Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

ISSN NO. 2320-5407

Article DOI:10.21474/IJAR01/11818
DOI URL: http://dx.doi.org/10.21474/IJAR01/11818

RESEARCH ARTICLE

MULTIPLE LINEAR REGRESSION APPROACH FOR SHORT-TERM FORECASTING OF ELECTRIC ENERGY CONSUMPTION IN TOGO

Comlanvi Adjamagbo, Akim Adekunle Salami, Yao Bokovi, Djamil Gado and Ayite Senah Akoda Ajavon
ResearchLaboratory in Engineering Sciences (Larsi), Regional Center of Excellence for the Control of Electricity (CERME), Department of Electrical Engineering of the National Superior School of Engineers (ENSI), University of Lome (UL), BP 1515, Lome, Togo.

Manuscript Info

Manuscript History
Received: 05 August 2020
Final Accepted: 10 September 2020
Published: October 2020

Key words:-
Load Forecasting, Multiple Linear Regression, Choice Of The Model Inputs Data

Abstract

A Linear Multiple Regression approach is used to model the energy consumption of electricity in Togo. This model is developed from the load data recorded at the electric power source stations in Togo during the period from 2016 to 2017. This model predicts four input parameters (Day of the week, the type of day (working day). or not), Hours in the day and Load data of the same time of the previous day) is used to predict the electrical energy consumption data for the period of 2018 with a MAPE of 4.4964% and a correlation coefficient R2 equal to 95.5889%.

Introduction:-

The fundamentalfunctions of modern energy management systems are based on an accurate short-termprediction model of the electricalload [1]. The precision of the prediction model leads to savings and increasedsecuritymeasures in the operation of systems for generating and transmission of electricalenergy [2]. Large predictionerrorscan lead either to toocareful or toorisky planning, whichcanalso lead to heavyeconomiclosses [3].

Statisticalapproachesrequire an explicit mathematical model whichgives the relationshipbetween the load and several input factors [4]. Severalclassicalmodels are applied for loadpredictions, such as regression-basedmethods for example [5] [6] and time seriesmethods [7].

To predictelectricalload, regressionmethods are usuallyused to model the relationshipbetweenloadconsumption and otherfactorssuch as weather conditions [8], type of day, and customercategory. Engle et al. [9] presentedseveralregressionmodels for predicting the nextday'sload.

This paperdescribes the experienceswe have gainedduring the development of a short-termprediction model of the electric charge of the next 48 half-hours per day for all year 2018 in Togo with a LinearRegressionmethod. Multiple.

Our goal is to predict the load data withvariouscombinations of explanatory variables to determinewhich configuration case gives the best results.

The problemsencountered and the solutions proposed are discussed. The developed model shouldprovidedailyload profile forecasts for the nextsevendays. The forecastresults for all year 2018 are alsopresented

Corresponding Author:- Comlanvi Adjamagbo
Address:- ResearchLaboratory in Engineering Sciences (Larsi), Regional Center of Excellence for the Control of Electricity (CERME), Department of Electrical Engineering of the National Superior School of Engineers (ENSI), University of Lome (UL), BP 1515, Lome, Togo.

Data presentation:

Electricityload or consumption data istakenat the various source substations in Togo during the periodfrom 2016 to 2018. The readings are made in 30-minute steps over the day, whichmakes 48 data per day or 52,608 data. Figure 1 gives an overview of the loadstatements in Excel.

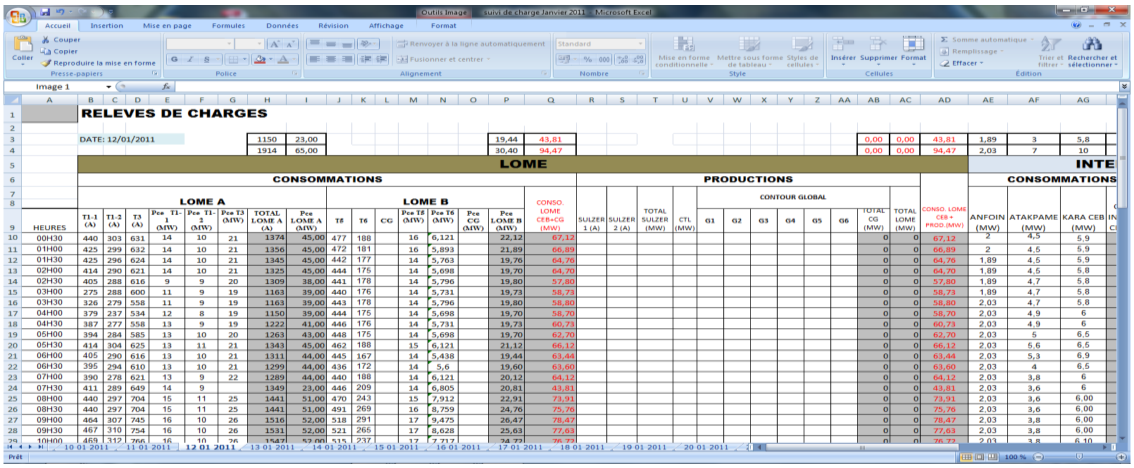

Figure 1:- Presentation of loadstatements in Excel.

The evolution of electricityconsumption in Togo from 2016 to 2018, shown in Figure 2, shows an increasing trend in electricityconsumption. Extrapolatingalinear trend estimates the annual rate of increase in electricalenergyconsumption to bearound 6.5183%.

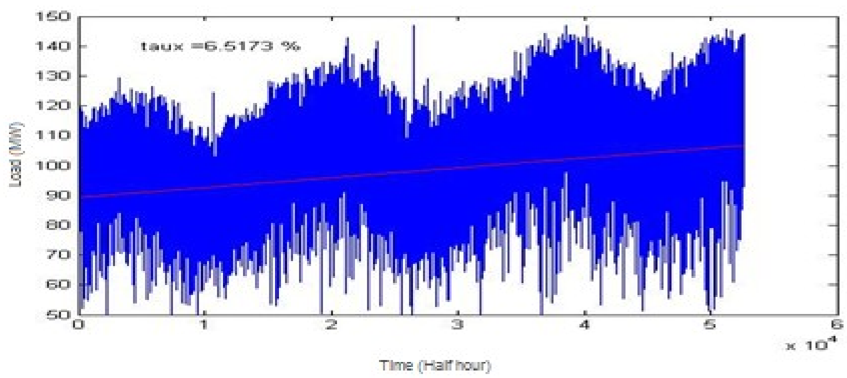

Figure 2:- Evolution of Togo'selectricalloadfrom 2016 to 2018 (MW).

Figure 3 shows the electricalenergyconsumption over a period of one (01) week (fromMondayJanuary 25 to SundayJanuary 31, 2016). Weseeadaily profile appear (7 patterns per week) reflecting a dailycyclical variation.

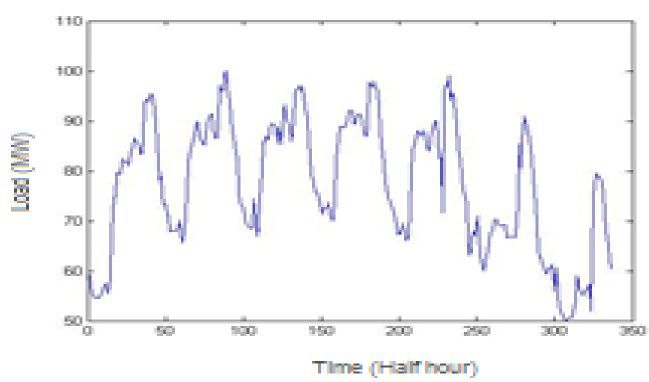

Figure 3: Evolution of Togo'selectricalload over a period of one week (MW).

Multiple linearregressionmethod :
Multiple linearregressionanalysis relies on descriptive analysis of data to observe the relationshipsbetween a quantitative dependent variable and n quantitative independent variables. Anymethodusingregressionsisbased on the acceptance of the foundingassumptions of parametricstatistics and the notion of least squares fit. The concept of least squares consists in minimizing the sum of the residualsraised to the power of twobetween the observed value and the extrapolated one [10].

The descriptive equation for multiple linearregressionis as follows (Equation (1)) [11], [12], [13]:

$$y = X\beta + \varepsilon \quad (1)$$

where :
$y$is the vector of responses;
$X$is the matrix of explanatory variables;
$\beta$is the vector of the model parameters;
$\varepsilon$is the vector of errors.
It istherefore a question of calculating the vector of the estimators$\hat{\beta}$ which is the solution, in the "least squares" sense. This vector of estimators$\hat{\beta}$ is defined by the equation ***(2)*** :

$$\hat{\beta} = (X * X')^{-1} * X' * y \quad (2)$$

This model canbeused to makepredictions. It istherefore a question of applying the relation defined by the equation***(3)*** :

$$\hat{y} = X * \hat{\beta} \quad (3)$$

where$\hat{y}$ is the vector of predicted responses.

Methodology:-
The choice and methodicalanalysis of the explanatory variables makeit possible to assess the influence of each input parameter on the output of the forecast model. Indeed, itisvery important, for the accuracy of the model, to chooseadequate input parameters. This stepisveryusefulbecauseitallowsyou to eliminatesome variables thatprovideverylittle or no information to describe the output, or to eliminateredundant variables. Wetookintoaccount the followingexplanatory variables (Table 1):

Table 1:- List of explanatory variables used.

Data types	Mathematical explanations	Data presentations
Day of the week	Monday = 2 ; Tuesday = 3 Wednesday = 4 ; Thursday = 5 Friday = 6 ; Saturday = 7 Sunday = 1	[1 … 7]
Working day or not	$\begin{cases} if\ working\ day\ then & 1 \\ if\ not\ then & 0 \end{cases}$	[0 ou 1]
Half hour in the day	$\frac{1}{2}h$	[1 … 48]
Load data for the same time of the previous day	$L_{\frac{1}{2}h-48}$	-
Load data for the same time of the previous week	$L_{\frac{1}{2}h-336}$	-
Load data for the same time of the previous year)	$L_{\frac{1}{2}h-17520}$	-

Average charges of the last 24	$Mean(\sum_{i=1}^{48} L_{\frac{1}{2}h-i})$	-
$L = data\ of\ loads$		

The data preprocessingisobtained by the MATLAB software

Either the following nomenclature adopted for the parameters :
A = Day of the week ;
B = Working day or not;
C = Hours in the day;
D = Load data for the samehour of the previousday;
E = Load data for the samehour of the previousweek;
F = Charging data for the samehour of the previousyear;
G = Average charges for the last 24 hours.

Our goal is to predict the load data withvariouscombinations of theseexplanatory variables in order to determinewhich configuration case gives the best results. Wetesteddifferent configuration cases which are summarized in Table 2, for a total of 7 configuration cases.

Table 2:- Summary of simulation cases in MATLAB.

Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
[A B C D E F G]	[A C D E G]	[A B C D G]	[A B C G]	[A B C D]	[A B C]	[B C D E F G]

We have dividedour data intotwo groups. The data for the years 2016 and 2017 are used for learning, thatis to say for the determination of the coefficients of the estimators$\hat{\beta}$ of the model and the data for the year 2018 are used for validation (for the test of the prediction). For each of the configurations adoptedpreviously, weapplied the modelingmethoddescribed in section 3. Thus, wecalculated the coefficients of the vector of the estimators$\hat{\beta}$ from equation (2). Once these coefficients wereknown, wethenperformed the prediction of the new load data by equation (3).

To evaluate the performance of eachprediction model, weused as measures: the average value of the absoluteerrors in percentage (%) (MAPE: MeanAbsolutePercentageError, [2]) committed, expressed by equation (4), the histogram of the absoluteerrors, as well as the correlation coefficient (R^2) between the predicted data and the real load data.

$$MAPE = \frac{100}{T} * \sum_{t=1}^{T} \left|\frac{y_t - \hat{y}_t}{y_t}\right| \qquad (4)$$

where
y_t the real value ;
$\hat{y}_t$ the predicted value ;
T the total number of samples.

Results and Discussions:-

In this part, wewill first discuss the choice of the prediction model and thenpresent the results of the prediction for the year 2018 of the electricalload of the energy system of Togo.

Choice of prediction model:

Tables 3 and 4 respectivelyrepresent the summaries of the coefficients of the estimators$\hat{\beta}$, the R^2 and MAPE for the seven case configurations thatwe have chosen for the input of each model.

Table 3:- Summary of estimators $\hat{\beta}$calculated.

Cases	Coefficients $\hat{\beta}$
Case 1	[-0.6273 ; 4.0738 ; 0.2045 ; 0.0882 ; 0.4469 ; 0.1347 ; 0.2832]

Case 2	[-0.4866 ; 0.1750 ; 0.1220 ; 0.5746 ; 0.2805]
Case 3	[-0.8856 ; 9.0544 ; 0.3628 ; 0.4168 ; 0.4568]
Case 4	[-0.4568 ; 7.1599 ; 0.6230 ; 0.8038]
Case 5	[-0.5324 ; 13.7348 ; 0.1594 ; 0.8710]
Case 6	[7.1766 ; 34.3660 ; 1.4236]
Case 7	[3.9927 ; 0.2157 ; 0.0877 ; 0.4682 ; 0.0980 ; 0.2689]

Table 4:- Effects of input parameter configurations on the prediction model during training.

Cases	MAPE (%)	R^2(%)
Case 1	6.0353	87.2980
Case 2	5.8618	88.5588
Case 3	7.3431	83.4300
Case 4	8.6288	78.8005
Case 5	7.7871	80.3996
Case 6	20.0856	61.5796
Case 7	6.0878	86.9836

We have 7 scenarios. Case 1 composed of all the explanatory variables givesacorrelation coefficient of 87.298%. Case 2 composed of only 5 explanatory variables gives the coefficient of 88.5588%, whichis the highest value. Case 6 composed of 3 explanatory variables gives the lowestcorrelation coefficient (61.5796%). However, all the other cases giveresults of more than 78%. Note alsothatthis case 2 whichquitelogicallypresents the smallest MAPE whichis 5.8618% and also case 6 the largest MAPE of 20.0856%. Thusfromtheseresultswecanexclude case 6 of the seven configurations thatwe have proposed. However, giventhat the choice of a model is not basedonly on itsprecisionduringits training, but also and especially on itsprecisionduring the validation tests, wecarried out validation tests for the seven cases of configurations including the results on R^2 and MAPE are shown in Table 5.

Table 5:- Effects of input parameter configurations on the prediction model during validation tests.

Cases	MAPE (%)	R^2(%)
Case 1	5.6879	90.0178
Case 2	5.5871	89.8176
Case 3	5.3032	92.4043
Case 4	8.8646	77.2865
Case 5	4.4964	95.5889
Case 6	22.3584	61.9213
Case 7	5.7205	89.7302

The results in Table 5 show that all modelsexcept the Case 4 model (becauseits R^2 decreased, 77.2865% vs. 78.8005% during training) fit the validation test data. Indeed, the MAPE and R^2 measurementswereimprovedduringthese tests. The model of case 1 allows us to predict the electric charges with an R^2 coefficient of 90.0178%. The model of case 2 gives us an R^2 coefficient of 89.8176%, whichis no longer the highest value, since the model of case 5 gives us an R^2 coefficient of 95.5889% whichis the largest. The model of case 6 alwaysgives us the lowestcorrelation coefficient (61.9213%). Note alsothatitis the model of case 5 whichpresents the smallest MAPE (4.4964%), whichismoreoverlogicalsinceitscorrelation coefficient R^2 is the highest (95.5889%). Thus the model of case 5 with the highest coefficient R^2 and the lowest MAPE during the validation tests ischosen for the prediction of the electricalload of the energy system of Togo.

Followingthisworkwewillpresent and discuss the prediction of the electric charge for the year 2018 with the model chosen, i.e. the model of case 5.

Prediction for the year 2018:

Figure 8 shows the result of the prediction of Togo'selectric charge for eachhalfhour of the year 2018. Wecan observe a linearcorrelationbetween the measured and predictedload data.

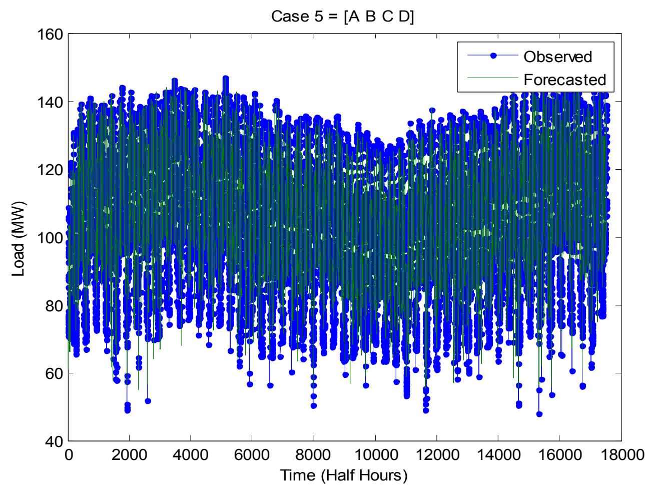

Figure 8:- Prediction of the load in MW for eachhalfhourin 2018.

This correlationbetween the measured and predictedload data iswellobserved if wevisualizethisresult for one week. Figure 9 shows the loadpredictionresultfromSunday 07 to Saturday 13 January 2018. Fromthis figure (Figure 9) wecan observe the predicted and measuredload data for eachday (delimitedaccording to Table 6) fromthisweek.

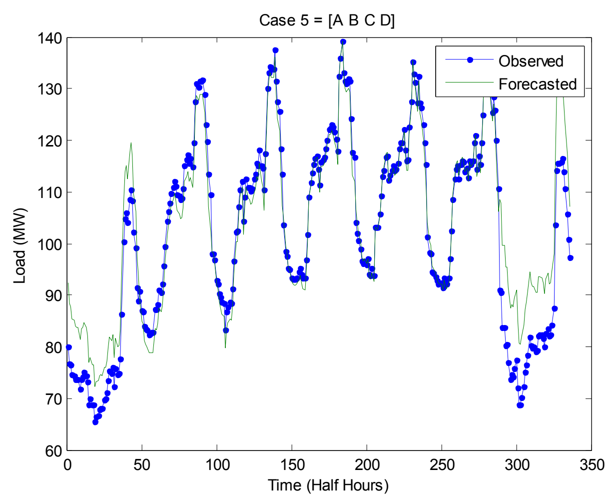

Figure 9:- Prediction of the load in MW for eachhalfhour of Sundayfrom 06 to 12 January 2018.

In Figure 9, we note a strongcorrelationbetween the measured and predictedload data for the sevendays (fromJanuary 07 to 13, 2018), howeverwe observe a very large differencebetween the curves of the measured and predictedload data for twodays. (Sunday 07 and Saturday 13 January 2018). This observation led us to measure the accuracy of the prediction of the electric charge for eachday of the week of the year 2018, the results of which are shown in Table 6.

Table 6:- Electric charge predictionprecisionmeasurements for eachday of the weekin 2018.

Days of the week	MAPE (%)	R^2(%)
Sunday	9.2746	99.1417
Monday	2.6603	99.6350
Tuesday	2.0647	99.4225
Wednesday	1.6415	99.4168
Thursday	1.6535	99.4155
Friday	1.3678	99.3635
Saturday	12.8566	99.0620

Annual	4.4964	95.5889

Thus the results of Table 6 show that there is a strong correlation between the measured and predicted load data for the seven days of the week in 2018 since the average of the R^2 for each day is greater than 99%. We also observe a large average of the MAPEs between the curves of the measured and predicted load data for Sundays (9.2746%) and Saturdays (12.8566%). From Monday to Friday, the model of case 5 retained for the prediction of the electric charge of the energy system of Togo presents good performances (MAPE < 2.67% and R^2 > 99%) on the prediction of the electric charge of each day.

Conclusion:-

This paperpresents the short-termprediction of the electricalload of Togo'senergy system by the Multiple LinearRegressionmethod. Amongsevenmodelsused (each of whichdiffersfrom the other by the nature of these input parameters), we have chosen for the prediction of the electricload a model having four input parameters (Day of the week, the type of Day (working or not), Hours in the day and Load data of the same time of the previousday). Our choicewasjustifiedthanks to the performances obtainedduring the validation tests of this model, sincethislinear multiple regression model allowed us to predict the electricalload of Togo'senergy system for the year 2018 with a MAPE of 4.4964% and a correlation coefficient R^2 equal to 95.5889%.

References:-

1. M. STAHEL et M. MOREL, (1967) : La prévision des consommations d'électricité. Journal de la société statistisque de Paris, Tome 108, ,pp 223-237.
2. HAHN HEIKO, SILJA MEYER-NIEBERG, and STEFAN PICKL, (2009): Electric loadforecastingmethods : Tools for decisionmaking. European Journal of OperationalResearch 199, pp 902 - 907.
3. Dominique LAFFLY, (2006) : Regression multiple : principes et exemples d'application. Université de Pau et des Pays de l'Adour.
4. Ricco RAKOTOMALALA, (2010) : Pratique de la Regressionlineaire Multiple : Diagnostic et sélection de variables. Université Lumière Lyon 2, France
5. G. GROSS, and F. D. GALIANA, (1987): Short-termloadforecasting. Proceedings of the IEEE 75, no. 12 pp 1558-1573.
6. H. S. HIPPERT, C. E. PEDREIRA, and R. C. SOUZA, (2001): Neural Networks for Short-TermLoadForecasting : A Review and Evaluation. IEEE Transactions on Power Systems vol. 16 no. 1, pp 44-55.
7. E. KYRIAKIDES, and M. POLYCARPOU, (2007): Short Termelectricloadforecasting: A tutorial. Edited by K., Wang,L. (Eds) Chen, Trends in Neural Computation, Studies in Computational Intelligence 35, pp 391-418.
8. FEINBERG, E. A., and D. GENETHLIOU, (2005): LoadForecasting, In : Chow, J.H.; Wu, F.F.; Momoh, J.J.; (Eds.), AppliedMathematics for Restructured Electric Power Systems : Optimization, Control and Computational Intelligence, Power Electronics and Power Systems, pp 269-285.
9. R. F. ENGLE, C. MUSTAFA, et J. RICE, (1992): ModellingPeakElectricityDemand, Journal of Forecasting 11, n° 3 pp 241 - 251.
10. G. E. P. BOX, and G. M. JENKINS, (1976): Time SeriesAnalysis: Forecasting and Control, Revisededition, Time Series and Digital Processing, Holden Day, 575 p.
11. C. L. HOR,S. J. WATSON, and S. MAJITHIA, (2012): Analyzing the impact of weather variables on monthlyelectricitydemands, IEEE Transactions on Power Systems 20(4), (2005) pp 2078-2085.
12. The MathWorks, Inc. www.mathworks.com
13. BhattiDhaval, AnuradhaDeshpande, (2020): Short-termloadforecastingwithusing multiple linearregression. International Journal of Electrical and Computer Engineering (IJECE) Vol. 10, No. 4, August 2020, pp. 3911~3917 ISSN: 2088-8708, DOI: 10.11591/ijece.v10i4.pp3911-3917.