International Journal of Computer Engineering & Technology (IJCET)
Volume 9, Issue 3, May-June 2018, pp. 233–240, Article ID: IJCET_09_03_028
Available online at
http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=9&IType=3
Journal Impact Factor (2016): 9.3590(Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6367 and ISSN Online: 0976–6375

DETECTION OF NON-TECHNICAL LOSSES IN POWER DISTRIBUTION USING REGRESSION ANALYSIS

T. Jothi
Research Scholar, Manonmaniam Sundaranar University, Tirunelveli

S. Arumugam
Dean, Nandha Engineering College, Erode - Perundurai Main Road

K. Senthamarai Kannan
Professor, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli

ABSTRACT

Indian electric power distribution system is constantly facing considerable loss of energy both technical and non-technical. A non-technical loss (NTL) in power utilities is termed as any consumed energy or service that is not billed by some type of anomaly. Unaccounted energy due to NTL is a barrier in the growth of Indian economy. Several attempts have made to minimize this problem, however the problem has persisted. Thus, this paper presents a framework for detecting NTLs using regression analysis. Sensible smart meter, energy consumption data, customer registration data are used by this model. The simulation results demonstrate the efficiency of proposed methodology applied for reliable detection of fraudulent electricity customers.

Keywords: Non-technical loss, R square analysis, F statistics analysis, P value analysis, data mining

Cite this Article: T. Jothi, S.Arumugam and K. Senthamarai Kannan, Detection of Non-Technical Losses in Power Distribution Using Regression Analysis. *International Journal of Computer Engineering & Technology*, 9(3), 2018, pp. 233–240.
http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=9&IType=3.

1. INTRODUCTION

Power system losses can be divided into two categories: technical losses and non-technical losses. Technical losses are naturally occurring losses and consist mainly of power distribution in electrical system components such as transmission lines, power transformers, measurement systems, etc. Technical losses are possible to compute and control, provided the power system in question consists of known quantities of load [13], [16],[17].

Non-technical losses (NTL), on the other hand, are caused by actions external to the power system or are caused by loads and conditions that the technical losses computational failed to take into account. NTL are more difficult to measure because these losses are often unaccounted for by the system operators [15] and thus have no recorded information. The most probable causes of NTL are:

- Electricity theft
- Non-payment of customers
- Errors in technical losses computation
- Errors in accounting and record keeping that distort technical information

However, NTL can also be viewed as undetected load; customers that the utilities don't know exist. When an undetected load is attached to the system, the actual losses increase while the losses expected by the utilities will remain. The increased losses will show on the utilities accounts, and the costs will be passed along to the customers as transmission and distribution charges. The load that exhibits patterns similar to a housing load, as implied by the house that load activities increase and decrease. The reason this load is chosen that most NTL causes in reality are found in residential areas or light industrial areas [2],[5],[6]. The main aim of the research work is to detect non-technical losses in the database of srivilliputhur town. The main task is to detect customers with anomalous drops in their consumed energy [8][9]. The algorithm is based on a regression analysis on the evolution of the consumption of the customer. The main aim is to search strong correlation between the time and the consumption of the customer. The regression analysis makes it possible to adjust the consumption pattern of the customer by means of a line with a slope.

2. PROPOSED METHOD

The following figure demonstrates the regression analysis for non-technical loss in power distribution.

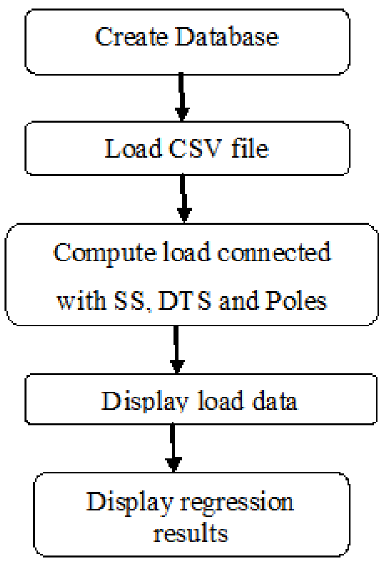

Figure 1 Diagrammatic representation for detecting NTL loss

Here, in this paper, the dataset of srivilliputhur town of Tamilnadu state is taken for the detection of non-technical loss using regression analysis. The dataset consists of the consumers in the srivilliputhur town in all tariffs, substations (SS), distribution transformers (DTs), poles and the load connected to each consumer, load connected to distribution transformer, and load connected to the poles[3].

The proposed methodology is based on regression analysis for DT. The load estimation is exploited to detect the evidences of non-technical loss. A measure of overall fit of the estimated values of customer data aggregated at the DT is used to localize the electricity usage irregularity [12]. Purely, the algorithms are based on a regression analysis of the connected load with the consumers. Regression analysis is widely used for prediction and forecast, where its substantiation overlaps with the field of machine learning. Regression analysis is an important statistical method for the analysis of electrical data. It enables the identification and characterization of relationships among multiple factors. Regression analysis employs a model that describes the relationships between the dependent variables and independent variables in a simplified mathematical form [10].

The statistical relationship between the variables i.e., poles, DTs and consumers has been expressed in regression analysis. In a linear regression model, it can be represented as,

$$y=ax_1+bx_2+ r \tag{1}$$

Where b is the slope

a is the intercept

y is the dependent variable

x1 & x2 are the independent variables

Also, x1 is the poles per DT

x2 is the consumers per DT

y is the load per DT

r = Correlation error

The aim is to search strong correlation between the following regression points.

Regression points have three parameters. They are:

1. Poles, DT, SS Connection Structure
2. Consumers per DT – This can be represented by the pattern of the poles per DT, Consumers per Poles and Consumer per DT respectively
3. Load per DT – This can be represented by the load consumption pattern Connected with DT and Poles.

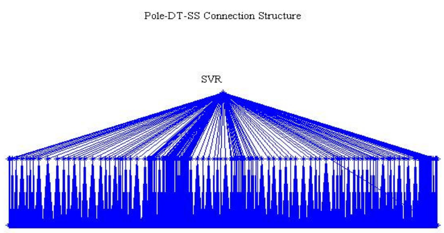

Figure 2 Poles DT-SS Connection Structure

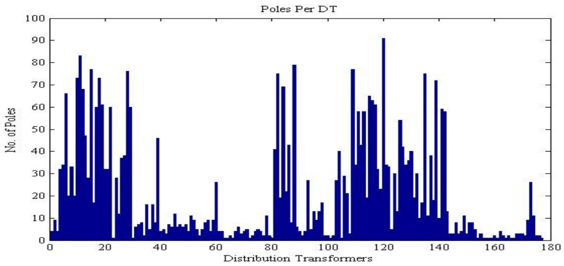

Figure 3 Poles per DT

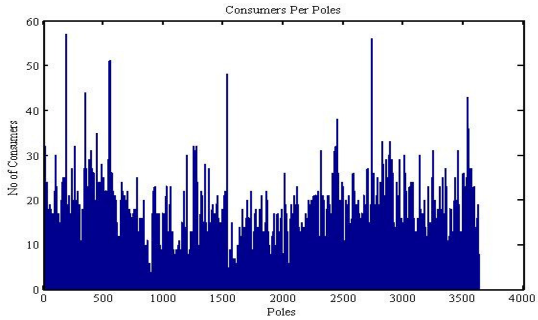

Figure 4 Consumers per poles

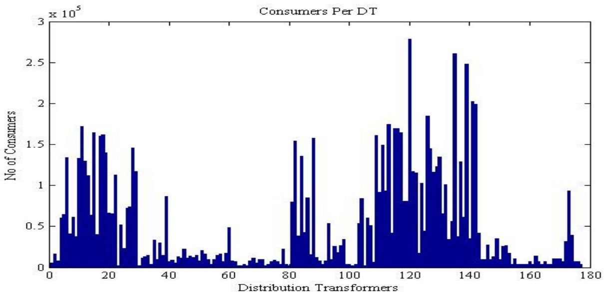

Figure 5 Consumers per DT

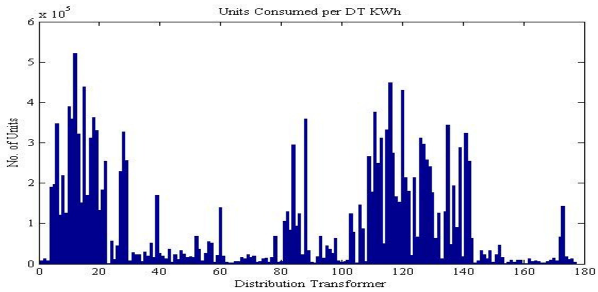

Figure 6 Load consumption per DT

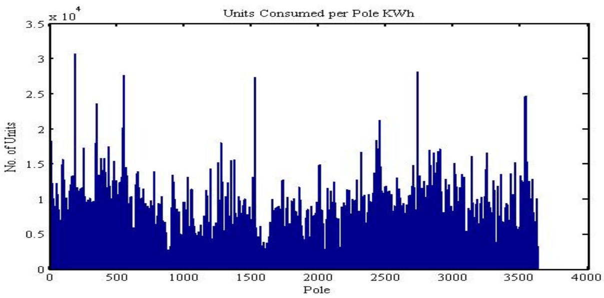

Figure 7 Load consumed per pole

Here, Poles, Distribution Transformers and substation Connection structure is displayed in Fig(1) and Fig(2), (3), (4) displays the pattern of the poles per DT, Consumers per Poles and Consumer per DT respectively, and Fig 5,6 shows the load consumption pattern connected with DT and Poles.

Algorithm

Following steps are used for estimating the regression analysis.

1. Plot the data y vs x1 and x2
2. Here y = connected load per DT

The points formed by all y and x pairs appear to fall to fall in straight line. So we go for linear pattern.

3. Estimate model parameters
4. Present data and the model in the same graph
5. Calculate SSE (sum of square error). SSE can be calculated by using

SSE = € $(y\text{-}y)^2$

6. Test the significance of the overall regression using an F-test.

$$F = \frac{\text{regression sum of square (SSR)}}{\text{Error sum of square (SSE)}}$$

This F-statistic regression will determine if a significant proportion of total variability in the data can be attributed to the relationship between variables.

The larger the F-value the more significant the regression. This ratio increases as the errors SSE decrease when data points move closed to the model line.

7. Determine the co-efficient of correlation(r). r measures the amount of linear relationship between y and (x1,x2).
8. Determine R Square static regression. R Square measures the amount of variation of the dependent variable i.e., connected load with the DT. R2 varies between 0 and 1.
9. Carry out an analysis of the residuals to verify if the ordinary least squares. The residuals can be calculated by the difference between the observed value, the dependent variable, connected load and the predicted value.
10. Regression plane – represents the projection of connected load into the Euclidean subspace spanned by the independent variables poles per DT and consumers per DT.
11. Outliers – a point that falls outside the dataset is classified as a minor outlier while one that falls outside the main cluster data is classified as major outliers. Depending on their location may have major impact on the regression line.

3. RESULTS AND DISCUSSIONS

Regression analysis results are displayed by:

1. Regression points
2. Regression plane
3. Regression residual
4. Outliers
5. NTL estimate

These analyses are discussed in fig 7,8,9,10,and 11 respectively.

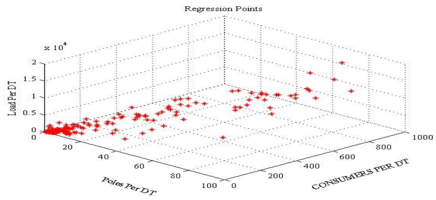

Figure 8 Representation of Regression points

Basically, from Fig (7) the regression analysis can be used to get point estimates. This is a statistical approach to forecasting change in a dependent variable on the basis of change in one or more independent variables known as curve fitting or line fitting because a regression analysis equation can be used in curve fitting or line to data points, in a manner such that the differences in the distances of data points from the curve or line are minimized.

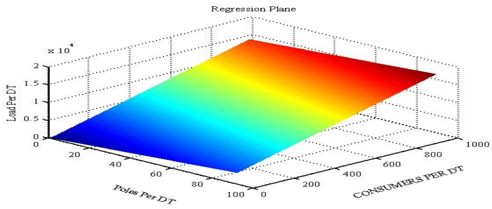

Figure 9 Representation of Regression plane

Fig 8. Denotes the regression plane. All the points do not perfectly lie on a single plane, they give the impression of being relatively clustered around a plane. This plane is called the regression plane. We can now define the regression plane as the plane that describes the best shape of the data. Specifically, we define the regression plane as the plane which minimizes the sum of the squared deviations from the estimated values to the observed values.

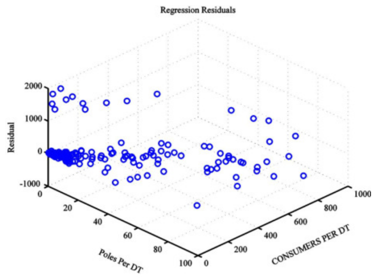

Figure 10 Representation of Regression Residuals

Here, Fig 9. gives the regression residuals. The data points usually fall exactly on this regression equation line; they are scattered around. A residual is the vertical distance between a data point and the regression line. Each data point has one residual. They are positive if they are above the regression line and negative if they are below the regression line. If the regression line actually passes through the point, the residual at that point is zero.

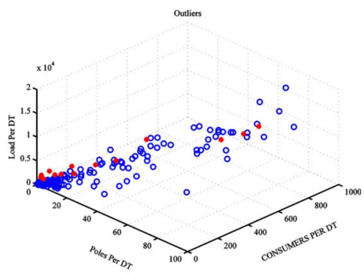

Figure 11 Representation of outliers

In Fig 10, the outliers are discussed. In statistics, an outlier is an observation point that is distant from other observations. An outlier may be due to variability in the measurement or it may indicate experimental error; the latter are sometimes excluded from the data set. An outlier can cause serious problems in statistical analyses. Outliers can occur by chance in any distribution, but they often indicate either measurement error or that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the distribution has high skewness and that one should be very cautious in using tools or intuitions that assume a normal distribution. A frequent cause of outliers is a mixture of two distributions, which may be two distinct sub-populations, or may indicate 'correct trial' versus 'measurement error'; this is modelled by a mixture model.

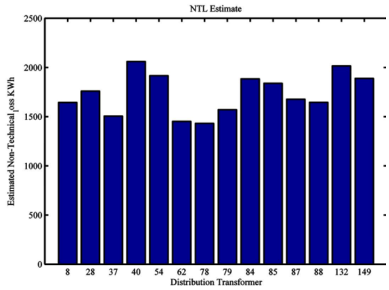

Figure 12 Representation of NTL Estimate

Non-technical loss estimate is represented in fig 11.

Also, for correct identification and detection of fraudulent non-technical losses as well as the statistical analysis the R square statistic regression, F-statistic regression and P-value regression can be calculated and displayed in the following Table.

RSquare Statistic of Regression	0.98576
F Statistic of Regression	12455.9418
p-value of Regression	1.4894e-164

R Square Statistic of Regression measures the strength and direction of a linear relationship between the poles per DT & consumers per DT with load per DT. Here, the value of r is closer to +1. That means more closely the variables are related. F Statistic regression value will range from zero to an arbitrarily large number. P – value >.05 indicates weak evidence against null hypothesis. So we can fail to reject the null hypothesis.

4. CONCLUSION

The prompt detection and isolation of possible contaminants could be a key issue to confirm the standard of life and safety. This paper introduces a completely unique meter data validation and estimation framework to reinforce the estimation and editing method to enhance the meter data quality. Based on the numerous behaviors of DSE and NTL, the feeder busses demand information may be detected. Observability analysis for the planned distribution system regression analysis is mentioned. The numerical observability approach could also be preferred. The test results have shown that the planned methodology could be a appropriate selection for the distribution system SE. Moreover, this study applies a planned framework for NTL analysis to notice and predict suspicious patterns of abnormal consumption behavior. This paper concludes that Non-Technical losses deduction and analysis could be done using regression analysis tools R Square Statistic of Regression, F Statistic of Regression and p-value of Regression.

Results obtained show that the planned framework may be used for reliable detection of fraudulent electricity customers. The strategy for fraud detection established to be terribly promising. It provides smart generalization ability for unseen information classification.

REFERENCES

[1] Shih-Che Huang, Yuan-Liang Lo, and Chan-Nan Lu, Fellow," Non-Technical Loss Detection Using State Estimation and Analysis of Variance "in IEEE TRANSACTIONS ON POWER SYSTEMS.

[2] P. Kadurek, J. Blom, J. F. G. Cobben, and W. L. Kling, "Theft detection and smart metering practices and expectations in the Netherlands," in Proc. IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT Europe), 2010.

[3] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in Proc. CRITIS 2009, 2010, vol.LNCS 6027, pp. 176–187.

[4] I. H. Cavdar, "A solution to remote detection of illegal electricity usage via power line communication," IEEE Trans. Power Del., vol. 19, no.4, pp. 1663–1667, Oct. 2004.

[5] A. Pasdar and S. Mirzakouchaki, "A solution to remote detecting of illegal electricity usage based on smart metering," in Proc. 2nd IEEE Int. Workshop Soft Computing Applications (SOFA 2007), Aug. 21–23, 2007.

[6] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, and S. K. Ahmed,"Non-technical loss analysis for detection of electricity theft using support vectormachines," in Proc. 2nd IEEE Int. Conf. Power and Energy,Dec. 1–3, 2008.

[7] D. Matheson, C. Jmg, and F. Monforte, "Meter data management for the electricity market," in Proc. 8th Int. Conf. Probabilistic Methods Applied to Power Systems, 2004.

[8] VEE Standard for the Ontario Smart Metering System, Issue 3.0, Mar.31, 2010.

[9] Release 2.0 of the NIST Framework and Roadmap for SmartGrid Interoperability Standards, NIST Special Publication 1108R2, Feb. 2012.

[10] I. Roytelman, "Real-time distribution power flow-lessons from practical implementations," in Proc. IEEE PES PSCE, Oct./Nov. 2006, pp.505–509.

[11] R. C. Dugan, R. F. Arritt, T. E. McDermott, S. M. Brahma, and K.Schneider, "Distribution system analysis to support the smart grid," in Proc. IEEE Power Energy Soc. Gen. Meeting, 2010, pp. 1–8.

[12] M. Baran and T. E. McDermott, "Distribution system state estimation using AMI data," in Proc. IEEE PES PSCE, Seattle, WA, Mar. 2009,pp. 1–3.

[13] M. E. Baran and A. W. Kelley, "State estimation for real-time monitoring of distribution systems," IEEE Trans. Power Syst., vol. 9, no. 3,pp. 1601–1609, Aug. 1994.

[14] C. N. Lu, J. H. Teng, and W. E.-H. Liu, "Distribution system state estimation," IEEE Trans. Power Syst., vol. 10, no. 1, pp. 229–240, Feb.1995.

[15] J. Wan and K. N. Miu, "Weighted least squares methods for load estimation in distribution networks," IEEE Trans. Power Syst., vol. 18,no. 4, pp. 1338–1345, Nov. 2003.

[16] Z. J. Simendic, V. C. Strezoski, and G. S. Svenda, "In-field verification of the real-time distribution state estimation," in Proc. 18th Int. Conf.Electricity Distribution (CIRED), Turin, Italy, Jun. 6–9, 2005.

[17] A. Monticelli, State Estimation in Electric Power Systems a Generalized Approach. Norwell, MA: Kluwer, 1999.