predict The future energy consumption of s supermarket in the U.K