JEAS

Journal of Engineering and Applied Sciences 7 (2011) 50-57

JOURNAL OF ENGINEERING AND APPLIED SCIENCES

Software approach to correlation analysis of natural gas production and disposal data recorded from oil companies' operations in Nigeria

G. C. Oguejiofor, [1] U. J. F. Ewurum

Department of Chemical Engineering,Nnamdi Azikiwe University, Nigeria
[1]Department of Management, University of Nigeria Enugu Campus, Enugu, Nigeria

Abstract
Within the context of operations research, an attempt aimed at evaluating and analyzing, the strengths of associations between the empirical data, generated from the operations of the Nigerian natural gas industry was undertaken. The data which spanned from 1965 to 2009, were subjected to simple correlation analysis, which employed a computer-driven approach, for the following pairings; flared-gas volume (X-variable) and utilized-gas volume (Y-variable); gas-production volume (X-variable) and gas-utilization volume (Y-variable); gas-production volume (X-variable) and flared-gas volume. The results from the computer analyses containing the evaluated coefficients of correlation and determination, test of significance of correlation coefficients and p-value, were presented as outputs 1, 2 and 3, detailed in this paper would show. In all of the results and interpretations reported, no cause-and-effect relationships could be inferred. These were pointers to the limitations of statistical correlation analysis.

Keywords: Operations research; empirical data; simple-correlation analysis; computer-driven approach; cause-and-effect relationship

1. Introduction

1.1 Background and problem statement

Nigeria has abundant proven reserves of natural gas estimated to be in the region of 150 trillion cubic feet (4250 billion cubic metres), which at the annual average production of 21 billion cubic metres will deplete in 202 years [Erinne, 1999:41]. The proven reserves are located in onshore and offshore wells in the Nigerian petroleum fields.

Natural gas productions record, 1961-2009, (see Appendix) shows that oil companies engaged in petroleum production have been disposing natural gas production in two ways, namely gas flaring and gas utilization. From the viewpoint of operations research the following pertinent questions are deductible from the natural gas production record (see Appendix) and they are:

- Is there a correlation between natural gas disposal by methods of flaring and utilisation in Nigeria's petroleum industry?
- Is there a relationship between Nigeria's natural gas production and disposal by flaring approach?
- Is there a relationship between natural gas production and disposal by utilisation approach in Nigeria's petroleum industry?

The answers to these questions will be sought from correlation analysis. Mason et al [1999:426] defines correlation analysis as a group of techniques to measure the strength of the correlation between two variables. The basic idea of correlation analysis is to report the strength of association between two variables.

In this study therefore, the tool of simple correlation analysis will be employed to measure precisely the relations between the underlisted variables in the natural gas production and disposal in Nigeria's petroleum industry.

- Gas flaring and gas utilisation
- Gas production and gas flaring
- Gas production and gas utilisation

1.2. Motivation for the study

The growing use of computers and the availability of statistical software systems such as MINITAB, SAS, CBS, SPSS and Microsoft Excel have revolutionized operations research. Spurred by these developments, this work will employ a software package named

MegaStat for Excel as study tool for analyzing Nigeria's recorded data for natural gas production and disposal.

Nigerian petroleum reserves are located in onshore and offshore wells. Whether in offshore or onshore location, natural gas occurs with oil in the petroleum wells as associated gas. During production, gas and oil are extracted from the wells. The mixture is separated at the surface, and the oil is stored while the associated gas is flared or utilised. Despite Nigeria's government efforts at stopping gas flaring as an approach to gas disposal, wasteful gas flaring has continued in the Nigerian onshore and offshore petroleum fields. On the average, about half of Nigeria's annual gas production is unfortunately wasted by flaring in the petroleum fields, while the rest is utilised for electricity generation and other industrial uses.

However, Shell and other upstream oil producing companies give the following as reasons for gas flaring in Nigeria.

- Nigerian petroleum occurs in remotely located offshore and onshore fields where the geography of the fields makes it very difficult for gas gathering and harnessing.
- There is limited market for gas in Nigeria and overseas, because gas is not a popular source of energy as crude oil.
- Gas gathering, processing and storage equipment are highly capital intensive, resulting in inadequate gas harnessing and storage facilities in Nigeria.

In order to contribute to the build-up of the literature that would enhance the management of gas production and disposal in Nigeria, this study would undertake a content analysis of the natural gas production record, 1961-2009, by employing the tool of computer correlation analysis.

2. Framework review

The outline for the review of framework are:

- Nigeria's gas production and disposal time series
- Simple correlation analysis
- Test of significance of the correlation coefficient, r
- Interpreting p-value in hypothesis testing
- Sample regression analysis

2.1. Gas production and disposal time series

The contents of Appendix is the study specimen covering gas production, utilization and flaring from 1961 to 2009. A cursory review of the records of gas production, utilisation and flaring as depicted in figure would reveal some general trends. The time series plot of the study specimen is shown below for gas production, utilization and flaring. First, on the average, there has been an expansion in the volume of gas produced over 1961-2009 period (Fig.1). This may be because of growth in petroleum production activities. Second, on the overall basis, the record of gas utilisation over 1965-2009 period showed a growth trend (Fig.1). Third, the record of gas flaring over 1965-2009 time frame indicated on the average a reduction in the flared gas volume (Fig.1). These may be explained by the expansion over 1965-2009, in the industrial usage of natural gas for electric power generation, liquefied natural gas production, cement manufacture in Obajana plant, Aluminium Smelter plant, and other chemical manufacture.

However, the contents analysis of Appendix will be undertaken by employing the statistical tool of Simple Regression and Correlation Analyses.

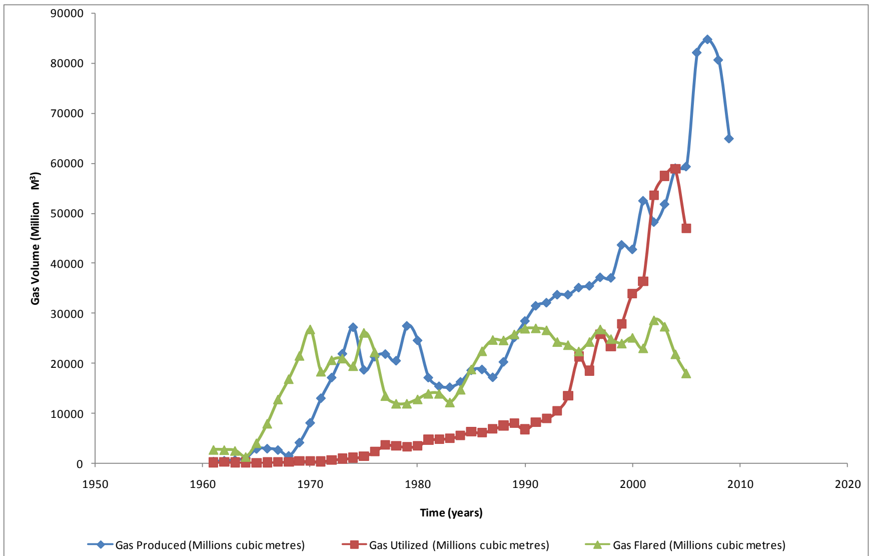

Fig. 1. Graphs of annual volumes of gas produced utilized and flared in Nigeria (1961 – 2009)

2.2. Simple correlation analysis

To understand the strength of the association between the study variables, namely; gas utilisation and gas flaring; gas production and gas flaring; gas production and gas utilisation; a correlation analysis will be reviewed for correlation coefficient determination using the raw score formula, or the deviation from the mean formula. Subsequently, the coefficient of determination will be reviewed.

The numerical value of the coefficient of correlation (r) can be determined from the raw score approach with the formula below (Mason et al, 1999:429)

$$r = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n(\Sigma X^2) - (\Sigma X)^2\right]\left[n(\Sigma Y^2) - (\Sigma Y)^2\right]}} \quad (1)$$

where the symbols are defined in the Notation Section. Eqn(1) is the Pearson Product Moments Correlation Coefficient from raw score approach. Alternatively, the deviation from mean approach to the Pearson Product Moment Correlation Coefficient is given by Ojih, (1996:130) and Nwabuokei (1986:319) as;

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{(\Sigma(X - \bar{X})^2 \; \Sigma(Y - \bar{Y})^2}} \quad (2)$$

where the symbols are defined in the Notation Section.

The coefficient of determination according to Mason et al (1999:430) is the proportion of the total variation in the dependent variable Y that is explained, or accounted for, by the variation in the independent variable X. This coefficient is computed by squaring the coefficient of correlation, r.

2.3. Test of significance of correlation coefficient, r

Is the correlation in the population from which the sample was selected zero? Mason et al (1990:507) states that the test statistic for this question is t, and the formula to answer the question is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (3)$$

where the number of degrees of freedom is n – 2.

2.4. Interpreting p-value in hypothesis testing

Mason et al (1999:316-317) writes that the p-value gives additional insight into the strength of the decision for hypothesis testing.

Accordingly, if the p-value is less

(a) 0.10, we have same evidence that H_o is not true.

(b) 0.05, we have strong evidence that H_o is not true.

(c) 0.01, we have very strong evidence that H_o is not true.

(d) 0.001, we have extremely strong evidence that H_o is not true. (Mason et al, 1999:317).

2.5. Simple regression analysis

The general form of the simple regression equation is given by Mason et al (1999:437) and Nwabuokei (1986:293) as:

$$Y_x = a + bX$$

where b, the slope of the regression line and a, the intercept on the Y-axis are given by Mason et al (1999:437) and Nwabuokei (1986:296-297) as:

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)} \quad (5)$$

$$\text{and, } a = \frac{\sum Y}{n} - b\frac{\sum X}{n} \quad (6)$$

3. Materials and methods of evaluation

This section will be reported under the following highlights.

- Evaluation of correlation coefficients, r
- Computation of the coefficients of determination, r^2
- t-test for correlation coefficients, r
- Calculations of the regression coefficients

3.1. Evaluation of correlation coefficients, r

The megaStat for Excel, a computer software package by J. B. Orris for enhancing the power of Microsoft Excel in statistical analysis was used as the evaluation tool. Using eqn(1), three separate correlation coefficients, r were computer evaluated from the natural gas production and disposal data, covering 1965-2009, for the underlisted categorization of dependent and independent variables.

- Gas utilisation data (Y-variable) and gas flaring data (X-variable) for 1965-2009 period.
- Gas production data (Y-variable) and gas utilization data (X-variable) for the 1965-2009 time frame.
- Gas production data (Y-variable) and gas flaring data (X-variable) for the 1965-2009 period.

3.2. Computation of coefficients of determination, r^2

Subsequently, three separate coefficients of determination, r^2 were computed for the above-listed three categories of paired observations by employing the MegaStat software described above.

3.3. t-test for correlation coefficients, r

The null and alternate hypotheses were stated as:

H_o: Þ = 0 (the correlation in the population is zero)

H_1: Þ ≠ 0 (the correlation in the population is unequal to zero)

The test was two-tailed, given the way H_1 was stated. Also, the t-test was driven by the megaStat software for Excel which exploited eqn(3).

3.4. Estimation of the coefficients of regression

Given the fact that megaStat is inbuilt with Eqs.(5) and (6), the regression coefficients were determined automatically.

4. Results and interpretation

The results obtained from the software-driven evaluations of the correlation coefficients, r and r^2; the test for correlation efficient, t; and the regression coefficients, a and b; are the computer results presented below as Output 1, Output 2, and Output 3, respectively.

The work is intended to estimate the values of r, r^2 and t-statistic. But the megaStat software did more than that, as the Outputs show. Interestingly, it evaluated the coefficients of regression, the standard error of estimate, confidence intervals, p-value (probability value) and variance analysis (see Outputs 1-3). However, the interpretation and discussion of the computer outputs will focus on the aspects of correlation analysis displayed in the computer report sheets.

4.1. Correlation between gas flaring (X) and gas utilization (Y)

Output 1: Computer-report sheet for Analysis of Gas Utilization (as Y-variable) and Gas Flaring (as X-variable)

MegaStat

Regression analysis

0.206 r_1^2

0.454 r_1

14720.997 std. error of estimate

45 observations

1 predictor variable

Y dependent variable

					Confidence interval	
Variables	Coefficients	std. error	t_1(df=43)	p-value	95% lower	95% upper
Intercept	a = -6,163.8106					
X_1	b = 0.9571	0.2866	3.34	.0017	0.3792	1.5350

ANOVA table

Source	SS	df	MS	F	p-value
Regression	2,417,266,384.4351	1	2,417,266,384.4351	11.15	.0017
Residual	9,318,433,786.7649	43	216,707,762.4829		
Total	11,735,700,171.2000	44			

The Peason product-moment correlation coefficient, r_1 is 0.454 (see Output 1). First, it is positive, which is a sign that there is a direct relationship between the volumes of natural gas flared and the volumes of natural gas utilized in the Nigerian petroleum fields. Second, the value of 0.454 (see Output 1) is not close to 1.00, so it could be concluded that the strength of association between the volumes of gas flared and utilized is moderate. Putting it in another way, it could be said that a 54.6 percent increase in the volume of gas flared in Nigeria would likely lead to 54.6 percent more utilized gas volumes in Nigeria.

However, the terms, weak, moderate and strong, do not have precise meaning for describing the coefficient of correlation. A measure that has a more easily interpreted meaning is the coefficient of determination, r_1^2. From Output 1, the coefficient of determination, r_1^2 is 0.206. This is a proportion, or a percentage. Therefore, it could be interpreted that 20.6 percent of the variation in the volumes of gas utilized in Nigeria is explained, or accounted for, by the variation in the volumes of gas flared.

The computed t_1 is 3.34 (see Output 1). Using the 0.05 level of significance, the decision rule states that if the computed t_1 falls in the area lying between plus 2.306, and minus 2.306 then the null hypothesis is not rejected. Since the computed t_1 = 3.34 falls in the rejection region, H_{01} is rejected at 0.05 significance level. This means that the correlation in the population is not zero.

The test of hypothesis will now be interpreted in terms of the p-value. From Output 1, the p-value is 0.0017 at 0.05 significance level. It could be recalled from the Framework Review Section 2.4, that if the p-value is less than 0.01, then there is very strong evidence that H_{01} is not true. This statement supports and agrees with the earlier decision drawn from hypothesis test for t, that H_{01} is rejected at 0.05 significance level, confirming that the correlation in the population is not zero. In the context of this study, it indicates that definitely there is correlation in the population of oil producing companies in Nigeria, with regard to volumes of flared natural gas and utilised natural gas. This may be one of the findings from this study.

4.2. Correlating gas production (X) with gas utilization (Y)

Output 2: Computer report card for Analysis of Gas Production (as X-variable) and Gas Utilization (as Y-variable)

MegaStat

Regression analysis

0.890 r_2^2

0.943 r_2

7063.968 std. error of estimate

45 observations

1 predictor variable

Y dependent variable

					Confidence interval	
Variables	Coefficients	std. error	t_2(df=43)	p-value	95% lower	95% upper
Intercept	a = 16,257.1531					
X_2	b = 1.2145	0.0652	18.63	3.31E-22	1.0830	1.3461

ANOVA table

Source	SS	df	MS	F	p-value
Regression	17,311,194,819.1417	1	17,311,194,819.1417	346.92	3.31E-22
Residual	2,145,684,759.7640	43	49,899,645.5759		
Total	19,456,879,578.9057	44			

The Pearson product-moment correlation coefficient, r_2 is 0.943 (see Output 2). First, the positive nature of the figure is indicative of a direct relationship between the quantities of natural gas produced in Nigeria and the quantities of natural gas utilized in Nigeria.

Second, since the value of 0.943 (see Output 2) is very close 1.00, it could be inferred that the strength of the association between the quantities of gas produced and consumed in Nigeria is strong. This is to say that a 5.70 percent increase in the quantity of gas produced in

Nigeria, would likely lead to 5.70 percent more consumed gas quantities in Nigeria.

Regarding the coefficient of determination r_2^2, Output 2 shows that its value is 0.890. Therefore, it could be interpreted that 89 percent of the variation in the quantities of gas consumed in Nigeria is explained, or accounted for, by the variation in the quantities of gas produced in Nigeria.

From Output 2, the computed t_2 is 18.63. At 0.05 significance level, $t_2 = 18.63$ falls in the rejection region, so the null hypothesis, H_{02} is rejected. This implies that the correlation in the population is not zero. Also, from Output 2, the p-value shows 3.31E-22 at 0.05 significance level. A recall from the Framework Review Section 2.4, shows that if the p-value is less than 0.001, then there is extremely strong evidence that H_{02} is not true. This statement agrees with the earlier hypothesis testing outcome that it is not true that the correlation in the population is zero. From the practical standpoint, the statement above suggests that definitely there is correlation in the population of Nigeria's oil producing companies, regarding the quantities of natural gas consumption and production in Nigeria. This may constitute another finding from this study.

4.3. Correlating gas production (X) with gas flaring (Y)

Output 3: Computer report file for Analysis of Gas Production (as X-variable) and Gas Flaring (as Y-variable)

MegaStat

Regression analysis

0.522 r_3^2

0.722 r_3

14713.831 std. error of estimate

45 observations

1 predictor variable

Y dependent variable

					Confidence interval	
Variables	Coefficients	std. error	t_3(df=43)	p-value	95% lower	95% upper
Intercept	a = -6,110.3897					
X_3	b = 1.9562	0.2857	6.85	2.15E-08	1.3799	2.5324

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	10,147,515,878.3625	1	10,147,515,878.3625	46.87	2.15E-08
Residual	9,309,363,700.5432	43	216,496,830.2452		
Total	19,456,879,578.9057	44			

The computer-evaluated correlation coefficient, r_3 is shown to be 0.722 (see Output 3). The positive value of r_3 suggests a direct relationship between the volumes of gas production and flaring in Nigeria's petroleum field. Second, with the value of $r_3 = 0.722$, which is more than half of 1.00 but not upto 1.00, it could be concluded that the strength of the association between the volumes of gas production and flaring is strong. In other words, this could be expressed that a 27.8 percent increase in the volume of gas production in Nigeria, would likely lead to 27.8 percent more flared gas volume in Nigeria.

Also, in Output 3, the coefficient of determination, r_3^2, is depicted as 0.522. To this end, it could be interpreted that a 52.2 percent of the variation in the volumes of gas flared in Nigeria is explained or accounted for, by the variation in the volumes of gas produced in Nigeria.

The computed t_3 is shown to be 6.85 (see Output 3). At 0.05 significance level, t_3 with the value of 6.85 falls in the rejection region of the normal distribution plot. Therefore, the null hypothesis, H_{03} is rejected. This means that the correlation in the population is not zero. Also, from Output 3, the p-value is depicted as 2.15E-08 at 0.05 significance level. A recall from the Framework Review Section 2.4, reveals that if the p-value is less than 0.001, then there is extremely strong evidence that H_{03} is not true. This statement is consistent with the earlier hypothesis testing outcome that it is not true that the correlation in the population is zero. Putting statistics in action, the statement above is indicative that definitely there is correlation in the population of oil producing companies in Nigeria, in view of the volumes of natural gas produced and flared in Nigeria's petroleum fields. This may be counted as another finding from this study.

4.4. Limitations of correlation analysis

In correlation analysis, no cause-and-effect relationship can be inferred. It seeks to find out if a correlation exists, but it does not suggest that the variations in say flared gas volumes are caused by the variations in gas production volumes, or vice versa. Therefore, there should be caution in the use of correlation analysis outcomes.

5. Conclusion and recommendation

The problems of natural gas production and disposal by flaring have persisted for over 50 years of petroleum production in Nigeria. Within the context of operations research, this endeavour exploited correlation analysis as the tool for the study. At the end, satisfactory insights were gained about the degree of associations between the variables in the gas production and disposal data. Interestingly, some of the major insights and findings obtained from the study are:

1. That there is a direct and moderate association between gas flaring and utilisation as approaches to gas disposal in Nigeria's petroleum industry. Also, another finding from the study indicates that definitely there is correlation in the population of oil producing companies in Nigeria regarding the volumes of flared natural gas and utilized natural gas.

2. That there is also a direct and strong relationship between Nigeria's gas production and disposal by utilisation approach. From practical stand point, an insight from the study suggests that definitely there is correlation in the population of Nigeria's oil producing companies, with regard to the quantities of natural gas consumption (utilisation) and production in Nigeria.

3. That there is also a direct and moderate association between Nigeria's gas production and disposal by flaring method. Also, an insight from the study shows that definitely there is correlation in the population of oil producing companies in Nigeria, in view of the volumes of natural gas produced and flared in Nigeria's petroleum fields.

These insights, no doubt have increased the depth of understanding of the data generated from the operations of the natural gas industry in Nigeria. However, the limitations of correlation analysis call for caution in its application. To this end, a multiple correlation analysis is recommended as a precaution and also a means of furthering the study with the view of comparing the outcomes. Also, a software capable of producing correlogram plot is recommended as a means of advancing the study.

Notation

a	The Y-intercept; The estimated value of Y where the regression line crosses the Y-axis when X is zero.
b	The slop of the line, or the average change in Y_x for a unit change in the independent variable X.
H_0	The null hypothesis
H_1	The alternate hypothesis
n	The number of pair observations
r	The coefficient of correlation
r^2	The coefficient of determination
r_1	The coefficient of correlation for gas utilisation and gas flaring
r_2	The coefficient of correlation for gas production and gas utilisation
r_3	The coefficient of correlation for gas production and gas flaring
t	The test statistic for the significance of r
t_1	The test statistic for the significance of r_1 (the correlation coefficient for gas utilisation and gas flaring)
t_2	The test statistic for the significance of r_2 (the correlation coefficient for gas production and gas utilisation).
t_3	The test statistic for the significance of r_3 (the correlation coefficient for gas production and gas flaring).
X	The X variable
$\bar{X}$	The mean of the X variable
Y	The Y variable
$\bar{Y}$	The mean of the Y variable
Y_x	The predicted value of the Y-variable for a selected X-value
$\sum X$	The X variable summed
$(\sum X^2)$	The X variable squared and the squares summed
$(\sum X)^2$	The X variable summed and the sum squared
$\sum XY$	The sum of the products of X and Y
$\sum Y$	The Y variable summed
$(\sum y^2)$	The Y variable squared and the squares summed
$(\sum Y)^2$	The Y variable summed and the sum squared
Þ	The correlation in the population from which the natural gas production and disposal records were obtained.

References

Erinne, N. J., 1999. Imperatives of process engineering in the midstream petroleum industry. Proceedings of the 29th Annual Conference of the Nigerian Society of Chemical Engineers (NSChE) held in Presidential Hotel, Port Harcourt, November 11-13.

Nigerian National Petroleum Corporation, NNPC, 2009. Annual Statistical Bulletin, Draft Edition, p.17. Accessed on August 24, 2010, from www.nnpcgroup.com/cpdd.

Mason, R. D., Lind, D. A., Marchal, W. G., 1999. Statistical Techniques in Business and Economics, Tenth Edition. Irwin/McGraw-Hill, Boston, p.429, pp.316-317, p.437, p.317 and p.507.

Nwabuokei, P. O., 1986. Fundamentals of Statistics. New Edition with Answers. Enugu, Koruna Books, p.319, p.293, pp.296-297.

Orjih, J., 1996. Business Research Methodology. Meteson Publicity Company, Enugu, p.130.

Orris, J. B., 1999. Megastat Software for Excel, Software that enhances the power of Excel in Statistical Analysis.