J. ENVIRON. SYS., Vol. 4(2), Summer, 1974

REGRESSION ANALYSIS AS AN AID IN MANAGING A MARINE ENVIRONMENTAL PROTECTION PROGRAM[1]

S. AWERBUCH
Research Associate
Program in Public Management
H. SMITH, JR.
Dean, School of Management
W. A. WALLACE
Director, Program in Public Management
Rensselaer Polytechnic Institute
Troy, New York

ABSTRACT

The U.S. Coast Guard has been given responsibility for protection of the marine environment within navigable waters. These responsibilities have resulted in the need for an environment management information system. The work reported on this paper is one aspect of such a system, assessing the value of regression analysis in allocating resources to the operations of field personnel in the Coast Guard.

Introduction

The Federal Water Pollution Control Act, as amended by the Water Quality Improvement Act of 1970, greatly increased the responsibilities of the Coast Guard in the protection of the marine environment. The new responsibilities most pertinent to this paper are that the U.S. Coast Guard (USCG) has been:

1. designated as the "appropriate agency" for the purpose of receiving the notice of discharges of oil or any hazardous substance in violation of the Act;
2. authorized to issue and enforce regulations establishing methods and procedures for removal of discharged oil; and

[1] Supported in part by U.S. Coast Guard, Department of Transportation under Contract No. DOT-CG-23876-A.

doi: 10.2190/732G-78ML-9PVB-769G
http://baywood.com

3. authorized to assess civil penalties of up to $5,000 for discharging oil in violation of the Act and $10,000 for failure to report the discharge.

Although the foregoing tasks build upon the traditional activities and strengths of the Coast Guard, they also represent major new endeavors, requiring new management systems.

The Coast Guard established a Marine Environmental Protection Program (MEP) to carry out its new responsibilities. Its first efforts in designing management procedures and systems focused on the development of an information system for collecting discharge information, the Pollution Incident Reporting System (PIRS), and promulgating performance standards detailing desired levels of field activity in marine environmental protection. Additional development was found necessary in order to provide a timely, accurate reporting system for management needs and information demands generated by sources external to the Coast Guard, and to ensure effective managerial planning and control by relating resource allocation to field operations. The purpose of this paper is to report on one aspect of the study commissioned [1] to meet the foregoing goal: the assessment of regression analysis [2] to relate resources to field operations.

The MEP program is a staff function reporting to the Chief of Staff of the Coast Guard. The activities of the program are performed by field personnel at the Captain-of-the-Port level (COTP), e.g., COTP/Seattle, and the District level, e.g., Ninth District comprising the Great Lakes. The program staff is not only responsible for developing and promulgating instructions for the field personnel such as percentage of all tank vessels that must be boarded for inspection, but must budget for the necessary men and materials to carry out its assigned mission. Therefore, it was necessary for the program staff to develop reporting systems that would permit identification of resources utilized by the field units.

In order to accomplish this objective, the use of USCG resources in MEP activities, i.e., inputs, to the results of those activities, i.e., outputs, had to be related. For example, activities such as vessel boardings, or waterfront facility inspections should relate to outputs, such as number and volume of pollution incidents.

Regression analysis was tested as a possible analytical tool for determining the relationship between various program activities (i.e., controllable variables) and pollution incidents. Its value to the management of the MEP program was assessed: 1) in measuring cost-effectiveness of the activities of the program; 2) as a tool for use within the "management by exception" framework; and 3) assisting in the budgetary process by demonstrating effective dollar allocations.

Each of these applications will be discussed in the material to follow.

Data Collection

The data used in the regression analysis consisted of the following:

1. *Independent Variables:* Members of the study team retrieved data on USCG activities related to MEP. The following items were obtained:
   a. number of tank ships boarded during the reporting period,
   b. number of barges boarded during the reporting period,
   c. number of waterfront facilities inspected,
   d. hours of port safety promotion,
   e. number of pollution incidents reported or observed,
   f. number of pollution investigations performed and man-hours expended in retrieving these data, and
   g. hours of harbor patrol.
2. *Dependent Variable:* The response variable in all cases is the volume of pollution incidents for a district. These data were obtained from PIRS. The volume data were used in two ways, "raw" and "corrected." The raw data consisted of the yearly spill volume for a given district, while the corrected spill data set was constructed by deleting all spills of 20,000 gallons or more from the raw data. Spills over 20,000 gallons were determined to be accidents–not in the regular controllable system.
3. *Weighting Factor:* The purpose of the regression analysis was to explain pollution incidence as a function of the activities performed by COTP's and other reporting units. In some of the analyses districts were put on an "equal basis" by weighting their activities and the volume of spills experienced. It was intuitively felt, at the outset, that the weighting factor should be the number of bulk liquid transfer operations in a district during a given year. The reason for this is that spills are felt by Coast Guard personnel to be related to the number of operations since each operation, whether large or small, involves hook-up, disconnect, judgment in filling tanks, etc. However, the total number of liquid bulk transfer operations by district was not available. In its place, a surrogate measure–the volume of liquid bulk handled by the district–was used; these volume figures were obtained from Part V, Waterborne Commerce of the United States [3].

Models of various forms were analyzed with the objective of yielding the following results:

1. District comparisons: this will enable the development of control charts, based on 3-sigma confidence intervals, to identify unusual performance using a "management by exception" concept.
2. Time-series: prediction of expected spill volumes and appropriate prediction intervals is the objective.
3. Inter-variable comparisons: this analysis will show the relative importance of the several MEP related activities on reduced spill incidence providing a cost/effectiveness capability with regard to specific mission components.

The following discussion uses examples which are based on currently available data. Recommendations for data augmentation made as a result of the study are currently being implemented and will permit more detailed and precise analyses in future applications.

MANAGEMENT CONTROL MODEL

A model for predicting the quarterly average volume of pollution incidents per district was developed. The data used were district oil discharge volumes corrected for major discharges (defined to be greater than 20,000 gallons). The variables studied in this model formulation were:

R_i = response: average district quarterly oil discharge volume, excluding major discharges for eight quarters, $i = 1, 2, \ldots, 8$;
U_{1i} = average number of boardings per district by quarter;
U_{2i} = average number of facility inspection per district by quarter;
U_{3i} = average number of harbor patrol hours per district by quarter;
U_{4i} = average number of promotion hours per district each quarter.

The best fitting predictive model found was

$$R_i = 46{,}846.89 + 383.13(U_{1i} - 615.88) + 10.48(U_{4i} - 1031.25),$$

where

46,846.89 = average quarterly district oil discharge volume in the years 1971-72.

The overall regression was significant at the 5% level, all points are within 95% confidence bands, and $R^2 = 76.55\%$, the amount of variability in the data explained by the model.

Using the prediction model, three-sigma control chart bands (i.e., the probability that a district's expected performance will fall outside the confidence bands due to random fluctuations is about .02%) can be placed on the expected district quarterly oil spill volume. These bands are shown in column 1 of Table 1. The actual district quarterly oil discharge volumes can now be compared to these norms. These are shown in the columns headed by District Numbers. The symbol H indicates the district is out of limits on the high (H) side. If this tool had been available earlier, action would have been taken to attempt to improve the performance of these districts. Usually the action needed would be a function of the characteristic of the district and its unique problems, such as more manpower and boats assigned to a district with a large geographic area.

Another way to look at this control procedure is shown on Figure 1. If the cyclical variations are neglected for the moment, a gradual improvement in spill volume is discernible by the line "Expected Average." This line is a plot of predicted yearly average spill volumes.

Table 1. District Control Chart

Predicted	*District*									
3 Sigma Bands	*1*	*2*	*3*	*5*	*7*	*8*	*9*	*11*	*12*	*13*
Quarters 1971										
First - 5,030-161,374		H				H				
Second - 14,303-111,105										
Third - 9,734-103,606						H				
Fourth - 27,212-122,458						H				
Quarters 1972										
First - 662-162,408			H							
Second - 32,614- 88,766		H	H							
Third - 58,277- 83,391										
Fourth - 36,915- 94,803						H				

Any district showing a quarterly volume below the upper boundary would not be subject to management concern. However, any district falling outside the line would be required to justify its performance and indicate steps taken to improve it.

The dotted lines (1973) indicate the targets, as given in Table 2, that management might deem reasonable. The targets also follow the cyclical pattern but show a general decrease in oil spill volume and also, a tightening up of allowable variances from norm. The dotted lines show one possible way of setting targets and tolerance bands. In reality, since this is a management decision based upon many factors, it would only be set after careful thought and discussion. For example, the control limits might no longer be 3 sigma limits, but say, 2 sigma limits. This means districts would be tagged somewhat more often for variances than would have been done previously.

The predictive model uses only two predictor variables, boardings and promotion. Thus, the model is conditioned by the present method of operation of the Coast Guard with respect to facility inspections and harbor patrols. Any district policy or scheduling changes in these two variables might seriously alter the ability of the chosen model to predict oil spill volume.

A MODEL FOR DETERMINING COST/EFFECTIVENESS

A model for explaining annual discharge volume was found useful in determining the effectiveness of several MEP related activities in reducing pollution incidence. This model attempts to eliminate inter-district differences by weighting all variables by the total volume shipped and received in the district. Data for ten districts in 1971 were used in this example.

The variables studied were:

Y_i = yearly volume of discharges in district i (i = 1, 2, . . ., 10),

X_{1i} = average quarterly number of tanker and barge boardings in district i,

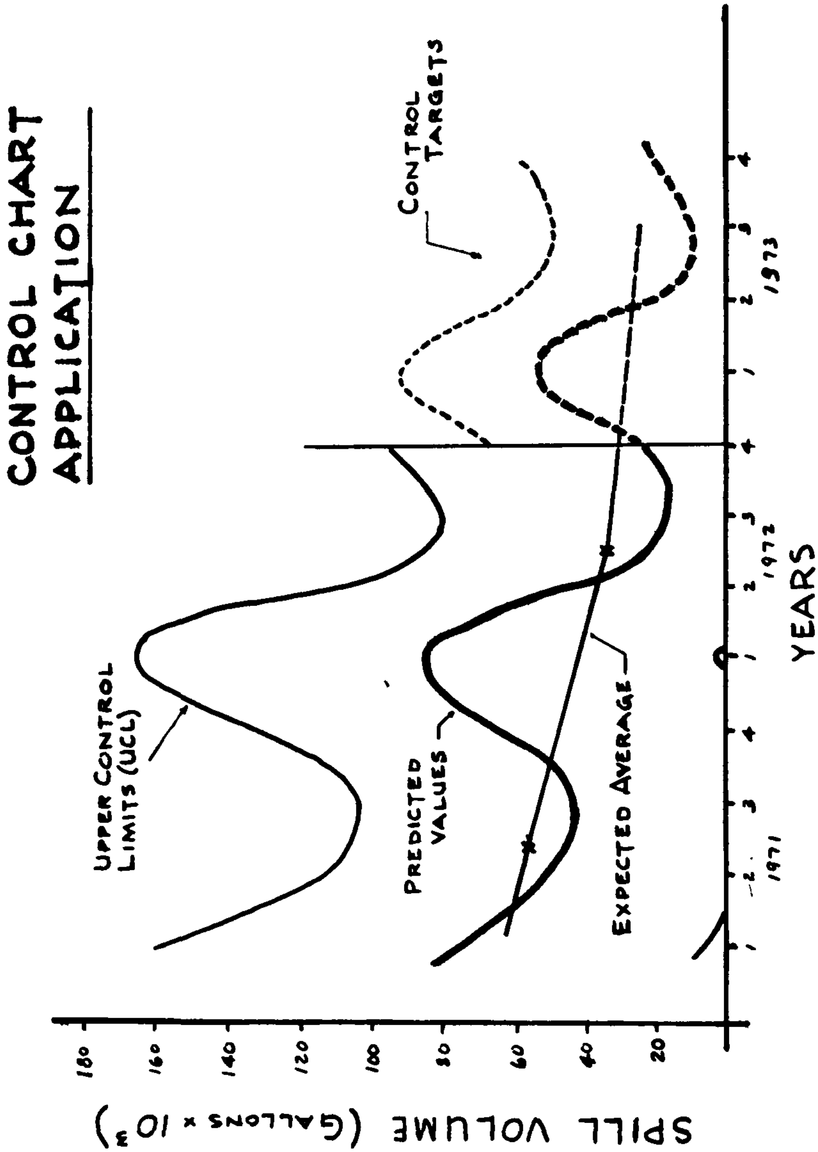

Figure 1. Control Chart Application.

Table 2. Predicted Values and "Targets"

Quarter	*1971*	*1972*	*% Change*	*Target 1973*
First	83,202	81,535	-2	50,000
Second	48,401	28,076	-42	20,000
Third	46,936	12,557	-73	10,000
Fourth	47,624	26,441	-44	20,000
Yearly	56,541	37,152	-34	25,000

X_{2i} = average quarterly number of waterfront facilities inspected in district i,
X_{3i} = average quarterly number of hours of harbor patrol in district i, and
X_{4i} = average quarterly number of hours of port safety promotion in district i,

with

W_i = weighting factor: total tonnage of petroleum and petroleum products (in thousand of tons) shipped and received in district i in 1970.

The best fitting explanatory model using all four variables was

$$\begin{aligned} Y_i = 21.426 - & [.19791 \times 10^8/(X_{1i}W_i] \\ + & [.89203 \times 10^8/(X_{2i}W_i] \\ + & [6.3465 \times 10^8/(X_{3i}W_i] \\ - & [1.4580 \times 10^8/(X_{4i}W_i] . \end{aligned}$$

The R^2 of the model using four variables was 91.80% indicating that about 92% of the variability of the data was explained by the model.

Both the weighted number of boardings and the weighted hours of harbor patrol were found to have a high correlation with the response, the weighted number of gallons discharged; the weighted number of facilities inspected and the weighted hours of port safety promotion had a low correlation with the response variable. The reason for the low correlation was due to the fact that waterfront facilities are "mixed," since data permit separation of cargo vessels from tank vessels and barges, but not differentiation between liquid bulk facilities and dry cargo facilities; a report system was proposed to distinguish between the two types of facilities. Using such data, the correlation between liquid bulk facility inspections and spill incidents may be improved.

The results of the correlation analyses permit the construction of an "effectiveness" graph as shown in Figure 2, a plot of weighted spill volume versus the reciprocal of the weighted hours of harbor patrol. It can be seen from Figure 2 that as the hours of harbor patrol are increased, the spill volume

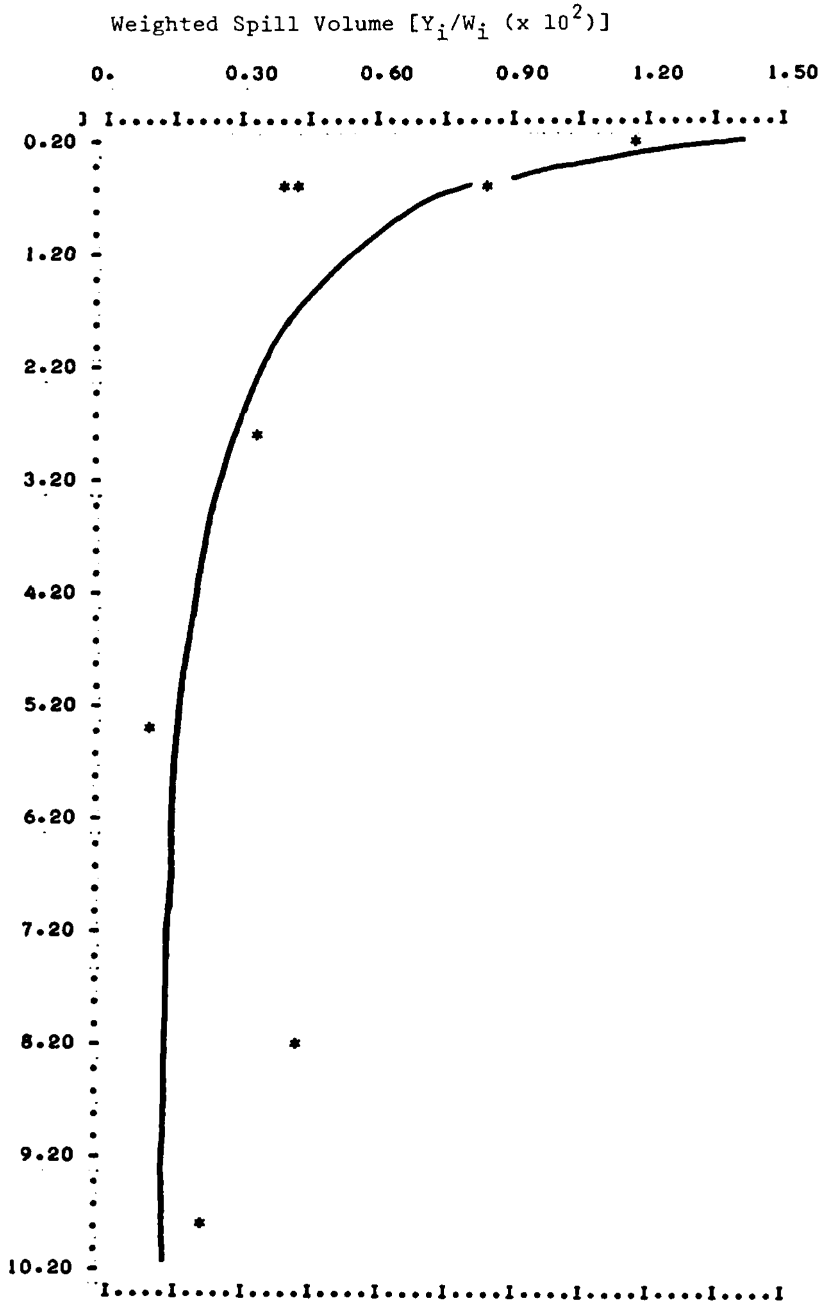

Figure 2. 1/Weighted hours of harbor patrol [$1/X_{3i}\,W_i$ (x 10^7)] volume spilled versus hours of harbor patrol.

decreases. While this group is useful, it should be remembered that it represents the effect of changing one variable only, and does not take into account the intercorrelation between variables. Using a bi-variate model as a predictor, in conjunction with USCG cost standards based upon recovery costs [4], a cost effectiveness capability can be developed.

A TIME SERIES MODEL

Regression analysis was also applied as a means for analyzing district performance over time. The following model was developed:

$$\overline{Y}_i = .4913 + 174.2(X_{1i}/W_i) + 44.81(X_{2i}/W_i)$$

where

$$\overline{Y}_i = Y_i/W_i.$$

This model had an R^2 of 66% and was significant at the 5% level. The predicted responses can be plotted for both years (Figure 2). Spill volumes have been weighted by the total volume of product handled in the district. Further, all major spills are deleted. Figure 2 is therefore a meaningful representation of the relative performance of the various districts, and their performance over the two years for which data were available.

The following comments should be made. First, as can be seen from Figure 3, the districts seem to fall into two clusters with the First, Seventh, and Eleventh Districts having a distinctly higher predicted response, i.e., volume of non-major spills ÷ total tonnage handled, than the remaining districts. Assuming good data and reasonable fit, this would imply that management undertake an inquiry to determine the cause of poorer performance by these three districts.

Second, all districts except for the First District and the Fifth District show an essentially constant or downward trend in the response. Again, this indication should bring further investigation. It is possible that the First and Fifth Districts had a significant increase in volume of product handled during 1972. Such a significant increase would not be reflected in the weighted response to this example since the weighting factor used is the 1970 total volume handled. If no sizable increase, e.g., one considerably larger than the average increase of all districts, in volume handled can be found, then events such as unusually poor weather or personnel cutbacks could be the cause.

Conclusions

Regression analysis using the available data led to the enlargement of the Pollution Incident Reporting System to include data on what were called Type I and Type II environmental (or noncontrollable) variables defined as:

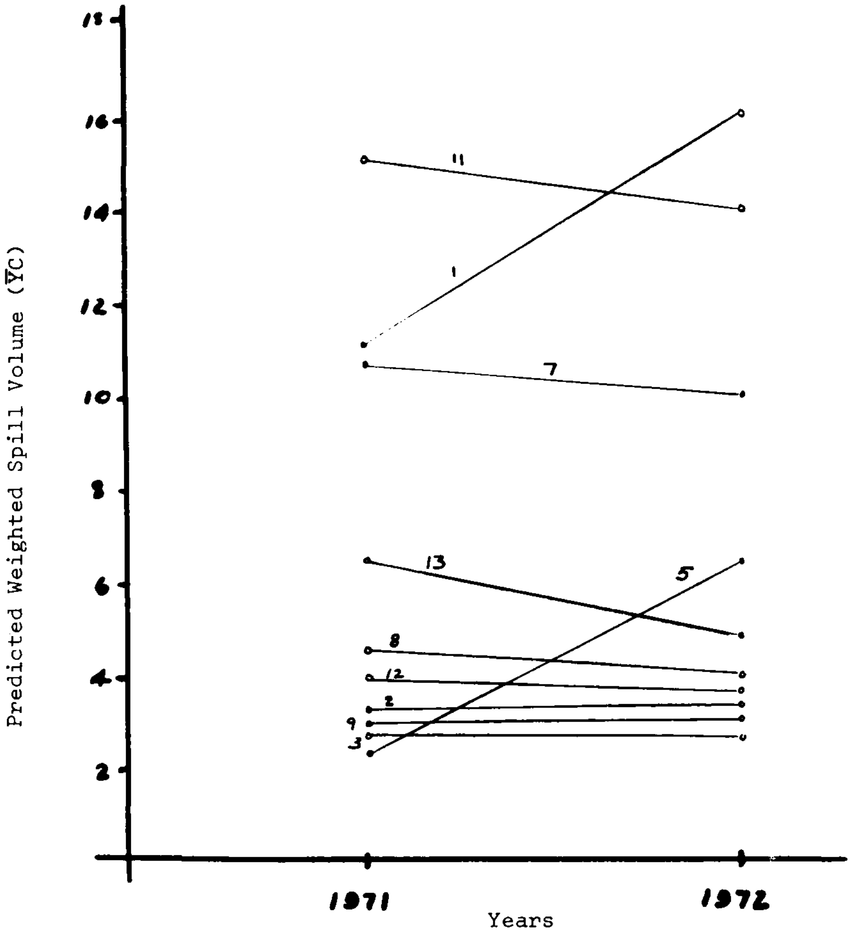

Figure 3. Predicted spill volume by district.

1. Type I–those that affect the level of performance of USCG at a particular organizational unit. These variables are characteristics that separate and differentiate the various operating units;
2. Type II–those that affect USCG's performance in a given operational context. These variables are characteristics that identify the uncontrollable circumstances surrounding a particular occurrence.

Examples of these variables are listed in Table 3 and 4. Data on these variables will be investigated for their ability to predict pollution incidence and further explain the performance of the operating units. In addition, a reporting system was initiated to gather more detailed data on the inputs, the resources utilized in carrying out MEP activities.

Table 3. Type I Environmental Variables

1. Climate	6. Type and size of vessels using vessels
2. Type of water bodies	7. Number and type of transfer operations
3. Miles of shore line	8. Number and type of facilities
4. Location of USCG facilities	9. Type of material handled
5. Amount of tonnage handled	10. Responsibility for recovery operations

Table 4. Type II Environmental Variables

1. Season	9. Source of spill
2. Time of day	10. Wind speed
3. Location of incident	11. Wind direction
4. Type of water body	12. Velocity of current
5. Type of material spilled	13. Direction of current
6. Quantity spilled	14. Responsibility for On-Scene-Coordinator of recovery
7. Quantity being transferred	
8. Affected area	

Statistical routines are being developed for use by USCG personnel to implement the management control model, cost/effectiveness model and time series model. Summary reports to MEP program management as well as detailed analyses will be the result.

Regression analysis was found to be a valuable aid in screening existing data for poor responses due to lack of information or inaccurate responses to the questionnaire and designing new reporting systems; this work assisted in the development and implementation of an augmented Pollution Incidence Reporting System and our new joint Port Safety and Security and Marine Environmental Protection Program Quarterly Report on resources. More importantly, the rigor of this technique forced MEP personnel to focus on specific items of program effectiveness, i.e., those quantitatively defined in terms of resources (inputs) and activities (outputs).

REFERENCES

1. Program in Public Management, "Measures of Program Effectiveness Report to the U.S. Coast Guard Marine Environmental Protection Program," Rensselaer Polytechnic Institute, Troy, N.Y., July 1973.
2. N. Draper and H. Smith, Jr., *Applied Regression Analysis*, John Wiley and Sons, New York, N.Y., August 1966.
3. Corps of Engineers, *Waterborne Commerce of the United States*, U.S. Army, 1971.
4. Department of Transportation, "COMMANDANT NOTICE 7302, Standard Rates For The Recovery of Costs For Pollution Incidents," U.S. Coast Guard, May 22, 1972.