6-1-1974

Application Of Statistical Techniques To The Selection Of An Optimal Pollution Treatment Program

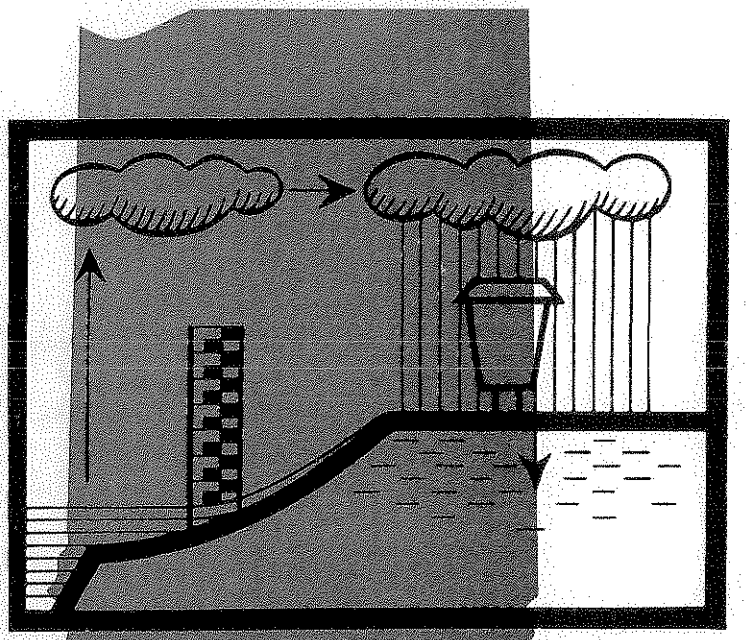
A. B. Whinston

J. R. Marsden

D. E. Pingry

TECHNICAL REPORT NO. 51

APPLICATION OF STATISTICAL TECHNIQUES TO THE SELECTION OF AN OPTIMAL POLLUTION TREATMENT PROGRAM



by

A. B. Whinston

J. R. Marsden

D. E. Pingry

June 1974

PURDUE UNIVERSITY
WATER RESOURCES RESEARCH CENTER
WEST LAFAYETTE, INDIANA

APPLICATION OF STATISTICAL TECHNIQUES TO THE SELECTION OF AN OPTIMAL POLLUTION TREATMENT PROGRAM

by

A. B. Whinston
J. R. Marsden
D. E. Pingry

The work upon which this report is based was supported in part by funds provided by the United States Department of the Interior, Office of Water Research, as authorized by the Water Resources Research Act of 1964 (PL 88-379 as amended).

Krannert Graduate School of Industrial Administration
Purdue University

A completion report of OWRR Project No. A-24-IND
Period of Investigation: July 1, 1971 - June 30, 1974

Technical Report No. 51

Purdue University Water Resources Research Center

West Lafayette, Indiana

June, 1974

Foreword

The study reported in the document was supported by the U.S. Department of the Interior, Office of Water Resources Research as an annual allotment grant of the Purdue University Water Resources Research Center. Project personnel were Dr. Andrew Whinston, Principal Investigater, and Dr. James Marsden and Dr. Dave Pingry. Acknowledgements of support is given to Dr. Daniel Wiersma, Directer of the Indiana Water Resources Center.

The authors are indebted to members of the Indiana Stream Pollution Control Board for assistance at numerous stages of the project. In particular, we cannot fail to mention Dr. T. P. Chang and Mr. Steve Kim. At Purdue, we were fortunate to have the advice of Dr. J. Etzel and Dr. R. Squires.

Table of Contents

Abstract

This report represents a broad range of research directed towards the application of statistical and related techniques to the selection of an optimal pollution treatment system. The work attempts to develop several techniques which can be used in analyzing data obtained from waste treatment plant operations. Using the concepts of cost and production functions, we explore the possibility of developing useful models for the operation of waste treatment plants. The Federal Water Control Act and the 1972 amendments have given to the states the problem of developing and monitoring regional plans for improving water quality. State pollution agencies are turning to the use of computers and modelling techniques for analyzing these problems. It is hoped that the work reported in here will eventually lead to techniques that help in this task.

Chap. 1 Economics of Wastewater Treatment: The Role of Regression

The recent interest in water quality has increased the concern with the costs of constructing and operating waste treatment plants. As evidence of this trend, we note the proliferation of articles on this subject in the various journals concerned with waste water treatment.[1] The majority of these articles have been written by engineers with a great deal of technical experience in the area of concern.

In the field of Economics there has been a long historical development in cost and production theory and applications. Several industries have been studied in detail with concentration on questions such as input interaction and substitutability, returns to scale, and optimal input allocation.[2] In an attempt to encompass increasing technological possibilities the functional forms employed in these studies have been expanded continuously since the pioneering work of Cobb and Douglas in 1928 [2][3].

As might be expected, however, the work done by economists has usually been of a general nature and often did not reflect a sophisticated view of the underlying technical relationships. It is the purpose of this paper to indicate to individuals possessing the technical skills some of the results of more general discussions of cost function and production function estimation.

Previous work by engineers in this area, while displaying technical experience, has suffered from myopic consideration of alternatives. Consideration has been centered on two dimensional log-log plots [e.g. Smith [11]] and two variable linear regression such as Michel [8]. Shah and Reed [10] present regressions of construction costs on design flow and design BOD levels. Rasmussen and Pisano [9] attempted to explain operating and maintenance cost using somewhat more involved functions, but, as they

note, they failed to consider treatment level or efficiency as explanatory variables. In general, the results presented thus far have been extremely limited with respect to the functional forms employed and the consideration of input interactions and process treatment levels. Several authors also have made definite policy proposals and forecasts based on inappropriate extensions of regression results.

This paper presents some initial work on developing operating cost functions for activated sludge waste treatment plants in a framework allied to the developments referred to in Economics. Section I details some specific alternative functional forms which might be employed, noting some of the specific implications and restrictions of each. Section II presents initial cost estimation results using data generously provided by the Board of Health of the State of Indiana. The last section is devoted to a consideration of the use of such results for policy decision-making-noting what information is provided by such results, what limitations regression estimation results possess in general, and what alternative methods may be used to provide the details necessary for policy planning decisions.

I

The following are three of many possible alternative functional forms which may be used in characterizing cost and production data:

i) General Exponential (GE)

$$C = f_1(x_1 \ldots x_n) = e^{\alpha_0} \prod_{i=1}^{n} x_i^{\alpha_i}$$

$$x_i > 0$$

ii) Generalized Quadratic (GQ)

$$C = f_2(x_1, \ldots, x_n) = A \frac{\left(\prod_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n} x_i^4} \sum_{i=1}^{n} \frac{\alpha_i}{x_i}$$

$$x_i > 0,\ A > 0$$

iii) Constant Cost Elasticity of Substitution (CCES)

$$C = f_3(x_1 \ldots, x_n) = A\left[\sum_{i=1}^{n} \alpha_i x_i^{-B}\right]^{-1/B}$$

$$x_i \geq 0,\ A > 0,\ B > -1,\ \sum_{i=1}^{n} \alpha_i = 1$$

$$\alpha_i > 0 \quad (i=1, \ldots, n)$$

Where C is some cost measure, the x_i are cost factor measures, e is the exponential constant, and α_i, B, and A are the parameters of the varying functions[4].

Each of these functional forms has definite "economic" implications which must be considered in using them to represent a particular production process. Both the general exponential, (GE) and generalized quadratic (GQ) form require all cost factors(x_i's) considered to be at positive levels for cost to be non-zero. The GE form implies increasing (decreasing) returns to scale if $\sum_{i=1}^{n} \alpha_i$ is less than (greater than) one. As its

name implies, the CCES forms implies constant cost elasticity of substitution of the x_i's across the entire output range in addition to constant returns to scale.[5] Each of the three functional forms outlined here allows cost factor interactions and substitutability. None rule out the inclusions of cost factors having a negative relationship to costs, which may prove useful in particular applications. For example, such a cost factor in waste treatment plants might be BOD in the effluent. For a given influent flow and BOD concentration , the lower the BOD in the effluent, the higher the total cost.

Using differing time period data, tests have been developed for considering such questions as shifts in technology, productivity, etc. For reasons relating to theoretical limitations on regression results detailed in section III of this paper, we present only estimation results here.

While its flexibility is somewhat limited in comparison to the other two, the GE class is the easiest of the three to deal with for estimation purposes. It is easily transformed for estimation purposes to log-linear form, and direct ordinary least squares can be applied. Functions in the other two classes may be estimated using non-linear least squares or one of several step-wise procedures, but these techniques require details and explication beyond the scope of this paper. For this reason, we concentrate detailed attention on general exponential cost functions, presenting some sample regression results and considering their implications.

II

The data provided by the Board of Health of the State of Indiana gives cost and operating data for activated sludge plants throughout Indiana for years varying from 1965-1969.[6] The dependent variables employed in this study were payroll, electricity and water, and total expenditures (payroll plus electricity and water plus "other expenses").[7] The operating data available provided information on million gallons treated, influent BOD, and effluent BOD. The cost data was deflated to 1965 dollars using the Production Workers (Manufacturing Industries-Indiana) index on payroll cost, the Electricity Power Index on the electricity and water costs, and the Wholesale Price Index on "other" costs.[8] The initial formulation was to consider the cost variables as functions of flow (million gallons treated), treatment levels, and per cent of design capacity at which each plant was operated. Treatment level variables can be introduced in many ways. Per cent removal (BOD) may be used, or influent BOD and effluent BOD may both be entered, or some composite work load measure may be formulated. In attempting to formulate such a composite work measure, several engineering facts were kept in mind. The difficulty of treatment for a given flow and influent quality increases at an increasing rate as the effluent BOD content becomes lwoer and lower, Using this, the variable employed here to measure treatment work load took the form: $\frac{BOD_I - BOD_E}{(BOD_E)^2}$

Where BOD_I is average mg/ℓ influent BOD and BOD_E is average mg/ℓ effluent BOD [9].

Thus, the cost functions considered are in the general form:

$$C = f (MG, TL, PC) \qquad (1)$$

Where MG is million gallons treated, TL is treatment load, and PC is percent of design capacity utilized by the plant. A priori, we expect million gallons treated to be the most significant variable in determining payroll (P), electricity and water (E&W), and total costs (TC). Per cent capacity should enter as a negative interaction influence on cost - i.e., we would expect higher cost for 40 MG treated in a plant operated at 50 per cent of design capacity than in a plant operated at 95 per cent design capacity (assuming the same treatment levels).[10] The treatment load should have a positive influence on cost, but we expect this influence to be greater on electricity and water than on payroll costs. Treatment load effects on total cost a priori are questionable since they may be overwhelmed by the other two factors considered.

Table 1 gives the results of the regression analysis using general exponential cost function forms. The sample size was 13 covering plants with all necessary data available. The R^2's reported here are somewhat higher than those of previous studies probably due to the entrance of additional variables and interactions, and all the signs of the coefficients follow a priori expectations.

Table # 1

$C_p = e^{6.227} \; MG^{.592} \; TL^{.173} \; PC^{-.254}$

$R = .954 \quad R^2 = .911$

			T - RATIO
STANDARD ERROR OF	α_0	= .539	11.557
	α_1	= .070	8.445
	α_2	= .098	1.772
	α_3	= .252	-1.006

$C_{E\&W} = e^{7.235} \; MG^{.254} \; TL^{.228} \; PC^{-1.88}$

$R = .876 \quad R^2 = .769$

			T - RATIO
STANDARD ERROR OF	α_0	= .974	7.425
	α_1	= .127	2.006
	α_2	= .177	1.288
	α_3	= .456	-4.125

$C_{Total} = e^{7.262} \; MG^{.513} \; TL^{.151} \; PC^{-.567}$

$R = .924 \quad R^2 = .854$

			T - RATIO
STANDARD ERROR OF	α_0	= .664	10.936
	α_1	= .086	5.946
	α_2	= .121	1.256
	α_3	= .311	-1.826

Estimation Form

$$\ell n\, C_j = \alpha_0 + \alpha_1 \ell n\, MG + \alpha_2 \ell n\, TL + \alpha_3 \ell n\, PC.$$

The R values ranges from .876 to .954 indicating good fit for all three cost equations. Each estimated equation, in addition, indicates increasing returns to scale in the historical data - i.e. decreasing per unit treatment cost as "scale" is increased. The most significant variable influencing cost appears to be million gallons treated. Additional variations on the reported regressions indicate the importance of interactions and inclusion of all three variables.

III

The regression results presented in Section II gave reasonably good fits of the historical operating cost data to the general exponential form. This section is devoted to possibilities for and prohibitions on the use of such regression results.

First, regression results do provide a characterization of historical data which may be useful as an initial screening device for proposed new waste treatment facilities or as a means for comparing the costs of existing facilities to some "norm". For instance, suppose that a new activated sludge plant is proposed where estimated construction and operating cost fall a great deal above the amount the regression results would predict for the specified flow and treatment loads. Unless the additional expenditures can be directly justified (e.g., additional treatment necessary for particular influent, new and costly legal requirements, etc.), the proposed expenditure should be prohibited.

Similarly, regression results may be used as benchmarks for comparing current operating costs of waste treatment plants of the same type. If a plant's expenses are high above the benchmark, regulatory agencies may seek justification from those in control of the plant for the additional costs - e.e., why the costs, are they necessary, are there any off-setting benefits?

Secondly, regression results may provide some rough basis for comparisons of variations over time. Tests of varying hypotheses may be formulated within the regression framework considering such questions as productivity changes and indications of technological change. Results of tests of this nature need to carefully considered, however, keeping in mind the rough basis for their formulation.

The use of regression is limited to a far greater degree than indicated in previous papers-especially those dealing directly with waste treatment. For this reason it is also useful to point out the weaknesses of regression analysis. There are two major areas where regression is <u>not</u> appropriate:

1) Planning future facilities.

2) Finding efficient rules for operating existing facilities.

Shah and Reed [10], for example, suggest that these regression results could be used for future planning. However, there are several basic problems with this approach. As noted above, regression yields a historical description of existing plants. The question, however, is whether we wish to construct similar plants or something considerably

different. It is most likely the latter for the following reasons:

1) relative prices of inputs have changed requiring a different mix of inputs for producing a particular level of clean effluent at least-cost;

2) technological changes which can substantially reduce cost have been introduced;

3) existing plants are likely to be an inefficient mix of technologies embodied in a series of additions or alterations which were made in response to earlier price and technology changes or to quality or quanitity adjustments in the input or output;

4) existing plants are not likely to be cost minimizers, since they are not operated for profit.

The regression techniques cannot handle the planning problem. No matter how good a fit, the regressions provide no more than a summary of historical data. This is sufficient for cost estimates for future plants only if we want to continue what we are currently doing - efficient or inefficient.

An appropriate model for future planning would be based on technical relationships (engineering, biological and chemical) that allow all possible production alternatives to be searched. The fitting of cost functions may provide guide lines as indicated above, but it is not a substitute for good models of the underlying phenomenon combined with an algorithm for searching for efficient solutions.

The second area where regression analysis is not appropriate is in the development of rules for operating existing plants more efficiently. If the operation and maintenance costs of a particular plant are found to be "out of line" using the size of the regression residual as the criteria, there is no way to tell from the regression equation how the cost might be reduced. For example the information that the cost varies directly with MGD or kilowatts of electricity does not help a

treatment plant supervisor reduce the operation costs of his plant.[11]

In summary, the fitting of cost and production data may provide helpful benchmarks and good historical summaries of cost or production data. However, for the purposes of planning or efficient operation of existing plants there is no replacement for good modeling of the underlying system.

The danger lies in the extension of the regression technique beyond its theoretical limitations. One reason for the common inappropriate extension of this technique is the relatively easy and inexpensive calculations involved as compared to detailed system modeling and analysis. Short-run savings, however, may be overwhelmed by long-run losses due to inefficient and outdated plants which may be constructed and justified by referral to historical data summaries.

An alternative approach to the development of cost and production functions was discussed in a paper by the authors [7]. The technique makes use of the underlying mass balance conditions and rate equations which are not considered in the usual approach of economists as discussed above.

Finally, it is of interest to note that although the literature concerning the economics of waste treatment facilities focuses on the estimation of cost functions, the estimations of these functions assumes some underlying productions functions or relations between inputs and the level of output. The production functions are independent of the relative or absolute price levels of the inputs. Any implication with respect to planning or operations rules for existing plants will have to come from a model of these underlying relationships. So

far the literature on waste treatment plants has very few examples of attempts to integrate existing technical sub models of particular processes into a productions function appropriate for these questions, an exception being the work by Fan and Erickson [4].

Footnotes

1. See [8], [9], [10] and [11]

2. See [3] and [6]

3. In addition to a large series of articles, at least one complete book has been devoted to the subject, see Johston [5]

4. The term "cost factors" is used here instead of "inputs". The former seems more appropriate since factors directly affecting cost may not be inputs in the production process sense.

5. Full discussion of these concepts is not presented here due to the nature of this paper. Interested readers may refer to [1],[3], and [5] for explicit definitions and lengthier discussions.

6. We wish to thank Steve Kim at the Board of Health of the State of Indiana for his help and advice in collection of the data. Individual plants for the sample were selected on the advice of Dr. J.E. Etzel of the Department of Civil Engineering, Purdue University, based on reliability and completeness of data from each plant.

7. "Other Expenses" as used here is a conglomerate of miscellaneous expenses, but does not include maintenance and bond costs. There appeared to be varying distinctions across plants between maintenance and capital costs, and, on the advice of those involved, we chose to consider the costs as indicated.

8. The sources for these indices were the Statistical Abstract of the United States, 1965-1969, especially 1967, tables 337, 499; 1968 tables 334; and 1970, tables 344 and 519. The base year for all indices was changed to 1965.

9. "Work load" is not used here in any usual technical engineering sense, but merely as an approximate measure of difficulty. We employed this measure after several discussion and interviews implied that influent quality did not vary extremely between identical type plants within the same general area. If variations in influent quality are extreme this measure should be reconsidered and possibly another should be employed. The regression results below seem to indicate some "explanatory" usefulness for the measure employed, but they can in no way justify it as the one, exact, proper measure,

10. This argument is presented in reference to the usual observed variations in per cent capacity utilized, say 50% -130% of design capacity. Extremes at either end are disastrous to functioning of the system - i.e. either the "bugs" die from lack of nourishmentor the system is "washed out". Hence, we present the argument here with this in mind, accordingly limiting the implications.

11. Though examples in this paper deal with estimation of cost equations, regression estimation of production functions fails to provide any sound basis for optimal planning decisions or operation rules.

References

1. Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," Review of Economics and Statistics, Vol. 43, No. 3, August 1961.

2. Cobb, C. W., and P.H. Douglas, "A Theory of Production," American Economic Review, Vol. 17, No. 1, Supplement, March 1928.

3. Dhrymes, Phoebus and Modecai Kurz, "Technology and Scale in Electricity Generation", Econometrica, Vo. 32, No. 3, July 1963.

4. Erickson, Larry E. and Liang-tseng Fan, "Optimization of the Hydraulic Regime of Activated Sludge Systems," Journal Water Pollution Control Federation, Vo. 40, No. 3, Part 1 March 1968.

5. Johnston, J., Statistical Cost Analysis, McGraw Gill Book Company, 1960.

6. Johnston, J., "Statistical Cost Functions in Electricity Supply", Oxford Economic Papers, N.S., Vol 4, 1951.

7. Marsden, J. R., D. E. Pingry and A. Whinston, "Production Function Theory and the Optimal Design of Waste Treatment Facilities," Applied Economics, Forthcoming.

8. Michel, Robert L., "Costs and Manpower for Municipal Wastewater Treatment Plant Operation and Maintenance 1965-68," Journal Water Pollution Control Federation, Vol. 42, No. 11, November 1970.

9. Rasmussen, Jon and Mark A. Pisano "Reassessment of Operation and Maintenance Cost Functions for Waste Facilities," Unpublished Manuscript, January 1971.

10. Shah, Kanti L. and George W. Reid, "Techniques for Estimating Construction Costs of Waste Treatment Plants", Journal Water Pollution Central Federation Vol. 42, No. 5, Part 1, May 1970.

11. Smith, Robert, " Cost of Conventional Advanced Treatment of Wastewater", Journal Water Pollution Control Federation, Vol. 40, No. 9, September 1968.

12. U.S. Bureau of Census, Statistical Abstract of United States, Washington, D.C.

Chap. 2 Environmental Data Management - The Identification of Outliers

Introduction

The last few years have evidenced large scale increases in the emphasis placed on water quality standards. The Federal Water Pollution Control Act and its 1972 amendments have served to bring new and strict controls into play and to increase the burden on state agencies for developing comprehensive regional plans.[1] As the tasks increase, the agencies are turning to high speed data handling and storage procedures and to sophisticated modeling and advanced quantitative techniques for analyzing the problems (e.g., the STORET program now being used in state agencies such as the Indiana Stream Pollution Control Board). At the same time, however, detailed consideration of the accuracy and reliability of the data itself has been ignored. Though little has so far been accomplished, the Environmental Protection Agency itself is currently pursuing these topic areas as evidenced by the following excerpts from the March 15, 1974 E.P.A. Water Quality Strategy Paper, "A Statement of Policy for Implementing the Requirements of the 1972 Federal Water Pollution Control Act Amendments":

> E.P.A. in conjunction with the States will develop a system for conducting compliance monitoring. Elements include:
> a. Integration of self-monitoring report analysis, and facility inspections. Self-monitoring reports will involve validation of data, followed initially by manual screening and subsequently computer screening of violations . . .
>
> . . .
>
> A major task will be to introduce procedures for quality control over data input, with the Regions conducting spot-check screening for this. (p. 38)

This paper is centered on problems associated with checking the accuracy of a large data set and designing appropriate editing procedures. As such, it is aimed at current problems such as these outlined by E.P.A., but, further, it is hoped that it will spur interest in a too-long neglected

problem area. Necessary caveats in the employment of highly aggregated data are discussed in the context of usefully employing large scale modeling techniques with particular attention being focused on implications for sensitivity analysis.

I -- The Data --Some Background

In a recent paper, [3], we made use of municipal treatment plant operating and cost data obtained from the State of Indiana in considering the use of regression results relating to treatment cost functions. As explained in that paper, the sample data was "pre-screened" for accuracy and reliability with the help of individuals with both the technical skill and extensive familiarity with individual plants and operators. This procedure, while subjective and far from perfect, still provided a means to employ more accurate and reliable data than earlier studies. Previous authors, [5], [7], [8], used entire data sets gathered from survey or mail questionnaire procedures ignoring consideration of the accuracy and reliability of the data. As long as sample sizes were "adequate" (e.g., see p. 776-779 [8], p. 1883-1884 [5]) the quantitative techniques were applied and the results analyzed in standard fashion.

One fundamental difficulty with data used in analyses of treatment plants and water quality conditions is that the data is supplied independently by those being "monitored". Monthly reports are submitted in standard forms, but each report is supplied by a different "first-level" interpreter of what is requested and of the appropriate data and required sampling results on pollutants. Each reporter is different-- some well trained, some poorly trained. Each records and reports using differing methods which may be at wide variance, indeed.

But what can be done? Do we create traveling monitoring committees? Do we require expensive monitoring equipment? The expenses rapidly progress above feasible limits, and, hence, we center our attention here on suggestions concerning relatively cheap and rapid data analysis to help identify "outliers" for which more concentrated, concerted effort can be employed in analyzing their specific conditions and data collection procedures. We use "outlier" here in the sense of a plant submitting data which, in one or more sections, diverges sharply from what is "expected". In the next section, we take the abstract terms "outlier", "diverges sharply", "expected", etc., and interpret them in specific terms by suggesting the employment of quantitative techniques such as discriminant analysis in quite novel ways.

II

We begin by suggesting the compilation of a "training sample" of accurate, reliable data relating to all aspects of the operation of varying size and type treatment plants. This initial phase would require sampling technique knowledge in conjunction with expertise in the physical operations of the plant and familiarity with the operators. The information gathering could be an expensive ingredient in the small program, but without this accurate beginning, advanced analysis would not progress beyond where we are now. The sample gathering task, as we propose, is thus removed from the usual channels and placed as a joint responsibility of engineers, analysts and economists, and state employees familiar with individual plants and operators. Thus, the initial "sampling" procedure is a concentrated effort on plants which are designated "most reliable" by those directly involved in research and regulation.

What we suggest, then, may be looked at as removing a certain subsection of the population for detailed sampling and data collection concentration. The expense involved is set against benefits derived from the employment of accurate, detailed data and the avenues it opens as discussed below.

Once the initial sampling task is properly completed, the researchers are able to employ the "training sample" as grounds for analyzing the continuing inflow of data from each treatment plant--including the required monthly reports. We now consider some of these possibilities making use of general quantitative techniques such as multiple discriminant analysis, linear and non-linear programming.

Setting Benchmarks--The Identification of Outliers

In a recent paper [4], we discussed the possibility of using discriminant analysis as a means to identify "outliers" to be studied further. We now detail that initial suggestion and present a basic format making use of the "training data".

The procedure in discriminant analysis provides an estimation of an individual unit's position, using its values on m characteristics (or "test scores"), on a line that "best" separates classes. One "best" line may not use up all possible information available from the characteristic measures and, hence, additional, mutually orthogonal discriminant functions may be calculated.

Three basic underlying assumptions are generally employed:

1) the groups under analysis are identifiable and discrete

2) each observation (sample) in each group can be described by values or measurements on m characteristics

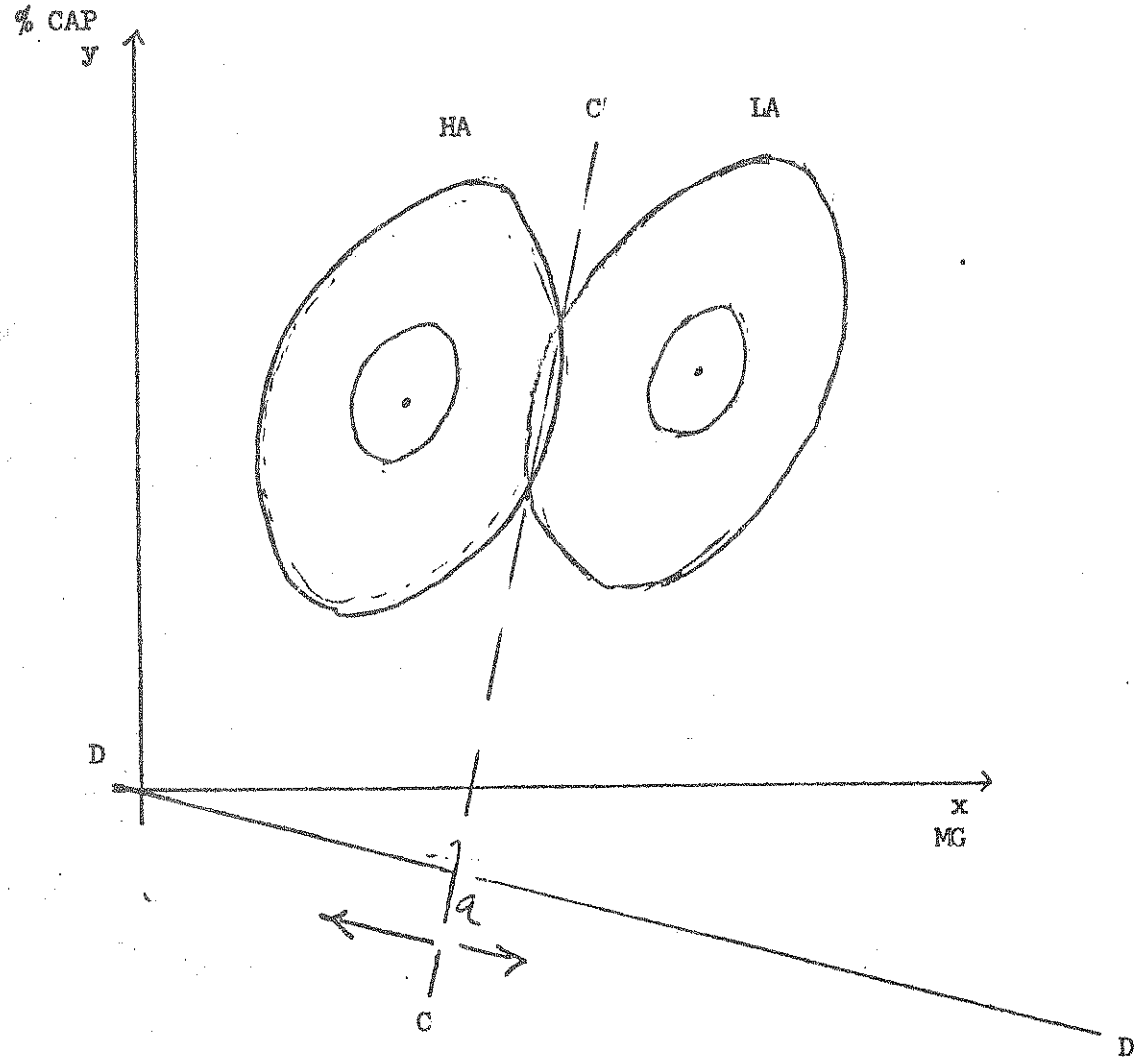
3) the m characteristic variables are assumed to possess a multivariate normal distribution in each population under study.[2]

A two group, two variable case should help clarify the technique and advance a straightforward geometric interpretation. Consider figure I. Values on the x and y axes represent measures on two characteristics, (e.g. million gallons treated and % capacity utilized), for the groups being studied, say, for groups LA (low average cost) and HA (high average cost). The ellipses represent equal density for the groups - i.e. the inner ellipse for each group might be a 50 centile contour within which 50 per cent of the group's members lie while the outer ellipse might be a 90 centile c Using the two points at which corresponding centile contours intersect, we can construct a line such as CC and then a second line DD which is perpendicular to CC. If the points in two dimensional space are projected onto DD, the two group overlap will be smaller than for any other possible line.

Thus, what the discriminant function does is to transform the measures on multiple individual characteristics into one score (i.e., a location along DD). That score (or location along DD) indicates probable membership in one group or another. In figure I, for instance, the point a where CC intersects DD, divides the line into two segments, one representing probable membership in group LA and the other probable membership in group HA.

The basic framework employs two matrices, say G and W, which serve to summarize the data measures on the m characteristics in terms of products of deviations from specific means or "average values." The G matrix summarizes variation between groups (e.g. average cost groups) while the W matrix summarizes variation within groups. Thus, if the groups are separable using the m characteristics, we would expect the between group variations

FIGURE 1



to be large relative to within group variations. Discriminant analysis uses this basic approach in attempting to maximize the ratio of between group to within group variations, in addition employing the sums of squares result that $v_i'Gv_i$ and $v_i'Wv_i$, where v_i is a vector, are both scalars. (See [1] [10] for detailed specifics.) Thus, the basic approach is to use the characteristics to group the individuals in such a way as to maximize the ratio of between group variations (sums of squares) to within group variations (sums of squares) or, more formally:

$$\underset{v_i}{MAX} \quad \frac{v_i'Gv_i}{v_i'Wv_i} \qquad i = 1,\ldots,K$$

where K is the lesser of the number of groups minus one and the number of characteristics employed.

In order to obtain a unique solution, the normalization constraint that $v_i'Wv_i = 1$ is added and the Lagrangian problem takes the form:

$$\underset{v_i}{MAX} \quad L_i = v_iGv_i - \lambda(v_i'Wv_i - 1)$$

which after taking the partial derivatives of L with respect to v_i and λ_i, yields the optimality conditions:

$$(A - \lambda_i W)v_i = 0 \qquad (1)$$

$$v'_iWv_i - 1 = 0 \qquad (2)$$

Equation (1) is usually written as:

$$(W^{-1}A - \lambda_i I)v_i = 0 \qquad (1a)$$

Equation (1a), having a non-trivial solution, (i.e., $v_i \neq 0$), is equivalent to the following holding:

$$| W^{-1}A - \lambda_i I | = 0 \qquad (1b)$$

Thus, the solution of the maximization problem reduces to solving for eigenvalues (λ_i) and corresponding eigenvectors (v_i). We could solve the above problem format K times ($i = 1,\ldots,K$), K being the lower bound of the number of groups less one and the number of characteristics employed for the

K eigenvalues and eigenvectors, but often the additional incremental values of the λ_i rapidly became small. A rough measure of how far to go, or "the percentage of total discriminating power of the characteristics contained in the i^{th} discriminant function" is represented by:

$$100\left(\frac{\lambda_i}{\sum_{i=1}^{K}\lambda_i}\right)$$

The actual classification procedure usually employs either a χ^2 or maximum probability method to fix the one-dimensional locations and provide class or group assignments for each individual. Though both of these classification procedures employ assumptions about the distributions of the variables involved, (see below), our initial work in this paper avoids significance tests and thus, basic difficulties associated with these assumptions.

Employing the training data and discriminant analysis, yields a discriminant function for use in classifying the remaining plants. For instance, in [4], the objective was classification of waste treatment plants into "average cost level" groups using variables including million gallons treated (MGT), per cent design capacity utilized (% Cap), per cent BOD removal (%BOD), per cent suspended solids removal (% SS), work load BOD (WLBOD), and work load suspended solids (WLSS).[3] In the initial results reported, the "hit-ratio" was relatively high (78%-93%) for the "reliable" sample, though training sample size was limited and results were reported only as "initial indicators". Also, in [4], the group classifications considered were by average cost since a main thrust of the paper was the consideration of the economies of scale and benefits to regionalization questions. It probably is of interest, however, to consider groupings according to most of the treatment variables--cost, treatment level, efficiency,....

The discriminant function obtained using the "training data" provides a discriminant function which can be used to identify and segment "outliers". Suppose we are considering average cost groupings of waste treatment plants and are using three groups--low, medium, and high average cost per million gallons treated. A discriminant function value may then be calculated for each plant in the population and the plant "assigned" to a group. These assignments can then be compared to the actual cost group for each plant and a hit-miss table set up. Mis-classified plants would then be candidates for further checks by control boards. Of special interest would be plants mis-classified by two groups--e.g. actual high average-cost plants classified as low average cost and vice-versa. Plants of these types might be labeled "outliers" as indicated earlier, and control agencies might seek to discover why the plants deviate from general classification schemes constructed using the sample data. Study of these outliers should prove helpful in several ways, including:

i) correcting misunderstandings and misinterpretations of data forms and thus providing improved and more accurate data for future research.

ii) correcting source reasons for the plants which have low average cost characteristics but which operate at high average cost.

iii) providing helpful operating hints from actual low average cost plants which have characteristics similar to high average cost plants in the "training sample" and are classified accordingly.

iv) a means for checking the consistency of data submitted and possibly discovering sources of intentional inaccuracies.

We consider each of these in order.

Each operator or data collector for each individual plant interprets the data requests somewhat differently. When an interpretation of this nature is at wide variance from that anticipated and designed for, use of the data submitted will continually lead to possible bias.

The monthly report forms themselves may lead to many misunderstandings which might easily be cleared up. These would include such things as incorrect measures (e.g. weight instead of parts per million), varying definitions of cost calculations, or varying means for sampling water pollutant content. These may further be compounded by lack of information or training on the part of the individuals involved. Identifying where these outliers occur give those responsible an opportunity to clarify the forms, other misunderstandings, or aid the data collector in his sample gathering tasks.

Plants which have low average cost characteristics, but which actually operate at a high average cost rate would also be singled out. Control authorities would then be able to study these plants further in an attempt to correct, if possible, operating procedures which might lead to the high cost. Further, it is possible that special conditions such as geographical or differing pollutant influent, are the cause of additional costs. If so, this additional information would be an aid to control boards in cost estimates and regulatory functions.

Low average cost plants with high average cost characteristics would also be marked for further study. They might provide operating procedures which would help decrease costs at other plants. In addition, they may possess system modifications which might be introduced at other plants to lower costs.

Not actually separate from the above illustrations is the checking of outliers for possible intentional incorrect data reporting. Detailed analysis of the outliers will serve as a means for checking data irregularities as outlined above and, in addition, as a means for detecting data intentionally misreported or "faked". In several conversations with control board employees and others familiar with the plants and operators, the problem of data reports filed without the completion of required sampling kept coming up. One plant's monthly report invariably listed per cent BOD removal at "94%"--influent quality varied, flow varied, weather conditions varied, chemicals used varied, other expenses varied, but per cent BOD removal always remained at 94%.

The studying of outliers should help to bring problems of this type to light, but it will not necessarily identify all data misrepresentations. Faked or false data which is consistent with the training sample observations will be correctly classified and hence the plant not identified as an outlier. As conditions change, selecting faking data consistent with the classification techniques used should become more difficult, but there is no guarantee that this will imply identification of an outlier.

The techniques sketched above can, however, be expanded in an attempt to deal with problems of this types. Up to this point and in the following example, we have concentrated on average cost, but there is no requirement that this is as far as we go. The same techniques can be applied classifying plants into groups by any of the pertinent variables the EPA might specify--treatment load (many possible types), per cent BOD or SS removal, levels of other pollutants in effluent discharge, sludge production, etc. Once the training sample data is properly gathered and useful

discriminant functions computed, application to further plants and identification of outliers with respect to any of several variables can be handled easily and rapidly, providing further means to accomplish one of the four tasks mentioned earlier.

To help outline and clarify the procedure set out above, we present the results of an example using a training sample of data from waste treatment plants selected as indicated above with much expert help and advice.[4] The full data set was generated using the basic data set from [4] and supplementing it with some hypothetical plant operations as suggested by several of our colleagues involved directly in the analysis and regulation of treatment plants in the State of Indiana. The results reported are not presented here as bases for possible policy recommendations. They are meant for descriptive and pedagogical purposes only - to aid in the presentation of the overall approach suggested above.

Two computer programs were employed - one for deflating the cost data to base year 1967 and calculating the necessary variables from the basic raw data, and the other for computing the discriminant function using the training sample and for performing the classifications on both the training sample and the entire data set. The first program was comprised of straightforward arithmetic operations and was written specifically for the present application. The discriminant program employed to perform the second stage above was the Cooley and Lohnes Multiple Discriminant Analysis program. The variables used in developing the discriminant function were: 1) work load, suspended solids (average) (WLSS) 2) work load, BOD (average) (WLBOD) 3) million gallons treated (annual) (MGT) 4) percent capacity utilized (annual) (%C) 5) percent suspended solids removed (average) (%SS) and 6) percent BOD removal (average) (%BOD).

The last four are self explanatory while the first two are of a type introduced in [3]. WLSS was measured by $\frac{SS_I - SS_E}{(SS_E)^2}$ and WLBOD by $\frac{BOD_I - BOD_E}{(BOD_E)^2}$ where:

SS_I - influent suspended solids (ppm)

SS_E - effluent suspended solids (ppm)

BOD_I - influent BOD (ppm)

BOD_E - effluent BOD (ppm)

These measures are employed to take into account that the difficulty of treatment for a given flow and influent quality increases at an increasing rate as effluent suspended solids and BOD reach lower and lower levels. The "work load" concept is not used here in a technical engineering sense, but as an approximate measure of difficulty[5] (see [3.]).

Average cost per million gallon treated was calculated using a total cost figure which included the sum of payroll, electricity and water, and other expenses, but not including maintenance and loan charges.[6]

The programs were run first using two average cost groups (less than $50 per MG treated and more than $50 per MG treated) and then using three cost groups ($0-$30, $30-52, and $52 and up per MG treated). Table I gives the results for the two cost-class case while Table II presents those for the three cost-class case. All runs were made on the CDC 6500 system at Purdue University.

TABLE I

a) Training Sample

Number of Cases Classified Into Group

Actual Group	LOW	HIGH
LOW	13	0
HIGH	2	3

correct classifications: 88.9%

b) Entire "Sample"

	LOW	HIGH
LOW	62	9
HIGH	26	18

correct classification: 69.6%

TABLE II

a) Training SAmple

Number of Cases Classified Into Group

Actual Group	LOW	MEDIUM	HIGH
LOW	11	0	0
MEDIUM	3	1	0
HIGH	1	0	2

correct classifications : 77.8%

b) Entire "Sample"

	LOW	MEDIUM	HIGH
LOW	31	6	2
MEDIUM	18	12	10
HIGH	4	20	12

correct classifications : 47.8 %

Using Tables I-b or II-b, "outliers" are easily identified for further study. The off-diagonal elements in these tables represent plants which were mis-classified into average cost groups. For instance, in Table I-b, the nine actually low average cost plants exhibiting high average cost characteristics (in terms of the variables employed here) might be studied for techniques which enable such relative efficiency or for possible misunderstandings or mistakes in data reporting. Likewise, for the twenty-six plants exhibiting low cost characteristics but actually belonging to the high average cost group. Further investigation might yield helpful expert advice aiding cost efficiency and, on the other hand, might serve to clear up data irrefularities. The same type of use can be made of the information presented in Table II-b. Here, however, we might wish to

center expensive individual attention on the upper righthand and lower lefthand "corner elements" - i.e., mis-classification by two groups or plants that "diverge sharply" from what is indicated by their characteristics. The procedures would be analogous to the two class case - identify and investigate the outliers to increase the accuracy and hence usefulness of the data gathered.

In pursuing the above example we might consider classifications using groupings other than those based on average cost per MG treated. Several such have been suggested by Dr. James Etzel of the Department of Civil Engineering at Purdue University - each of which might enable sharper and sharper delineations of outliers. Further, we might pursue the problem using an increased number of classes or variations on class boundaries. Here, however, our purpose is exposition and, for brevity, we present the above example in its short form. We present what appears to us as the basics: the foundations for what we term "outlier identification". Each particular problem will need separate detailed analysis and various appropriate modifications. The above can serve as a base for these, a base that allows rapid and relatively cheap identification of "data inconsistencies".

Summary

In the above we have centered attention on an approach for identifying data inconsistencies and identifying outliers. The actual problem in the collecting of accurate and useful data, however, is much more fundamental for it lies in the current structure of the collection and data management system itself. For instance, the first level data interpreter or data collector at the individual treatment plant has little or no direct

positive incentive to check the accuracy of the data he submits or passes on to the next level. He is required to submit a monthly report to the state and from all indications, it is just assumed that he does so accurately.

At the next level the state agencies enter to act in a data gathering capacity for individual plant data. Here again the emphasis and positive incentive is on the gathering and storage of complete data sets rather than on validity checks of the accuracy of the data.

At the federal level, the EPA data gathering moves one more step up the aggregate ladder and one step further away from the possibility of verifying the data observations of the first level reporters. Researchers involved then often use highly aggregated data but without any means of validating it.

At each step, the emphasis is on collection, organization, and labeling rather than verification. Current final use of the data does little to press for changes since the end intention has been use as bases for policy decisions. There is a limited amount of time and the pressing need for an "answer". The data is taken as given and policy implications are discerned. Caveats are added in standard form about possible data deficiencies, but it seems to stop there.

If the policy maker is challenged, he may shrug it off and refer to the researchers' efforts. The researchers in turn may present their techniques and back up their recommendations by the "this is what the data given us indicates" route. The EPA shrugs "all we do is collect the data from the states", while the states point to the plants and the plants point to the monthly reports that they have diligently filed and note the completion of their assigned task.

Through all of this, analysis of the validity and accuracy of the data is lacking. There is no strong incentives for accurate reporting nor for checking the data received from the previous level. What we are suggesting is a "tailor-made", relatively cheap and rapid processing procedure for identifying outliers so that expensive in-depth study may be concentrated on analyzing the reasons why problems exist and what can be done to correct them. The product of the in-depth study will not be limited to data analysis, but will include at the same time an analysis of the plant operations by experts working to bring the outlier plant up to its highest obtainable efficiency level or to study the high efficiency plants for helpful procedures and ideas to pass on to other plant operators. Through this concentrated use of expert ability the procedure provides the means to both upgrade the data collection and accuracy and to study and spread efficient operation techniques. Emphasis is on why problems exist and "where from here". Is there a need for additional system training, for better information flow procedures, for more direct contact? Are data outliers occurring due to operational variations or data irregularities and mis-reporting?

Further, the technique also provides a means for handling one of the difficulties posed by the 1972 Amendments to the Federal Water Pollution Control Act - the determination of what constitutes best practicable treatment. The term best practicable treatment is at best sketchy and poses the possibility of an agency having to make a specific interpretation for each individual industrial and municipal treatment facility. To avoid the overwhelming complexity of this, several administrators have suggested stratifying the treatment facilities by type and size and then using some subset of efficient plants as a basis for determining approxi-

mate measures of current best practicable treatment for each group. But this follows almost directly the process set out above, a process which would help identify outliers to be studied and directed to bring treatment within the bounds of the "best practicable" set.

The approach we have set out in this paper is but one possible one for handling some of the problems incurred in large scale data use. It provides a relatively cheap and easily adaptable means for identifying outliers and formulating benchmarks. It is no panacea, but it may well be a fruitful beginning.

Footnotes

[1] See for instance, the broad spectrum of requirements outlined in the Federal Water Pollution Control Act Amendments of 1972.

[2] Since we do not deal with significance tests here, this assumption is not as troublesome as it might appear.

[3] These variables are detailed below.

[4] Dr. James Etzel, Department of Civil Engineering, Purdue University, and staff members of the Stream Pollution Control Board of the State of Indiana were extremely helpful, especially in these regards.

[5] In expanding the basic approach of this paper, we are currently expanding work load measures taking into account sludge volume indices, mixed liquid solids, etc., and interrelations. As Dr. Etzel has pointed out, the work load measures we employ are somewhat naive, but are quite convenient at this initial stage.

[6] See [3]. Methods of reporting maintenance and loan costs vary widely over plants; though work is now under way by the Indiana Stream Pollution Control Board to correct this.

[7] Though the three average cost class training sample has relatively few entries in the medium and high average cost groups, our pedagogical and expository purpose alleviates the usual problems associated with "significance" tests. We are outlining a technique, rather than presenting an applied result for policy use.

References

1. Cooley, W. W., and Lohnes, P. R., "Multivariate Procedures for the Behavioral Sciences," John Wiley and Sons, New York, (1962).

2. Cooley, W. W., and Lohnes, P. R., "Multivariate Data Analysis," John Wiley and Sons, New York, (1971).

3. Marsden, J. R., Pingry, D., and Whinston, A., "Regression Analysis Applied to the Wastewater Treatment Field," Jour. Water Poll. Control Fed., October, 1973, Vol. 45.

4. Marsden, J. R., Pingry, D. and Whinston, A., "A Consideration of the Large-Scale, Regional Plant Hypothesis Using Classification Techniques," (Mimeo, Purdue Univ., 1974).

5. Michel, R. L., "Costs and Manpower for Municipal Wastewater Treatment Plant Operation and Maintenance 1965-68." Jour. Water Poll. Control Fed., 42. 1883 (1970).

6. Morrison, D. F., "Multivariate Statistical Methods," McGraw-Hill, New York, (1967).

7. Rasmussen, J., and Pisano, M. A., "Reassessment of Operation and Maintenance Cost Functions for Waste Facilities." Unpublished (January 1971).

8. Shah, K. L., and Reid, G. W., "Techniques for Estimating Construction Costs of Waste Treatment Plants." Jour. Water Poll. Control Fed. 42, 776 (1970).

9. Smith, R., "Cost of Conventional Advanced Treatment of Wastewater." Jour. Water Poll. Control Fed., 40, 1516 (1968).

10. Van de Geer, J. P., "Introduction to Multivariate Analysis for the Social Sciences," W. H. Freeman and Company, San Francisco, (1971).

Chap. 3 A Consideration of the Large-Scale, Regional Plant Hypothesis Using Classification Techniques

Water Quality has been the topic of much recent legal and journalistic endeavor, much of which includes reference to possible benefits occurring from regionalization of treatment facilities. Economies of scale and more efficient operation in larger plants are usually cited as reasons for consolidation and centralization in one larger plant. In this paper we center our attention on one technique -- discriminant analysis -- which has so far been overlooked in these initial considerations and which might serve as a quite useful and easily employed "first look" at the relevant questions.

I

Since their inception, much controversy has surrounded the Federal Water Pollution Control Act Amendments of 1972, especially sections 303 (especially subsection e) and 208 which deal with basin and regional planning of desired water quality and treatment procedures. Though individual authors, studies, and administrators have varied on interpreting the intent and range of the law, few contest that accomplishing the intent of the law in any direction will require large scale data presentation and summary analysis for policy purposes. One method to deal with such problems which has largely been ignored in this area is a classification technique commonly referred to as discriminant analysis. Its use holds the potential of rapidly compiled and readily interpretable large scale data analysis especially in dealing with some of the questions brought up by recent laws such as the 1972 Amendments referred to earlier.

Discriminant analysis has often been employed in both predictive and inferential frameworks. In this paper we focus our attention on the predictive uses of discriminant analysis, on the development of classification procedures for use in considering hypotheses dealing with economies

of scale and relative large scale efficiency. What we are interested in analyzing are hypotheses such as one often put forth in support of regionalization, i.e. that large scale (regional) plants possess certain characteristics which provide lower average cost benefits. If this hypothesis holds, then we should be able to classify treatment plants into average cost groups using measures on these characteristics to formulate classification procedures. Cooley and Lohnes [2] present the approach more formally as:

> Classification methods make it possible for the researcher who has data from two or more populations in a common measurement space to demonstrate the validity of the measurements for predicting membership in the populations.

In general, there are three underlying assumptions of discriminant analysis:

1. the groups under analysis are identifiable and discrete
2. each observation (sample) in each group can be described by values or measurements on m characteristics
3. the m characteristic variables are assumed to possess a multivariate normal distribution in each population under study.[1]

The problem studied below employs classes of treatment plants separated by average cost of treatment. The basic hypothesis under review is whether lower average cost plants exhibit the characteristics of larger scale in terms of million gallons (MG treated) and in addition, have higher treatment loads (levels) and operate at higher per cent capacity utilized. One argument often posited is that economies of scale exist and hence, average cost is lower the higher the MG treated given the same approximate treatment levels (load) and percent capacity utilized.

The procedure in discriminant analysis provides an estimation of an individual unit's position, using its values on the m characteristics

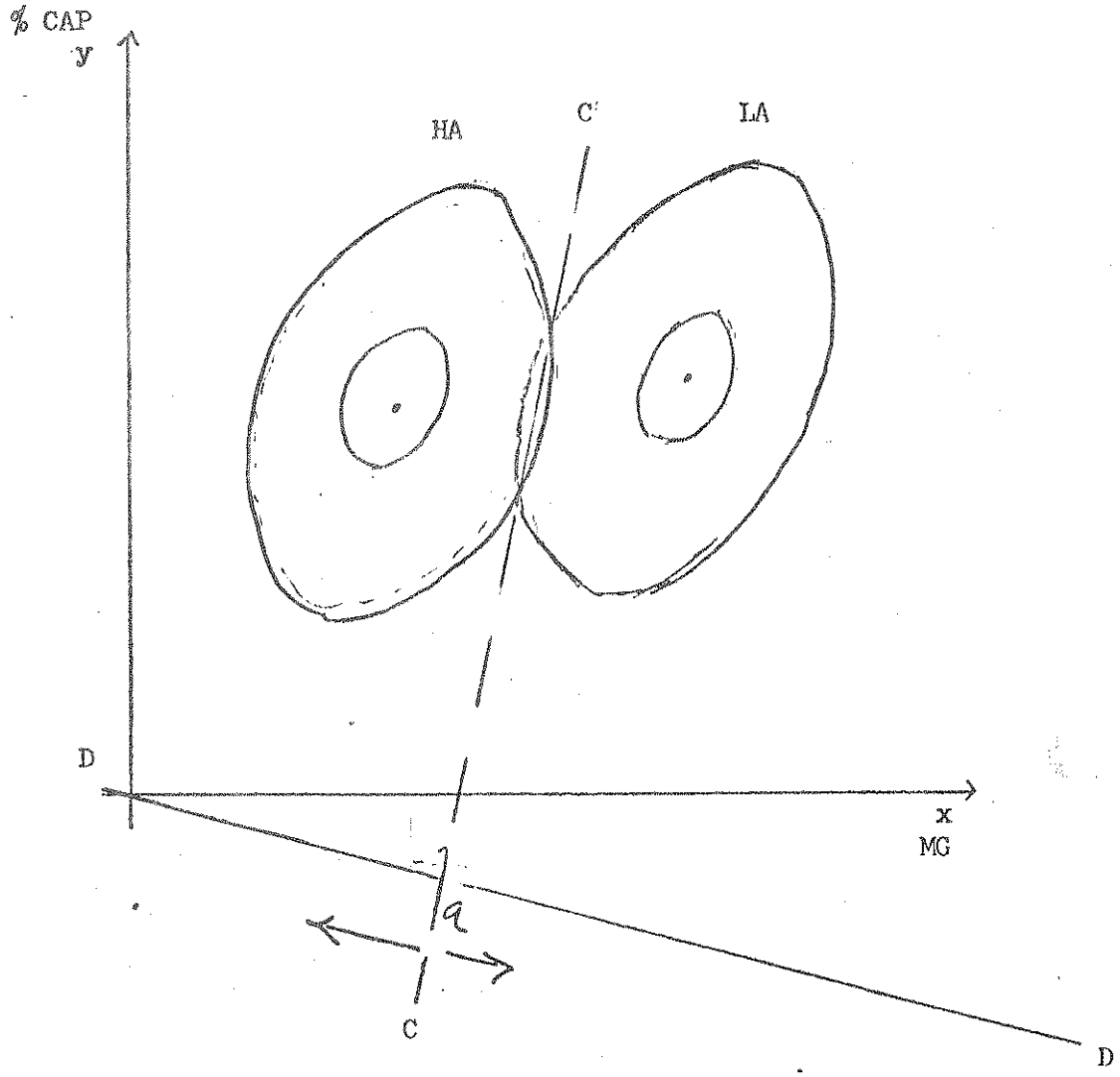
(or "test scores"), on a line that "best" separates classes. One "best" line may not use up all possible information available from the characteristic measures and, hence, additional, mutually orthogonal discriminant functions may be calculated.[2]

A two group, two variable case should help clarify the technique and advance a straight forward geometric interpretation. Consider figure I. Values on the x and y axes represent measures on two characteristics (eg. MG treated and % capacity), for the groups being studied, say groups LA (low average cost) and HA (high average cost). The ellipses represent equal density for the groups - i.e. the inner ellipse for each group might be a 50 centile contour within which 50 per cent of the group's members lie while the outer ellipse might be a 90 centile contour. Using the two points at which corresponding centile contours intersect, we can construct a line such as CC and then a second line DD which is perpendicular to CC. If the points in two dimensional space are projected onto DD, the two group overlap will be smaller than for any other possible line.

Thus what the discriminant function does is to transform the measures on multiple individual characteristics into one score (i.e., a location along DD). That score (or location along DD) indicates probable membership in one group or another. In figure I, for instance, the point a where CC intersects DD, divides the line into two segments, one representing probable membership in group LA and the other probable membership in group HA.

The basic framework employs two matrices, say G and W, which serve to summarize the data measures on the m characteristics in terms of products of deviations from specific means or "average values." The G matrix summarizes variation between groups (e.g. average cost groups)

FIGURE 1



while the W matrix summarizes variation within groups. Thus, if the groups are separable using the m characteristics, we would expect the between group variations to be large relative to within group variations. Discriminant analysis uses this basic approach in attempting to maximize the ratio of between group to within group variations in addition employing the sums of squares result that $v_i'G\,v_i$ and $v_i'\,Wv_i$, where v_i is a vector, are both scalars. (See [1] [5] for detailed specifics.) Thus, the basic approach is to use the characteristics to group the individuals in such a way as to maximize the ratio of between group variations (sums of squares) to within group variations (sums of squares) or, more formally :

$$\underset{v_i}{MAX}\ \frac{v_i'Gv_i}{v_i'Wv_i} \qquad i = 1,\ldots,k$$

where k is the lesser of the number of groups minus one and the number of characteristics employed.

In order to obtain a unique solution, the normalization constraint that $v_i'Wv_i = 1$ is added and the Lagrangian problem takes the form:

$$\underset{v_i}{MAX}\ L_i = v_i'Gv_i - \lambda(v_i'Wv_i - 1)$$

which after taking the partial derivatives of L with respect to v_i and λ_i yields the optimality conditions:

$$(G - \lambda_i W)v_i = 0 \qquad (1)$$

$$v_i'Wv_i - 1 = 0 \qquad (2)$$

Equation (1) is usually written as:

$$(W^{-1}G - \lambda_i I)v_i = 0 \qquad (1a)$$

Equation (1a) having a non-trivial solution, (i.e., $v_i \neq 0$), is equivalent to the following holding:

$$|W^{-1}G - \lambda_i I| = 0 \qquad (1b)$$

Thus, the solution of the maximization problem reduces to solving for eigenvalues (λ_i) and corresponding eigenvectors (v_i). We could solve the above problem format K times ($i = 1,\ldots,K$), K being the lower bound of the number of groups less 1 and the number of characteristics for the K eigenvalues and eigenvectors, but often the additional incremental values of the λ_i rapidly became small. A rough measure of how far to go , or "the percentage of total discriminating power of the characteristics contained in the i^{th} discriminant function" is represented by:

$$100\left(\frac{\lambda_i}{\sum_{i=1}^{K}\lambda_i}\right)$$

The actual classification procedure usually employs either a χ^2 or maximum probability method to fix the one-dimensional locations and provide class or group assignments for each individual. Though both of these classification procedures employ assumptions about the distributions of the variables involved, (see above), our initial work in this paper avoids significance tests and thus basic difficulties associated with these assumptions.

In the example in Section II below, there are six plant characteristics considered: 1) work load suspended solids, 2) work load BOD, 3) per cent removal suspended solids, 4) per cent removal BOD, 5) million gallons treated and 6) per cent capacity utilized. The plants are separated into groups or classes according to their annual average cost per million gallons treated. Two discriminant functions are estimated, one using three average cost groups or classes and the other using two average cost groups. In the first, K is equal to two (i.e., the lesser of the number

of groups less one or 3-1 or 2, and the number of characteristics) for those cases where 2 or more of the six characteristics were used. When only MG treated was used in the three group case and for all two group runs, K was equal to unity.

The above is a short sketch to aid readers unfamiliar with discriminant analysis. It is not in any way meant as a detailed discussion such as may be found in [2], [4], or [5].

The data we employed in the following example was obtained from the Indiana Stream Pollution Control Board to whom we are deeply indebted. The data details influent and effluent quality measures and operating statistics (physical and monetary) for municipal treatment plants throughout the state. The usable sample was reduced to 22 (and to 19 if certain variables are used -- see below) by data gaps and reporting failures. Further, the sample was divided into two parts based on "reliability" judgments after conversations with Dr. J. E. Etzel of the Department of Civil Engineering, Purdue Univers ty. The BIOMED STEPWISE DISCRIMINANT ANALYSIS and the Cooley and Lohnes [1] multiple discriminant analysis programs were employed. The runs were made on a C.D.C. 6500 at Purdue University.

II

The data employed was for activated sludge plants varying in design size from .2 to 60 million gallons per day. Cost and operating data were obtained for 1965 where available. Plants cost data from years other than 1965 was deflated to constant 1965 dollars.

As indicated above the variables considered for use in developing the discriminant function were: 1) work load, suspended solids (average)

(WLSS) 2) work load, BOD (average) (WLBOD) 3) million gallons treated (annual) (MGT) 4) percent capacity utilized (annual) (%C) 5) percent suspended solids removed (average) (%SS) and 6) percent BOD removal (average) (%BOD). The last four are self explanatory while the first two are of a type introduced in [3]. WLSS was measured by $\frac{SS_I - SS_E}{(SS_E)^2}$ and WLBOD by $\frac{BOD_I - BOD_E}{(BOD_E)^2}$ where:

SS_I - influent suspended solids (ppm)

SS_E - effluent suspended solids (ppm)

BOD_I - influent BOD (ppm)

BOD_E - effluent BOD (ppm)

These measures are employed to take into account that the difficulty of treatment for a given flow and influent quality increases at an increasing rate as effluent suspended solids and BOD reach lower and lower levels. The "work load" concept is not used here in a technical engineering sense, but as an approximate measure of difficulty[3] (see [3]).

Average cost per million gallon treated was calculated using a total cost figure which included the sum of payroll, electricity and water, and other expenses, but not including maintenance and loan charges.[4] The programs were run on both the "reliable" and total samples with the plants first grouped into low, medium and high average cost categories and then into only two groups, low and high. When three groups were used, they were divided by average cost according to the following:

low - less than $30 per MG treated

medium - $30 to $50 per MG treated

high - greater than $50 per MG treated

When two groups were used, they were divided using:

low - less than $40 per MG treated

high - $40 or greater per MG treated

The BIOMED STEPWISE DISCRIMINANT ANALYSIS performs a stepwise multiple discriminant analysis at each step of which the statistically most significant (using F value) variable is entered. If a variable's F value falls below a pre-set variable, that variable is deleted. The Cooley and Lohnes program provides both chi-square and maximum probability for classification.

We first considered classification techniques making use of the data which had been indicated as very reliable, initially dividing the data into three cost groups as detailed above. Using the stepwise approach, the first variable to enter was MC treated and the classification matrix was as follows:

TABLE I

Number of Cases Classified into Groups

Actual	Low	Medium	High
Low	2	3	0
Medium	1	1	3
High	1	0	3

When all six variables were employed in the multiple discriminant analysis program, the classification results took the form:

	Number of Cases Classified into Group		
Actual	Low	Medium	High
Low	4	1	0
Medium	0	4	1
High	0	1	3

Variables: MGT, WLSS, WLBOD, % CAP, % SS, % BOD

Now, in this case the hit ratio (11/14) is .786 and no plant is classified "two groups away." There are three misclassifications, one in each group, and, upon rechecking the data, it was found that each was close to a class boundary (about $3 away in the two lower cases and $6 away for the high cost group).

Next we attempted to increase the sample size by employing the other available, but indicated as unreliable, data.

Here diagonal elements are "hits" (correct classifications) and off diagonals are "misses." Although the hit ratio is less than 50% (6/14), it is interesting that only one plant was misclassified two groups away, i.e. only one actual high cost group plant was classified as low cost and no low cost group was classified as high cost. Also interesting here is the high percentage of medium cost group members that were misclassified (4/5).

As more and more variables enter the analysis, we would expect better results -- hit/miss ratios, less "two-group away" misclassifications, etc. In Table II, for the interested reader, we reproduce some of the classification matrices, noting the variables used in each analysis.

Table II

Number of Cases Classified into Groups

Actual	Low	Medium	High
Low	2	1	2
Medium	1	3	1
High	0	2	2

Variables: MGT, WLBOD

	Low	Medium	High
Low	3	1	1
Medium	0	3	2
High	1	1	2

Variables: MCT, WLSS, WLBOD, % CAP

	Low	Medium	High
Low	3	2	0
Medium	1	3	1
High	0	1	3

Variables: MCD, WLSS, % CAP, % SS

When all six variables were employed in the multiple discriminant analysis program, the classification results took the form:

Number of Cases Classified into Group

Actual	Low	Medium	High
Low	4	1	0
Medium	0	4	1
High	0	1	3

Variables: MGT, WLSS, WLBOD, % CAP, % SS, % BOD

Now, in this case the hit ratio (11/14) is .786 and no plant is classified "two groups away." There are three misclassifications, one in each group, and, upon rechecking the data, it was found that each was close to a class boundary (about $3 away in the two lower cases and $6 away for the high cost group).

Next we attempted to increase the sample size by employing the other available, but indicated as unreliable, data.

Table III summarizes the results obtained:

Table III

Number of Cases Classified into Group

Actual	Low	Medium	High
Low	2	4	0
Medium	1	3	4
High	1	0	7

Variables: MGT

	Low	Medium	High
Low	2	3	1
Medium	1	4	3
High	1	2	5

Variables: MGT, % CAP, % BOD

	Low	Medium	High
Low	4	1	1
Medium	1	5	2
High	0	2	3

Variables: MGT, WLSS, WLBOD, % CAP Note: Sample for this case was 19 since no data was available on SS for three of the additional plants.

	Low	Medium	High
Low	5	1	0
Medium	0	6	2
High	0	3	2

Variables: ALL SIX - MGT, WLSS, WLBOD, % CAP, % SS, % BOD Note: Here again, the sample size is 19.

In considering the results presented in Table III, it is interesting that the addition of two variables (from 4 to 6) does not aid in the classification procedure. The best hit ratio was .684 (13/19).

Next, we switched our attention to considerations of two group analysis -- low and high -- dividing plants as noted earlier. The best classification result for the "reliable" sample is given in Table IV.

Table IV

Number of Cases Classified into Group

Actual	Low	High
Low	7	1
High	0	6

Variables: MCT, % SS, % BOD

The introduction of additional variables did not increase the high hit ratio $\left(\frac{13}{14}\right)$ of .929 which we were able to obtain using the three variables - MGT, % SS, % BOD - as reported in Table IV. Where the data set was expanded (to either 19 or 22) by employing the "unreliable" data, the results were as reported in Table V.

Table V

Number of Cases Classified Into Group

Actual	Low	High
Low	8	2
High	1	11

Variables: MGT, % BOD Sample: 22

	Low	High
Low	7	3
High	1	8

Variables: MGT, % CAP, WLSS, WLBOD Sample: 19

	Low	High
Low	8	2
High	1	8

Variables: MGT, % SS, % BOD Sample: 19

Adding additional variables did not help in upping the classification hit ratio, the highest being obtained using MGT and % BOD for the sample of 22 (hit ratio = 19/22 or .864) and using MGT, % SS, and % BOD for the sample of 19 (hit ratio = 16/19 or .342). As would be expected, the hit ratios for all samples increased as the number of groups was decreased to two. In addition, the hit ratio remained the highest when only "reliable" data was employed - .929 against .864 and .842 for the others.

III

The classification techniques, including that illustrated in the previous section can provide a cheap and easily handled method for initial considerations of questions concerning the advantage of large scale or regional plants. The results presented there ran according to the "lowering average cost" arguments, i.e., the higher rate plants yield lower average costs. The average values of the variables by groups are given in Tables VIa and VIb.

Table VIa - Three Groups

Average Cost Category

	Low	Medium	High
MGT	7267.88	3250.44	2119.31
WLSS	1.47	1.85	1.08
WLBOD	1.09	.65	.99
% CAP	.91	.76	.88
% SS	.93	.94	.93
% BOD	.90	.83	.86

Sample: "Reliable" - 14

	Low	Medium	High
MGT	6361.58	3161.75	1748.98
WLSS	1.37	1.20	.91
WLBOD	1.13	.49	.77
% CAP	.87	.69	.84
% SS	.93	.86	.90
% BOD	.91	.77	.84

Sample: including "unreliable" - 19

	Low	Medium	High
MGT	6861.58	3161.75	1230.67
% CAP	.87	.69	.87
% BOD	.91	.77	.77

Sample: All 22

Table VIb - Two Groups

Average Cost Category

	Low	High
MGT	6336.40	1729.60
WLSS	1.52	1.46
WLBOD	.97	.82
% CAP	.82	.88
% SS	.93	.92
% BOD	.90	.81
Sample - "Reliable" - 14		
MGT	5669.64	2056.89
WLSS	1.32	1.01
WLBOD	.92	.61
% CAP	.78	.79
% SS	.92	.86
% BOD	.89	.76
Sample - Including "unreliable" - 19		
MGT	5669.64	1634.37
% CAP	.78	.82
% BOD	.89	.74

Sample - All 22

Thus, as these tables indicate, lower average cost plants also tend to average higher MGT, higher % capacity utilized, and higher levels on the work measure units (WLSS, WLBOD, % SS, % BOD). In almost all cases, for those groups the values are highest for the low cost group, though there is some ambiguity on treatment measures between the medium and high cost groups (Table VIa).[5] When two groups -- high and low -- are considered, the values are highest for the low cost group except for % cap (Table VIb). Thus efficiency, at least in terms of the measures used here, appears higher for the larger, lower average cost plants--adding strength to large scale, regional treatment plant proposals.

The above is meant only as an initial look into a critical planning problem. Results using the technique employed seem to agree with theoretical statements concerning cost and treatment advantages enjoyed by centralization into larger plants. But what we have presented is done only as an initial step, as an indication of the potential usefulness of classification techniques. Before their role can be realized, however, the data on existing facilities must be increased in volume and concentration and upgraded in quality and accuracy. The results we have presented suffer from the sparseness of the sample, but as yet it is the most we are able to garnish. The use of outside expertise in the field helped us separate reliable and unreliable data supplies. Interestingly, the classification results degenerated somewhat each time "unreliable" data was added to the "reliable" data set.

What we now suggest is further useful employment of classification techniques such as discriminant analysis as a means of screening large amounts of data for consistency.

Suppose, using expertise (e.g., Dr. Etzel's and others) and familiarity with a certain set of operating plants we were able to formulate a large sample of reliable data and construct an accurate classification

technique (in cost terms, as above, or possibly in terms of a treatment variable). Once constructed, this procedure could serve as a means for selecting "outliers." For instance, using the discriminant function whose parameters are estimated using the reliable data, it would be possible to easily feed in data corresponding to additional plants and check for correct or incorrect classification. Plants incorrectly classified, especially those misclassified into groups far divergent from their correct one might then be checked more carefully for one of several possible difficulties such as: 1) data collector does not understand collection forms, 2) data generating techniques (BOD sampling, etc.) are not conducted properly, 3) data reporting is inconsistent due to mislabeling (i.e. calling maintenance costs "capital improvement" or "labor," etc.), 4) possible intentional misrepresentations. Certain misclassified plant's data would probably possess none of these problems, but regulatory agencies might find it quite useful to study these "divergers" in an attempt to find out whether they are outliers due to personnel capabilities, equipment capabilities, influent or geographic conditions, etc.

IV

Concluding Remarks

The questions of potential cost and treatment advantages from regionalization and plant consolidation form a major consideration for planning decision makers. Using some initial results, we have attempted to show how classification techniques can provide a relatively easy and inexpensive means for beginning to answer these questions. In addition, with sharper data collection and analysis, these techniques can provide a basis for indicating outliers which need to be studied.

FOOTNOTES

[1] We avoid use of significance tests here, thus reducing the necessity of this last assumption.

[2] The number of functions which can be so calculated is limited by the lesser of two numbers: 1) the number of groups less one, and 2) the number of characteristic measures used.

[3] In expanding the basic approach of this paper, we are currently expanding work load measures taking into account sludge volume indices, mixed liquid solids, etc., and interrelations. As Dr. Etzel has pointed out, the work load measures we employ are somewhat naive, but are quite convenient at this initial stage.

[4] See [3]. Methods of reporting maintenance and loan costs vary widely over plants, though work is now under way by the Indiana Stream Pollution Control Board to correct this.

[5] The distinction between Medium and High Cost plants is not as pronounced as it is between the Low Cost plants and the other two. For the former two the relative efficiency measures vary considerably with no apparent set pattern as between Low Cost plants and the others. This may be due in some part to the selection of cutoff values for the Low, Medium, and High Cost plants, though these levels were suggested by the data collected and by individuals involved directly with regulation.

REFERENCES

1. Cooley, W. W., and Lohnes, P. R., *Multivariate Procedures for the Behavioral Sicences*, John Wiley and Sons, New York, (1962).

2. Cooley, W. W., and Lohnes, P. R., *Multivariate Data Analysis*, John Wiley and Sons, New York, (1971).

3. Marsden, J. R., Pingry, D., and Whinston, A., *Regression Analysis Applied to the Wastewater Treatment Field*, JWPCF, October, 1973, vol. 45

4. Morrison, D. F., *Multivariate Statistical Methods*, McGraw-Hill, New York, (1967).

5. Van de Geer, J. P., *Introduction to Multivariate Analysis for the Social Sciences*, W. H. Freeman and Company, San Francisco, (1971).

Chap. 4 Production Function Theory and the Optimal Design of Waste Treatment Facilities

Section I

This paper is divided into two sections, the first contrasting the theoretical production function framework of the current study with the historical considerations of production functions so often presented. The second section presents an application in the development of theoretical micro-production and cost functions for water treatment plants.

R. W. Shephard chooses the following to begin his treatise on cost and production functions [1]:

> In economic theory the production function is a mathematical statement relating quantitatively the purely technological relationship between the output of a process and the inputs of the factors of production, the chief purpose of which is to display the possibilities of substitution between the factors of production to achieve a given output. The distinct kinds of goods and services which are usable in a production technology are referred to as the factors of production of that technology, and, for any set of inputs of these factors, the production function is interpreted to define the maximal output realizable therefrom. (N.B. Underlining is added here for emphasis.)

Shephard continues [1, p.4]:

> The production function is regarded here as a mathematical construction for some well defined production technology. The technology consists of a family of conceivable and feasible engineering arrangements, not restricted necessarily to particular realizations found in practice and possibly spanning historical changes in the application of the technology. Once defined, the technology implies a certain set of factors of production and no limitation will be put upon the inputs of these factors both as to type and amount available. Thus the production function will be taken to describe the unconstrained technical possibilities of a technology without limitation to any existing or realized production units.

We note that much of the previous work dealing with the development of production functions fails to comply with the Shephardian notion of this concept. Several of these papers (see [2], [3], [4]) begin with a hypothetical "production functional" form which is then fit to some set of empirical data. These relationships are neither analytically derived nor tested, but merely used in a procedure of fitting historical empirical data and hence are of a descriptive rather than analytic nature. The assumptions of efficiency and cost-minimization are generally imposed in what really are treatises on the technological level currently employed, rather than production functions in the sense Shephard points to.

What is presented here is not meant as a critical review of prior literature nor as a deification of a chosen (i.e., Shephard's) explication of production functions. Rather we wish to present an alternative approach which, under certain circumstances, will prove more appropriate than simple fit or descriptive procedures, and which also is developed in such a way as to fairly closely satisfy the Shephardian criterion for a production function.[1] We wish to study efficient production processes emphasizing their importance and usefulness for planning purposes. The employed notion of a production function is not related to what people currently do, but rather a function relating inputs to outputs summarizing efficient production methods. The technique set out below has no necessary correspondence to actual historical occurence nor are any of several 'usual' properties (i.e., decreasing or increasing returns to scale, decreasing

marginal returns, etc.) initially assumed, though, in fact, one or several may occur within a particular production process.

Consider some ℓ-stage process K, which for convenience we term a production process, acting on an m x 1 input vector, I, in such a way as to obtain an n x 1 output vector, O.[2] We thus consider the ℓ-stage process K as a mapping from m space to n space, the particular mapping being given by the exact specification of the process under consideration. The stages of the process are not necessarily time-order linked (i.e., ordering of the stages may be arbitrary in certain processes and set by necessity in others). If an exact specification of the production process is achievable, then the mapping from I to O is fully determined. It should be noted that the specification sought here is not an empirical or estimated relationship, but rather a hypothetical engineering specification. It is not limited to currently employed processes and is restricted only by engineering infeasibility. Complete specification of the process allows selection of alternatives based on maximization criteria rather than a selection limited to some set of historically employed alternatives.

Given a complete specification of the process, a production frontier composed of maximal output levels obtainable from differing combinations of inputs is deriveable. A conceptual difficulty arises here from the introduction of the term 'maximal output' in the case where n is greater than unity.[3] In any particular study, a concrete interpretation of this term, dependent upon the process, is necessary.

The production frontier forms the basis for the development of a production function for the process. No information is available from the production function that is not available directly from the full specification of the process system. The production function is introduced for the purposes of compacting the system through the elimination of non-optimal alternatives and allowing a simplified tool for planning purposes. The production function is here treated as a convenient intermediate step in the determination of a general 'least-total cost' output relationship.

Theoretically, the inclusion of all relevant variables should allow the possibility of complete specification of the production process. In curve fitting, the continuing introduction of more and more explanatory variables into the estimating equation enable the fit to be improved, possibly approaching an almost perfect fit. But how far does one go? Transparently, all variables cannot be explicitly dealt with in a complex process. Some delineation must occur. But what criteria does one employ in the selection of variables to be explicitly treated as inputs? Should variables be lumped together into 'capital' and 'labor' classes shunning the individual input identity and hence clouding the specification of the system? Or should a specific selection process be followed in order to single out variables for explicit consideration?

Since we are employing the production function as an intermediate step in the attainment of useful total cost relationships, the

unexplained variance in the latter function appears a 'relevant' criterion. The approach we follow for limiting variables thus centers on reduction of the unexplained variance of the final cost estimates, and selection of variables explicitly considered as inputs in the engineering model (used in developing a production frontier) is based on the amount they add to total cost.[4]

In order to deal with the specification problem we now introduce the notion of tolerance limits. The development of tolerance limits rests on the consideration of the relative importance of individual inputs to the total cost of the process as discussed above. Tolerance limits can be established so as to require possible fluctuations in total cost due to variations in inputs to remain below some chosen level, say some percent of total cost. The inputs which exceed the determined tolerance level are termed 'tolerance chosen major' (TCMA) and the remaining inputs are termed 'tolerance chosen minor' (TCMI). The selection of tolerance limits does not necessarily imply the unique selection of variables to be labeled TCMI, but it does specify ranges of quantitative fluctuation which may be used in determining the separation of inputs in TCMI and TCMA categories. We point out here, that, though certain inputs affect total cost only to some minor extent, their influence on variations in the output considered may be quite significant. Similarly, variables affecting total cost greatly, may actually have little or no influence on output over wide ranges of input levels employed. A programming problem can be constructed in

which the TCMI inputs are variables and the TCMA inputs are parameters. The determination of the solution to this problem over various combination of TCMA variables yields the points on the production frontier necessary to calculate the least-total-cost-output relationship.

In the development of the production function outlined above, relative cost of input factors plays a specific, defined role. The result is somewhat different from, yet not necessarily contradictory to, the usual interpretation of a production function presented by economists. Input costs enter indirectly into the development of the production function by serving as a basis for grouping the inputs. The production function, when used for planning purposes, is employed in a cost-minimization study. The employment of tolerance limits basically separates out inputs with little effect on total cost (though they may or may not have a similar small influence of total output) and treats them as variables in the maximization solution. The procedure follows the Shephard approach with one explicit minor alteration - this being the separation of inputs. The selection of tolerance level determines the maximal possible error in the total-cost minimization solution. We now present a somewhat more formal consideration of the methodological concepts involved and presented above, and then proceed to Section II and a detailed example concerned with waste water treatment plants.

The production function relating input factors to output is usually presented as a function 'giving the maximal output obtainable

from a given set of inputs'. Following this, using I_i to denote TCMA variables, a_i to denote TCMI variables, and using '0' to denote output, we have:

$$0 = \tilde{f}(I_1, I_2, \ldots, I_n, a_1, \ldots, a_K) = \tilde{f}(I, a)$$

The problem may then be formulated as:

$$\begin{aligned} &\underset{I,a}{\text{Max}} \quad \tilde{f}(I, a) \\ &\qquad C(I) \leq \bar{C} \\ &\qquad\quad a \in A \\ &\qquad\quad I \in \Omega \end{aligned}$$

where $C(I)$ represents the cost of the inputs $I_1, I_2, \ldots, I_n$, $\bar{C}$ a possible cost constraint, and A and Ω are possible constraints on the input variables a and I respectively.

We may also view the problem in a "two-stage" framework of the following form:

$$\begin{aligned} &\underset{I}{\text{Max}} \left\{ \underset{a \in A}{\text{Max}} \tilde{f}(I, a) \right\} \\ &\qquad C(I) \leq \bar{C} \\ &\qquad\quad a \in A \\ &\qquad\quad I \in \Omega \end{aligned}$$

The "inner" maximization problem yields a function of the form $f(I)$ and the problem may be rewritten as:

$$\begin{aligned} &\underset{I}{\text{Max}} \quad f(I) \\ &\qquad C(I) \leq \bar{C} \\ &\qquad\quad I \in \Omega. \end{aligned}$$

Informally, the dual of this problem may be specified as:

$$\underset{I}{\text{Min}}\ C(I)$$

$$f(I) \geq \bar{0}$$

where $\bar{0}$ is some set bound on output.

A difficulty with this approach is that the dichotomy between the variables $I_1, I_2, \ldots, I_n$ and $a_1, \ldots, a_K$ is not clear. There appear in practice to be few variables which relate to the production process and yet do not affect cost in some, though possibly minute, way. The division of variables into sets I and a may be easily accomplished in certain processes where production expense is centered in certain inputs while other inputs are minute cost items in comparison. In other processes the division of inputs may be an extremely difficult process where, if C^T is the total cost function, the a_i would be chosen according to:

$$\left| \frac{\partial C^T}{\partial \bar{a}} |a_i - \bar{a}_i| \right| < \varepsilon, \ \forall_{i, i=1,\ldots K}$$

where $|a_i - \bar{a}_i|$ is the maximal feasible variation around a typical point $\bar{a}_i$.[5]

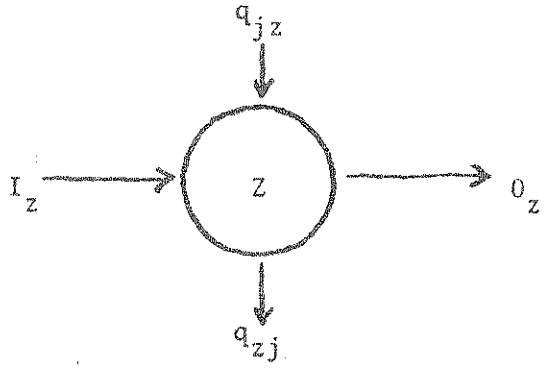
Thus the effect on values of total cost, C, for variations of the elements of 'a', is small (some chose 'epsilon' region), and the possible error involved in using the approach laid out above is hopefully held within an acceptable bound.

Section II

Let K_ℓ be particular ℓ-stage production process providing a mapping from I to O, as outlined above. At any of the ℓ stages of

process K_ℓ, there are $\ell_i (i = 1, \ldots, p)$ alternatives for moving to the next stage. We allow that various sets of stages might be alternatives for other such sets, and that some of these different combinations of stages may provide identical transformations from I to 0.

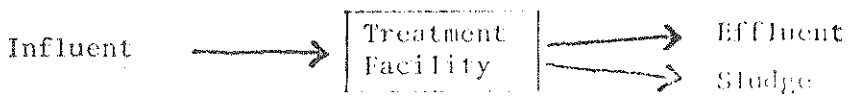
Schematically, the z^{th} stage of process K_1 may be characterized as:



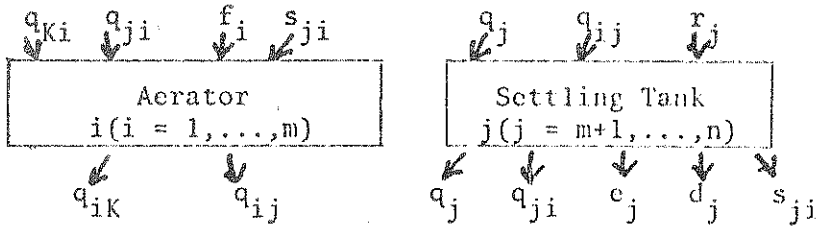
where I_z is the set of inputs directed into stage z, $q_{jz} (j=1,\ldots\ell)$ is the transfer flow from stage j to stage z, q_{zj} is the transfer flow from stage z to stage j and 0_z is the flow from stage z to final output.[6] Constraints are placed on the flow loads in the above schema by the technology and engineering infeasibility of certain flow movements. Constraints may also be of the joint form across transformations as in a staircase linear or non-linear programming structure to bring in the interrelatedness of the stages in the final production transformation.

The process of waste-water treatment is of the nature of process K_1. This section presents a detailed example of the generation of a production frontier for waste-water treatment facilities using the technique described in Section I.

The inputs into the waste-water treatment process are the waste water to be treated and the different elements (i.e., aeration, settling tanks, chemicals, etc.) used in treating the influent water. The influent waste stream may be characterized as to chemical and biological content in addition to temperature and volumetric flow. Concentration is centered on the BOD level since it serves as a basis for determining water quality levels.[6] Output from the production process is in the form of treated effluent and sludge waste for disposal. The effluent stream may be characterized in the same manner as the influent stream. The general process for secondary waste treatment plants makes use of two major basic units: aerators (where flocculation and micro-biological actions occur) and settling tanks (where treated water and sludge are separated). Aerators vary both in type and dimension while settling tanks generally differ with respect to dimensions. The main types of aerators are activated sludge (AS) and trickling filter (TF).[8,9] Waste water enters the production process and is channeled to some series of aerators and settling tanks for treatment (i.e., reduction of biological oxygen demand). Treated water passes from the system as effluent while waste sludge is fed to a sludge digester for ultimate disposal. Thus present analysis is centered on optimal selection of the variables in a system characterized generally as follows:



Inflow into aerator i(i=1,...,m) may originate from another aerator, k(k=1,...,m)(q_{Ki}), from a settling tank j(j=M+1,...,n)(q_{ji}), from initial inflow into the system channeled to the aerator directly (f_i), or from sludge recycling from settling tank. Outflow from aerator i goes either to another aerator (q_{iK}) or to a settling tank (q_{ij}), where indexing is as above and the q's are flow variates. Inflow into settling tank j comes from one of three sources: an aerator tank (q_{ij}), another settling tank, or from the initial influent into the system f_j. Outflow from settling tank j may go to another settling tank ($\ell = m+1,...,n$)($q_{j\ell}$), an aerator (q_{ji}), to the sludge digester (d_j), to an aerator in the form of recycled sludge digester (d_j), to an aerator in the form of recycled sludge (S_{ji}), or out of the system as effluent (e_j). Using vector and matrix notation, the general system may be outlined as follows:



K = 1,...,n = m+1,...,n

$$F = \begin{bmatrix} f_1 \\ \\ f_m \\ \hline f_{m+1} \\ \\ f_n \end{bmatrix} = \begin{bmatrix} F_A \\ \hline F_S \end{bmatrix}$$

- "Disposal of Initial Inflow Vector"

$$S = \begin{bmatrix} s_{m+1} & \cdots & s_{m+1,m} \\ \vdots & & \\ s_{n,1} & \cdots & s_{n,m} \end{bmatrix}$$ - "Rycycle Flow Matrix"

$$D = \begin{bmatrix} d_{m+1} \\ \vdots \\ d_n \end{bmatrix}$$ - "Digester Vector"

$$E = \begin{bmatrix} e_{m+1} \\ \vdots \\ e_n \end{bmatrix}$$ = "Effluent Flow Vector"

$$Q = \left[\begin{array}{ccc|ccc} q_{11} & \cdots & q_{1m} & q_{1'm+1} & \cdots & q_{1,n} \\ & & & & & \\ q_{m1} & \cdots & q_{mm} & q_{m,m+1} & \cdots & q_{m,n} \\ \hline q_{m+1m1} & \cdots & q_{m+1,m} & q_{m+1m+1} & \cdots & q_{m+1,n} \\ & & & & & \\ q_{m,1} & \cdots & q_{n,m} & q_{n,m+1} & \cdots & q_{n,n} \end{array}\right]$$

$$= \left[\begin{array}{c|c} Q_1 & Q_2 \\ \hline Q_3 & Q_4 \end{array}\right]$$ - "Inter-Unit Flow Matrix"

If ϵ_p is a unit vector of p components then the system flow constraints are of the form:

1. $Q_1^T \epsilon_m + Q_3^T \epsilon_{n-m} + S^T \epsilon_{n-m} + F_A - Q_1 \epsilon_m - Q_2 \epsilon_{n-m} = 0$

2. $Q_4^T \epsilon_{n-m} + Q_2^T \epsilon_m + F_s - Q_3 \epsilon_m + Q_4 \epsilon_{n-m} - S \epsilon_m - D - E = 0$

Equation 1 refers to the flow balance constraints for the aerator and 2 refers to the same for settling tanks. Equation 1 and 2 are simply constraints of the form 'volumetric inflow = volumetric outflow.'

Additional constraints are necessary for each unit concerning the micro-organism and nutrient (BOD) conservation. Let r_{11} be the organism growth rate and r_{12} be the nutrient consumption rate in AS type aerators and r_{21} and r_{22} the same in the TF type. The organism steady-state balance equation around the aerators take the form:

3. $Q_1^T q_A^o + Q_3^T q_S^o + S^T s^o + \bar{F}_A I_A^0 - \Delta_A Q_1 \epsilon_m - \Delta_A Q_2 \epsilon_{n-m} + V_A R_A = 0$

where

$$q = \begin{bmatrix} q_A^o \\ \hline q_S^o \end{bmatrix} = \begin{bmatrix} q_1^o \\ \vdots \\ q_m^o \\ \hline q_{m+1}^o \\ \vdots \\ q_n^o \end{bmatrix}$$

- Vector of Concentrations of micro-organism in flow leaving unit

$$\bar{F}_A = \begin{bmatrix} f_1 & 0 & \cdots & 0 \\ 0 & f_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & f_m \end{bmatrix}$$

- Diagonal Matrix of Influent Flows

$$V_A = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & V_m \end{bmatrix}$$

- Diagonal Matric of volumes of aerators

$$\Delta_A = \begin{bmatrix} q_1^o & 0 & \cdots & 0 \\ 0 & q_2^o & & \\ \vdots & & & \\ 0 & \cdots & & q_m^o \end{bmatrix}$$

- Diagonal Matrix of concentrations of organisms leaving Aerators.

$$s^o = \begin{bmatrix} s_{m+1}^o \\ \vdots \\ s_n^o \end{bmatrix}$$

- vector of concentrations of recycle sludge from settling tanks

$$I_A^o = \begin{bmatrix} P \\ \vdots \\ P \end{bmatrix}$$

- An mxl vector of a constant p where p is the initial microoranism concentration of the influent.

and $$R_A = \begin{bmatrix} R_{11} \\ \vdots \\ R_{11} \\ R_{21} \\ \vdots \\ R_{21} \end{bmatrix}$$

- mxl vector of m_1 R_{11}'s (referring to the m_1 A.S. aerators) and m_2 R_{21}'s (referring to the m_2 T.F. Aerators) such that $m_1 + m_2 = m$.

The second steady-state balance equation, referring to nutrient concentration, is given by:

$$4.\quad Q_1^T q_A^B + Q_3^T q_S^B + S^T s^B + \bar{F}_A I_A^B - \Delta_A^B Q_1 \epsilon_m - \Delta_A^B Q_2 \epsilon_{n-m} + V_A R_A^B = 0$$

where all symbols are defined corresponding to above with terms referring to nutrient (BOD) concentration substituted in as indicated by the superscript B, $R_{12} = 0$ and $R_B' = [0\ 0\ 0\ \ldots\ 0\ R_{22}\ R_{22}\ \ldots\ R_{22}]$. These material balance equations basically state that the outflow concentration from a unit in the process must be equal to the inflow concentration plus what happens within the unit.

In addition to the above, we have the following constraints:

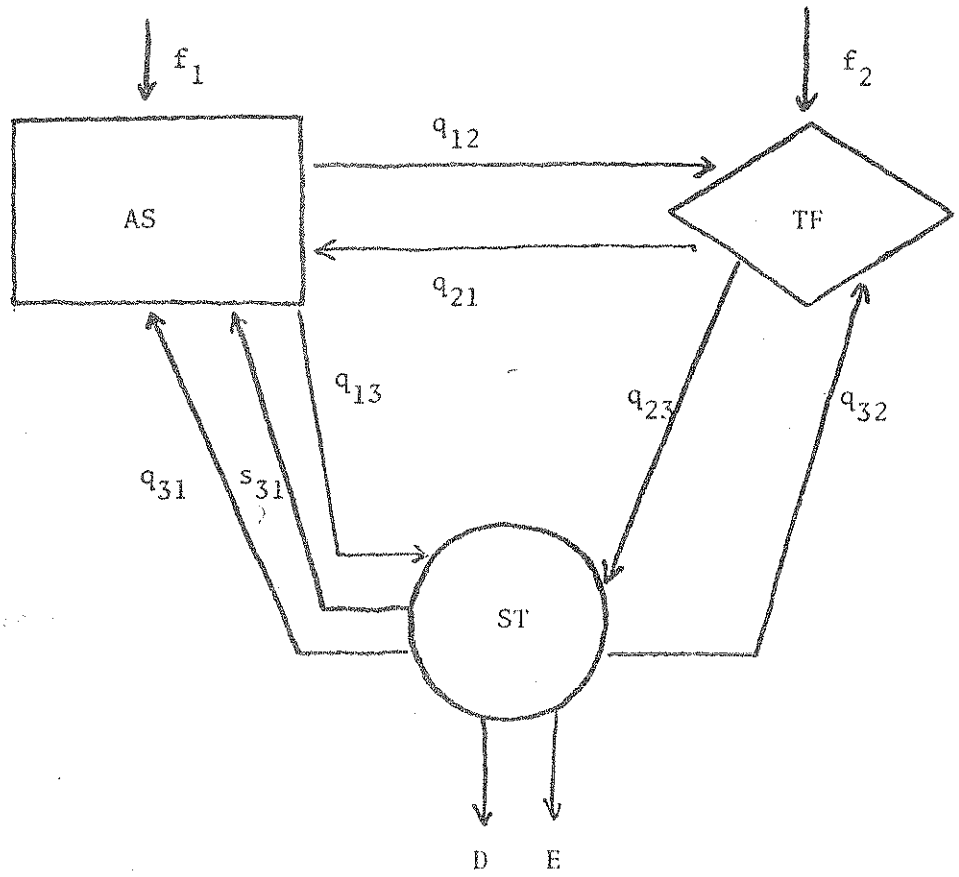
5. $F^T \epsilon_n = I_F$ - Total influent flow must be equal to total distributed initially to aerators and settling tanks.

6. $I_F = D\epsilon_{n-m} + E\epsilon_{n-m}$ - Total influent flow must equal effluent flow plus sludge digester flow.

The notion of the 'steady-state' balance equation as outlined above is similar to general equilibrium system analysis, where the state being sought is the optimal convergent state of the system.

The general model of the production process for secondary waste treatment plants is the basis of an in depth study currently being conducted by the authors. We now present a detailed small scale example illustrating the specification of a system and the generation of points on a production surface.

The flow network of the examples is illustrated below.



The specification of the general model is complete with the explicit exposition of the micro-organism growth rate and nutrient consumption rate for each type of aeration unit.

The rates used in the present study are those presented in Erickson and Fan [5] and Rich [6] and found or suggested in much previous work in the area. A complete study must take into consideration the very formulation of these equations. Little in the way of large scale testing seems to have been done previously. The rates for the (AS) units are given by:

$$7. \quad R_{11} = \frac{(GRR)(BOD)(ORG)}{(GRHF + BOD)} - (DELL)(ORG)$$

$$8.\quad R_{12} = \frac{(GRR)(BOD)(ORG)}{YBR(GRRH + BOD)}$$

where,

BOD = Nutrient concentration of influent (mg/ℓ)

ORG = Micro-organism concentration of influent (mg/ℓ)

GRR = Maximum growth rate when the nutrient concentration is not a limiting factor (hr^{-1})

GRRH = Concentration of nutrients at which the specific growth rate observed is one-half the maximum value (hr^{-1})

DILL = Micro-organism attrition rate (hr^{-1})

The consumption rate of nutrients in the (TF) units is given by:

$$9.\quad R_{22} = 10^{a(DTF)^b/(FITF/ATF)^c}$$

where,

DTF = Depth of filter (ft)

FITF = Total flow coming into filter (ft^3/Sec)

ATF = Surface area of filter (ft^3/Sec)

a, b, c = Estimated parameters

The micro-organism level is assumed to be unchanged in the (TF) unit. The percent of the organisms and nutrients settled out in the settling tank is assumed to be an exponential function of the holding time.

A nonlinear programming model can now be formulated which will enable the minimum BOD level possible under the above restrictions to be found for any possible combinations of inputs. The nonlinear problem can be solved using the algorithm described in [7]. The solutions to a large number of these problems can be used as observations to estimate the production surface. Again we note that the production

surface yields no more information than the specification of the production process itself.

The variables of the nonlinear programming formulation are the TCMI inputs. In the waste water treatment process these variables are assumed to be the volumetric flow rates between the AS, TF and ST units. The TCMA variables are assumed to be the variables associated with the size of the AS, TF and ST units. Other parameters such as the micro-organism growth rate are also given.

The small example model above was solved under the following conditions:

FI = Influent Flow = 1.0 cfs

BI = Influent Flow BOD = 100.0 mg/l

RI = Influent Flow ORG = 100.0 mg/l

TCMA Inputs

DAS = Depth of AS Unit = 10 ft

AAS = Area of AS Unit = 440.0 ft^2

DTF = Depth of TF Unit = 10.0 ft

ATF = Area of TF Unit = 440.0 ft^2

DST = Depth of ST Unit = 10.0 ft

AST = Area of ST Unit = 440.0 ft^2

Engineering and Biological Constants

SMST = Rate Parameter for ST Unit = 0.001

GRR = 0.1

GRRH = 100.0

YBR = 0.6

DILL = 0.002

$a = .00371$

$b = 0.67$

$c = 0.5$

The solution to this maximization problem was to use the TF and ST units and recycle half of the effluent from the ST unit back through the TF unit. This resulted in 95% of the BOC being removed.

The procedure of this paper yields points on the production frontier which may be used in the developemnt of a production function as suggested in section I. A further step is necessary for the utilization of the results for planning purposes. In this step the cost functions of the individual processing units would be the objective funtion of a nonlinear programming problem to be minimized subject to the constraints as set out above and an additional inequality constraint. The additional constraint would require that the output be of some present level. In terms of our example the BOC level would be required to be less than some given level. The use of this approach for planning purposes differs widely from the employment of fit, prechosen production functions.

The work involved in fully specifying a production process is voluminous, but the implications for planning purposes may well prove the initial input very productive. This approach allows the consideration of all available and forthcoming technology and is a break-away from the sterile curve fitting approach.

FOOTNOTES

1. The 'criterion' referred to here is the initial verbal explication Shephard presents, not the properties of the mathematical criterion he proceeds to.

2. The presentation here is simplified to a great extent in deference to possible length of the paper. Expansion and detail will be reported as part of a project currently being conducted by the authors.

3. This problem is not extensively treated in this paper not because of its lack of importance, but rather because of the chosen direction of the paper. The consideration of proper specification of output for an n-output production process will be considered in detail later in the current project. In this initial presentation emphasis is centered on BOD removal, both because of its importance in waste water treatment and because it provides a basis for comparison to previous work in the area.

4. As noted further on below, the problem may be formulated as a programming problem employing differing sets of inputs as fixed and maximizing output for these fixed sets using the 'non-selected' inputs as variables over which maximization occurs.

5. Though perfect specification is not attained, the approach laid out should serve to bound the possible error due to the lack of a 'perfect specification' rather than simply leaving the error to be assumed away and the specification clouded. Maximal variation levels of the a_i's clearly may be suggested by the process under analysis and the influence on total cost easily obtained. In some cases, however, this specification may by no means represent a trivial task. This needs to be kept in mind when making the explicit selection of the a_i's.

6. It is also possible that recycling may occur as in the waste treatment process set out below.

7. See [6] and [8].

8. Activated sludge aerators are generally of two types - plug flow and continuous mixing. The latter is used in the example presented.

9. The processes considered are limited in the present analysis for pedantical ease. Several other techniques are in different stages of technological development and/or actual use. A fully general approach would need to consider these recent and upcoming innovations. This forms a part of a larger study, with the presentation here centered on an easily extended methodology.

REFERENCES

1. Shephard, Ronald W., Theory of Cost and Productions, Princeton University Press, 1970.

2. Dhrymes, Phoebus J. and Kurz, Mordecai, "Technology and Scale in Electricity Generation," Econometrica, Vol. 32, No. 3, July 1964.

3. Borts, G. H., "Production Relations in the Railway Industry," J. P. E., Vol. 62, 1954.

4. Johnston, J., "Statistical Cost Functions in Electricity Supply," Oxford Economic Papers, N.S., Vol. 4, 1952.

5. Erickson, Larry E. and Fan, Liang-Tseng, "Optimization of the Hydraulic Regime of Activated Sludge Systems," Journal of Water Pollution Control Federation, Vol. 40, No. 3, Part 1, March 1968.

6. Rich, Linvil G., Unit Processes of Sanitary Engineering, John Wiley & Sons, N. Y., 1963.

7. Graves, G. W., Pingry, D. E. and Whinston, A., "Application of a Large Scale Nonlinear Programming Algorithm to Pollution Control," Proceedings Fall Joint Computer Conference, 1971 (Forthcoming)

8. Clark, John W. and Viessman, Warren Jr., Water Supply and Pollution Control, International Textbook Company, Scranton, Pennsylvania 1965.

Chap. 5 Engineering Foundations of Production Functions

The development of production theory within the field of Economics has centered around two classic papers - one by Cobb and Douglas [2] first appearing in 1928 and the other by Arrow, Chenery, Minhas, and Solow [1] appearing in 1961. These original papers combined theoretical requirements and empirical evidence to develop functional forms. Economists have continued to present generalizations of these two functions with little development or consideration of the underlying principles, e.g., [3], [7], [10].

This paper details an alternative approach which ties the derivation of the production function form to reaction formulas common in chemical engineering and biology.[1] The basis of the functional forms is shifted from "desirable" economic properties to physical aspects of the process under consideration. Instead of imposing strictures, i.e., homogeneity of degree one, perfectly competitive factor markets, etc., on the representation of the process, emphasis is placed on employing an aggregated representation obtained

---

*Research supported by O.W.R.R. The authors are indebted to their colleague, Professor Robert Squires, of the Chemical Engineering Department at Purdue University. Any errors remain the responsibility of the authors.

from the reaction properties of the process. The validity of the representation hinges on the accuracy of the reaction equations used rather than on the validity of assumptions concerning economic conditions such as cost minimization, constant returns to scale, and perfect competition.

We use the following format in this initial consideration:

I) A review of the assumptions and procedures used in the development of the Cobb-Douglas, CES, and related production functional forms emphasizing the link of the developments of these functions to properties other than technical process description.

II) A presentation of an alternative production function derivation technique beginning with the introduction of several well-known engineering relationships and employing this background to derive production functional forms linked directly to physical aspects of the production process. Several of the forms derived correspond to Economic functional forms often used, but are free of the "desirable economic properties" assumptions.

III) A consideration of the problems of obtaining micro-production functions.

I

The Cobb-Douglas presentation introduces the production function, P', of labor and capital alone and the authors select the particular form using the following reasoning:

1) The theory referred to (due to J.B. Clark, Wicksteed et al.) states that Production, Labor, and Capital are so related that if we multiply both Labor and Capital by a factor m then Production will be increased m times, that is Production is a first degree homogeneous function of Labor and Capital. Now, P' is taken to be such a function.

> 2) Among such functions the further restriction is placed upon P' that it should approach zero as either L or C approaches zero.
>
> Among functions with these properties (1) and (2) let us make a definite choice and examine the consequence of that choice, reserving the right to make other choices if we wish. Let us choose the function
>
> $$P' = bL^{k}C^{1-k}$$
>
> and find such numerical values of b and k that P' will "best" approximate P in the sense of the Theory of Least Squares. (p. 151-152 [2])

In addition, the following two provisos are added so that certain mathematical properties of the authors' chosen function are easily derived:

> Let us choose the following assumptions and let their justification rest on what we deduce from them.
>
> A) The Physical Volume of Production is proportional to the Volume of Production due to manufacturing alone.
>
> B) Any departure of P from P' may be represented by a change in value of the coefficient of $L^{3/4}\ C^{1/4}$ so that always
>
> $$P = bL^{3/4}C^{1/4}$$
>
> where the value of b is independent of L and C.
>
> These two assumptions are made in accordance with a general policy to ignore the quantitative effects of any force for which we have no quantitative data. The coefficient b is thus made a catch-all for the effects of such forces. (p. 155 [2])[2]

Throughout the paper the authors refer to the historical nature of their approach and the limitations on their analysis.

In the ACMS paper, the CES production function is developed by employing the following:

1) an observed "good fit" regression result;
2) an assumption that the production function is homogeneous of degree one;
3) an assumption that the labor and product markets are competitive.

The regression result observed was the "good" least squares fit obtained for the following equation:

$$\log V/L = \log a + b \log W$$

where

V = value added in thousands of U. S. dollars

L = labor input in man-years

W = money wage-rate (total labor cost divided by L in dollars per man year).

Using the assumption of homogeneity [i.e. $V/L = F(K/L,1)$], the authors rewrite the production function, $V = F(K,L)$ as follows:

$$y = f(x)$$

where

$y = V/L$

$x = K/L$ {K is capital input}.

The assumption that the product markets are competitive is in the form:

$$W = f(x) - x f'(x)$$

where w is the wage rate with output as numeraire. Finally, solution of a differential equation yields the CES functional form:

$$V = L\,(\beta K^{-p} L^{p} + \alpha)^{-1/p}$$

or

$$V = (\beta K^{-p} + \alpha L^{-p})^{-1/p}.$$

More recently, several extensions and generalizations of the CES production function form have been presented [3], [7], [10]. Modifications have been introduced to allow for such economic properties as changing returns to scale over the function [3], positive and diminishing marginal productivities for certain values of the parameters [10], and variable factor shares [10].

Our basic disagreement with these formulations and developments of production function forms is that they are based on varying combinations of desirable economic properties and assumed market parameter relationships. Further, though many applications have been on the micro-levels, the development of the functional forms has generally been on a highly aggregated level. Breaking from this approach and viewing the production function as a technical relationship representing the "maximal output achievable from a given set of inputs," we now proffer an alternative derivation method based on the general reaction principles of many processes. The primitives are thus shifted from the economic properties to the engineering relationships used. The validity of the final production functional forms rests on the accuracy of the engineering relationships used rather than on the preciseness of the economic properties assumed. What we present is directly linked to the notion of a production function as a technical relation and, as such, reverts back to several

very basic notions presented in such earlier works as Smith [9] and Shepard [8]. While differing from the present study in approach and general concentration, Smith's monograph does introduce the direct engineering approach and several examples in Chapters II and VI. In formulating the production function, the procedure presented below does not require use of any price data nor information on market structures. Because of this it provides a method for direct technical analysis rather than indirect analysis the validity of which rests on the reliability of assumptions not tied to the technical process itself. The functional form is derived from technical relations which may be tested for reliability rather than from assumptions such as perfectly competitive markets which provide little hope for testability.

II

A large class of production processes can be characterized by a system of unit reactors. The most general models which describe these systems are based on the fundamentals of mass and energy balance around each unit reactor. For all processes (exclusive of muclear reactions) the steady-state mass balance relationship can be described as::

[input of material m_1 to reactor] - [output of material m_1 from reactor] = [amount of material m_1 reacted].

The study of these balances is called *stoichiometry* in chemical engineering and is the basis for describing chemical reaction systems

If the material of concern is conserved, that is the amount of material reacted is zero, then the mass balance around a reactor is

easily calculated since no production or decay occurs. On the other hand if the material is not conserved, then the reaction rate of the material in the reactor much be known. For a continuous reactor, with negligible volumetric expansion, chemical engineers use the more explicit form of the mass balance equation:

(1) $C_{10}Q-C_1Q = \bar{r}_1V$

where

C_{10} = initial concentration of material m_1, moles m_1/ft^3

C_1 = final concentration of material m_1, moles m_1/ft^3

Q = volumetric flow entering and leaving reactor, ft^3/time

V = volume of reactor, ft^3

$\bar{r}_1$ = average reaction rate of material m_1 between 0 and x, moles m_1 formed/time - ft^3

The reaction rate is written more specifically as:

(2) $r_1 = r_1 (C_1, C_2, \ldots, C_n, P, T)$

where

$C_1,\ldots,C_n$ are the reactants or products of reaction

P = Pressure

T = Temperature

For practical purposes the implicit function, r_1, often takes the following explicit form:

(3) $r_1 = k(P,T)\, C_1^{\alpha_1} C_2^{\alpha_2} \ldots C_n^{\alpha_n}$

a functional form similar to a generalized Cobb-Douglas. K(P,T)

in (3) is commonly referred to as the "rate constant" though, in fact, it is a function of temperature and pressure.

The description of any particular unit reactor usually consists of a system of differential equations which can be solved for the concentrations of the various materials. There is a long history of these types of models in chemical engineering. Consider a chemical reaction of the form:

$$m_1 + m_2 \xrightarrow{k_1\ (T,P)} m_3$$

$$m_1 + m_3 \xrightarrow{k_2\ (T,P)} m_4$$

where $k_1\ (T,P)$ and $k_2\ (T,P)$ are the rate constants associated with the two reactions. One standard linear model of this system is as follows:

$$r_1 = - k_1 C_2 - k_2 C_3$$

$$r_2 = - k_1 C_2$$

$$r_3 = k_1 C_2 - k_2 C_3$$

$$r_4 = k_2 C_3$$

where the symbols are as defined earlier.

With the above brief background in mind and employing the fundamental of mass balance with the kinetics used in practice by chemical engineers, we can derive production functional forms which are comparable to those commonly used by economists but which require

very stringent technical restrictions to fall in such commonly used classes of functions as Cobb-Douglas or CES.

For illustrative purposes consider a reactor which is producing material m_1 at a rate which is a function of the concentration of m_1 and the concentration of a second material, m_2.

Assuming the rate equation to be:

(4) $$r_1 = k_1 C_1^{\alpha_1} C_2^{\alpha_2}$$

Using the differential form and integrating the resulting differential equation over the reactor length, the following equation results:

(5) $$C_1 = \left\{ k_1 t_H (1-\alpha_1) C_{20}^{\alpha_2} + C_{10}^{(1-\alpha_1)} \right\}^{1/(1-\alpha_1)}$$

If $t_H = V/Q$, temperature and pressure fixed, then equation (5) is a variable elasticity of substitution production function where C_{20} and C_{10}, the initial concentration of m_1 and m_2, are the inputs and C_1, the final concentration of m_1, is the output of the production process. The elasticity of substitution[3] is:

(6) $$\sigma = \frac{C_1^{1-\alpha_1} + k_1 t_H \alpha_2 C_2^{\alpha_2}}{(1-\alpha_2) C_1^{1-\alpha_1} + \alpha_1 \alpha_2 K_1 t_H C_2^{\alpha_2}} .$$

The condition for the form to be CES is

(7) $$1-\alpha_2 = \alpha_1$$

which implies

$$\sigma = \frac{1}{1-\alpha_2} .$$

The production functional form in (5) is derived from a differential equation which is dependent strictly on the underlying engineering restrictions and has nothing to do with the market behavior or conditions on economies of scale. This is in direct contrast to the derivations of the Cobb-Douglas and CES forms which make direct use of such restrictions for their derivation.

It is of interest to note that the functional form derived in (2) describes one of the most simplistic production processes- a single reactor and two inputs. However, the production functional form obtained has the property of variable elasticity of substitution which is comparable to the more complex of the economic production functional forms generally employed.

If the temperature or pressure were considered as inputs, or a sequence of reactions such as

$$m_n \xrightarrow{k_{n-1}} m_{n-1} \rightarrow \cdots \xrightarrow{k_1} m_1$$

were assumed, then the appropriate functional form becomes more complex. For example, if three materials were assumed to react sequentially in a continuous reactor where

$$r_i = k_i C_i^{\alpha_i} C_{i+1}^{\alpha_{i+1}} \qquad i = 1,2 \tag{8}$$

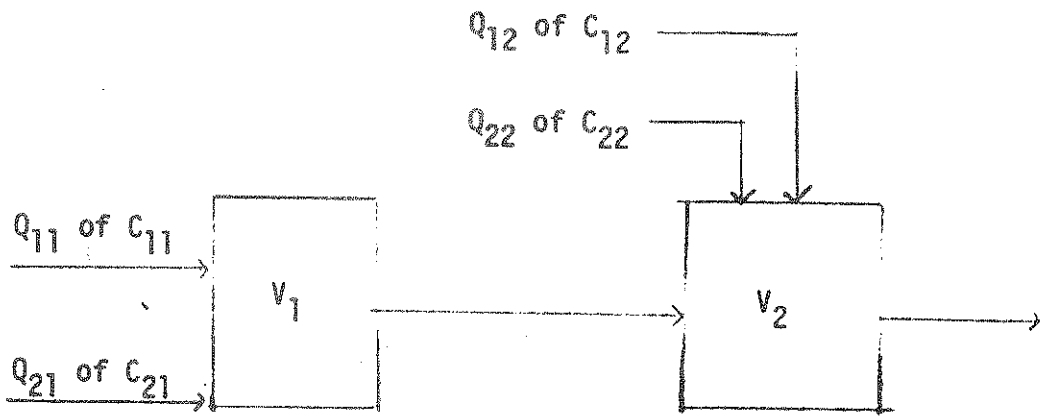
and the concentration of m_1 is the output of interest, then assuming an initial concentration of m_3 of C_{30}, the appropriate functional form is

(9)
$$C_1 = \left\{ A_1 (1-\alpha_1) \left\{ A_2 (1-\alpha_2) C_{30}^{\alpha_3} + C_{20}^{(1-\alpha_2)} \right\}^{\alpha_2/(1-\alpha_2)} + C_{10}^{(1-\alpha_1)} \right\}^{1/(1-\alpha_1}$$

where A_1 and A_2 are constants and the other symbols are as used earlier.

Another complication in developing the appropriate functional form can arise when the process occurs in a series of reactors with inputs entering intermediate reactors. For example, consider two continuous flow reactors of the type considered above with two reacting materials. This system is outlined in figure 1.

Figure 1: Two Reactor System with Intermediate Inputs



The production function for material m_1 in the first unit is:

(10)

$$C_1^1 = \left[\frac{A_1}{\sum_i Q_{i1}} \left(\frac{C_{21}Q_{21}}{\sum_i Q_{i1}}\right)^{\alpha_2} + \left(\frac{C_{11}Q_{11}}{\sum_i Q_{i1}}\right)^{1-\alpha_1}\right]^{1/(1-\alpha_1)}$$

$i = 1, \ldots, m$ (m = number of materials)

$j = 1, \ldots, n$ (n = number of reactors)

where, $A_1 = k_1 V_1 (1-\alpha_1)$, Q_{ij} is the volumetric flow of material i into reactor j, C_{ij} is the concentration of material i entering reactor j, and V_i is the volume of reactor i.[5]

Note that if the volumetric flows are assumed to be fixed that (10) can be rewritten as:

(11)

$$C_1^1 = \left[\beta_1 C_{21}^{\alpha_2} + \beta_2 C_{11}^{(1-\alpha_1)}\right]^{\frac{1}{1-\alpha_1}}$$

where

$$\beta_1 = \frac{A_1 Q_{21}^{\alpha_2}}{(\Sigma Q_{i1})^{\alpha_2+1}}$$

$$\beta_2 = \left(\frac{Q_{11}}{\sum_i Q_{i1}}\right)^{1-\alpha_1}$$

The production form in (11) is also a variable elasticity form. The production function for the combined system is

(12)

$$C_1^2 = \left[\frac{A_2 \left(\sum_j C_{2j}Q_{2j}\right)^{\alpha_2}}{\left(\sum_i \sum_j Q_{ij}\right)^{\alpha_2+1}} + \left(\frac{C_1^1 \left(\sum_i Q_{i1}\right) + C_{12} Q_{12}}{\sum\sum_{ij} Q_{ij}}\right)^{1-\alpha_1}\right]^{1/1-\alpha_1}$$

or by substituting (11) into (12),

(13)
$$C_1^2 = \left[\frac{A_2 \left(\Sigma C_{2j} Q_{2j}\right)^{\alpha_2}}{\left(\sum\sum_{ij} Q_{ij}\right)^{\alpha_2+1}} + \left[\frac{A_1 \left(C_{21} Q_{21}\right)^{\alpha_2}}{\left(\sum_i Q_{i1}\right)^{\alpha_2+1}} + \frac{C_{11} Q_{11}}{\sum_i Q_{i1}} \right]^{1-\alpha_1} \right]^{\frac{1}{1-\alpha_1}} \left(\frac{\sum_i Q_{i1}}{\sum\sum_{ij} Q_{ij}} \right) + \frac{C_{12} Q_{12}}{\sum_i \sum_j Q_{ij}} \Bigg]^{\frac{1}{1-\alpha_1}}$$

Again, if the volumetric flows are fixed (13) can be rewritten

(14)
$$C_1^2 = \left[\left(\sum_j C_{2j} \phi_{2j} \right)^{\alpha_2} + \left[\left(\phi_1 C_{21}^{\alpha_2} + \phi_2 C_{11}^{1-\alpha_1} \right)^{\frac{1}{1-\alpha_1}} \phi_3 + \phi_4 C_{12} \right]^{1-\alpha_1} \right]^{\frac{1}{1-\alpha_1}}$$

where the ϕ_j are appropriately defined.

Another system which can be modeled using the prescribed procedure is river water quality. A fairly simplistic, but often employed, bi-dimensional output model has the form

(15)
$$r_B = -k_1 B$$

$$r_D = k_1 B - k_2 D$$

where

B = Biochemical oxygen demand

D = Dissolved ozygen deficit

and

k_1 = Deoxygenation rate coefficient

k_2 = Reaeration coefficient

The solution of this model is:

(16)
$$B = B^0 e^{-k_1 t}$$

$$D = \frac{k_1}{k_2 - k_1} \left(e^{-k_1 t} - e^{-k_2 t}\right) B^0 + C^{-k_2 t} D^0$$

A unit reactor is assumed to be a section of the river of some given length over which the volumetric flow is fixed. If temperature and pressure are also assumed to be fixed then the production process can be represented by

(17) $$\begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} \delta_1 & 0 \\ \delta_2 & \delta_3 \end{bmatrix} \begin{bmatrix} B^0 \\ D^0 \end{bmatrix}$$

where δ_1, δ_2 and δ_3 are appropriately defined.

A more realistic and complex model of water quality which includes temperature as a variable has recently been solved by Pingry and Whinston in [6]. This model has the form

$$r_T = k_3 T$$

$$r_B = -k_1 B$$

$$r_D = k_1 B - k_2 D$$

$$k_1 = k_1^{20}\, \theta_1^{T_A - 20}$$

$$k_2 = k_2^{20}\, \theta_2^{T_A - 20}$$

$$T_A = T + T_E$$

where

$T = T_A - T_E$ = Deviation of actual temperature (T_A) from equilibrium temperature (T_E).

k_1^{20} = rate constant at 20°C

k_2^{20} = rate constant at 20°C

B = Biochemical oxygen demand

D = Dissolved oxygen deficit

θ_1, θ_2 = parameters of the rate function.

The type of analysis considered above can be directly applied to chemical and biochemical processing plants such as oil refineries and wastewater treatment plants. But the discussion above is by no means limited simply to processes of this type. A large number of processes fall directly into easy representation by a series of reactors, while a further set of processes are representable to varying degrees by functional forms derived as outlined above. For example, an assembly plant could be characterized by a series of assembly units ("reactors") and the process could be described in terms of a rate equation. If a rate model is appropriate for a basic unit of the production process it might be interpreted as:

$$\frac{dy}{dt} = f(x_1, x_2, \ldots, x_n) \tag{18}$$

where the x_i's represent inputs such as electricity and varying forms of labor.

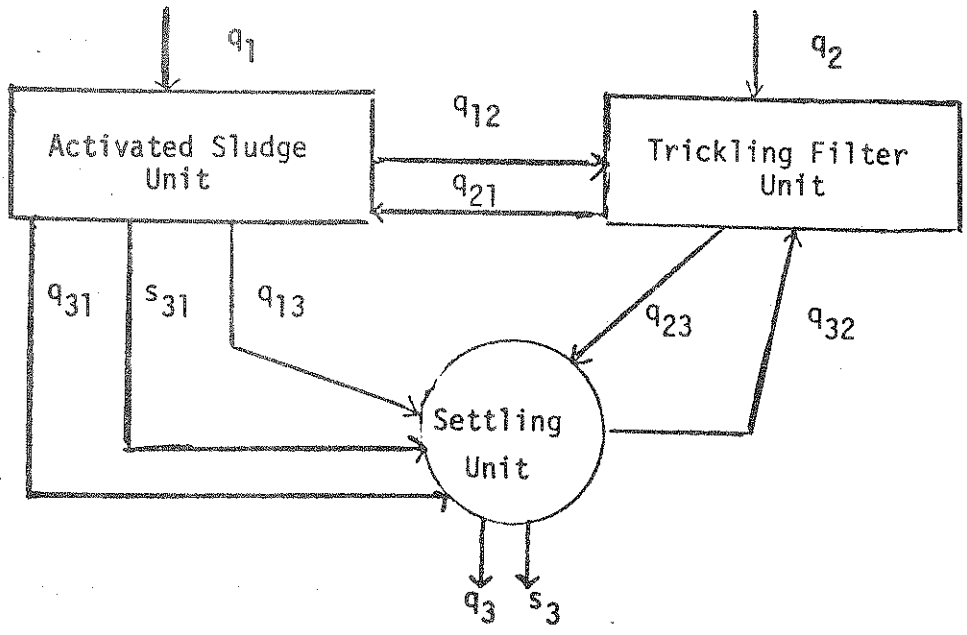
The analysis above also provides a framework for examining the short run vs. the long run. Long run variables are relatively easy to isolate in this framework where the analysis is directly technically linked. For example, in chemical processing, unit reactor volumes would appear as long run variables since it is difficult to alter these variables in any short period of time once a plant is constructed. In

this case, long run analysis would treat the reactor sizes as variables.[5]

III

The previous section of this paper has presented a case for the development of production function forms based on the underlying engineering principles. The question arises, however, regarding the feasibility of the suggested procedure for a particular production process. Many production processes consist of an assortment of reactors each of which contains a reaction which can be described by a large system of simultaneous non-linear differential equations. The process could then be further complicated by a number of recycle loops. A wastewater treatment plant is typical for such a system. Figure 2 is an illustration of a model for a wastewater plant used by Marsden, Pingry and Whinston [5].

Figure 2: Wastewater Treatment Plant



q's = Wastewater flows s's = Sludge flows.

Such a system which is typical of many chemical plants, may prove to be nearly impossible to solve. Suppose this is the case in attempting to solve a system for an explicit production functional form. What functional form would one use for estimation? The available functions are the traditional forms used by economists and forms constructed by assuming a simplification of the complex system and using the procedure described in section II. In both of these cases the parameters of the system have no physical interpretation. In fact there is no theoretical reason to believe that any one of these forms would be more appropriate than any other. The problem is now strictly an empirical question with the traditional economic forms and the forms derived from the mass balance conditions having neither more nor less meaning than a polynomial of some order.

Since many systems are difficult to solve it is of interest to isolate some of the systems which are relatively easy to solve. The class of linear systems meet this criteria. These systems have the following important property.

Superposition

If C_1 and C_2 are outputs of the system associated with inputs C_1^0 and C_2^0 then $C_1 + C_2$ is the output associated with the input $C_1^0 + C_2^0$.

This property allows a complicated solution for the equations to be built up as a linear sum of simpler solutions. This property is found

in the water quality model above. The order equation model does not have this property and, therefore, describes a nonlinear system.

Since the linear systems have the property of superposition and are therefore, easy to solve, it is possible in this case to write a production function for a complex production process. The parameters of these complex linear systems will have specific interpretations as functions of the rate parameters of the individual reactions associated with each reactor.

Although the linear systems may seem overly restrictive, in fact there are a large number of systems which can be described using these linear systems. A large number of processes from electrical engineering, chemical engineering and biology can be adequately described using linear systems, an excellent example of which is the water quality model in section II.

On the other hand there are a number of systems which are non-linear which can be explicitly solved for a production functional form. These forms do not have the property of superposition. However, if the reactors are located in a cascade and the solutions of each reactor can be found, then simple substitution, using mass balance around the reactors, will yield the appropriate production function. Equation (9) is an illustration of a production functional form derived from this sort of procedure.

The problem with the use of the traditional functional forms is that the parameters have no physical interpretation and no foundations in commonly accepted scientific principles. Therefore the use of these functional forms, especially for prediction or planning is very risky.

The situation is not altered by driving functional forms from mass balance conditions and rate equations which are not appropriate for the system. In fact, one might argue in this case for the use of the economic based equations since they do have certain interesting economic properties with respect to elasticity of substitution and scale. Whenever feasible, however, we argue for the derivation of the explicit production functional form as the preferable method since it provides a basis for important direct technical analysis.

FOOTNOTES

[1] In [1], Smith presents a detailed analysis of production and investment theory containing several descriptions and analysis of production functions describing engineering processes and emphasizing the "embedded" problem of optimal capital replacement policy. The analysis presented in this paper has some similar beginnings to Smith's work, but differs in depth, scope, and emphasis. Unlike Smith's work market conditions do not enter the production function analysis. At the end of Section II of this paper several areas of recently developed interest are pointed out where the current approach is directly applicable.

[2] The exponents of 3/4 and 1/4 in "$L^{3/4}C^{1/4}$" stem from the explicit example used in the Cobb-Douglas paper. The statements made by the authors, however, are of general form and there would seem to be no confusion if "α" were read for "3/4" and "$1-\alpha$" for "1/4". The original form is used here only because of the direct quote.

[3] Employing the definition of elasticity of substitution as found in R.G.D. Allen's *Mathematical Analysis for Economists* (MacMillan, 1966), p. 504:

$$\sigma_{rs} = \frac{X_1 f_1 + X_2 f_2 + \ldots + X_n f_n}{X_r X_s} \cdot \frac{F_{rs}}{F}$$

$(r, s = 1,2,\ldots n)$

where $F = \begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{12} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & & & \\ f_n & f_{1n} & f_{2n} & \cdots & f_{nn} \end{vmatrix}$

$f_i = \frac{\partial f}{\partial x_i}$, $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$., Frs is the cofactor of f_{rs} in F, and f is the production function.

[4] Note that $V_i / \sum_j Q_{ji}$ is the holding time of reactor i.

[5] The distinction between long and short run formulations is considered in more detail in [5]. The division of variables is somewhat arbitrary depending on the particular problem under consideration.

REFERENCES

1. Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *The Review of Economics and Statistics*, Vol. 43, No. 3, August 1961.

2. Cobb, C. W. and P. H. Douglas, "A Theory of Production," *American Economic Review*, Vol. 17, No. 1, Supplement, March 1928.

3. Lu, Yoo-chi and Lehman B. Fletcher, "Generalization of the CES Production Function," *Review of Economics and Statistics*, Vol. 50, No. 4, 1968.

4. Marschak, Jacob "Economic Measurements for Policy and Predict Studies in Econometric Methods," edr. William C. Hood and Tjalling C. Koopmans, New York, 1953.

5. Marsden, J. R., D. E. Pingry and A. Whinston, "Production Function Theory and the Optimal Design of Waste Treatment Plants," *Applied Economics*, Forthcoming.

6. Pingry, D. E. and A. Whinston, "A Regional Model for Water Quality Control," Unpublished, Mimeograph, Purdue University.

7. Sato, Ryuzo and Ronald F. Hoffman, "Production Functions with Variable Elasticity of Factors Substitutions: Some Analysis and Testing," *Review of Economics and Statistics*, Vol. 50, No. 4, 1968.

8. Shepard, Ronald W., *Theory of Cost and Production*, Princeton University Press, 1970.

9. Smith, Vernon L., *Investment and Production*, Harvard University Press: Cambridge, Mass., 1961.

10. Tsang, Herbert and Patrick Yeung, "The CES-VES Family of Production Functions: Further Generalization and Empirical Analysis," presented at 46th Annual Meeting of the Western Economic Association, Vancouver, British Columbia, Canada, August 1971.

Chap. 6 Process Analysis: A Comment

Several recent articles have considered alternative techniques in characterizing production processes. Smith, Miller, and Golladay centered their attention to production of medical services while Griffin applied his techniques to the petroleum industry. Each of these papers, as noted by the authors, is concerned with short-run analysis characterizing currently existing and currently employed production processes. The purpose of this note is twofold - 1) critically review Griffin's process analysis, and 2) suggest a generalized production theory alternative which overcomes several problems inherent in Griffin's process analysis. It is shown that Griffin's process analysis, while possibly having some computational advantages, suffers from many of the same difficulties which generate criticism for standard statistical fitting. The generalized production theory alternative is based directly on engineering relationships and is designed to be capable of dealing with long-run planning problems. Further, as a special case, it is shown that properly placing restrictions on the optimization problem in this latter approach yields short run results similar to Griffin's though not restricted by the assumptions of a linear programming framework. After a few short comments on data from existing plants and on standard statistical fitting techniques, section I is devoted to a critical consideration of several of Griffin's claims for process analysis. Section II sketches a generalized production theory approach.

I

In order to compare the effectiveness of production function generating techniques one must have an idea of the type of data which is available. Observations on existing plants are likely to have one or more of the following characteristics:

1) The observations will be closely grouped and only represent a single or few alternative technologies.

2) The firms will be using a non-production frontier technology because of additions or alterations to existing plants in response to relative price changes or because the firms are not profit maximizers.

3. The firms will be using a production-frontier technology but will be operating in the short-run off the frontier.

A. Statistical Technique

Given the data restrictions, the best we can hope for this technique is good historical description. There are no short run operational rules implied by a particular fitted functional form and there is no reason to expect that the fitted functional form will correspond to the production frontier. There is no solid basis for planning decisions nor can questions on such subjects as returns to scale or convexity of the production process be adequately handled. Even if the observed firms were operating on the frontier the data would likely be very localized and extrapolation of any functional form would be extremely risky.

B. Process Analysis

Griffins recent article proposed an alternative approach to statistical cost functions. It is the authors' contention that Griffin's process analysis and the statistical cost function approach have basically the same disadvantages.

1) "This paper demonstrates the potential of the process analysis approach for deriving the short-run properties of cost curves... In both cases (single and joint product) the results here substantiate the classical assumptions about short run cost functions, i.e., that marginal costs slope upward and average costs are U-shaped." (p. 55, emphasis ours).

2) "...changes in costs can be linked directly with the limited capacity of the process units and the substitution between the various processes (p. 51).

3) "...the complete range of the cost curve can be investigated rather than a limited range of actual observations." (p. 51)

4) "...the effect of technological change as reflected in new processes and product mixes can be explicitly built into the cost function." (p. 51)

We take these claims in order:

1. It appears from these statements that Griffin views his petroleum industry example as a test of the classical cost function assumptions. The fact is that Griffin obtains an increasing marginal cost curve as a direct implication of the convexity of the linear programming - activity analysis model (as he notes on page 50). This is no different than assuming a Cobb-Douglas functional form for statistical estimation and then claiming an empirical result that substantiates the classical assumptions about the homogenaity of production functions. If one is to test the hypothesis that the cost curves are shaped in the classical manner at the very least there should be a criteria to accept or reject the hypothesis. The question of convexity of the production process cannot be dealth with in Griffin's framework since he allows no possible alternatives. Griffin's upward sloping

marginal cost curves do not imply convexity - they are an implication of his technique. The study of convexity requires the use of techniques such as those set out in Afriat [a,b] and Hanoch and Rothschild for empirical production data. For the general production problem the technique suggested in Marsden, Pingry, and Whinston is an appropriate one.

2. The marginal cost curve implied by the linear programming-activity analysis model is a step-function (as noted by Griffin in footnote 10, p. 49) rather than a positively sloping continuous curve. The inverted-L shaped curve as discussed by Griffin is simply a special case where only output. The appropriate procedure for obtaining the marginal cost curve would be to use parametric programming and generate the step-function over the feasible output range. Griffin, however, insists on connecting the points he generated with straight lines to further enhance his claim that his paper is a test of the classical cost function characteristics. His procedure eliminates the information which he claims the process analysis approach provides - the relation between the increasing marginal cost and the limited capacities. One could view Griffin's procedure as a first-order approximations of a convex nonlinear technology. However, this does not eliminate the convexity assumption and its implication of an upward sloping marginal cost curve implicit in his approach

3. Griffin's claim that the entire range of the cost function can be investigated is conditional on one's acceptance of the activity analysis formulation as a non-local representation. Activity analysis can be viewed as a first-order approximation of a set of non-linear production processes. The a_{ij}'s in this model do not have an engineering interpretation, but are simply the coefficients of the linear representation of the process. Since this is the case, then one would not expect the activity analysis formulation

to be valid for levels of output some distance away from observed values.

4. The claim of the possiblilty of incorporating new technological processes into the cost function depends on being able to represent these new processes with activity analysis. Since the A_{ij}'s are not direct engineering coefficients but simply estimated coefficients obtained for a "representative plant," incorporating new technology is not possible unless a model of the system exists. If this is the case, why not use all of the information available instead of a first-order approximation?

The only apparent advantage of Griffin's process analysis is a set of short run operation rules for restricted output levels. The technique is not directly applicable to planning problems, nor can it be used to deal directly with questions relating to returns to scale or convexity of the production process. Further, it may even be true that fitting a non-linear statistical cost function will yield a more accurate estimation of cost than the linear activity analysis, although the former does not imply short run operation rules.

II.

As an alternative, consider some ℓ-stage production process, K, acting on an m x 1 input vector, I, yielding an n x 1 output vector, O. K is thus a mapping from I to O is fully determined.[1] It should be noted that the specification here is not an empirical or estimated relationship, but rather a hypothetical engineering specification. It is not limited to currently employed processes and is restricted only by engineering feasibility. Complete specification of the process allows a selection of alternatives based on optimization criteria rather than a selection limited to some set of historically employed

alternatives.

Given a complete specification of the process, a production frontier composed of maximal output levels obtainable from differing combinations of inputs is deriveable. Each point, O_i^*, in this production frontier is associated with a particular technology which can be defined in terms of process sequencing or types of processes employed.

Mathematical approximation of the production frontier yields a production function for the process. No information is available from the production function that is not available directly from full specification of the process system. The production function is introduced for the purpose of compacting the system through the elimination of nonoptimal alternatives and allowing a simplified tool for planning purposes. The production function is treated here as a convenient intermediate step in the determination of a general "least total cost" output relationship.

Now consider a division of the input variables into two groups - a division that may be accomplished based upon several criteria. For the purpose of this exposition, we simply base the division on "long-run" vs. "short-run" variables.[2] Once the division is completed, the optimization problems may be formulated as a two-stage nonlinear programming problem. The non-linear problem can be solved using a non-linear algorithm such as the one described in Graves, Pingry and Whinston.

More formally, using I_i to denote "long-run" inputs, a_i "short-run" inputs, we specify the problem as

$$\underset{I,a}{\text{MAX}}\ f(I_1, I_2, \ldots, I_p, a_1, a_2, \ldots, a_s) = \underset{I,a}{\text{MAX}}\ f(i,a)$$

$$\text{S.T. } c \leq \bar{c}$$

$$a \in A$$

$$I \in \Omega$$

where c represents input costs, $\bar{c}$ a possible cost constraint, and A and Ω possible constraints on the input sets a and I respectively. The function f(I,a) is based on engineering relationships, and implicit here are constraints such as mass-balance, steady-state balance, and engineering restrictions.[3,4]

Viewing the problem in a two-stage framework yields the following form:

$$\underset{I}{\text{MAX}} \left\{ \underset{a}{\text{MAX}} \; (I,a) \right\}$$

$$\text{S.T.} \quad c \leq \bar{c}$$
$$a \in A$$
$$I \in \Omega.$$

The inner maximization problem will yield a function of the form $\tilde{f}(I)$ enabling the problem to be re-written as:

$$\underset{I}{\text{MAX}} \; (I)$$

$$\text{S.T.} \quad c \leq \bar{c}$$
$$I \in \Omega$$

Informally, a reciprocal of the two-stage problem is:

$$\underset{I}{\text{MIN}} \left\{ \underset{a}{\text{MIN}} \; c(I,a) \right\}$$

$$\text{S.T.} \quad f(I,a) \geq \bar{O}$$
$$a \in A$$
$$I \in \Omega$$

Where $\bar{O}$ is some bound on output, and constraints such as those noted above (e.g., material-balance, steady-state balance, engineering restrictions) are again implicit in the formulation. The inner minimization problem in this formulation is a more general formulation of Griffin's process analysis but overcomes several of the disadvantages of that technique.

The non-convex optimization technique has the following advantages:

1. does not prejudge convexity and therefore does not necessarily imply the traditional rising marginal cost curve;
2. allows changes in cost to be related to changes in technologies and capacities of various process elements;
3. is based on engineering relationships and can be extended over the full range of cost curve;
4. allows for the inclusion of new technologies using engineering relationships;
5. provides an inner-optimization problem which is appropriate for deriving short run operation rules which are independent of the validity of any convexity assumption.

The technique selected by an individual research depends on the problem faced. The statistical technique is more appropriate for a historical description. The process analysis approach is a short-run, first-order approximation which yields tentative short-run operation rules for convex technologies. This technique may prove a useful and inexpensive method to provide better short-run operation rules for specific plants. Application of the generalized production theory approach is more complex and hence more costly than the others. But it is the only technique broadly applicable to planning problems and to a detailed short-run analysis.

Footnotes

1.) A degree of randomness in a production process may preclude complete(exact) specification-i.e., certain random variables (such as growth rates of micro-organisms as discussed in Marsden, Pingry and Whinston) may be present and require adjustments in studying the process.

2.) This division of input variables is used here for purposes of comparison with Griffin's approach. Many alternatives are available. A detailed example of the application of generalized production theory is presented by the authors in Marsden, Pingry and Whinston where divison of the input variables is based on variability of total cost with respect to each inputs utilization.

3.) For a detailed example which explicitly considers such constraints, see Marsden, Pingry and Whinston.

4.) The approach outlined here is applicable to multi-product cases using vector representations. This problem is largely ignored here because of the chosen direction of this note.

References

Afriat, S.N. "The Output Limit Function in General Convex Programming and Theory of Production." Econometrica 39 (May 1971): 309-39(a).

Afriat, S. N. "Efficiency Estimation of Production Functions." Paper presented at the August, 1971, meetings of the Econometric Society (b).

Graves, G. W., D. E. Pingry and A. Whinston, "Application of a Large Scale Nonlinear Programming Algorithm to Pollution Control." Proceedings Fall Joint Computer Conference, 1971.

Griffin, J. M. "The Process Analysis Alternative to Statistical Cost Functions: An Application to Petroleum Refining." American Economic Review LXII (March 1972); 46-56.

Hanoch, G., and M. Rothschild, "Testing the Assumptions of Production Theory: A Nonparametric Approach." Journal of Political Economy 80 (Part II May, June 1972): 256-275.

Marsden, J. R., D. E. Pingry, and A. Whinston, "Production Function Theory and the Optimal Design of Waste Treatment Facilities." Applied Economics (forthcoming)

Smith, K., M. Miller and F. L. Goladay: "An Analysis of the Optimal Use of Inputs in the Production of Medical Services." Journal of Human Resources, VII (Spring 1972): 208-225.

Chap. 7 Mass Balance and Economic Models

Mass Balance and Economic Models

I. Introduction

Several recent articles in production theory, planning models and environmental economics have focused attention on the importance of understanding the physical relationships, and in particular on mass balance. (See Ayres [1], Ayres and Kneese [2], Noll and Trijonis [7], Marsden, Pingry and Whinston [6]). The basic position of these papers appears to be that in order to predict the effects of a particular policy we must know the underlying mass balance conditions.

In this paper we discuss mass balance in more general terms as a procedure to construct models in all disciplines, including economics. Section II will examine this more generalized approach to mass balance. The following sections will present examples from these disciplines. The relation between the generalized mass balance approach and two traditional problem forms will be presented in Section IV followed by a discussion of the implications in Section V.

II. General Mass Balance

Mass balance is in the limit not a sophisticated scientific theory but simply a method of accounting. For example, the models of Leontief [5] and Ayres and Kneese [2] simply account for all the flows between sectors of an economic unit during a certain period of time. In these linear, steady state models it is assumed that the economy consists of a series of sectors which if given a unit of input convert it into output at some predetermined proportion. The proportions are estimated from observed set of inputs and outputs. That is, the reaction trans-

formation in the sector is assumed to be linear and the parameters are estimated by examination of a particular set or sets of accounts. There are two problems of interest. One is the simple accounting for the materials which enter and leave the reactor and the other is the model of the reaction rate, production coefficient or transformation function. In general, the mass balance equation for material m_1 around any unit reactor, using the terminology of chemical engineering, will be of the form of equation (1).

$$\Delta m_1 = m_1(\text{INPUT}) - m_1(\text{OUTPUT}) + m_1(\text{SOURCE}) - m_1(\text{SINK}) \qquad (1)$$

Equation (1) is a tautology. The relation in equation (1) is of no particular interest to the theorist. It holds for any reactor, such as a closed economy, a firm, a chemical reactor or a river segment. The interesting aspect of this equation, and the key to the predictive power of a model, is the specification of the sources and sinks of the material in the reactor of interest. Since conceptually the sources and sinks for material m_1 can be a dynamic nonlinear function of the levels of the other materials the model can be extremely complex.

At this stage of generality the model explains everything and, therefore, nothing. The interesting questions are:

1) Is there a set of systems which are applicable to all disciplines or are disciplines characterized by certain reactor types?
2) Under what conditions do traditional modeling techniques correspond to the generalized mass balance approach?

The generality of the mass balance approach to modeling can be illustrated by examining three examples. The first two examples will

come from chemical engineering and sanitary engineering respectively, and the last example from economic growth theory, an area in economics not usually associated with mass balance models.

The first example taken from Marsden, Pingry and Whinston [6] is a chemical reactor of length Δx and cross sectional area A. The reactor is assumed to be producing material m_1. The rate at which m_1 is being produced is a function of the concentration of m_1 and some other material m_2. The mass balance conditions around the reactor are stated in equations (2) and (3).

$$\overbrace{V\Delta C_1}^{\text{Input}} = \overbrace{Q_1 C_1 \wedge t - Q\left(C_1 + \frac{\partial C_1}{\partial x}\wedge x\right)\wedge t}^{\text{Output}} + \overbrace{K_1 C_1 C_2 V \wedge t}^{\alpha_1\alpha_2 \text{ Source}} \qquad (2)$$

$$\overbrace{V\Delta C_2}^{\text{Input}} = \overbrace{QC_2\Delta t - Q\left(C_2 + \frac{\partial C_2}{\partial x}\Delta x\right)\Delta t}^{\text{Output}} \qquad (3)$$

The notation is defined as follows:

C_1 = Concentration of material #1 $\left(\frac{\text{Mass}}{\text{Volume}}\right)$

C_2 = Concentration of material #2 $\left(\frac{\text{Mass}}{\text{Volume}}\right)$

K_1 = Rate Constant $\left(\frac{1}{\text{Time}}\right)$

Q = Flow entering and leaving reactor $\left(\frac{\text{Volume}}{\text{Time}}\right)$

U = Velocity of Flow $\left(\frac{\text{Length}}{\text{Time}}\right)$ $U = Q/A$

V = Volume of Reactor (Volume) $V = A\Delta x$

x = Length (Length)

Equations (2) and (3) can be transformed into equations (4) and (5) by multiplying by $V\Delta t$ and letting the $\wedge$ terms go to zero.

$$\frac{\partial C_1}{\partial t} = - U\frac{\partial C_1}{\partial x} + K_1 C_1^{\alpha_1} C_2^{\alpha_2} \tag{4}$$

$$\frac{\partial C_2}{\partial t} = - U\frac{\partial C_2}{\partial x} \tag{5}$$

If we assume the steady state $\left(\text{e.g., } \frac{\partial C_1}{\partial t} = \frac{\partial C_2}{\partial t} = 0\right)$ and the initial conditions of $C_1 = C_{10}$ and $C_2 = C_{20}$ at $x = 0$, then we can solve (4) and (5) for C_1 and C_2.

$$C_1 = \left[\frac{K_1 x}{U} (1-\alpha_1) C_{20}^{\alpha_2} + C_{10}^{(1-\alpha_1)} \right]^{1/(1-\alpha_1)} \tag{6}$$

$$C_2 = C_{20} \tag{7}$$

Note that function (6) is of CES form, assuming that $1-\alpha_2 = \alpha_1$, and is derived from engineering assumptions.

The second example comes from biology. Assume that we want to predict the divergence of the dissolved oxygen concentration from the oxygen saturations level increasing as a function of the demand for the oxygen caused by the injection of unoxidized waste. This problem can also be placed in a mass balance framework. The reactor of interest is a river segment of width Δx and cross sectional area A. The mass balance conditions are

$$V\Delta D = QD\Delta t - Q\left(D + \frac{\partial D}{\partial x}\Delta x\right)\Delta t + (K_1 B - K_2 D)V\Delta t \tag{8}$$

$$V\Delta B = QB\Delta t - Q\left(B + \frac{\partial B}{\partial x}\Delta x\right)\Delta t - K_1 BV\Delta t \tag{9}$$

Where

D = Dissolved oxygen deficit $\left(\frac{\text{mass}}{\text{Volume}}\right)$

B = Biochemical oxygen demand $\left(\frac{\text{mass}}{\text{Volume}}\right)$

The solution to this model assuming again the steady state and the initial conditions of $D = D_o$ and $B = B_o$ at $x = 0$ is

$$D = \frac{K_1 B_o}{K_2 - K_1}\left[e^{-K_1 x/U} - e^{-K_2 x/U}\right] + D_o e^{-K_2 x/U} \qquad (10)$$

$$B = B_o e^{-K_1 x/U} \qquad (11)$$

The third example model is the growth model presented by Solow [8]. The model was originally presented in the form

$$Y_t = C_t + I_t$$

$$C_t = (1-s)Y_t$$

$$I_t = \frac{dK_t}{dt} + \varphi K$$

$$Y_t = F(L_t, K_t)$$

$$L_t = L_o e^{nt}$$

The model could be formulated in mass balance terms assuming the reactor is a closed economy of length calendar time Δx.

The mass balance conditions around the reactor for capital and labor are:

$$V\Delta K = \overbrace{QK\Delta t}^{\text{Input}} - \overbrace{Q\left(K + \frac{\partial K}{\partial x}\Delta x\right)\Delta t}^{\text{Output}} + \overbrace{sF(K,L)V\Delta t}^{\text{Source}} - \overbrace{\varphi K V\Delta t}^{\text{Sink}} \qquad (12)$$

$$V\Delta L = \overbrace{QL\Delta t}^{\text{Input}} - \overbrace{Q\left(L + \frac{\partial L}{\partial x}\Delta x\right)\Delta t}^{\text{Output}} + \overbrace{nL}^{\text{Source}} \qquad (13)$$

Where

K = Capital

L = Labor

s = Marginal property to save

φ = Depreciation rate

Assuming the steady state, the initial conditions of $K = K_o$ and $L = L_o$ at $x = 0$, and a Cobb-Douglas production function of the form $Y = \beta K^{\alpha} L^{1-\alpha}$ the solution for K and L yields

$$K = \left[\frac{s\beta L_o^{1-\alpha}}{(\alpha-1)(\varphi+n)}\left[e^{(1-\alpha)nx/U} - e^{(\alpha-1)\varphi x/U}\right] + K_o^{(1-\alpha)} e^{(\alpha-1)\varphi x/U}\right]^{1/1-\alpha} \quad (14)$$

$$L = L_o e^{nx/U} \quad (15)$$

Since the velocity is the time it takes to travel one calendar time unit it is equal to one.

The similarities of the formulation of the three models are striking and can be seen by examining Table 1. Table 1 is constructed by writing the sources and sinks of the models in the notation of the first example.

Table 1: Model Reaction Forms

Model	Material	Form
Chemical	C_1	$K_1 C_1^{\alpha_1} C_2^{\alpha_2}$
Reactor	C_2	0
River	$D = C_1$	$K_2 C_2 - K_1 C_1$
Quality	$B = C_2$	$-K_2 C_2$
Economic	$K = C_1$	$K_1 C_1^{\alpha} C_2^{1-\alpha} - K_3 C_1$
Growth	$L = C_2$	$K_2 C_2$

In fact, each model form from one of the disciplines could easily be interpreted in terms of the other two. For example, let C_1 and C_2 be the capital stock and labor supply and interpret the reactor as a closed economy. Model #1 can be interpreted as an economy with a generalized Cobb-Douglas production, constant labor force and no capital depreciation. Model #2 is an economy with the production of capital linear in labor, a depreciation of capital at the rate K_1 and labor supply decreasing at rate K_2, or if $K_2 < 0$ as an increasing labor force.

As is illustrated above the models in three different areas as formulated under the mass balance approach are very much the same. One can also think of a large number of models from other areas of study which would be similar. For example inventory models from business management and learning models from educational psychology. In fact, one can conceive of models from all areas of study being constructed around a particular reactor or set of reactors. These models may be static, dynamic, steady-state, non-steady state, linear or nonlinear. Models of any of these characteristics could be written in general mass balance form.

III. Generalized Mass Balance and Two Traditional Problems

A) Mass Balance and Optimization.

In the previous sections we have seen how problems from three diverse areas can be approached as mass balance problems. One type of general problem is conspicuously missing from the examples. Optimization models or mathematical programming models are very important in the current literature of most disciplines. Is the mass balance approach ~~constraint~~ consistent with these models?

The determination of the reactor of interest for a particular discipline or problem area will place bounds or constraints on the researcher. The specification of the reactor determines the constraint set. Usually, however, the research will not simply have the goal of modeling a particular reactor, but of providing some information for policy making about the trade-offs involved in producing some output or restricting some input. That is, some criterion function is defined over the outputs, inputs or both. The reactor model is simply the constraint set for this criterion function

Consider the examples given above. For the chemical reactor one is not likely to be interested in the reactor itself but in some criterion such as maximizing net benefit where benefit is defined over the output and losses are a function of the inputs. The water quality model can be used to determine the least damaging effluent discharge patterns. The growth model could be used to pick an optimum growth path according to the collective preferences of the members of the economic reactor.

B) Mass Balance and Aggregation

One particular area of interest in economics has been the aggregation problem. (See F. M. Fisher [3].) For example, under what conditions can aggregate production function be used to represent an economy? This aggregation problem is simply stated in the reactor approach as, under what conditions can a set of reactors be represented by a smaller set of reactors. For example, the models presented in W. Fisher [4] are exactly along this line They try to examine the condition under which the costs of aggregation are minimized. Or in other words, given input and output accounts from a large number of reactors, what is the optimal number of reactors which can be used to represent this large amount of

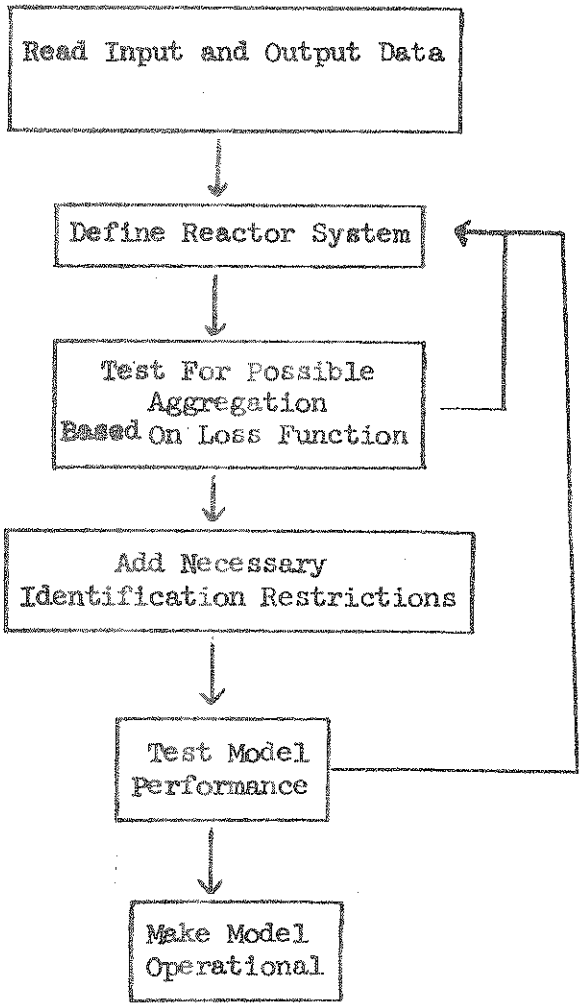
detailed data. This is particularly clear in the input-output models which are linear-steady-state reactors. The purpose of aggregation in this context as presented in W. Fisher [4], is to combine industries into single classifications without losing information. It might be of interest to examine the typical unit reactors and their sequencing for the industries which are classified together. One might find they share the same general mass balance form.

V. Implications for Future Research

The isomorphic nature of all discipline under the mass balance approach leads one to some fascinating implications. Under the current approach each science builds its own reactor models, many of which overlap as illustrated. However, if these models are isomorphic under the mass balance formulation one would expect to easily adopt these models across disciplines. This leads one to further speculate that a computer language (MASBAL) could be written to store these models in mass balance form to be adopted by the individual research to his particular reactor. The critical point is that although the particular models may not be isomorphic in their specific form, they are isomorphic in the more general mass balance framework. Information could be read in concerning the reactor and observed inputs and outputs. Specific reactor models of varying complexity could then be generated by MASBAL appropriate to the input and output data available. Information concerning the identification of the mass balance models could also be stored. If the model is over identified, additional restrictions could be suggested.

A flow diagram of the generalized mass balance language is illustrated in Figure 1.

Figure 1: Flow Chart of MASBAL



The advantages of such a system would be numerous.

1. Disciplines could be clearly defined based on the specification of their reactor.
2. Models could easily be adopted across disciplines since their forms would be easily compared.
3. Interdisciplinery work would require learning only the new reactors and not the new models allowing easy flow by researchers among areas (i.e., reactors).
4. An area of model building and testing could be well defined so that certain individuals would concentrate on augmenting the model bank with models of "nice" properties.
5. Studying a particular discipline could be clearly divided between the study of MASBAL and defining the appropriate reactors of interest.

The benefit from the construction of such a system would appear to be very great.

References

[1] Ayres, Robert U., "A Materials-Process-Product Model," Environmental Quality Analysis, Allen V. Kneese and Blair T. Bower (eds.), Resources for the Future, the John Hopkins Press, Baltimore 1972.

[2] Ayres, Robert U. and Allen V. Kneese, "Production, Consumption and Externalities," American Economic Review, June 1969.

[3] Fisher, Franklin M., "The Existance of Aggregate Production Functions," Econometrica, Vol. 37, No. 4, October 1969.

[4] Fisher, Walter D., Clustering and Aggregation in Economics, The John Hopkins Press, Baltimore 1969.

[5] Leontief, W., "Environmental Repercussions and the Economic Structure: An Input-Output Approach," Review of Economics and Statistics, August 1970.

[6] Marsden, James R., David E. Pingry and Andrew Whinston, "An Alternative Approach for Deriving Production and Cost Functions," Proceedings of the Sixth International Conference on Systems Sciences, Honolulu, Hawaii, January 1973.

[7] Noll, Roger G. and John Trijonis, "Mass Balance, General Equilibrium, and Environmental Externalities," American Economic Review, September 1971.

[8] Solow, R. M., "A Contribution to the Theory of Growth," Quarterly Journal of Economics, 1956.