

Applying the Path Analysis Method to determine the significance of input parameters on the output of Derbendikhan power station

R. A. Saeed¹, M. M. Fage¹ & F. M. Khoshnaw²

¹*Sulaimani University, Iraq*

²*Loughborough University, UK*

Abstract

Derbendikhan power station works under different load conditions, because the power demand of the electric distribution system varies throughout the day. It is usually higher during daylight and lower at night. The power P generated by a hydropower plant depends on head H , discharge Q and efficiency η . Hence it is necessary to understand how these quantities (H , Q , η and P) vary. Model test data for the operational parameters rotational speed n_m , head H_m , discharge Q_m , load P_m and efficiency η_m can be converted to the values of those quantities in the prototype.

In this study a statistical method – Path Analysis – is proposed for the first time to determine the effect of head H , discharge Q and efficiency η on the power output P of the prototype turbine, and also the behavior of these quantities with respect to one another. The results shows that P mainly depends on H , Q and η , and the total effect of Q on P is more than the total effect of either H or η .
Keywords: Path Analysis Method, hydropower station, operational parameters.

1 Introduction

The hydraulic turbine is usually designed for particular values of net head H , discharge Q , output power P and efficiency η which are known as the design conditions. Relation between any two of the above quantities is called a characteristic of the machine. The Q and H are the physical independent variables. It is, therefore customary to study the variation of all other quantities with respect to the variation of any one of the independent quantities. The power



output from the hydro power plant depends on three factors which are head, efficiency and discharge [1-5].

2 Derbendikhan hydropower station

The Derbendikhan hydropower station was constructed in Iraq on the Diala river 10 km downstream of the confluence of the Tanjero river and about 230 km northeast of Baghdad. In order to obtain complete information regarding the characteristics and performance of a water turbine, it is usual to carry out full laboratory tests on reduced scale mode. Therefore, the behavior of the reduced scale model was studied [6].

2.1 Model testing

The model test for the water turbine of the Derbendikhan hydropower station, were performed by the experimental laboratory of TAKASAGO Technical Institute, Mitsubishi Heavy Industrial, Ltd. The model test data included the rotational speed n_m , head H_m , discharge Q_m , load P_m and efficiency η_m (total of one hundred tests were performed which repeat two tests for each case). A sample of the model test data is shown in table 1.

Table 1: Sample of operation model test data [1].

Test No.	n_m (rpm)	H_m (m)	Q_m (l/s)	η_m (%)
1	1187.2	30.22	692.4	90.55
2	1186.8	30.16	689.4	90.97
3	1192.7	30.47	666.7	88.03
4	1192.6	30.46	644.9	91.85
5	1188	30.24	582.4	93.98

2.2 Derived results of the prototype turbine

Model test data can be used to determine the quantities of efficiency η , output P , discharge Q , and net head H of the prototype turbine by using the converting equations [1]

Table 2 shows the sample of design parameters values of the prototype turbine, which are derived from the model test data.

Table 2: Sample of the design parameters of prototype turbine [1].

No.	η (%)	H (m)	P (kw)	Q (m ³ /s)
1	94.07734	80.66541	89656.91	121.0638
2	94.34057	80.55953	89471.76	120.5799
3	92.49796	80.58434	84681.99	116.0327
4	96.22704	80.61011	77053.98	101.7621
5	94.89209	80.5714	83785.92	112.2481

Design parameter values of the prototype turbine are used in the Path Analysis Method to determine the effect of head H , discharge Q and efficiency η on the power output P , and to study the behavior of these quantities with respect to one another.

3 The normality test

The Chi-square test was used to test if a sample of data came from a population with a specific distribution. The Chi-square test is defined for the hypothesis [7]:

H_0 : The data follow a specified normal distribution.

H_1 : The data do not follow the specified normal distribution

Test Statistic: For the Chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is defined as

$$\chi^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i \quad (1)$$

where O_i is the observed frequency for bin i and E_i is the expected frequency for bin i . The expected frequency is calculated by

$$E_i = N(F(Y_u) - F(Y_l)) \quad (2)$$

where F is the cumulative Distribution function for the distribution being tested, Y_u is the upper limit for class i , Y_l is the lower limit for class i , and N is the sample size.

4 Multiple regressions

Multiple regressions can also use more than two predictors. There is really no limit to the number of predictors that can be included. At the very least, you must have more subjects than predictors. Usually, people recommend that there be 5 or 10 cases per predictor to have sufficient power. This is a very general rule, because, as it is obvious, power depends on many things. A careful power analysis is needed to truly determine the sufficient sample size. With more than two predictors, the regression plane becomes a regression cube. Geometrically speaking, instead of three dimensions, there would be four or five dimensions, depending on how many predictors there are.

The general (population) equation is usually written as follows:

$$y = a + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (3)$$

where β_k represents the last slope and x_k represents the last predictor variable. k represents the number of predictors [8, 9].

We have an overall test of the amount of variation in y that we can account for by the predictors. For this, we are essentially testing the amount of shared variation, the R -square, and how effectively we can predict the y scores overall [10].

Table 3: Summary of the F-test.

Source of Variation	D.F.	SS	Mean Square	F-test
Regression	k	$\sum (\hat{y} - \bar{y})^2$	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Residual(or error)	n-k-1	$\sum (y - \hat{y})^2$	$MSE = \frac{SSE}{n - k - 1}$	
Total	n - 1	$\sum (y - \bar{y})^2$		

4.1 The coefficient of multiple determination

The coefficient of multiple determination, R^2 , is defined as the proportion of the total variation in Y “explained” by the multiple regression of Y on X_1 and X_2 , and it can be calculated by the following equations [11]

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2} = \frac{\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2}{\sum y^2} \quad (4)$$

$$\text{Explained variance} = \sum (\hat{y}_i - \bar{y})^2 / (k - 1) = RSS / (k - 1) = \sum \hat{y}_i^2 / (k - 1) \quad (5)$$

where k is the number of estimated parameters

$$\text{Unexplained variance} = \sum (y_i - \hat{y}_i)^2 / (n - k) = ESS / (n - k) = \sum e_i^2 / (n - k) \quad (6)$$

5 Path Analysis – definition

Path Analysis is a special use of multiple regressions to help understand and parcel out the sources of variance. Many applications that have used stepwise multiple regressions may have really been looking to do a Path Analysis and not realized it [12]

1. Variable 1 is exogenous.
2. Variables 2, 3 and 4 are endogenous.
3. The P_{ij} represent path coefficients from variable j to variable i .
4. The e_i represent errors or residuals for variable i .



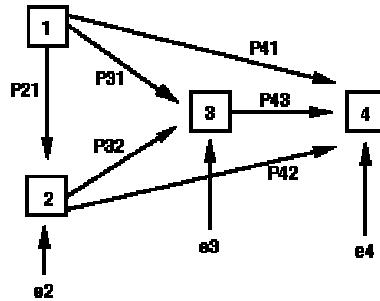


Figure 1: Path Analysis diagram.

5.1 Decomposition of correlations:

Each correlation can be decomposed into one or more of the following four types of effects, as shown in the Figure 2 [12]

- Direct Effect (DE) - path coefficient from one variable to another, e.g., P_{21} .
- Indirect Effect (IE) - sequence of paths through one or more intermediate variables, e.g., $P_{32}P_{21}$.
- Unanalyzed Effect due to correlated causes (U) - correlation of variable with cause of the second, e.g., $P_{42}P_{31}r_{12}$.
- Spurious Effect due to common cause (S) - variable that causes both first and second variable, e.g., $P_{41}P_{31}$.
- Sum of Direct Effect (DE) and Indirect Effect (IE) = total causal part of the correlation between two variables.
- Sum of Unanalyzed effect (U) and Spurious effect (S) = total non causal part of the correlation between two variables.
- Effects relating variables 1 and 2

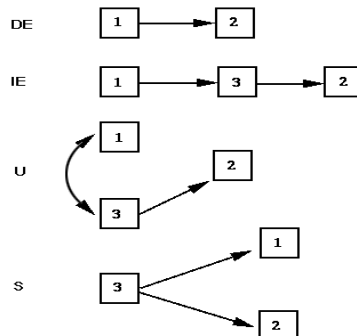


Figure 2: Direct and indirect effect.

6 Results and discussion

In this study a number of computer applications are used for the statistical analysis like MS Excel, Static Graph and Matlab. For each statistical analysis study, it is necessary to test data for normality, which depends on the residual (where residual is difference between real value and estimate value) ($e_i = Y - \hat{Y}$).

Figure 3 represents the normality of the data by using the Chi – square test, it is shown that the data is normal. The hypothesis test is as follows:

$$\begin{aligned} \text{Hypothesis} \quad & H_0: \text{Residual} \sim \text{Normal}(0, \sigma^2) \\ & H_1: \text{Residual} \sim \text{Normal}(0, \sigma^2) \end{aligned}$$

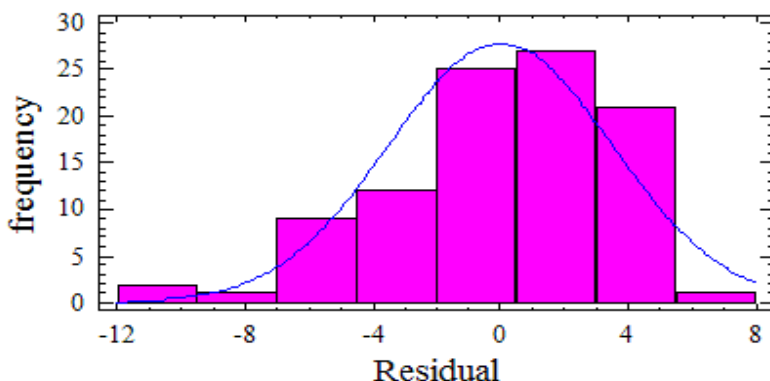


Figure 3: Histogram for residual.

The Chi-Square test can be used for this analysis depending on the P -value. By comparing this value with ($\alpha=0.05$), it can be noted that the P -value of the Chi-Square test is greater than α which means that H_0 is acceptable:

$$H_0: \text{residual} \sim \text{normal}(0, \sigma^2)$$

6.1 Estimation of regression equations

Regression equations show the relation between power production P with Discharge Q , Head H and Efficiency η , as shown below:

$$P = -48.8 - 0.252 \eta + 0.727 H + 0.875 Q \quad (7)$$

Standard Deviation (SD) is one of the measures of dispersion and variation. The SD value shows whether the data is homogenous or not. When SD is small in value, it means that homogenous does exist, in other words the data is

dependable. As shown in table 4, the SD value is small in value ($SD=3.569$), so that data is homogenous.

Table 4: Parameters value, standard deviation and T-test.

Predictor	Coefficient	St. Dev.	T-test	P-Value
Constant	-48.800	2.195	-22.24	0.000
H	-0.25242	0.03614	-6.98	0.000
H	0.72676	0.02158	33.68	0.000
Q	0.87453	0.01811	48.30	0.000

$SD = 3.569$ $R\text{-Sq} = 98.7\%$

We can depend on the regression equation above because the value of R -square is high (98.7%), it means that the effect of Q , H , and η factors on the P are (98.7%), and the remaining (1.3%) occurs through other factors that are not mentioned in this study.

Table 5: Analysis of variance (ANOVA).

Source	DF	SS	MS	F	P- Value
Regression	3	94148	31383	2464	0
Error	94	1197	13		
Total	97	1197			

From table 5, and by comparing the value of F between F -calculated and F -table, we conclude that the F -calculated is greater than F -table. This gives an indication that F is significant, i.e. there is a relation between the depended and independed variables.

The Path Analysis method depends on presenting the causal models, which assumes that the linear relation between cause and effect exist. The correlation matrix that is calculated from the data is shown below:

	η	H	Q	P
η	1.00	0.434	0.756	0.706
H	0.434	1	0.362	0.686
Q	0.756	0.362	1	0.914
P	0.706	0.686	0.914	1

Figure 4 presents the relation (direct and indirect) between cause P and effects H , Q and η

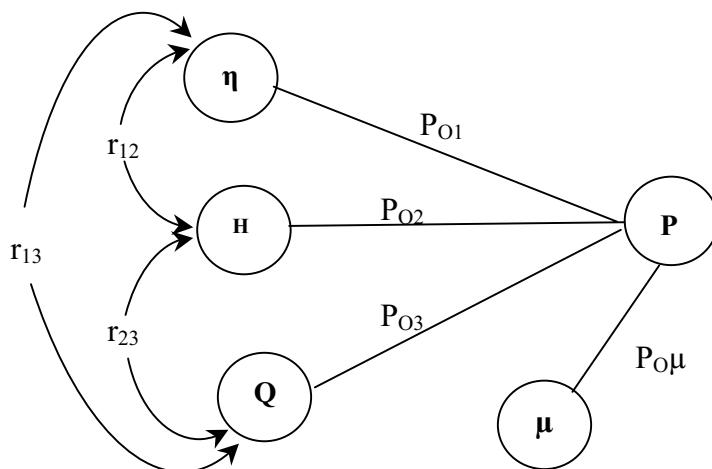


Figure 4: Path diagram.

The universe matrix is derived from the correlation matrix, and the results are shown below:

$$\begin{bmatrix} 2.5024 & -0.4502 & -1.7288 \\ -0.617 & 1.2339 & -0.0976 \\ -1.7247 & -0.1063 & 2.3423 \end{bmatrix} \begin{bmatrix} P_{01} \\ P_{02} \\ P_{03} \end{bmatrix} = \begin{bmatrix} 0.706 \\ 0.686 \\ 0.914 \end{bmatrix}$$

The estimated value of direct effect (P_{01} , P_{02} and P_{03}) is calculated by using equation $P = R^{-1}r$, the result is shown below:

$$\begin{bmatrix} P_{01} \\ P_{02} \\ P_{03} \end{bmatrix} = \begin{bmatrix} -0.122 \\ 0.431 \\ 0.85 \end{bmatrix}$$

Table 6 shows that changing the efficiency η , as much as one standard deviation, causes a direct change in the rate of power output P , which is equal to -0.12, as well as indirect change through the variant head H and flow rate Q equal to 0.1853 and 0.6426. Consequently the total effect of the efficiency would be equal to the direct change and indirect change (0.7059).

Moreover, when the head H is changed as much as one standard division, it will lead to the direct change of the rate of power output P which is equal to 0.431, as well as indirect change through the variant efficiency η and flow rate Q equal to -0.0525 and 0.3077; consequently the total effect of the efficiency would be equal to the direct change and indirect change (0.6862).

When the flow rate Q is changed as much as one standard division, it will lead to the direct change of the rate of power output P which is equal to 0.850, as well as indirect change through the variant efficiency η and head H equal to -0.0922 and 0.1560; consequently the total effect of the efficiency would be equal to the direct change and indirect change (0.9138).

Table 6: Direct and indirect coefficient value of the Path Analysis result.

ID	Type effect	Coefficient Value
1	Effect η on P	
	a)Direct Effect= P_{O1}	-0.12
	b)Indirect Effect	
	$\rightarrow H = r_{12}P_{O2}$	0.1853
	$\rightarrow Q = r_{13}P_{O3}$	0.6426
	Total Effect	0.7059
2	Effect H on P	
	a)Direct Effect = P_{O2}	0.431
	b)Indirect Effect	
	$\rightarrow \eta y = r_{12}P_{O1}$	-0.0525
	$\rightarrow Q = r_{23}P_{O3}$	0.3077
	Total Effect	0.6862
3	Effect Q on P	
	a)Direct Effect= P_{O3}	0.850
	b)Indirect Effect	
	$\rightarrow \eta = r_{13}P_{O1}$	-0.0922
	$\rightarrow H = r_{23}P_{O2}$	0.1560
	Total Effect	0.9138

Table 6 shows that the total effect of discharge (0.9138) on the power output is more than the total effect of net head (0.6862) and efficiency (0.7059). This result agrees with Rai [4], who confirms that the power output P from hydro power plant depends on the head H , discharge Q and efficiency η , and the power generation mainly depends on the quantity of water available.

7 Conclusions

1. This paper presents a new method to analyze the significant effect of the input parameters on the output power using the Path Analysis technique.
2. The research allows one to compute the total effect of discharge on the power output which is 0.9138 more than the total effect of net head and efficiency which is 0.6862 and 0.7059.
3. The R-square value is %98.7, which means that we can depend on the independent variable. This model can be also used to predict power production.

References

- [1] Saeed R. A., "An investigation about preventing the mechanical failures that occur in Derbendikhan power station", M.Sc. Thesis, University of Sulaimani, 2003.
- [2] Jain, A. K., "Fluid Mechanics", Delhi, 1995
- [3] Fink, D. G. and Bealy, H. W., "Standard Handbook for Electrical Engineering", McGraw-Hill, 1999.
- [4] Rai, G. D., "An Introduction to Power Plant Technology", Delhi, 1995.
- [5] Franzini, J. B. and Finnemore, E. J., "Fluid Mechanics with Engineering Application", McGraw-Hill, 1997.
- [6] Catalog of Derbendikhan Hydropower Station, Mitsubishi Electric Corporation (Tokyo, Japan)
- [7] Bernad Ostel, "Statistics in Research", the Iowa state University, 1963.
- [8] Chatfield, Christopher, "Statistical for Technologe", 2nd.Ed., John Wiley and Sons, New York, 1978.
- [9] Draper, Norman & Smith, Harry, "Applied Regression Analysis" 2nd.Ed., John Wiley and Sons, New York, 1981.
- [10] Evgenia Vogiatzi, "Coefficient of multiple determinations and Partial Correlation-Coefficient", 2002.
- [11] Li.C.C. "Path Analysis – A primer", The Boxwood Press, CA, 1977
- [12] See this site: <http://www.gseis.ucla.edu/courses/ed230bcl/notes2/path1.html>

