ETM

ENGINEERING & TECHNOLOGY MANAGEMENT

Title: Multiple Regression for Daily Chlorine Use

Course:
Year: 1994
Author(s): V. Han

Report No: P94004

ETM OFFICE USE ONLY	
Report No.:	See Above
Type:	Student Project
Note:	This project is in the filing cabinet in the ETM department office.

Abstract: This project is to find if there were any correlation between the parameters, two main ones are BOD (biochemical oxygen demand) and SS (suspended solids), and the number of lb. of chlorine used daily in Columbia Boulevard Wastewater Treatment Plant. The goal is to have a regression equation that would predict the daily chlorine use with accuracy.

Multiple Regression for Daily Chlorine Use

Vu Han

EMP-P9404

EMGT 510
Research Methodology

TERM PROJECT

"Multiple Regression for Daily Chlorine Use"

Submitted to: Dr. Deckro
Submitted by: Vu Han

Portland State University
Engineering Management Program
March 11, 1994

TABLE OF CONTENTS

Introduction 1
Objective 2
Data Collection 3
Definition of Variables 3
Multivariate Analysis Techniques 4
Data Analyses 5
Interpretation 5
Conclusion/Recommendation 8
References 10

EXHIBIT I: Monthly Data Reports for 1993
EXHIBIT II: Data List for the Analyses

ATTACHMENT A: SPSS Printouts of Normal Regression
ATTACHMENT B: SPSS Printouts of Stepwise Method
ATTACHMENT B: SPSS Printouts of Factor Analysis

INTRODUCTION

The Columbia Boulevard Wastewater Treatment Plant (CBWTP) is located in North Portland. It has a design capacity of treating 100 million gallons per day (MGD) of wastewater, and serves a population of about 423,000 people.

The CBWTP receives wastewater and storm water (runoffs) from households, industries, institution, etc. through branches (pipes range from 4" to 12" in diameter). These branches then connected to collectors (pipes range from 12" to 48") and eventually merged to three large interceptors to the CWBTP.

When the wastewater is conveyed to the CBWTP, it has to be treated by many unit operations and processes. It has to be properly treated and disinfected before being discharged to the Columbia River.

Wastewater has many characteristics and there are many parameters used as measures, but the two most widely used parameters to measure wastewater characteristics are BOD (biochemical oxygen demand) and SS (suspended solids). BOD is a measure of dissolved oxygen used by microorganisms in biochemical oxidation organic matter [1], and SS is the nonfilterable solids through a filter with nominal size of 1.2 micrometers [2].

Each month, a Monthly Data Report must be submitted to DEQ (Department of Environmental Quality) by the CBWTP. In this data report, daily data on BOD and SS from the CBWTP are measured against the standards set by DEQ. It requires at least 85% of the plant influent BOD and SS must be removed daily, a discharge of less than 30 mg/L of effluent BOD and SS, and no more than 25,000 lbs of BOD and SS per day are allowed to discharge into the Columbia River. The standards are set for daily monthly average. Any violation of the standards will result in fines. The Monthly Data Reports are shown in Exhibit I.

A parameter in the Monthly Data Report that measures direct cost in complying to the DEQ standards is the amount of chlorine used daily to disinfect the plant effluent. The amount of chlorine use depends on the wastewater characteristics, flow, treatability of the plant, and many other factors. The purpose of this paper is to provide a model that would predict the daily amount of chlorine used base on some of the parameters selected from the Monthly Data Report.

OBJECTIVE

The objective of this paper is to find if there were any correlations between the parameters (selected from the Montly Data Report) and the number of lbs of chlorine used daily. If there correlations, then find the variables that are most correlated and also obtain a regression equation that would predict the amount of chlorine use daily. the goal is to have a regression equation that would predict the daily chlorine use with accuracy; that is, a coefficient of determination (R square) of at least 0.75.

considered significance except for X3, X4 and X11 because their Sig T values are less than 0.05.

Refer to Attachment B.

With using the stepwise method to introduce one variable into the regression equation at a time, and at a significance level of 95%, the R square of X8 comes out to be 0.083. This is equivalent to 34% of the R square for all the variables included in the equation.

When X5 was introduced to the regression equation along with X8, R square increased to 0.153. This means that with these two variables alone in the regression equation, they explain 15.3 % of the variance in X10. This is equivalent to 63% of R square of all variables in the regression equation.

X9 was next introduced next to the equation along with X8 and X5, the R square increased to 0.175. Again, it means that with X5, X8, and X9 alone in the regression equation, they explain 17.5% of the variance in X10. This is equivalent to 71% of R square of all variables in the regression equation.

None other variables got added to the regression equation after X9 because they were considered insignificant at 95% confidence level; therefore, they should be dropped from the equation.

Refer to Attachment C.

The correlation matrix in Attachment C shows that X1, X2, and X3 are highly positively correlated. X4 and X5 are correlated. X6, X7, X8, and X9 are highly positively correlated.

X11 has low correlation with all the variables. Hence, most of the variables in the regression equation are collinear. the collinearity between the independent variables could affect the prediction of the dependent variable X10.

Principle components technique was used to reduce the collinearity between the independent variables; hence, better prediction of the dependent variable X10.

In the Factor Matrix, three factors were extracted for the variables; other factors had been dropped because their Eigenvalues were less than 1. As one can see, X1 to X9 are heavily loaded on Factor 1. This factor could represent BOD and SS discharge. X1 to X3 are negatively loaded on Factor 2. This factor could represent the flow of sewage through the plant. X4 and X5 are loaded on Factor 3. This factor could represent the influent BOD and SS.

In the Final Statistics section, the three extracted factors significantly explained the variance in each of the variables (percent of variance explained are listed under communality), except for X10 and X11. Eigenvalues for the three factors are also listed in this section and they are all greater than 1 because that is how the factors were extracted in the first place. The Percent of Variance is the percent of total variance explained by each factor. For example, Factor 1 alone explained 36.8% of the total variance in the variables. All three factors explained 73.1% of the total variance in the variables.

In the Factor Score Coefficient Matrix, the factors have been rotated to reduce the collinearity between the independent variables. However, the factors remain 90 degree (orthogonal) with each other. These new rotated factors are called principle components. As one can see, the variables are no longer significantly loaded on any of the factors. The purpose of the principle components is to reduce the collinearity between the independent variables.

The principle components were then regressed on X10. The value of R square increased to 0.4239 compared to 0.245 without using principle components. This increase is equivalent to 73%.

CONCLUSION/RECOMMENDATION

The coefficient of determination (R square) of the independent variables regressed on X10 is 0.245. This R square is considered low, therefore, the prediction of X10 is low. There are some correlations between the independent variables and X10. The variables that are significantly correlated (at 95% significance level) with X10 are X5, X8, and X9. These three variables alone captured 71% of R square of all the variables.

When the variables were grouped into three factors, X1 to X3 were negatively loaded on Factor 2. This factor is the opposite of the other two factors. This could not easily be seen with the correlation matrix. The three factors then were converted to principle components, and hence, the collinearity between the variables was reduced. These principle components were then regressed on X10 and the R square increased to 0.4239.

REFERENCES

[1] George Tchobanoglous and Franklin L. Burton, Wastewater Engineering, McGraw-Hill Publishing Co., New York: 1991, p. 71.

[2] Ibid., p. 52.