MULTIPLE LINEAR REGRESSION ANALYSIS FOR PREDICTION OF BOILER LOSSES AND BOILER EFFICIENCY

Chayalakshmi C.L[1], D.S. Jangamshetti[2] and Savita Sonoli[3]

[1,2]Dept. of EIE, BEC, Bagalkot
[3]Dept. of ECE, RYMEC, Bellari

ABSTRACT

Calculation of boiler efficiency is essential if its parameters need to be controlled for either maintaining or enhancing its efficiency. But determination of boiler efficiency using conventional method is time consuming and very expensive. Hence, it is not recommended to find boiler efficiency frequently. The work presented in this paper deals with establishing the statistical model for boiler efficiency using major boiler losses. Collected data from an eminent industry shows that the loss due to dry flue gas, loss due to hydrogen content in fuel and loss due to moisture content in fuel are the major losses. The boiler efficiency depends mainly on these losses. Multiple regression analysis is used for building the model.

1. INTRODUCTION AND RELATED WORKS

Coal is one of the major fossil fuels used for generation of steam in any process industry. Steam or hot water is essential for many processes. The conversion of chemical energy of coal to thermal energy is carried out in boilers during combustion. The volume of steam is very much higher than the volume of water. Always there is a chance of boiler explosion, causing destruction of process plant and even life casualty if not treated. Hence, boilers are to be treated with at most care. Finding boiler efficiency is better way of understanding and identifying problems associated with the boiler and steps can be initiated immediately to improve its efficiency. Direct method and In-direct method are the two types of conventional method in finding the boiler efficiency. In almost all industries, in-direct method is followed, as it is concerned with various boiler losses. In-direct method is time consuming as it involves fuel analysis, flue gas analysis and complex mathematical calculations. In-direct method is expensive as the equipment used for fuel analysis and flue gas analysis is costly [1].

Regression analysis is widely used method of statistical technique for modeling the relationship between variables. Regression analysis has numerous applications in every field such as engineering, chemical science, economics, management, life and biological sciences, social sciences. Regression models are used for data description, parameter estimation, prediction, estimation and control. Multiple regression model involves more than one regressor variables [2]. M.R. Braun et al. used multiple regression analysis for prediction of energy consumption of a super market in UK, based on gas and electricity data for the year 2012 [3]. Room temperature in office building is modeled by Siyu Wu and Jian-Qiao Sun, using multi stage regression, based on

DOI : 10.5121/ijics.2018.8201

thermodynamic equations [4]. In Siberia, building sector energy consumption is 40% of the total energy consumption. In order to improve energy performance, refurbishment work is initiated in 62 public buildings. Regression analysis is used for investigating the energy savings [5]. The main aim of Joseph Al Asmar et al. is to determine the optimal cogeneration capacity to be installed in a factory with environmental constraints. Genetic algorithm is used for optimization and optimal result selection is performed using multiple linear regression [6]. Linear and non-linear regression analysis is used for heavy metals removal using Agaricus bisporus macrofungus, commonly known button mushroom by Boldizsar Nagy et al. For analyzing the experimental data, linear and non-linear regression models are used along with kinetic models [7]. Artificial Neural Network (ANN) modeling is considered for prediction of the oxide deposition rate on the water wall tubes of a coal fired boiler. Results of ANN predictions are validated with the data from plant and also compared with the regression fit between predicted and measured oxide scale deposition [8]. Monitoring or controlling temperature is one of the fundamental activities in many processes. Engine oil flows easily at operation to allow proper lubrication of engine parts. In cold weather, the power generation unit should be kept warm to avoid damage. Block heaters are used for the purpose of maintaining the temperature of power generation unit, so that the easy flow of oil is assured to provide lubrication. Based on the temperature of weather the block heater is controlled for the generation of power. Logarithmic regression is used for prediction of temperature based on the resistance of the sensor [9]. From the above discussion, it is clear that the applications of regressions are many and in varieties of areas. Further, its application can be exploited for prediction of boiler parameter.

The research work carried out on multiple regression analysis for prediction of boiler efficiency based on major boiler losses is presented in the following section.

2. METHODOLOGY

The loss due to dry flue gas, loss due to hydrogen in fuel and loss due to moisture in fuel are the major boiler losses which contribute for the overall boiler losses.

Table 1: Major boiler losses and boiler efficiency

Sl. No	**Loss due to dry flue gas (L_1)**	**Loss due to hydrogen in fuel (L_2)**	**Loss due to moisture in fuel (L_3)**	**Boiler efficiency**
1	4.7	5.09	1.81	85.777
2	4.6	6.03	1.75	84.928
3	4.7	6.66	1.73	85.287
4	4.3	3.60	2.10	87.406
5	5.4	4.48	2.47	84.891
6	3.1	4.40	2.79	87.056
7	4.3	3.60	2.10	87.406
8	4.5	8.01	2.23	80.776
9	4.5	8.01	2.23	83.592
10	5.2	7.97	2.22	82.941
11	4.9	8.29	2.31	82.874
12	5.4	8.65	2.41	81.844
13	5.6	9.23	4.10	79.698
14	5.7	9.05	4.61	79.330

15	5.5	8.54	3.75	80.750
16	5.4	8.92	3.69	80.676
17	5.3	8.28	3.38	81.792
18	5.0	8.16	3.20	82.440
19	5.0	8.16	3.20	82.440

Table 2: Summary of statistics

Variable	**Obs.**	**Min**	**Max**	**Mean**	**Std. deviation**
Boiler Eff.	19	79.33	87.406	83.258	2.558
L_1	19	3.10	5.700	4.900	0.622
L_2	19	3.60	9.230	7.112	1.938
L_3	19	1.73	4.610	2.741	0.848

Multiple linear regression is carried out to develop a regression model by considering the data from JK cements Pvt. Ltd., Muddapur, Lokapur taluk, Bagalkot district. The three boiler losses and boiler efficiency of this plant are recorded and are shown in Table 1. Initially the correlation between these boiler losses is studied by performing the statistical analysis and the minimum, maximum, mean, and standard deviation (summary of statistics) is as shown in Table 2.

Scatter plot for three boiler losses and boiler efficiency is shown in Fig. 1 indicates that linear relationship exists between boiler losses and boiler efficiency.

The correlation between boiler losses and boiler efficiency is listed in Table 3. Maximum correlation exists between boiler loss due to hydrogen in fuel L2 and boiler efficiency, as shown in Table 3(-0.932). The three boiler losses viz. L1, L2, and L3 are negatively related to boiler efficiency. As these boiler losses increases, boiler efficiency reduces proportionately.

Table 3: Correlation matrix

Variables	***L_1***	***L_2***	***L_3***	***Boiler Efficiency***
L_1	**1.000**	0.663	0.547	-0.752
L_2	0.6630	**1.000**	0.591	-0.932
L_3	0.5470	0.591	**1.000**	-0.732
Boiler Efficiency	-0.752	-0.932	-0.732	**1.000**

Table 4: Multicolinearity statistics

Statistic	***L_1***	***L_2***	***L_3***
Tolerance	0.524	0.486	0.608
VIF	1.910	2.056	1.645

Multicolinearity is a process in multiple regression, which provides information about the correlation between two or more predictor variables. One predictor variable can be predicted from the others with a significant degree of accuracy. In order to recognize the tolerance and variance of inflation factors, multicolinearity test is performed and the result is given in Table 4. If tolerance of any independent variable is less than 0.2, one predictor cannot be predicted from another. Variance inflation factor (VIF) measures how much the variance of the estimated regression coefficients are inflated as compared to the predictor variables. If variance of inflation factor is 1, predictor variables are not correlated, if it is greater than 1 and less than 5, then they are moderately correlated and if it is greater than 5 and less than or equal to 10, then they are highly correlated [10].

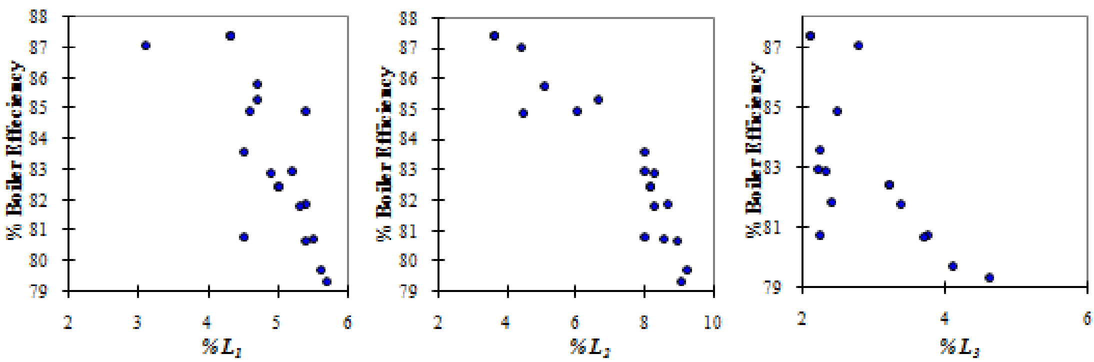

Fig. 1: Scatter plot for L_1, L_2, and L_3 with boiler efficiency

From Table 3, it is clear that the tolerance of variable is not less than 0.2 and does not causes any multi-colinearity problem. From the basics of statistics, if the variance inflation factor of variables varies between 1 and 5, shows that the correlation between predictor variables is moderate.

3. RESULTS AND DISCUSSION

Multiple regression analysis is performed on boiler efficiency based on boiler losses due to dry flue gas, hydrogen content in fuel and moisture content in fuel. The coefficient of determination (R^2) determines the goodness of fit. The range of R^2 varies from 0.0 to 1.0. If determination coefficient value is 0.0, knowing the value of X variable, Y variable cannot be predicted. This indicates that no linear relationship exists between variables and the best fit line is horizontal. The line passes through the mean of all Y values. If its value is 1.0, all points lie exactly on the straight line without scatter. If X variable is known, Y variable can be perfectly predicted.

Parameter values for goodness of fit are listed in Table 5. R^2 value for the multiple regression is 0.935 which shows that predictor variables are correlated to boiler efficiency by 93.5 % and adjusted R^2 is 0.922 for 19 observations and with -9.286 as AIC.

Significance or P_r value is the probability that the current observation has occurred by chance. P_r value must be less than 0.15 for the best fit. In this research work carried out, the significance or P_r value is less than 0.0001, which is very small and not even a single prediction occurs by chance, as listed in Table 6. Table 5 shows that Degree of freedom (DF) for the model is 3, as the three boiler losses are considered for forecasting boiler efficiency.

Table 5: Statistics of goodness of fit

Observations	19.000
R^2	0.935
Adjusted R^2	0.922
MSE	0.510
RMSE	0.714
MAPE	0.481
AIC	-9.286
SBC	-5.508

Table 6: Analysis of variance

Source	**DF**	**Sum of squares**	**Mean squares**	**F**	$P_r > F$
Model	03	110.171	36.724	72.010	< 0.0001
Error	15	7.650	00.510		
Corrected Total	18	117.821			

Table 7: Model parameters

Source	**Value**	**Standard error**	**T**	$P_r > \lvert t \rvert$	**Lower bound (95%)**	**Upper bound (95%)**
Intercept	95.072	1.385	68.662	< 0.0001	92.121	98.024
L_1	-0.714	0.374	-1.909	0.076	-1.511	0.0830
L_2	-0.894	0.125	-7.181	< 0.0001	-1.159	-0.629
L_3	-0.714	0.255	-2.805	0.013	-1.257	-0.171

The steps discussed in above section, are followed for the calculation of coefficients of the model and the coefficients are listed in Table 7. From these parameters listed in Table 7, the equation for the model is constructed. Boiler efficiency based on the major losses is predicted from the equation 5.

$$\textit{Boiler effeciency} = 95.072 - (0.714 \times L_1) - (0.894 \times L_2) - (0.714 \times L_3) \qquad (5)$$

With the help of the developed model, the boiler efficiency is predicted for 19 values of boiler losses and the residuals are calculated and listed in Table 8. Residual is the difference between predicted and actual values of boiler efficiency. Standardized residual is the ratio of residual to the standard deviation of the residual. Standardized residual provides the strength of difference between the predicted and observed values. Based on prediction and residuals, graphs are plotted as shown in Fig. 2-5.

If standardized residual is less than -2, the predicted value is less than the actual value and if greater than + 2, the predicted value is greater than actual value. From Fig. 3, the standardized residual lies between ± 1, this indicates that the predicted values are in the expected range.

Table 8: Predictions and residuals

Observation	**Actual Boiler Efficiency**	**Predicted Boiler Efficiency**	**Residual**	**Std. residual**
Obs1	85.777	85.874	-0.097	-0.135
Obs2	84.928	85.147	-0.219	-0.307
Obs3	85.287	84.527	0.760	1.064
Obs4	87.406	87.284	0.122	0.170
Obs5	84.891	85.448	-0.557	-0.780
Obs6	87.056	86.933	0.123	0.172
Obs7	87.406	87.284	0.122	0.170
Obs8	80.776	83.106	-2.330	-3.262
Obs9	83.592	83.106	0.486	0.681
Obs10	82.941	82.649	0.292	0.409
Obs11	82.874	82.513	0.361	0.506
Obs12	81.844	81.763	0.081	0.114
Obs13	79.698	79.894	-0.196	-0.275
Obs14	79.330	79.620	-0.290	-0.406
Obs15	80.750	80.833	-0.083	-0.116
Obs16	80.676	80.607	0.069	0.096
Obs17	81.792	81.472	0.320	0.448
Obs18	82.440	81.922	0.518	0.725
Obs19	82.440	81.922	0.518	0.725

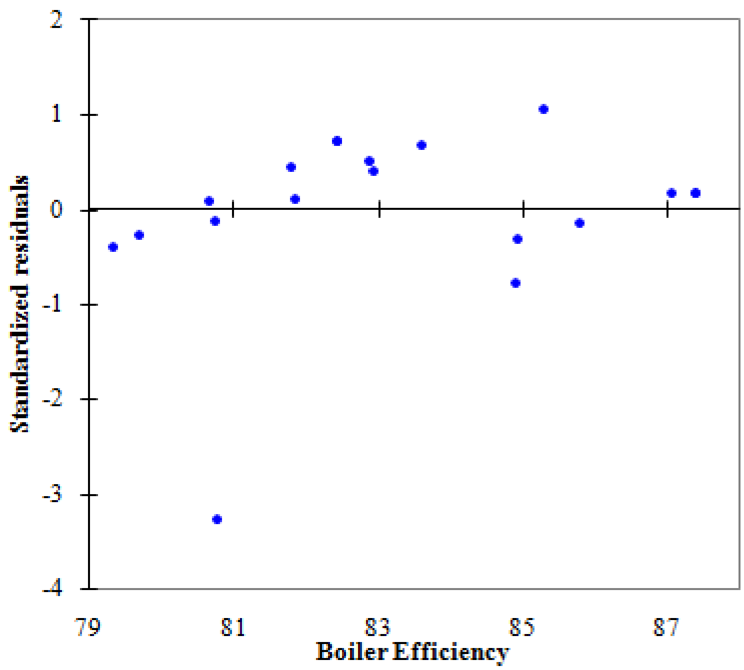

Fig. 2: Boiler efficiency and standardized residual

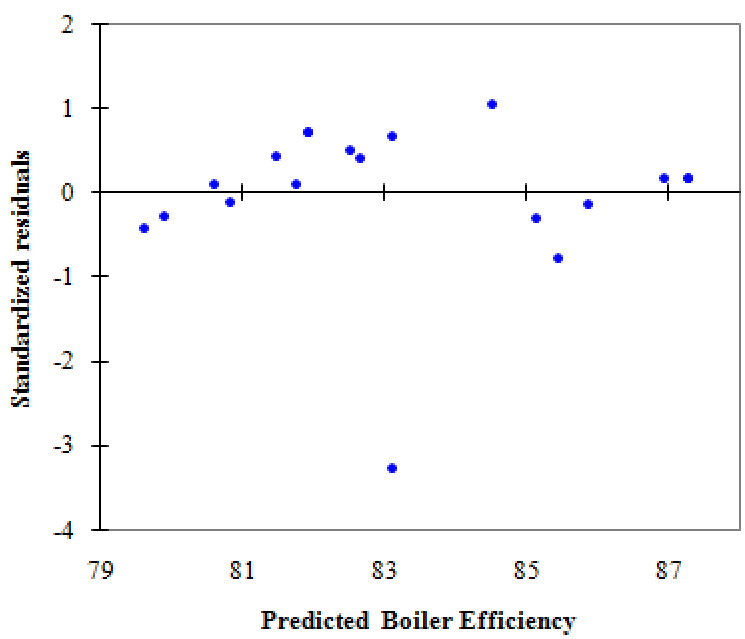

Fig. 3: Predicted boiler efficiency and standardized residual

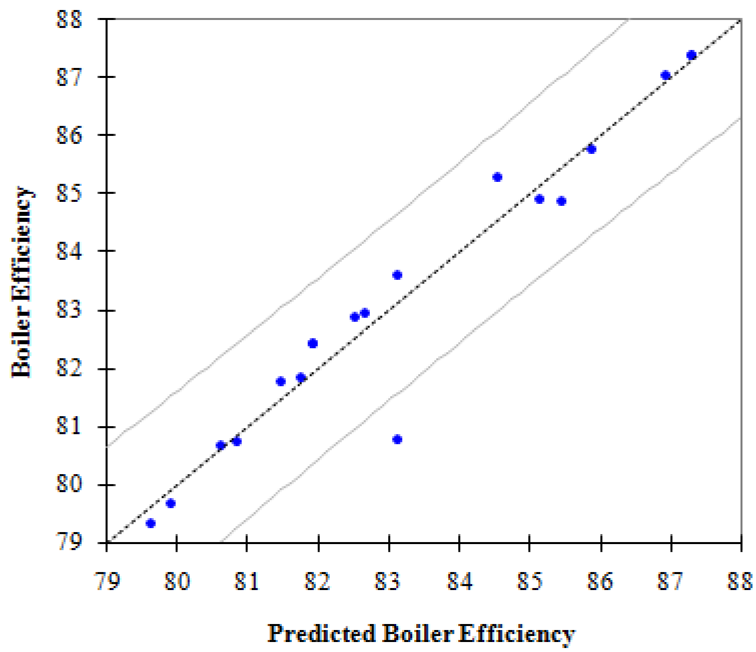

Fig. 4: Predicted boiler efficiency and boiler efficiency

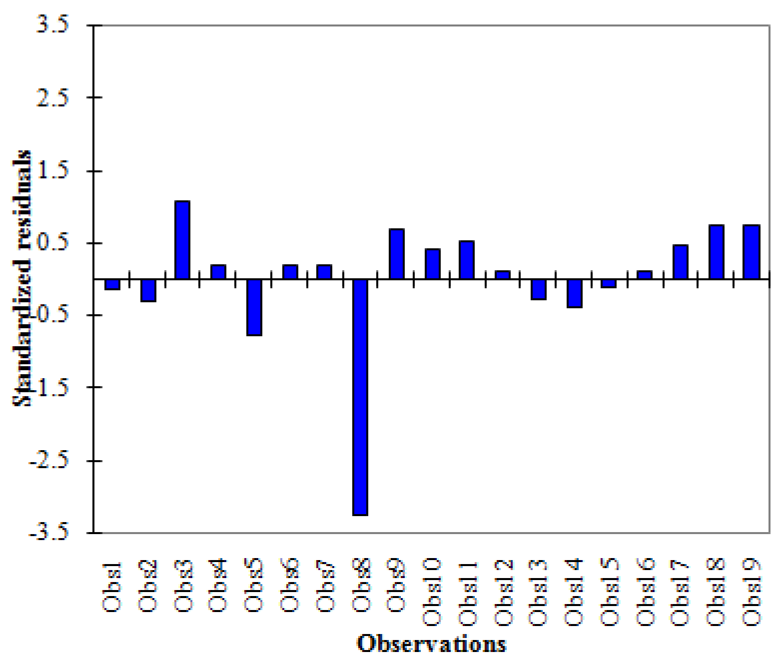

Fig. 5: Standardized residual for nineteen observations

The value of boiler efficiency is very much essential in understanding the problems related to boiler. But the conventional method of finding boiler efficiency is time consuming and expensive. In the research work presented in this section, a method of boiler efficiency prediction is proposed and implemented.

4. CONCLUSION

The contributors for over all boiler losses are loss due to dry flue gas and loss due to moisture in fuel along with loss due to hydrogen in fuel. Multiple linear regression is performed to assess the effect of major boiler losses on boiler efficiency.

- Multiple linear regression is considered for development of the model
- Boiler loss due to dry flue gas, hydrogen in coal, and moisture in coal are the major losses considered for the model development as these losses contribute more for overall boiler loss
- Model is tested and validated from the data obtained from a cement plant
- The variance inflation factor of variables is in between 1 and 5, which shows that the correlation between predictor variables is moderate

REFERENCES

[1] Tai Lv, Linghao Yu, Jinmin Song, "A Research of Simplified Method in Boiler Efficiency Test," International Conference on Future Electrical Power and Energy Systems, Elsevier's Energy Procedia, 17 (2012), pp. 1007–1013.

[2] Douglas C. Montgomery, Elizabeth A. Peek, G. Geofferey Vining, "Introduction to Linear Regression Analysis," 22nd edition, John Wiley and Sons, Canada, 2012.

[3] M.R. Braun , H. Altan , S.B.M. Beck , "Using regression analysis to predict the future energy consumption of a super market in the UK," Elsevier's Applied Energy, 130 (2014), pp. 305–313.

[4] Siyu Wu, Jian-Qiao Sun, "Multi-stage Regression Linear Parametric Models of Room Temperature in Office Building," Elsevier's Building and Environment, 56, 2012, pp. 69–77.

[5] Sanja Petrovic Be cirovic, Mileva Vasi, "Methodology and Results of Seberian Energy-Efficiency Refurbishment Project," Elsevier's Energy and Buildings, 62, 2013, pp. 258–267.

[6] Joseph Al Asmar, Raed Kouta, Salah Laghrouche, Joseph El Assad, Maxime Wack, "Cogeneration Systems for Industrial Sector: An Optimal Power," 2014 IEEE Conference on Control Applications (CCA), 978-1-4799-7409-2/14.

[7] Boldizsar Nagy, Carmen Manzatu, Andrada Maicaneanu , Cerasella Indolean, Barbu-Tudoran Lucian, Cornelia Majdik, "Linear and Nonlinear Regression Analysis for Heavy Metals Removal using Agaricus bisporus macrofungus," Arabian Journal of Chemistry, 2014, pp.1-11.

[8] Amrita Kumari1, S.K. Das, P.K. Srivastava, "A Neural network Model for Quantitative Prediction of Oxide Scale in Water Walls of an Operating Indian Coal Fired Boiler," Research and Reviews in Electro Chemistry: An Indian Journal, Vol. 5, Issue 6, 2014, pp.155-168.

[9] Nabeel Kadim Abid Al-Sahib, Hameed Salman Hameed, "Monitoring and Wireless Controlling of Power Generation by LabView," Journal of Control Theory and Informatics, Vol.4, No.2, 2014, pp.1-13. ISSN 2224-5774 (Paper) ISSN 2225-0492 (Online)

[10] Vrieze, Scott I., "Model Selection and Psychological Theory: A Discussion of the Differences Between Akaike's Information Criterion and the Bayesian Information Criterion," Psychological Methods, Vol. 17, No. 2, 2012, pp. 228-243.