ROBUST TECHNIQUES FOR DEVELOPING EMPIRICAL MODELS
OF FLUIDIZED BED COMBUSTORS

J. Gruhl
S.E. Tung
F.C. Schweppe

ROBUST TECHNIQUES FOR DEVELOPING EMPIRICAL MODELS
OF FLUIDIZED BED COMBUSTORS

J. Gruhl
S.E. Tung
F.C. Schweppe

Acknowledgment and Notice

None of these organizations, nor any person acting on behalf of these organizations: (a) makes any warranty or representations, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or (b) assumes any liabilities with respect to the use of any information, apparatus, method, or process described in this report.

1

Abstract

This report is designed to provide a review of those data analysis techniques that are most useful for fitting m-dimensional empirical surfaces to very large sets of data. One issue explored is the improvement of data (1) using estimates of the relative size of measurement errors and (2) using known or assumed theoretical relationships. An apparently new concept is developed, named <u>robust weighting</u>, which facilitates the incorporation of <u>a priori</u> knowledge, based upon the values of input and response variables, about the relative quality of different experiments. This is a particularly useful technique for obtaining statistical inferences from the most relevant portions of the data base, such as concentrating on important ranges of variables or extrapolating off the leading edge of the frontier of knowledge for an emerging technology. The robust weightings are also useful for forcing <u>a priori</u> known asymptotic behaviors, as well as for fighting biases due to shear size of conflicting data clusters, and for formulating separate models for conflicting clusters. Another new development has evolved from the two very different objectives of the empirical modeling in this project. The first objective is the usual requirement for the best possible predictive mechanism, and standard techniques are useful with their emphasis on model building, specifically the successive separation of trend techniques. In addition, a second objective involves the pursuit of high-dimensional, yet simple, models that could provide insight into analytic gaps and scientific theories that might govern the situation. For this second objective a new stepwise process was developed for rapidly sweeping the data base and producing crude quantitative measures of the next (or the first) most important m-tuple relationship to incorporate into the empirical model. These quantitative

2

guidelines have been named the fit improvement factors. Some of the standard statistical techniques reviewed include: graphical displays, resistant models, smoothing processes, nonlinear and nonparametric regressions, stopping rules, and spline functions for model hypothesis; and robust estimators and data splitting are reviewed as polishing and validating procedures. The concepts of setting, depth and scope of the validation process are described along with an array of about sixty techniques for validating models. Actual data from the recent literature about the performance of fluidized bed combustors is used as an example of some of the methods presented. Also included is a bibliography of more than 150 references on empirical model development and validation.

TABLE OF CONTENTS

Appendixes available
upon special request

I. INTRODUCTION AND CONCLUSIONS

A common problem in statistical analysis is the development of empirical models that fit a set of data. There is more involved in this problem than what may be considered an issue of surface fitting, in fact, empirical modeling is a four-step process. Step one, described in Chapter 2, The Data, encompasses the tasks of organizing the data, improving the data on the basis of known or suspected relationships and magnitudes of measurement errors, and focusing precisely on the crucial areas of the data base, using robust weighting. Step two involves the postulation of a model form and the rough fitting of model parameters and is discussed in Chapter 3, Model Building. The next step, Chapter 4, Model Calibration, requires choosing a measure of the quality of fit of the model, such as with a robust estimator, hypothesis checking on the model structure, and polishing of the model coefficients. Finally, Chapter 5, Model Validation, is a discussion of various types and depths of validation and the creation of probabilistic measures of the predictive quality of the empirical model. An illustrative example is displayed in Chapter 6, showing the use of these techniques on a emerging energy technology about which there is a large amount of data. Chapter 7 is a discussion of the various uses of empirical models.

1.1 Objectives

The goal of this research is to develop and demonstrate a methodology that will be useful in representing the current state of knowledge by fitting an m-dimensional, nonlinear, empirical model to a very large set of information of uneven quality, varying completeness, and various degrees of relevance to the current situation. Behind this immediate objective are a number of indirect objectives which have underlying motivating factors that

are related to the specific topic of this study, namely the area of

exploring advanced energy technology data. Some of these other objectives

include desires for:

1. a modeling mechanism that is fast and flexible enough to be operated many times, accepting new assumptions and criteria with relative ease;

2. the possibility of an automated, or at least a fairly systematic, update procedure as new batches of data become available;

3. achieving the right balance between theoretical presumptions and data implications, or optionally to develop models solely on experimental data so discrepancies between theory and observations can be defined for further research;

4. comparable measures of expected values and uncertainties associated with the economic, efficiency and environmental performance of several advanced energy sources;

5. the quantification of the risk that a particular technology or design will not be able to meet threshold pollutant emission standards, and related to this, a determination of what would be the strictest standards that could reasonably be met;

6. the operating parameters for the next batch of experiments that would have the greatest payoff in terms of reducing the most critical uncertainties; and

7. the design and operating parameters for the next pilot plant or commercial facility that would show the best performance.

1.2 Problem Formulation

As mentioned previously, there are several steps into which this

problem logically divides, and the first of these concerns the data itself.

The data consists of measurements of several different system performances

y_j, j = 1,2...,J, which will be called _response variables_. In some of the

cases investigated here, it is possible to look separately at each

individual response and in these cases there will be no subscipt on y. The

other types of observations included in the data base are the _input_

variables, x_i, which may include design parameters, operating conditions,

the facility used, and even the year and performer of the experiment.

Although the example displayed in Chapter 6 uses only 14 different input
variables, in follow-on applications there may be as many as 200 different
categories of input variables (Louis, Tung, 1977). Such excessive
dimensionality of the data vectors is a recurring problem in multivariate
analysis (Hawkins, 1976). Experimenters, often properly, tend to measure
as many variables as possible so as to avoid omitting any that may later
prove to be relevant. The inclusion of irrelevant variables, however,
markedly reduces the effectiveness of most data analysis techniques,
including multiple regression (Allen, 1973). Furthermore, if
"capitalization on chance" effects are to be avoided (Finney, 1974) the
sample sizes must be extremely large to accommodate high dimensionality of
data vectors. For the example in Chapter 6, the number of experiments, N,
is 369, and this would have to increase sharply to accommodate explorations
of 200-dimensional data. Aggrevating the need for large sample sizes are
the additional "capitalization on chance" effects that are caused by the
sparsity of the data matrix. It would indeed be a hopeless problem if
there were not a number of theories that have helped guide the experiments
and that can help guide the data analysis. In any event, the definition
can be made that

$$\text{data base} = \{(y_{jn}, x_{in}) \ j=1,2,\ldots,J; \ i=1,2,\ldots,I; \ n=1,2\ldots,N\} \quad (1.1)$$

The other concern with the data is that it is of very uneven quality.
Experiments performed in the 1950's, for example, may have little to
contribute to the current state of knowledge. A priori weightings have
been assumed and are used to highlight the most important ranges of the
x_i's and y. These robust weightings are combined in several different
ways, but essentially result in fractional numbers of experiments, that is,
instead of all experiments being counted as "one" observation, some count
only as one-half, or some other fraction, of an observation.

Given the robustly weighted observations, the model building involves a stepwise process leading to the hypothesis of a simple model with structure F(.) and parametric coefficients $\underline{p}$ such that

$$y_n = F(x_{in}, \underline{p}) + e_n, \qquad\qquad n = 1,2,\ldots,N \qquad\qquad (1.2)$$

where e_n is the error, or residual, associated with the prediction of the response of experiment n. F(.) is a point in the I-dimensional functional space, that is, the space of all functions of I variables. The vector $\underline{p}$ is the set of coefficients, or constants, in the hypothesized model. Previous scientific laws or intuitions about the structure of F(.) can be described as limiting F(.) and $\underline{p}$ to a subspace (Fo, Po).

1.3 Empirical Modeling

The model development process used in this research is aimed at the goal of attempting to postulate actual cause-effect relationships, as suggested in (Mandel, 1969, p.428), not just statistical inferences or high quality predictions. The laws of nature tend to be simple, and if the right variables are available, those laws rarely contain more than 4 or 5 variables in 2 or 3 separate terms. Thus, if 10 or 15 linear relationships can be replaced by one 4-variable nonlinear term, then the single term is greatly prefered for this type of expository empirical modeling.

With "exploration" thus being much more important than "building" a stepwise process was developed that advances by determining at each iteration which new term of m-tuple variables $(x_i, x_j,\ldots,x_r)$ is likely to most improve the empirical model if added at that point. A crude measure, called the fit improvement factor, FIF, is developed to direct the modeler toward these high priority areas. The scheme for developing the FIF $(i,j,\ldots,k)$ begins somewhat like the Hinich-Talwar procedure (Hinich, Talwar, 1975). The $(x_i, x_j,\ldots,x_k)$ space is divided into a number of

nonoverlapping cells. Statistics are developed for each cell which show

the means, or medians, which when viewed together as a series of step

functions, form a crude <u>cellular</u> <u>model</u>. The deviations, or H-spacings (see

Appendix C), of the data in each of these cells can be combined to act as a

lack-of-fit measure for this model. Dividing this lack-of-fit statistic

into the lack-of-fit statistic without this model yields a 1-to-∞ measure

of the potential improvement $FIF(i,j,\ldots,k)$ that results from using an

$(x_i, x_j,\ldots,x_k)$ term in the empirical model.

The calibration of the model proceeds via a robust error penalty

function, which is used to fine-tune the coefficients $\underline{p}$. For a robust

function such as the minimum absolute deviation criterion, this calibration

requires determining a $\underline{p}_{opt}$ with the characteristic that

$$\underline{p}_{opt} = \min_{\text{all } \underline{p} \epsilon P_o} \{ \sum_{n=1}^{N} |y_n - F(\underline{x}_n ,\underline{p})| \} \quad \text{given } F(.)\epsilon F_o \qquad (1.3)$$

Minimum absolute deviation has received the most acclaim as a generally

useful one-step nonquadratic error penalty function. A two-step process

that is particularly promising for this particular application begins with

the least squares $\underline{p}_{opt}$, that is

$$\underline{p}_{opt} = \min_{\text{all } \underline{p} \epsilon P_o} \{ \sum_{n=1}^{N} [y_n - F(\underline{x}_n ,\underline{p})]^2 \} \quad \text{given } F(.)\epsilon F_o \qquad (1.4)$$

Starting with the least squares opt a new search is begun with a

quadratic-constant error function (Handschin, Schweppe, Kohlas, Fiechter,

1975) that strictly limits the influence of the outlying data. A

quadratic-constant function by itself generally exhibits instability in the

convergence procedure, which is overcome with the initial least squares step. This two-step procedure is more appropriate for this application where the bad data tends to cluster causing multi-modal distributions of the residuals.

The validation procedure operates with respect to a specific, fixed model structure and a specific set of coefficients. To a great extent, the validity of a well-built regression model will depend upon the source of the data. Designed data, that is, developed from an experiment designed to produce parameteric data for the model (Snee, 1977), will generally result in a close fitting model, but it is difficult to predict that model's performance on a new batch of data. Historical data, on the other hand, is uncontrolled, with large sparse areas in the variable space, and is thus difficult to model. These models are, however, then relatively good predictors of new data. The empirical data from the emerging energy technologies lies between the extremes of designed and historical data. For this type of data, the four validation techniques recently presented in (Snee, 1977) are expanded to more than sixty techniques in Chapter 5. Also included are comments about the possible variations in scope and depth of the validation procedure that have resulted from research at the M.I.T. Model Assessment Laboratory.

The illustrative example presented in Chapter 6 is the beginning of an extensive empirical modeling of fluidized bed combustors that is on-going within the M.I.T. Energy Laboratory. As such, the results that have been displayed should be viewed as means of exercising some of the methodologies discussed, and at most may be considered first attempts at empirical models. For additional factual information about fluidized bed combustors (Gruhl, Teare, 1978) is a recent source that may be of interest.

Chapter 7 contains a short discussion on the uses of empirical modeling and Chapter 8 is an important bibliography of literature related to this research.

II. THE DATA

As mentioned previously, the data base is comprised of several response variables, y_j, several input variables x_i, and both sets of variables are measured for a number of experiments N:

$$\text{data base} = \{(y_{jn}, x_{in}); j=1,2,\ldots,J; i=1,2,\ldots,I; n=1,2,\ldots,N\} \qquad (2.1)$$

It is possible that for the majority of datum positions in this (J+I) times N <u>data</u> <u>matrix</u> there will not be any measurements available and this lack of information will be assumed recorded, such as with an "NA", "-", or "-1". An example of a data matrix is displayed in Appendix A; in this example J=6, I=14, and N=369 which thus creates a 20 by 369 data matrix.

2.1 Data Improvement

Without any additional information, of course, it is not possible to enhance the quality of the data. There are, fortunately, a surprising variety of types of information that can be useful for data base improvement. One of these involves the use of estimates of the relative size of the measurement errors that were likely to have been made while observing the various categories of variables. If $\underline{y}_n$ and $\underline{x}_n$ are the measured response and input vectors and $\hat{\underline{y}}_n$ and $\hat{\underline{x}}_n$ are the actual (unobservable) values, then $\underline{u}_n$ and $\underline{v}_n$ can be defined as the measurement errors with:

$$\underline{y}_n = \hat{\underline{y}}_n + \underline{u}_n, \qquad (2.2)$$

$$\underline{x}_n = \hat{\underline{x}}_n + \underline{v}_n. \qquad (2.3)$$

While it is not possible to precisely reconstruct the values of the individual measurement errors, $\underline{u}_n$ and $\underline{v}_n$, it is often possible to make well-founded speculations about the statistics of these errors. For example, if temperatures appear only as multiples of $50^{\circ}F$, or carbon monoxide is known to be particularly difficult to measure, then estimates

of the standard deviation of measurement errors for those variables is to some extent possible. On the other hand, cross-sectional areas of combustors, coal sources, and a number of other variables may, for practical purposes, be presumed to have been measured quite exactly.

For purposes of mathematical formulation, suppose the variances of the measurement errors can be estimated:

$$E(u_j^2) = \frac{1}{N} \sum_{n=1}^{N} u_j^2 \equiv \sigma_j^2 \qquad (2.4)$$

$$E(v_j^2) = \frac{1}{N} \sum_{n=1}^{N} v_j^2 \equiv \rho_j^2 \qquad (2.5)$$

These σ's and ρ's can now be useful in dividing up the responsibilities for discrepancies among the measurements. The size of such discrepancies can be determined by checking the known interrelationships among the various variables. These _a priori_ known relationships, previously defined as the structural and parametric set (Fo, Po) could include:

1. physical laws, such as mass or energy balances;

2. presumed scientific theories, such as thermodynamic theories or analytic hypotheses;

3. intuitive insight;

4. speculations resulting from dimensional analysis; and

5. other plausibility checks, possibly including consistency checks where all influencing factors are presumed know.

13

If discrepancies are uncovered within the data base, and the measurement errors are assumed to tend toward Gaussian distributions, then the optimal relegation of those discrepancies among the potentially contributing measurement errors is accomplished by distributing the discrepancies in proportion to the variances of the measurement errors.

These genuine improvements in the quality of the data base can be performed as an initial step in each separate modeling session. It is advisable, however, to carry out any improvements using indisputable interrelationships, such as mass balances, as a matter of course before entering new batches of data into the archival copy of the data base. Speculative interrelationships should not, of course, be allowed to alter the data base archives (See Figure 2-1).

2.2 Robust Weightings

As can be seen in Figure 2-1, there is one more task involved in making the data base ready for modeling, and this is the addition of the robust weightings. Data is generally known a priori to be of widely varying quality, and this is particularly true concerning the data available about emerging technologies. There is generally a progressive improvement in the importance of this kind of historical data: designs mature, sizes increase, appropriate operating ranges come into sharper focus, and so on. Thus, one would like to build a model that could extrapolate off the leading edge of the frontier of knowledge. The statistician, or rather the model builder, should be made aware of which are the more important regions of the data base.

For these maturing systems, the experimenters can generally convey information on the ranges of input variables that are still under active consideration for the eventual (perhaps commercial) design. Likewise,

14
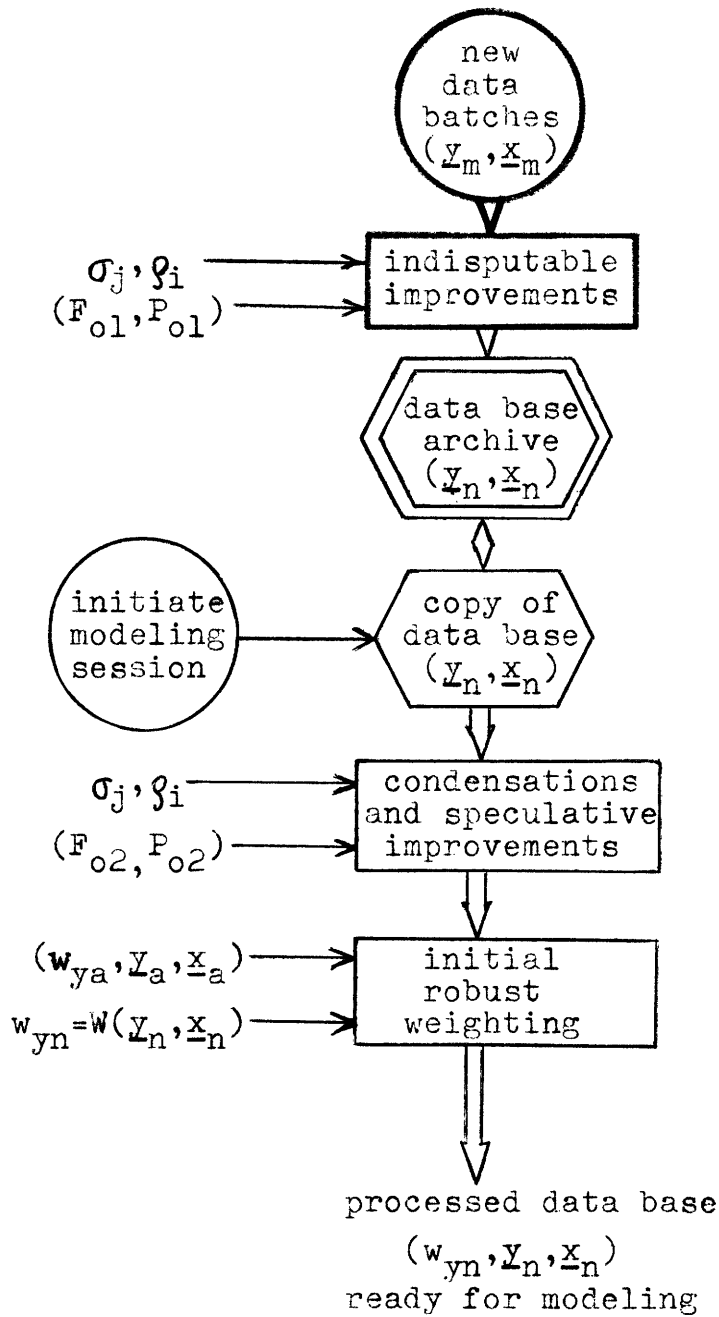
Figure 2-1 Flowchart of data base updating and processing procedures

response variables such as pollutant emissions can be expected to settle out at, or below, the threshold emission standards. The weighting functions that indicate whatever is known or supposed a priori about the relative value of information are here called robust weightings, and can generally be written as functions of the input and response variables for the nth experiment and will usually be different depending upon the response variables being modeled:

$$w_{yn} = w(\underline{y}_n, \underline{x}_n).$$ (2.6)

Now, instead of each experiment having a weight of one, experiments would (usually) be tagged with a zero-to-one factor w_{yn} that indicates whether the experiment is not to be considered, $w_{yn} = 0$, is to be fully considered, $w_{yn} = 1$, or fractionally considered, $0 < w_{yn} < 1$. It appears that the most appropriate manner for generating the weighting functions would be through a composite of several zero-to-one weightings that were functions of single variables, $ws_{yj}(y_j)$ or $wr_{yi}(x_i)$. Some examples of possible separable weighting functions are shown in Figure 2.2.

The manner in which the separate weightings are combined to form a single function for the relative weighting of individual experiments is a matter for some judgment. Conceivable composites of the separate weights could include (the first type is the one used in the illustrative example in Chapter 6):

$$w_{yn} = w(\underline{y}_n, \underline{x}_n) = \prod_{j=1}^{J} ws_{yj}(y_{jn}) \prod_{i=1}^{I} wr_{yi}(x_{in})$$ (2.7)

$$w_{yn} = \min_{\text{all } i,j} [ws_{yj}(y_{jn}), wr_{yi}(x_{in})]$$ (2.8)

$$w_{yn} = \frac{1}{k} [\sum_{j=1}^{J} ws_{yj}{}^{m}(y_{jn}) + \sum_{i=1}^{I} wr_{yi}{}^{m}(x_{in})] \text{ for any } m,k$$ (2.9)
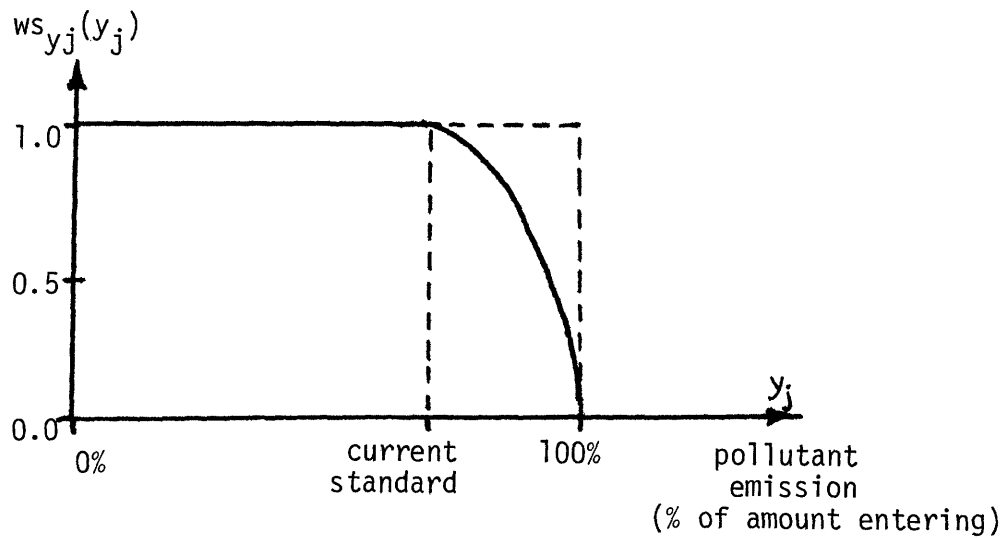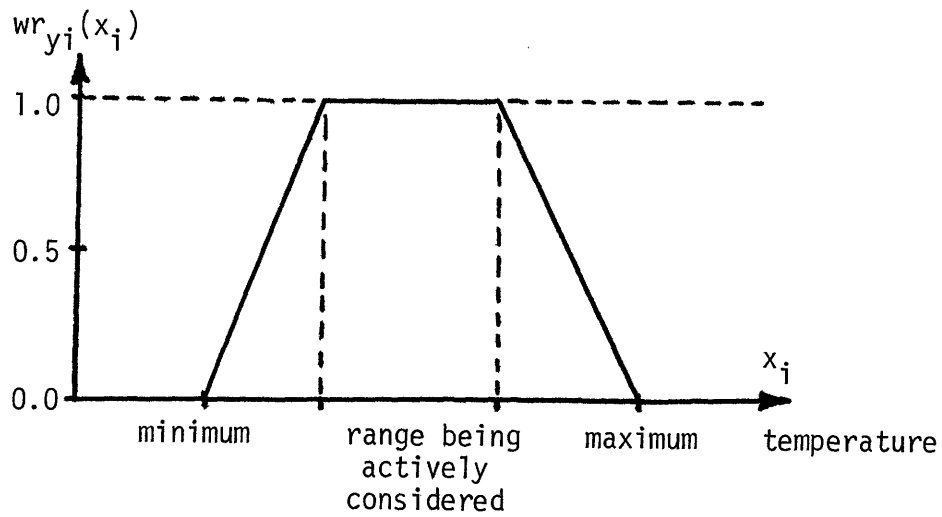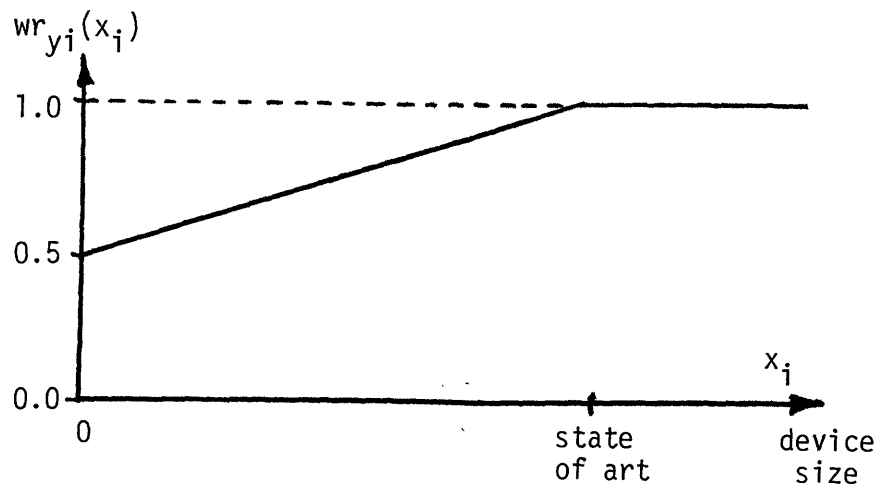
Figure 2-2 Various examples of a priori specified robust weighting
functions for fluidized bed modeling

$$w_{yn} = [\prod_{j=1}^{J} ws_{yj}^{m}(y_{jn}) \prod_{i=1}^{I} wr_{yi}^{m}(x_{in})]^{1/k} \text{ for any } m,k \qquad (2.10)$$

with arithmetic and geometric means being important special cases of types (2.9) and (2.10) (where $m=1$, $k=j+1$). In any of these cases, if the separate weighting functions are chosen as combinations of binary step functions, the entire weighting procedure degenerates to the standard technique of hatcheting out experiments where any of the variables is outside its range to be considered. Aside from automating the standard hatchet technique the robust weights allow for a full range of degrees of importance of experiments.

Aside from the issues concerning the development of these weightings, there is another notable issue regarding the choice of which portions of the modeling procedure these weights should be used in. The easiest tactic would be to initially develop the weight of each experiment, w_{yn}, and carry those weightings right through the entire modeling process. Again, this would be analogous to the standard data hatcheting technique. It is conceivable, however, that if one of the y_j's or one of the x_i's is not taking part in the modeling that it should not exert influence on the data to be studied. It is also conceivable that individual weightings might best be temporarily relaxed to get a fuller display of the spectrum of data. For example, with robust weightings forcing the modeling to focus on designs that are near or better than the SO_2 emission standard for fluidized bed data, a density scatterplot shown in Figure 2-3 results, and perhaps suggests a linear model. Temporarily recalling the data that was weighted out, as shown in Figure 2-4, may lead the modeler to an exponential model. Experience with the particular data base in Appendix A

```
                                          Ca/S Mole Ratio →
          -1. 0.     0.6   1.2   1.8   ?.4   3.0   3.6   4.?   4.8   5.4   6.0
          +-++-+--+--+-+--+--+-+--!-+--+-+--+--+--!-+--+--+-+--+--+-!--+--+--+
     100.+                          ...         ......            ......+100.  ↑
          !                          ...         ......            ......!
          +...                            ......      ...          ███      +
S         !...                            ......      ..           ███      !     S
0    92.+                         ...000       ...      ...         0U0      + 92. 0
2         !                         ...000       ...      ...       000      !     2
          +                       +++   ...+++   +++   ++++++...   ...███     +
R         !                       +++   ...+++   +++   +++++++...  ...███     !     R
e    84.+                    ......███000++++++    ...          ......     + 84. e
m         !                    ......██S000++++++    ...          ......     !     m
o         +              ...    +++000██████00U                    +        o
v         !              ...    +++000██████000                    !        v
a    76.+              +++000300██████000                  ......     + 76. a
l         !              +++000C30████████000                  ......     !     1
          +       ......   000+++++++++++000                    +
%         !       ......   000+++++++++++000                    !        %
     68.+         ...000000...+++...    ...                     + 68.
          !         ...000000...+++...    ...                     !
          +         ......+++...+++...000             ...          +
          !         ......+++...+++...000             ...          !
     60.-         ...+++...    ...                     + 60.
          !         ...+++...    ...                     !
          +     ......+++       ...                     +
          !     ......+++       ...                     !
     52.          .........   ...   .......            + 52.
          !         .........   ...   .......            !
          +                                             +
          !                                             !
     44.                                              + 44.
          !                                             !
          +                                             +
          !                                             !
     36                                               + 36.
          !                                             !
          +                                             +
          !                                             !
     28                                               + 28.
          !                                             !
          +                                             +
          !                                             !
     2C                                               + 20.
          !                                             !
          +                                             +
          !                                             !
     1:                                               + 12.
          !                                             !
          +                                             +
          !                                             !
     4.+                                              + 4.
          !                                             !
     0.+                                               + 0.
          !                                             !
          +-++-+--+--+-+--+--+-+--+--+-+--+--+-+--+--+-+--+--+-+--+--+-+--+--+
          -1. 0.     0.6   1.2   1.8   2.4   3.0   3.6   4.2   4.8   5.4   6.0
                                          Ca/S Mole Ratio →

     0-.5=       .5-1.5=...   1.5-2.5=+++   2.5-3.5=000   3.5-4.5=000   >4.5=███
                        ...           +++           000           000        ███
```
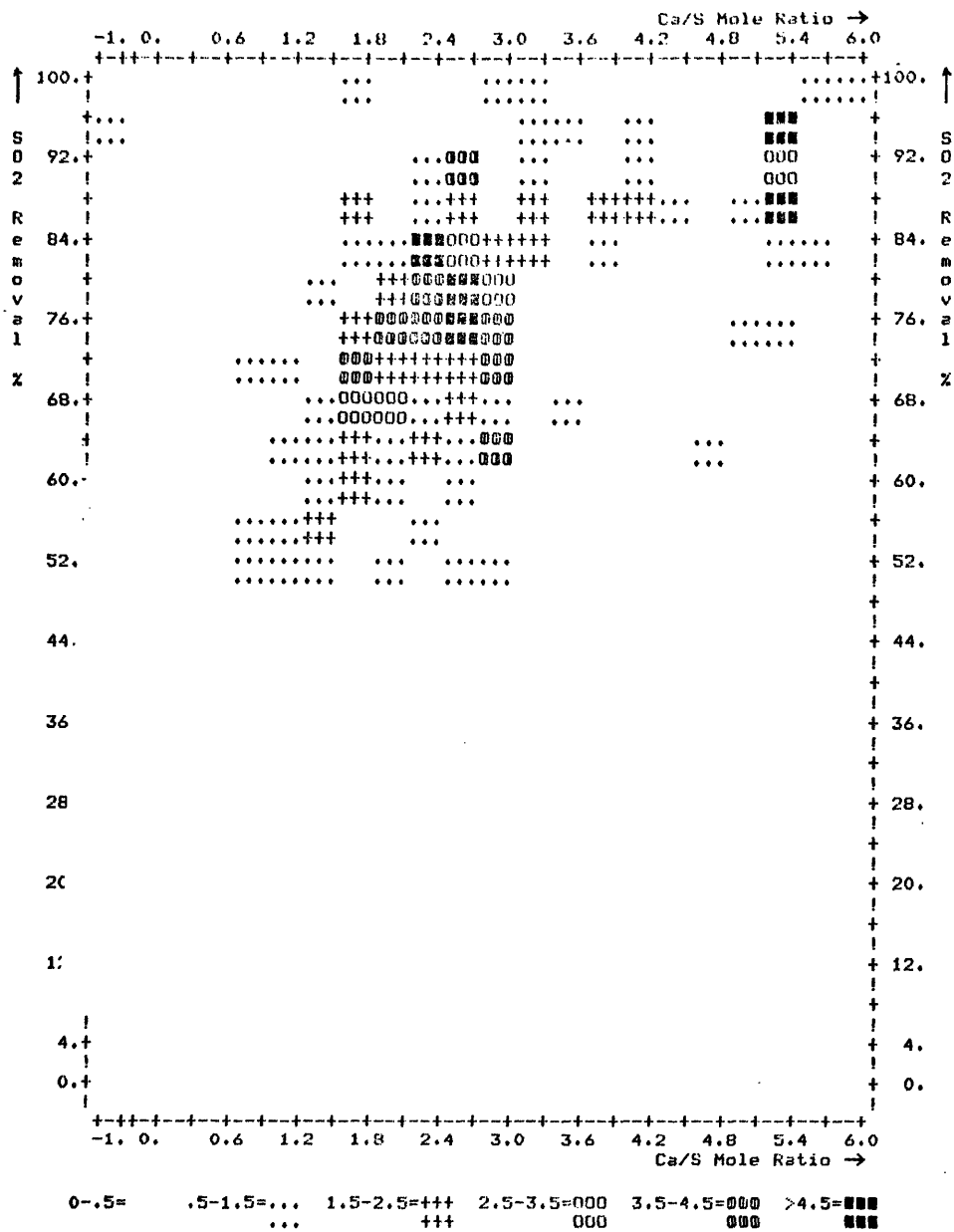
Figure 2-3 A density scatterplot of robustly weighted SO_2 emission data

from fluidized bed combustors
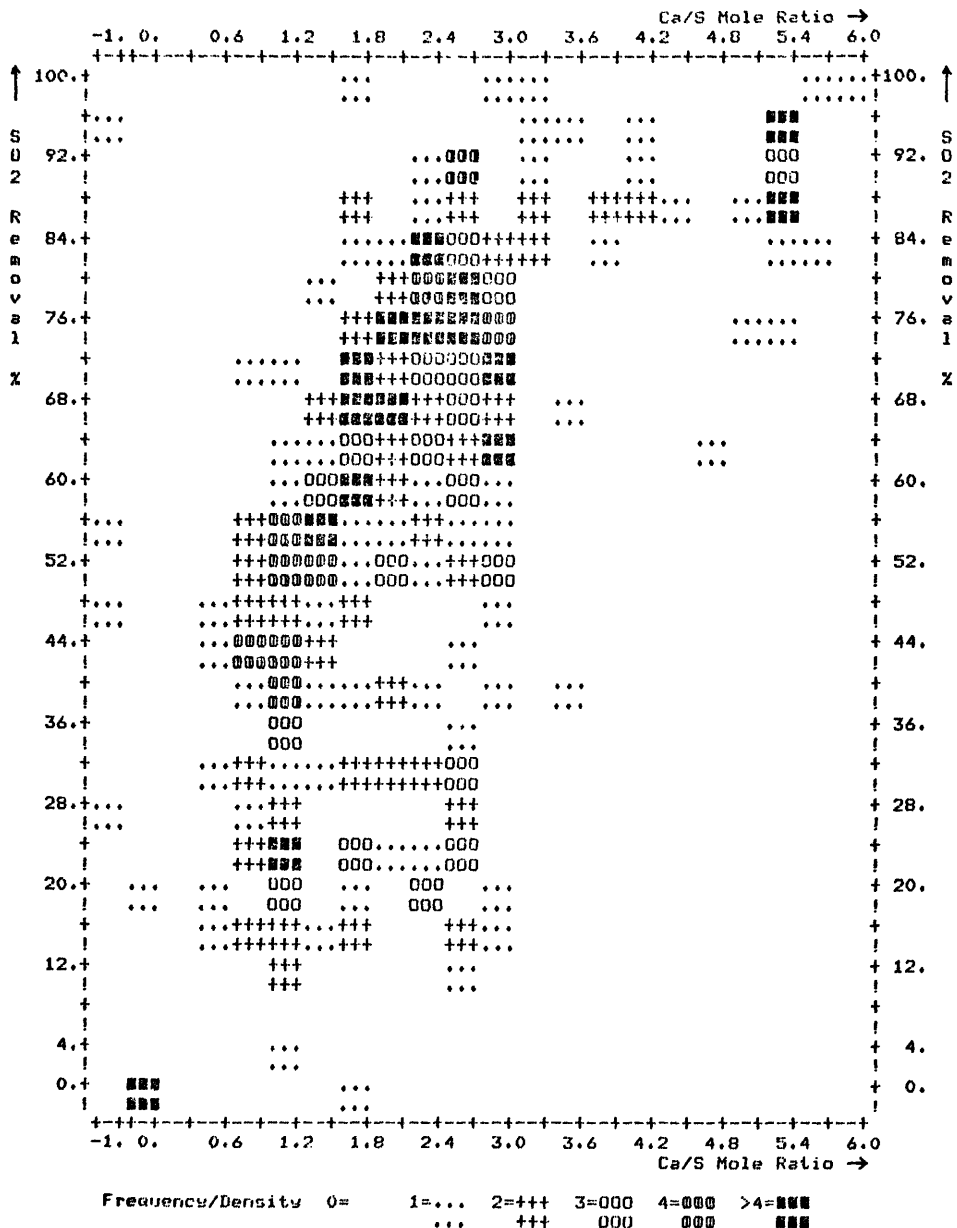
Figure 2-4   A density scatterplot of all fluidized bed SO$_2$ emission data

suggests that in the model building and validating procedures the robustly removed data should be differentiated but displayed. In the model calibration process it was helpful to leave in all weightings. Every case is, however, likely to be different and in the absence of contrary information, consistency would appear to be the best policy.

Two other utilizations of the robust weighting technique have proved useful in examples. New data can be generated that reflect known asymptotic behaviors, $(\underline{y}_a, \underline{x}_a)$, and it is then quite easy to force the modeling and calibration to incorporate these data by assigning to these points very large robust weights, such as $w_{ya}=1000$. Also, the robust weighting concept was very effectual in temporarily disregarding certain clusters of data during the process of developing models for conflicting data clusters. This particular application required the use of nonseparable weighting functions and is described further in section 3.3.

III. MODEL BUILDING

The free-form model hypothesis described in this chapter involves a graphically aided, quite systematic, investigation of the general functional types that are most compatible with the data. The standard model building strategies in the literature involve gradually particularizing the model (Mandel, 1969) by adding new structural components until the residuals have been reduced to a specific size or distribution. The method presented here is principally _exploratory_ with building occuring only as a last step. Preconceived models, such as linear or polynomial, are not used, instead the internal structure of the data is explored to uncover possible cause-effect relationships by searching for the simplest model that will do justice to the data. Keeping down the number of variables in the model facilitates interaction with analytic modeling research, and may even increase the accuracy of the model (Das Gupta, Perlman, 1974).

3.1 Mathematical Formulation

Reviewing the previous definitions:

n = 1,2,...,N label of observation or experiment number;

y_n = a specific response y measured in experiment n;

x_{in} = the ith input variable measured in experiment n;

$\underline{x}_n$ = the vector with x_{in} as entries;

$\underline{p}$ = the vector of coefficients or constants that represent the magnitude of the functional forms;

$F(.)$ = a hypothetical model relating the x_i to y_j and

e_n = the error or residual in the prediction of the nth experiment due to unobservable experimental or measurement errors; so

$$y_n = F(\underline{x}_n, \underline{p}) + e_n, \quad n = 1, 2, \ldots, N. \tag{3.1}$$

In addition, a set of robust weights, w_{yn}, are available as quantifications of the importance or believability of the various experiments from the point of view of the yth response. The variances of the measurement errors on y and x_i it will be recalled were defined in equations (2.4) and (2.5).

As formulated this problem is not in a classic linear regression category, as nonlinear forms are definitely to be considered. Nonlinear regression (Gallant, 1975) and group methods of data handling, GMDH, algorithms (Ivakhmenko, Todua, Fomicheu, 1974) require prespecified functional forms. In addition the GMDH algorithms and the best subset methods of multivariate analysis of variance, MANOVA, (Hawkins, 1976), are severely hampered by very large dimensionality. Multiple regression (Allen, 1974), stepwise regression (Efroymson, 1960) and Mahalanobis distance techniques (Rao, 1965, p.482), as well as the transformation methods such as principal component analysis, factor analysis, and canonical variate analysis are all hampered by excessive dimensionality, particularly the inclusion of irrelevant variables, distributional problems, and "capitalization on chance" effects (Hawkins, 1976). Reduction of dimensionality can be accomplished using the technique of interdependence analysis (Beale, Kendall, Mann, 1967), but this involves elimination of highly correlated variables and this elimination may be undesirable for this application. For example, combustor cross-sectional area may be eliminated because it correlates closely with the group of experimenters. One method geared for handling high dimensionality (Hawkins, 1976) involves a stepwise procedure for testing and eliminating irrelevant variables but apparently also requires an explicit assumption of a model functional form. The non-parametric estimation techniques (Clark,

1977), (Benedetti, 1977), and spline function methods do not require prespecified functional forms (except perhaps the family of convoluted forms of chosen kernels), yet have some difficulty with high dimensionality and are definitely not conducive to the development of _simple_ functional models.

To solve the problem as formulated, a method is described whose motivation and algorithmic aspects are allied closely to data sweeping techniques. The procedure is approximate, in that it is not guaranteed that a globally optimal model will be found. Its advantages are computational simplicity, ease of modeler interaction, and applicability to large batches of data of high dimensionality.

3.2 Fit Improvement Factors

The process described here is a method of quantifying the probable importance to the empirical model of various m-tuple relationships. This is accomplished by building very crude models, sometimes called _pseudo-models_, and comparing the tightness of fit with and without the pseudo-model.

The first step in this procedure is much like the Hinich-Talwar method (Hinich, Talwar, 1975); given the particular one, two-, or m-tuple relationship to be investigated, the range of the data in that m-space is divided into k nonoverlapping subsamples, or cells. A cellular pseudo-model, or just _cellular model_, is then developed from the statistics of the samples in each cell. Some of the possible cellular models could be based upon:

1. ordinary means;

2. robust means, that is using the weights w_n to count the weight of each sample;

3. partially robust means, here using the robust weightings only for the variables being examined;

4. alpha trimmed means;

5. ordinary medians;

6. robust or partially robust medians;

7. ordinary least squares estimates using relative measurement errors;

8. non-quadratic criteria estimates; or

9. robust estimates based upon median statistics, such as the Andrews' estimator (Andrews, 1974).

The final step in the development of the cellular model involves the possibility of smoothing the cell statistics. Keeping in mind the automation and rapidity that is desired, the blurring, or eye smoothing, and complex estimation techniques must be ruled out. Two powerful techniques that appear appropriate and useful are:

1. running methods, that is, either taking medians of overlapping cells, or more commonly taking the n-at-a-time median of the statistics of all neighboring cells; and

2. hanning methods, either the means of overlapping cells; or more commonly, the average of the statistics of neighboring cells (Tukey, 1977).

These techniques are often used repeatedly or alternately. Figure 3-1 shows some median rehanning and rerunning median-of-three statistics.

Once this cellular model has been developed it is a relatively easy task to measure the lack of fit with and without this model. The _fit improvement factor_ for making (or improving) a model using the x_i, x_j,...,x_k m-tuple is then:

FIF(i,j,...,k) = (lack of fit without model)/(lack of fit with

model) (3.2)

The FIF is then a statistic that is greater than one by the extent of importance of the m-tuple as a modeling component. Some of the obvious lack of fit measures include [most are available on TROLL (NBER, 1974)]:
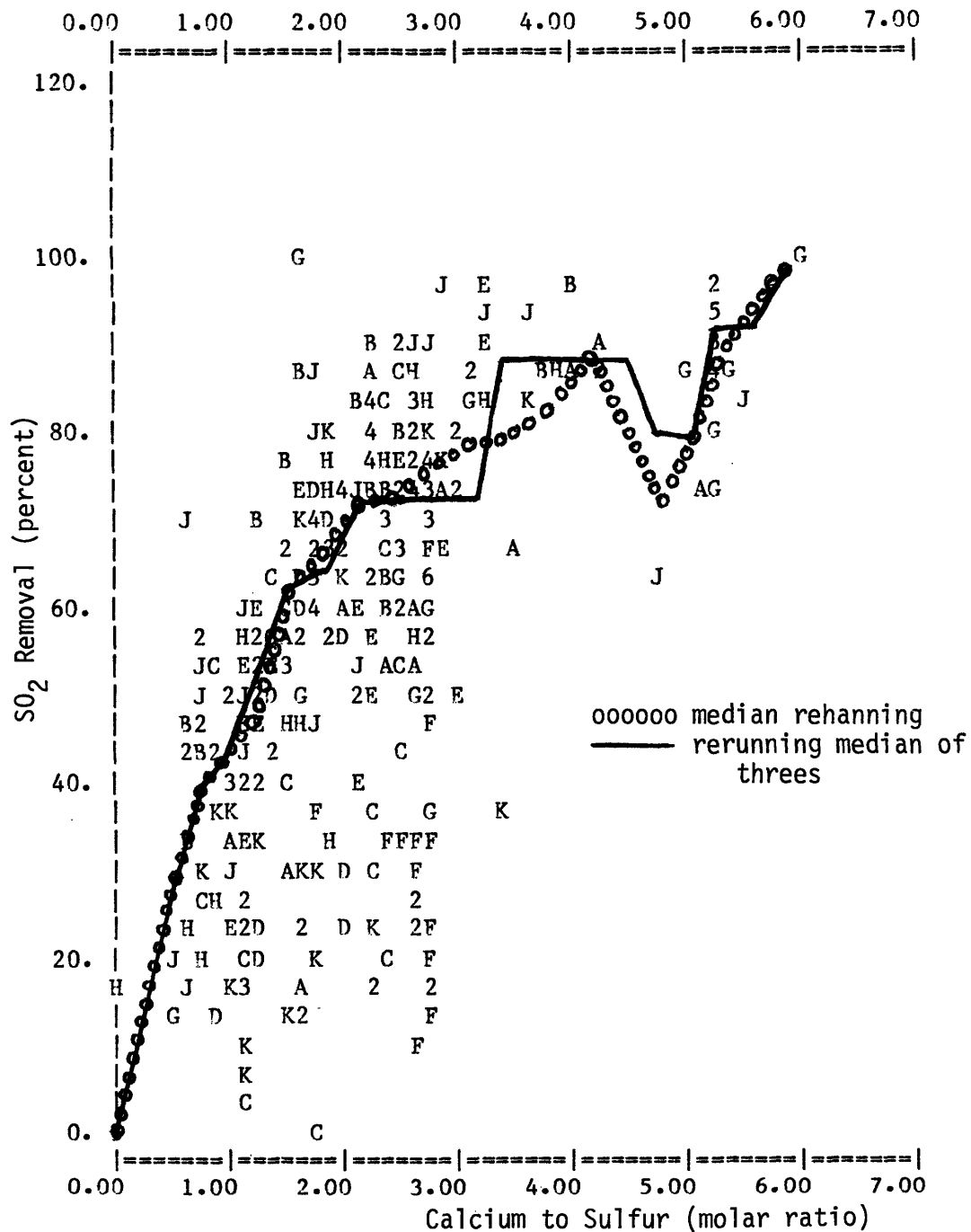
Figure 3-1 SO$_2$ removal versus calcium to sulfur ratio statistics are here smoothed using two very easy and common smoothing techniques, note the susceptibility of both techniques to chance points in sparse regions

26

1. ordinary variances or standard deviations;

2. robust or partially robust variances or standard deviations;

3. H-spreads or steps, see Appendix C;

4. robust or partially robust H-spreads or steps;

5. F-statistics or t-statistics;

6. squared partial correlation coefficients (Bendel, Afifi, 1977), and

7. the C_p statistic (Mallow, 1973).

Although the median statistics are more resistant to the effects of data of uneven quality, the mean statistics are easier to compute and more conveniently adapted to multivariate problems.

An issue that was not addressed earlier, but that is of paramount importance, is the choice of the number and shape of those nonoverlapping cells for the cellular model. It would seem obvious that the cells either have equal dimensions or contain equal numbers of samples. These two techniques are somehow related to the statistical philosophies of the mean, for the former, and the median, for he latter; but this need not constrain the choice. Equal cell dimensions is the scheme used in the examples of this report.

The number of cells to be used still remains an unresolved issue. There does not appear to be much statistical literature related to this topic. In grouping together normal data into cells for frequency distributions, the classical suggestion is that 10 to 20 occupied cells provide adequate detail for most purposes. A suggestion of Hinich and Talwar is that the number of cells be set equal to N/m where N is the number of observations and m the number of input variables x_i being studied. This formula, obviously, is intended for a very different application and would not work well here. There appear to be just a few

considerations that should be recognized in selecting the appropriate number of cells:

1. for discrete variables, it would be desirable to have cells for each discrete point, or equal numbers of discrete points in each cell;

2. if the cellular model is to be displayed to the modeler, convenience should be considered, that is, display of hundreds or thousands of occupied cells would be expensive and relatively indigestible;

3. smoothness should be a consideration, for example if all data points had their own cell the fit would look perfect but hardly be indicative of the fit of a smooth model; and finally,

4. detail must be included, that is, enough cells to give the cellular model a chance to pick out important fluctuations in the m-tuple interrelationships that may be present.

Some of the pieces of information that are available for making the decision about number of cells are:

1. the number of observations, N,

2. the number of x_i's being studied, m,

3. the sparsity of the data as meassured by the fraction of unoccupied cells (given a trial number and configuration of cells),

4. the smoothness of the cellular model, measured by, for example, the mean or median of the residuals (again given a trial number of cells).

This is a very difficult issue and one that is not pretended to be fully treated or resolved here. One additional complexity that should be mentioned is that there is obviously no reason why all dimensions x_i should be divided into the same number of intervals.

To close this discussion, a suggestion for the choice of the number of cells k, in the form of a strategy for splitting up the ranges of the x_i, is given. If N is the number (robust or ordinary) of observations, and the relationship to be studied is the m-tuple $(x_i, x_j, \ldots, x_k)$ with respect to y, then the range of each of the $x_i, x_j, \ldots, x_k$ is broken up into h parts,

28

where

$$h = 1 + \text{IFIX}(N^{1/m+1}) \tag{3.3}$$

where IFIX is a function that extracts the integer position of its

argument. The number of cells k is thus

$$k = h^m \tag{3.4}$$

Table 3-1 shows an example of the manner in which these formulas are

carried out.

Table 3-1. Example of the Choice of Number of Cells for

N=369 Observations

Number of Input Variables $x_i = m$	Number of Intervals in Range of Each $x_i = j$	Total Number of Cells = k
1	20	20
2	8	64
3	5	125
4	4	256
5	3	243
6	3	729

Just one comment on this example: there is little that can be gained

by divisions of ranges of x_i into only h=3 intervals (suggested for the

5-tuple relationship for 369 observations). It is felt, however, that this

limitation, which results from the small number of observations, is

reasonable so as to avoid entrapments of "capitalization on chance".

Having dealt with the difficult task of number of cells, it is now

appropriate to mention some potential fine-tunings of the lack-of-fit

measures used in FIF(i,j,...,k). If an unsmoothed cellular model is used

there is some compensation that can be made to avoid apparent good fits

that will show up due to large numbers of sparsely populated cells. For a

normal distribution the standard error of the mean that is due to lack of a large number of samples is:

$$\sigma_M = \sigma / \sqrt{n} \qquad (3.5)$$

where σ is the standard deviation of the ideal (infinite sample) distribution, and n is the number of samples. It would be possible to penalize the inappropriately good fit of cells with few samples by this $1/\sqrt{n}$ measure.

An example of just such a penalty factor has been tested on a set of data. The standard deviation computed for a particular cell was penalized by adding directly to the standard deviation a term of that standard deviation times $1/\sqrt{n}$. This is obviously a crude measure, and counts on the cell standard deviation underestimating the ideal deviation by about the same amount as the direct addition of deviations overstates the square root of sum of squared deviations. The exact formula for this FIF example involved using only those cells with more than three entries (or a collected robust weight greater than three):

$$FIF = \frac{SD_t + \dfrac{SD_t}{\sqrt{NW_t}}}{\dfrac{1}{NW_t} \sum_{\substack{all\ cells \\ with\ NW_i > 3}} [NW_i (SD_i + \dfrac{SD_i}{\sqrt{NW_i}})]} \qquad (3.7)$$

where FIF = fit improvement factor,

SD_t = standard deviation of total sample,

SD_i = standard deviation of cell i,

NW_i = number of observations in cell i, or total robust weight in cell i, and

30

$$NW_t = \sum_{\substack{\text{all} \\ \text{cells} \\ \text{with} \\ NW_i > 3}} NW_i \qquad\qquad (3.8)$$

Figure 3-2 shows how this particular FIF worked in practice. As can be seen the majority of variables do not show potential for usefulness for the modeling. Some variables show marginal usefulness, and two of the variables show very strong potential for one-dimensional models. (These two variables in the example are superficial velocity and bed area, with the response variable being dust loading in lb/hour. Complete details about this example are shown in Chapter 6 and Appendix B.)

In this type of interactive data exploration the next step is to display FIF's for two-at-a-time sets of input variables versus the response variable. If there are I different input variables, then there are Ix(I-1) two-at-a-time relationships. This may or may not be a manageable number of FIF's to display. Looking ahead, however, to the number of all possible sextuplets, namely I!/(I-6)!, it is clear that some kind of reduction of this number of FIF's will be required. What is proposed here is a guided search, and the case of I=20 is used as an example. It would not be impossible to consider the 20x19=380 different FIF's for all possible pairs of input variables. Examining 20x19x18=6840 triplets, however, would seem unreasonable, so at the time of examining all the FIF's of all pairs, those pairs must be separated into three possible categories:

1. those variables (or m-tuple relationships) that definitely will not play a part in the eventual empirical model;

2. those variables (or m-tuples) that definitely will play an eventual role; and

31

$$FIF = \frac{SD_t + \dfrac{SD_t}{\sqrt{NW_t}}}{\dfrac{1}{NW_t} \displaystyle\sum_{\substack{\text{all cells} \\ \text{with} \\ NW_i > 3}} \left[NW_i \left(SD_i + \dfrac{SD_i}{\sqrt{NW_i}} \right) \right]}$$
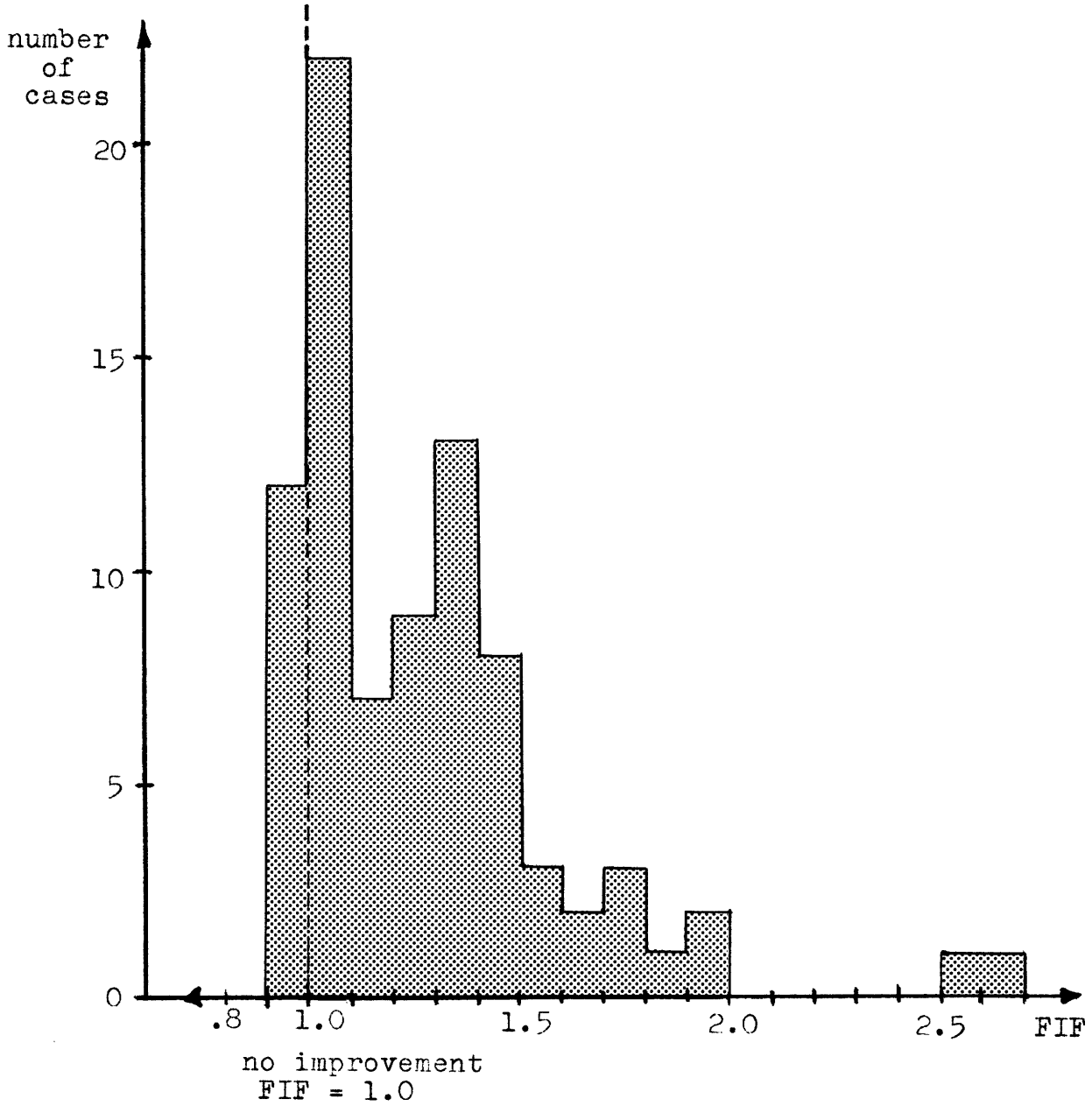
Figure 3-2   Fit Improvement Factors for cellular models for a
number of different variables

3. those variables (or m-tuples) that cannot definitely be ruled in or out of the model at this time.

(Some more complex categories, such as multiple choice from sets, mutual exclusion, and so on are also possible.) These decisions can be made on the basis of the FIF's, displayed statistics, displayed scatterplots, intuition, analytic information, or other grounds. Back to the example of I=20, if just one variable can be _excluded_ from the model, then there are about 15% fewer triplets to consider; or if just one variable x_i must be _included_, the 6840 FIF's drop to 342; or if two variables are required to be included, this reduces the number of triplets to 18. It is easy to compute at each stage of the process how many decisions must be made in order to keep the size of the problem tractable. In general, if I equals the number of variables x_i, p the number of variables that necessarily must be included, and q the number excluded, then the number of different m-tuples, N_m, is

$$N_m = \frac{(I-p-q)!}{(I-m-q)!}$$ (3.9)

This process can be viewed as a guided search toward the most important m-tuple relation(s), and is principally an exploratory scheme. It is equally possible to use these fit improvement measures in successive separation and modeling of trends. For example, all one-at-a-time relationships judged to be significant could be modeled. Then the residuals could be examined for further trends in one-, and new trends in two-, at-a-time data, and these can be added to the model. This process can proceed iteratively until a desired quality-of-fit level is reached. In this same vein, although it is difficult to envision a fully-automated modeling strategy that is largely exploratory, it is quite conceivable that

the FIF's could be used in a fully-automated successive separation and modeling of trends. In linear and canonical regression, measures of fit such as the standard F test (Draper, Smith, 1966) or Hotelling's T^2 test of additional information of an extra variable (Rao, 1965) and (Subrahmaniam, 1971) have been used in automated procedures. What would be required to automate the FIF procedure would be an FIF threshold that would identify a level of significance for including a new hypothesized m-tuple of variables. Similar thresholds, such as using C_p statistics (Mallows, 1973), have worked well in stepwise regression techniques, (Christensen, 1973), (Bendel, Afifi, 1977), (Draper, et al., 1971), and (Hawkins, 1976). Some branch-and-bound techniques, possibly applicable here, have been developed (Furnival, Wilson, 1974) that could be used to guide the search for important relationships.

3.3 Robust Weightings

A number of uses of robust weightings have been presented previously. There are two final uses that can arise during an interactive model building session. One use is to handle systematic biases or inaccuracies correlated with some variable that may be identified by examining the residuals during the modeling effort. Less emphasis could then be put on those experiments by using robust weightings. For example, if x_i is a discrete variable signifying the group of experimenters that performed the observations, biases on a particular output y due to apparent inaccuracies or biases due to shear numbers of observations performed by that particular group, can be eliminated by altering the robust weights w_{yn} in the data base (w_{yn}, $\underline{y}_n$, $\underline{x}_n$). This technique is somewhat similar to standard robust estimation methods.

34

Another possible application occurs when the cellular models demonstrate definite bi-modal or multi-modal behavior. In these cases, as mentioned briefly at the end of Chapter 2, functions $g_1(\underline{y},\underline{x})$, $g_2(\underline{y},\underline{x})$,..., $g_i(\underline{y},\underline{x})$ can be formulated to yield values of "one" in the region of the data cluster i and "zero" elsewhere. Thus i separate data bases, corresponding to the i clusters, are then readily available for the separate exploration:

$$\text{data base of cluster i} = (w_{yn}\, g_i\, (\underline{y}_n, \underline{x}_n), \underline{y}_n, \underline{x}_n) \tag{3.6}$$

3.4 Surface Fitting

It is not actually possible to separate the development of the fit improvement factors from the development of the model; some iteration and rechecking are advantageous. Likewise, it is not possible to separate the model calibration from the model development, fine-tuning of parameters along the way can help expose new structural possibilities. To show how the surface fitting stage thus assumes a central role in empirical modeling a flow chart of the process is shown in Figure 3-3. This figure begins where Figure 2-1 stops, with the improved and weighted data base, and includes several steps described in the preceding sections.

In the terminology previously presented, the general area of statistics known as surface fitting involves estimating the model structural form F(.) and the inital trial parameters $\underline{p}$. It should be recalled that formerly discussed conditions restrict those estimates to the subspace (F_o, P_o). At this point there is a particularly strong distinction between models that are to be essentially predictive and those also intended to be expository, that is, those intended to offer analytic insight. One possibility for building predictive empirical models is to draw their structure from many terms in a polynomial, trigonometric,

processed data base

$(w_{yn}, \underline{y}_n, \underline{x}_n)$

ready for modeling

$F_1 = 0$ or
initial
model idea

begin with
m = 1

m=m-1

determine
number and
size of cells

m=m+1,
guide search
to important
m-tuples

no

yes

back up
to improve
model ?

calculate
Fit Improvement
Factors on
residuals

eliminate or
definitely
include some
variables

FIF 'importance'
threshold as
function of m,
or number of
displays
desired

display statistics
of residuals for
important m-tuple
relationships

further
modeling
?

not at this point

yes

decisions on
functional forms
to use, or
automated
curve-fitting

fit curves to
data and add to
existing model

calibrate
?

not at this point

decision to or
not to include
robust weights

yes

robust error
penalty
function

gradient
search for $\underline{p}$
optimum parameters

performance
measure
threshold

quality
check

still unsatisfactory

satisfactory

empirical model

Figure 3-3 Block diagram outlining model development strategy

36

logarithmic, or other functional selection lists. Alternately, predictive models can use surfaces fitted automatically by nonparametric regressions or spline functions which can be structurally quite complex yet make no assumptions except those concerning the bandwidth, or smoothness, of the functions; these are the so-called "fully automated French curve techniques" (Benedetti, 1977), (Clark, 1977), (Gallant, Fuller, 1973), (Greville, 1969), (Priestley, Chao, 1972), (Reinsch, 1971), (Woodford, 1970), and particularly (Wahba, Wold, 1975).

On the other hand, expository empirical modeling should, whenever possible, be performed interactively as it relies heavily upon matters of good judgment and experience. It appears appropriate thus to avoid the use of those previously described sophisticated, structural identification computer algorithms for this energy technology problem which does not have unfathomable dynamics and which does not have the pressures of real-time, instantaneous modeling requirements. Clues for the choice of surfaces for the expository structure should come from examination of the input-to-response statistics and from graphical displays. An enormous number of summary statistics and graphs, intended to foster perception and appreciation of broad features as well as details, are available from any number of statistical computer packages, see (NBER, 1974) for example. Some of the more useful displays include density scatterplots, frequency scatterplots (see Figure 3-4), triple scatterplots where symbols or numbers show a third variable on the scatterplot, and rotating data cloud programs (NBER, 1974) that make possible the display of four variables. Graphical displays are easily obtained and their importance to the model building process cannot be overstated.

```
                                              Ca/S Mole Ratio →
        -1. 0.    0.6   1.2   1.8   2.4   3.0   3.6   4.2   4.8   5.4   6.0
        +-++-+--+--+--+--+-+--+--+--+--+--+--+--+--+--+--+--+--+--+-+--+--+
↑  100.+ .  .  .  .  .  .  . 1 .  .  . 1 1 .  .  .  .  .  .  . 1 1 +100. ↑
    !                                                                 !
   +  1  .  .  .  .  .  .  .  .  .  . 1 1 . 1 .  .  . 7 .  . +       S
S     !                                                             O
O 92.+ .  .  .  .  .  .  . 1 4 . 1 .  . 1 .  .  . 3 .  . + 92.      2
2     !                                                             
   +  .  .  .  .  .  .  . 2 . 1 2 . 2 . 2 2 1 . 1 5 .  . +          R
R     !                                                             e
e 84.+ .  .  .  .  .  .  . 1 1 7 3 2 2 . 1 .  .  . 1 1 . + 84.      m
m     !                                                             o
o  +  .  .  .  .  .  . 1 . 2 4 7 3 .  .  .  .  .  .  .  . +         v
v     !                                                             a
a 76.+ .  .  .  .  .  . 2 5 5 6 4 .  .  .  .  . 1 1 .  . + 76.      l
l     !                                                             
   +  .  .  . 1 1 . 5 2 3 3 5 .  .  .  .  .  .  .  .  . +           %
%     !                                                             
 68.+ .  .  .  .  . 2 4 5 2 3 2 . 1 .  .  .  .  .  .  . + 68.
    !                                                             
   +  .  .  .  .  . 1 1 3 2 3 2 7 .  .  .  .  . 1 .  .  . +
    !                                                             
 60.+ .  .  .  .  . 1 3 5 2 1 3 1 .  .  .  .  .  .  .  . + 60.
    !                                                             
   +  1  .  .  . 2 4 5 1 1 2 1 1 .  .  .  .  .  .  .  . +
    !                                                             
 52.+ .  .  .  . 2 4 4 1 3 1 2 3 .  .  .  .  .  .  .  . + 52.
    !                                                             
   +  1  .  . 1 2 2 1 2 .  .  . 1 .  .  .  .  .  .  . +
    !                                                             
 44.+ .  .  . 1 4 4 2 .  .  . 1 .  .  .  .  .  .  .  . + 44.
    !                                                             
   +  .  .  . 1 4 1 1 2 1 . 1 . 1 .  .  .  .  .  . +
    !                                                             
 36.+ .  .  .  . 3 .  .  .  . 1 .  .  .  .  .  .  .  . 36.
    !                                                             
   +  .  .  . 1 2 1 1 2 2 2 3 .  .  .  .  .  .  .  . 
    !                                                             
 28.+ 1  .  .  . 1 2 .  .  .  . 2 .  .  .  .  .  .  . 28.
    !                                                             
   +  .  .  . 2 5 . 3 1 1 3 .  .  .  .  .  .  .  . 
    !                                                             
 20.+ . 1 . 1 . 3 . 1 . 3 . 1 .  .  .  .  .  .  . 20.
    !                                                             
   +  .  .  . 1 2 2 1 2 .  . 2 1 .  .  .  .  .  . 
    !                                                             
 12.+ .  .  .  . 2 .  .  .  . 1 .  .  .  .  .  .  . 12.
    !                                                             
   +  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . 
    !                                                             
  4.+ .  .  .  . 1 .  .  .  .  .  .  .  .  .  .  . 4.
    !                                                             
  0.+ . 53 .  .  .  . 1 .  .  .  .  .  .  .  .  . 0.
    !                                                             
    +-++-+--+--+--+-+--+--+--+--+--+--+--+--+--+--+--+--+-+--+--+
        -1. 0.    0.6   1.2   1.8   2.4   3.0   3.6   4.2   4.8   5.4   6.0
                                              Ca/S Mole Ratio →
```
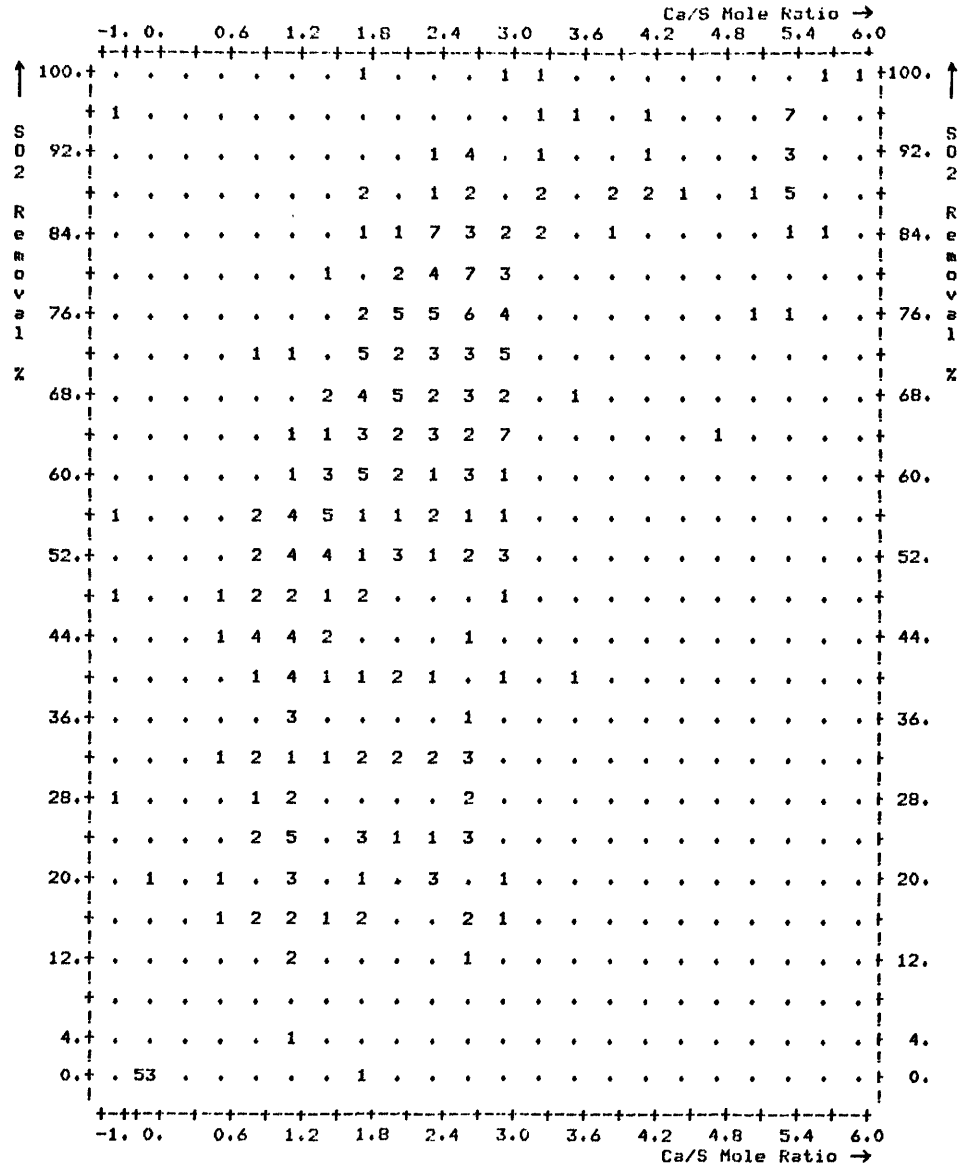
Figure 3-4   A frequency scatterplot of SO_2 removal in a fluidized bed

as a function of calcium to sulfur ratio

It is, unfortunately, not easy to enlarge the experience of the modeler in concocting the simplest functional forms that fit the displays. The natural laws that are to be mimicked are, however, generally composed of simple relationships: powers, exponentials, and sometimes trigonometric functions. Even when using natural laws, it is not uncommon, though, to have to treat one portion of the data with one curve while the remainder demands another. Under these conditions particular care should be taken to preserve continuity at the joints so that singularities or instabilities are not introduced into the model calibration or uses.

3.5 Stopping Rules

In Figure 3-3 of the modeling procedure, there are several points shown which require choices such as between:

1. more exploration or more modeling,

2. more modeling or calibration, and

3. examinations of higher dimensions or lower dimensions.

These decisions are all closely related to the criteria in statistics known as regression stopping rules (Bendel, Afifi, 1977). In all cases these rules involve considering one or more lack-of-fit, or quality, measures, such as those listed in section 3.2, and comparing them against a threshold either directly or indirectly (as through hypothesis testing). In fully automated procedures good stopping rules can mean the difference between excellent models or infinite loops within the computerized modeling process. In interactive modeling schemes, such as the one outlined in this report, automated stopping rules are not used, instead the strategy is to display one or more quality measures to the modeler who then is responsible for the decision at each branch point.

IV. MODEL CALIBRATION

Calibration of a model is a standard operation that aims at developing a set of parameters $\underline{p}$ that is optimum with respect to a specific quality measure, such as an error penalty function $J(.)$. Recall from equation (3.1) that the arithmatic error, or residual, is

$$e_n = y_n - F(\underline{x}_{in}, \underline{p}) \qquad n = 1,2,,...,N. \qquad (4.1)$$

The geometric error is also often of interest:

$$e_n = y_n / F(\underline{x}_{in}, \underline{p}), \qquad n = 1,2,...,N. \qquad (4.2)$$

Calibration can involve finding a $\underline{p}$ that minimizes

$$\underline{p}_{opt} = \min_{\text{all } p \varepsilon P_o} \left. \sum_{n=1}^{N} J(e_n) \right| \text{ given } F(.) \varepsilon F_o \qquad (4.3)$$

this $J(.)$, and in fact all $J(.)$'s, make possible a scalar measure with which the desirability or performance of alternative $\underline{p}$'s, or even alternative $F(.)$'s, can be quantified. The $J(.)$'s based upon the error measures in equations (4.1) and (4.2) lead to the usual mean-type of error functions, and these are the types discussed in the following sections. One should, however, not rule out the possibility of error functions based upon median statistics which could lead to calibrations, for example, that would be based upon

$$\underline{p}_{opt} = \min_{\text{all } p \varepsilon P_o} \left. \text{H-spread}(e_n) \right| \text{ given } F(.) \varepsilon F_o \qquad (4.4)$$

4.1 Ordinary Least Squares

The least squares best estimate of $\underline{p}$ is based upon equation (4.3) with the error function:

$$J(e_n) = e_n^2. \qquad (4.5)$$

This penalty function has many desirable characteristics that have led to its great popularity:

1. it gives equal weight to positive errors and negative errors,

2. small errors are relatively unimportant compared to larger errors, and

3. it is a quite simple function.

Solutions for the $\underline{p}_{opt}$ generally proceed through solution of the necessary, but not sufficient, optimality conditions:

$$\left. \frac{dJ}{d\underline{p}} \right|_{\underline{p}=p_{opt}} = 0 \tag{4.6}$$

$$\left. \frac{d^2 J}{d\underline{p}^2} \right|_{\underline{p} = p_{opt}} > 0 \tag{4.7}$$

Other times, and particularly in the event of discontinuous functions, the solution for $\underline{p}_{opt}$ results from a standard gradient search using:

1. a fixed set of step sizes,

2. Gauss-Newton gradient-calculated step sizes, or

3. Fibinocci step sizes, or for nonlinear problems

4. Hartley's modified Gauss-Newton method (Hartley, 1961), or

5. Marquardt's algorithm (Marquardt, 1963).

Given a set of robust weights w_{yn} that are to be used to focus the calibration effort, the ordinary least squares procedure becomes a search for $\underline{p}_{opt}$ such that

$$\underline{p}_{opt} = \min_{\text{all } \underline{p} \epsilon P_o} \left. \sum_{n=1}^{N} w_{yn} e_n^2 \right| \text{given } F(.)\epsilon F_o \tag{4.8}$$

In other words, this is equation 4.3 with

$$J(e_n) = w_{yn} e_n^2. \tag{4.9}$$

When using a gradient search procedure, it is probably desirable, in order to avoid to some extent local optimums, to use a two-step process where p_{opt} is first determined for unweighted ordinary least squares and this p_{opt} is then used as the initial guess for the robustly weighted search using equation (4.8). It is common practice (Mandel, 1969) to take the iteratively generated optimum parameter values p_{opt} and round them off according to the rule that the rounding errors should not affect the calculated value of the response by more than 1/10 of the observation error. This rounding technique prevents the number of significant digits in the final answer from reflecting undue accuracy.

4.2 Robust Estimators

There are a great number of penalty functions $J(.)$ and lack-of-fit statistics (Bendel, Afifi, 1977) that could, in fact, be used to calibrate the parameters p of the empirical model. It is even conceivable that the inverse of the FIF could be used as a quality measure, where

$$FIF^{-1} = \frac{(\text{lack of fit with model } F(.) \text{ and trial parameter } p)}{(\text{lack of fit without model})} \qquad (4.10)$$

This section will concentrate on the so-called robust estimators, including particularly the non-quadratic error penalty functions $J(e_n)$.

Robust estimation is the general practice of estimating parameters of a model while giving different weights to the observations based upon the apparent quality of those observations. Whereas in robust weighting the "quality" is determined as a priori functions of the input and response variables, in robust estimation "quality" is determined as an a posteriori function of the error. To a certain extent, robust estimation has long

been a practice in statistics, since the standard practice upon observing large errors has generally been to:

1. check the data,

2. set aside these points for special study, and

3. weight these points out of the process.

Elegant robust estimation schemes have only recently been developed. The first of these included ordered residual searches and grouped residual searches (Schweppe, Wildes, 1970), (Schweppe, Rom, 1970), (Schweppe, 1970), (Masiello, Schweppe, 1971), (Dopazo, Klitin, VanSlyck, 1972), and (Handschin, Schweppe, Kohlas, Fiechter, 1975). These techniques were aimed particularly at the detection of whether electric power system datum points were "good" or "bad". They involved searching the errors either in descending order of magnitude, or ascending order within a group of large residuals, to check against the hypothesized structure of the power system whether each datum was consistent or inconsistent with that structure.

Robust techniques based upon non-quadratic error penalty functions are more appropriate for empirical modeling where the data is not either good or bad but has all degrees of quality. Non-quadratic measures that have been used are shown in Figure 4-1, formulas for some of these can be found in (Handschin, Schweppe, Kohlas, Fiechter, 1975). Examples of the use of these types of criteria include: quadratic-square root (Merrill, Schweppe, 1971), quadratic-constant (Handschin, Glavitsch, Kohlas, 1973), quadratic-straight and quadratic-multiple segment (Andrews, et. al., 1972), and quadratic-discard (Hinich, Talwar, 1975). Some other specific proposals for robust methods can be found in (Huber, 1973), (Schlossmacher, 1973), (Beaton, Tukey, 1974), (Andrews, 1974) with reviews in (Huber, 1972), (Hampel, 1973), (Hill, Holland, 1977), and (Harvey, 1977).
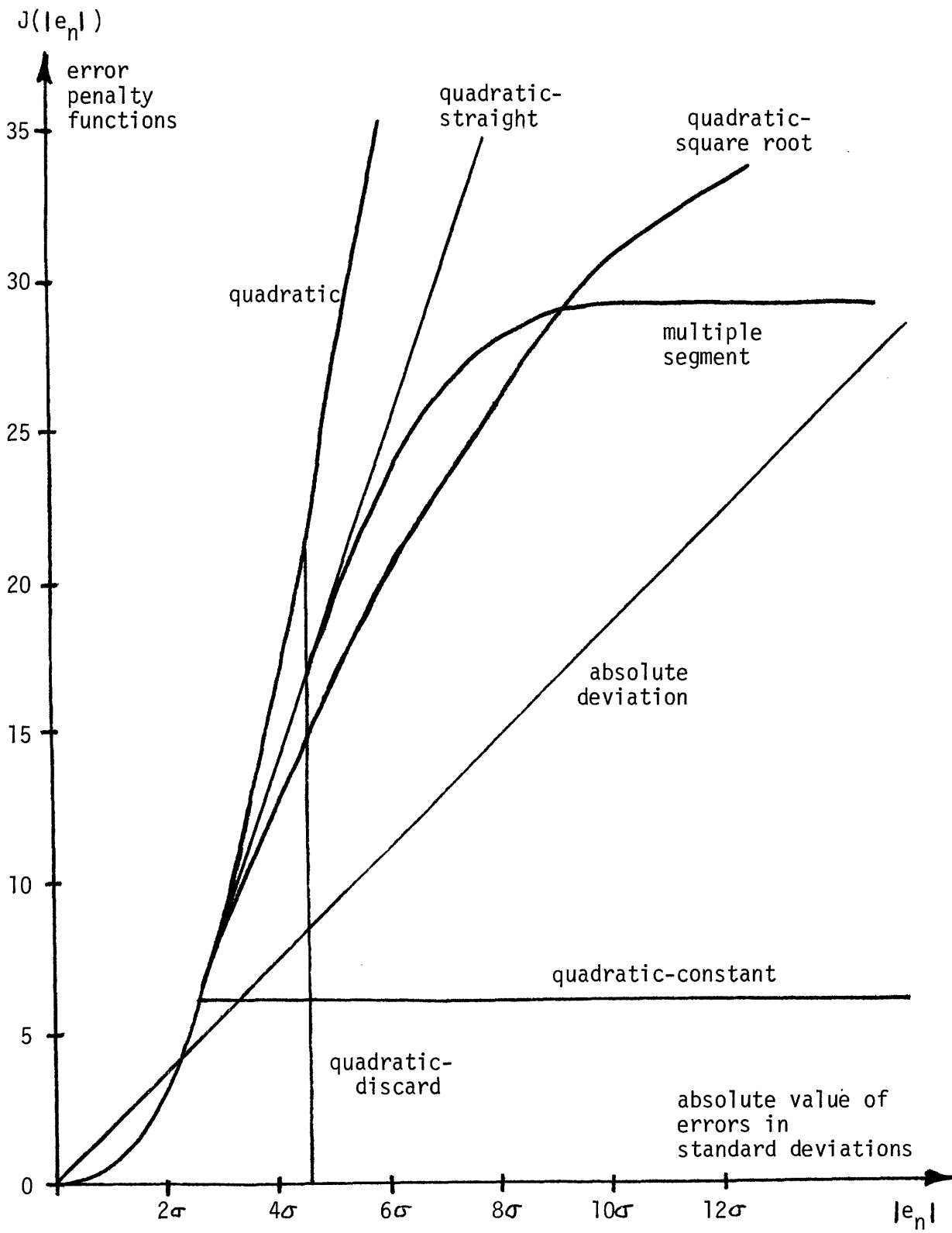
Figure 4-1 Various non-quadratic error penalty functions for use as mean-statistic type robust estimation criteria

For a number of applications, the minimum absolute deviation estimator has been judged (Harvey, 1977) to be superior to many other of the robust criteria, including some of the median statistic techniques and particularly including other one-step estimators of $\underline{p}$. This minimum absolute deviation, or MAD, estimator produces a $\underline{p}_{opt}$ such that

$$\underline{p}_{opt} = \min_{\text{all} \underline{p} \epsilon P_o} \left\{ \sum_{n=1}^{N} w_{yn} | y_n - F(\underline{x}_n, \underline{p})| \right\} \text{ given } F(.) \epsilon F_o \qquad (4.11)$$

Techniques such as the quadratic-constant and the quadratic-discard are inclined to be unstable at worst and at best largely dependent upon the starting conditions. To avoid these tendencies toward local optimums, it is generally recommended that non-quadratic criteria be used from the starting position $\underline{p}_o$, that is the optimum parameter selection from the ordinary least squares estimator. This then represents a two-step procedure.

V. MODEL VALIDATION

The process of model validation has been described (Wood, 1978) as the formulation and resolution or evaluation of questions and hypotheses about appropriateness of the model's logical structure and the validity of its implementation. As pointed out in (Snee, 1977) there is general agreement that validation of a model is a most important procedure, however, there is almost no literature available to explain validation techniques, and individual modeling efforts rarely explain which, if any, validation schemes have been used. This chapter will summarize the validation terminology and the methods available, a great deal of which is new to this report. Some of these ideas have evolved from initial, and not generally consistent, ideas from the parallel experience of the authors with dynamic models validation at the MIT Model Assessment Laboratory. The most important expressions characterized in this chapter are:

1. _setting_ of the validation,

2. assessment _depth_,

3. _scope_ of the verification,

4. validation _techniques_, and

5. measures of the _quality_ of predictions.

For each of these concepts, there are several available procedures and ideally one would follow all paths for all concepts. The time and resource constraints of reality, however, necessitate the evaluation of only the most cost-effective of the various possible schemes. Generally such an evaluation will depend upon size, structure, complexity, cost per run, and maturity of the model.

5.1 Setting, Depth, and Scope

As determined during the course of previous research at the MIT Model Assessment Laboratory (MIT, 1978), it has become apparent that it is initially most important to define the categories of setting and depth of the validation. Setting is defined as the combination of the model or set of models to be accessed along with the group or groups to perform the validation. There are particularly great differences in the expectations and potentials of the validation process depending upon the relation of the validator to the model, with different advantages to each of these different settings. The various possibilities for the setting of validation include one or several models being investigated by one or several:

1. model builders,

2. model sponsors or users, or

3. independent third parties.

Table 5-1 shows the more important of these settings in an array with different important depths. (It should be noted that each of the levels of depth include the previous categories.) For regression models that are quite simple, independent model audits are probably unnecessary and the ideal validation depth, examination of alternative methodologies, is quite possible. For very complex models it has been determined (MIT, 1978) that the independent audit may be perhaps the most cost-effective validation scheme and that examining alternative methodologies may well be impracticable. Thus, the appropriate, that is to say, cost-effective depth is inversely proportional to the complexity of the model. In other words, more complexity suggests less depth.

Table 5-1 Different important model assessment categories

ASSESSMENT SETTING ⟶

		Model Builder	Model Sponsor	Third Party	Several Model Builders	A Third Party on Several Models
A S S E S S M E N T — Overview	Evaluate documents for model results			interchanges and peer review through publications		classical lit.survey
	Evaluate documents for model struct.					methodol. lit.survey
	Examine comp. code				.	
Audit	Hands-off model exercising	self-audit	classical valid.	independ. audit		model forums
D E P T H — In-Depth	Hands-on model exercising -	classical valid.			model forums	
	Component structural changes					
	Recoding computer program					
	Alternate modeling methodologies					

48

Next, it must be determined exactly what there is that should be part

of the model assessment process. This bounding of the _cases_ and _aspects_ of

the investigation is called the _scope_. There have been some studies of

possible scopes of assessment, including (MIT, 1978), (Greenberger, 1977),

and particularly (Gass, 1976). The ideas about scope presented in

Table 5-2 are somewhat differently organized but essentially include all of

those previous concepts. The most noteworthy of these concepts of scope is

the specification of cases of interest, that is that it emphatically must

be defined exactly what ranges and types of cases are to be studied. For

empirical regression models, this scope would differentiate, for example,

between interpolating and extrapolating uses over certain ranges of

operating and design variables.

Table 5-2. Outline of Information Necessary for Defining the

Scope of the Assessment

1. __Specific Cases of Interest__

- particular issues or problems for which the model(s) is to be
  accessed for applicability and adequacy; requires specification
  of ranges of variables, degree of aggregation required, absolute
  value versus perturbation studies, and so on.

2. __Aspect(s) to be Assessed with Respect to Those Cases__

2.1 Documentation
- of the structure, validations performed, past uses, and
applicability of the model; and of the computer code (if any).

2.2 Verifiability
- extent to which model can be validated, specifically raising
issues of other comparable models, other data, complexity,
size, and cost difficulties.

2.3 Validity

2.3.1 Data or Input Validity
- empirical implementation; quality of updating procedure

2.3.2 Logical or Structural Validity
- behavioral and technical structure

2.3.3 Predictive or Output Validity

2.4 Operational Characteristics
- flexibility, extensibility, training required, and other ease
of use measures; program efficiency, in terms of time and cost
per simulation.

5.2 Validation Techniques

The choice of suitable validation techniques should be directed by the

type of assessment. These choices should generally be obvious upon

recognizing the limitations and emphases denoted in the setting, depth, and

scope of the evaluation. Concerning the various types of regression

validation techniques the November 1977 article by R.D. Snee (Snee, 1977)

is probably the definitive document-to-date, listing four different

methods. Using these four as a start, a number of additions are made to

this list by drawing upon similar dynamic validation lists, (Boshier,

Schweppe, Gruhl, 1978) and (M.I.T. 1978), computerized literature surveys,

and the suggestions offered by the symmetric framework into which these

techniques have been fit. Table 5-3 thus contains a descriptive listing of

these different evaluation schemes. It should be remarked that validity in

its true sense is an absolute term, and can only be approached to a greater

or lesser degree. The extent to which absolute validity can be achieved

will depend primarily upon whether the data is historical or designed,

designed in that variable ranges have been filled parametrically by

replicable experiments. To a reduced, but still important extent, the

degree of validity can be increased by the breadth and quality of the

evaluation processes that are conducted. These validation techniques

involve two steps - first some piece of the model is examined or changed -

then this action is evaluated with repect to some basis for comparison.

50

Although Table 5-3 outlines 18 possible actions times 6 possible

comparisons, only a little more than half of these 6x18=108 combinations

seem to be of interest (Boshier, Schweppe, Gruhl, 1978). Observed data is

historical or other data used to build the model; input and output are

those data associated with a particular predictive use of the completed

model.

Table 5-3 Outline of various evaluation schemes that may be useful

```
1.  Actions: examinations or changes
    OBSERVED DATA:
    1.1  Examinations of the observed, historical, or estimation data
    OBSERVATIONS-TO-STRUCTURAL:
    1.2  Observed data perturbation effects on structure and
         parameters
    1.3  Propagation of estimation error on structure and parameters
    1.4  Measures of fit of structure and parameters to observed data
    1.5  Effect of correlated or irrelevant observed or estimation
         data on structure and parameters
    1.6  Sensitivity analysis: quality of fit of structure and
         parameters to observed data for altered structure and
         parameters (includes ridge regression)
    OBSERVATION-TO-OUTPUT:
    1.7  Effects of correlated or irrelevant observed data on outputs
    INPUT:
    1.8  Base case or recommended input data examinations
    INPUT-TO-OUTPUT:
    1.9  Examine outputs with respect to base case input data
    1.10 Simplify, e.g. linearize, this relationship to provide
         understanding
    1.11 Develop confidence measures on outputs by propagating input
         error distributions through structural error distributions
    STRUCTURE:
    1.12 Structural form and parameter examinations
    1.13 Simplify, e.g. linearize, structural form analytically, or
         group parameters to provide better understanding
    1.14 Decompose structure physically or graphically
    STRUCTURAL-TO-OUTPUT:
    1.15 Examination of outputs for various structural and parametric
         perturbations and error distributions
    OUTPUT:
    1.16 Examination of outputs
    OUTPUT-TO-INPUT:
    1.17 Examination of optimal inputs (controls) to reach target
         outputs
    1.18 Contribution analysis, percentage changes in outputs due to
         changes in inputs
```

2. Bases for Comparison
   2.1 Comparison with other empirical models
   2.2 Comparison with theoretic or analytic models, or hand
       computations
   2.3 Data splitting on observed data, by time or region
   2.4 Obtain new estimation/prediction data with time, new
       experiments, or in simulated environments
   2.5 Examination of reasonableness and accuracy
   2.6 Examination of appropriateness and detail

5.3 Quality of Predictions

There is an obvious need for a probabilistic measure of the quality of

predictions from the model. A first attempt at developing such a measure

could come from any of the lack-of-fit statistics listed in section 3.2.

This proposal, however, means that the data used to validate the model will

be the same as that used to build the model and as such, it is suggested

here that that statistic will represent an upper bound on the quality of

new predictions. Data splitting techniques (Snee, 1977), (Simon, 1953),

(Savage, 1962), (Ezekiel, Fox, 1959), (Marquardt, Snee, 1975), (McCarthy,

1976), (Novick, et. al., 1972), (Stone, 1974), (Draper, Smith, 1966),

(Welsch, 1974), can be viewed as a means of obtaining a more realistic

estimate of the quality of prediction of new situations. It has seemed to

puzzle some investigators that these more elegant data splitting schemes

result in models that fit the data more "poorly". It is suggested here

that these apparently "poorer" quality measures are just truer measures of

quality than the previous upper limits. This concept has been illustrated

in (Snee, 1977), where the quality measure is 25% to 100% "poorer" with

data splitting but these models predict more accurately on new data. A

method described in (Welsch, 1974) results in a large number of data splits

and it is proposed here that the worst of the resulting quality measures

may be very close to the actual predictive quality of the model.

One final issue should be noted and that is the effect of model smoothness on the measure of predictive quality. It has been discussed in several places in the literature (Clark, 1977) that model building requires a conscious or unconscious tradeoff between smoothness and fidelity to data. Measures of predictive quality that are developed from the estimation data will tend to discourage smoothness and encourage contrived roughness. To a certain extent, this effect may be recognized through data splitting, but in the final analysis, a personal judgment of the reasonableness with respect to smoothness of the model's functional forms must be made.

VI. EXAMPLE OF FLUIDIZED BED COMBUSTORS

Although the data from (Strom, et al., 1976) does contain 369 experimental observations, this is only a very small sample of the extensive data base that is currently being developed on fluidized bed combustion (Louis, Tung, 1977). Thus, the example of modeling in this chapter, using the 369 observations, should be viewed as a small methodological exercise. The small number of input variables, 14, that are carried for this example is another limitation that has been made to facilitate the exploration of statistical techniques. Despite these restrictions, this example may be to-date the most complete fluidized bed empirical modeling effort. Appendixes A and B should be consulted for more complete information on the data base and statistical exercises.

6.1 Empirical Models of Emissions

In developing these empirical models there was a certain amount of guidance that could be received from published parametric series of experiments. The functional forms intimated by these parametric results were of limited usefulness, however, as none displayed relationships that were universally consistent with data from other experiments. The primary usefulness of these studies was to highlight the operating and design parameters that could exert influence on particular outputs. The use of Fit Improvement Factors, however, not only highlighted the important variables but to a certain extent provided quantification of the relative magnitude of their importance.

Atmospheric fluidized bed combustors, FBC, are being actively developed because aside from potential cost and efficiency advantages, they offer built-in opportunities for controlling SO_2 emissions. Empirical modeling of SO_2 is thus of interest for:

(1) defining the SO_2 removal capabilities of FBC's

(2) pointing toward improvements in designs, and

(3) directing the search toward more desirable operating conditions.

The dominant term that starts up the modeling exercise is the calcium-to-sulfur ratio (see Figure B-1, Appendix B) which shows a Fit Improvement Factor, FIF, of 1.778 (next best initial term is calcium top particle size with FIF = 1.167). The cellular statistics for calcium-to-sulfur versus SO_2 removal were displayed (see Appendix B) and the effect of this term was removed. At first the residuals were checked for the possibility of an additive term, however, the maximum FIF of 1.324 (for calcium top size) showed this avenue to be unspectacular. Considering the possibility of a multiplicative term, showed excellent potential with FIF = 1.932 for fluidizing velocity. After the entire process of adding single terms was completed, modeled, and finetuned the nonlinear combinations of terms were explored with the FIF techniques. No FIF's were developed for 3-tuples, although this should certainly be done once the large data base is available (Louis, Tung, 1977).

The best model for SO_2 removal fit only with the disappointingly large standard deviation of 13.2%. Using robust weighting techniques to focus modeling on 80%, or higher, SO_2 removals resulted in much closer, 3.46%, fit, with equations shown in Table 6.1-1 and scatterplot in Figure 6.1-1. Mean calcium size contributed very little to the fit and was not very frequently measured by experimenters. So mean calcium size was dropped from the model, increasing the number of experiments that could be modeled from 41 to 62, and the new (probably more reliable) model is shown in Table 6.1-2 with scatterplot in Figure 6.1-2. The original, 13.2% deviation, model over all of the data is shown in Table 6.1-3, with scatterplot in

Table 6.1-1 Empirical models of sulfur removal for observations where removal equalled or exceeded 80%

Symbols:

- S = sulfur removed, as percent
- C = calcium to sulfur mole ratio
- V = fluidizing velocity, in m/sec
- T = bed temperature in degrees centigrade
- D = static bed depth, in cm
- A = bed area, in square meters
- M = calcium mean size, in microns
- P = parameter for effect of coal source
- Q = parameter for effect of calcium source
- R = number of times sorbent reused, 1+number of recycles

Model from a limited number of parametric experiments:

$$S = 100.-(30.6/C)[(V/.3048)^{.25}-.30][1.+.429(\frac{1.8T-1492+45.5abs(2.35-C)}{127.3})^2] *$$

$$[2.24(\frac{D}{2.54})^{-.25}](R^{-.5})P$$

[fit=arithmetic standard dev. 5.43%]

Model on data base where all experiments with these parameters and S over 80%:

$$S = (100.- \frac{41.4}{C+.414})(.923+.0762V^{-1})(.997+.0247A)[1.23-.00015(1.8T+32)]*$$

$$(1.0058-.000013M)(.99337\ P\ Q)$$

[fit=arithmetic standard dev. 3.46%]
on 41 experiments

Tables of parameters:

coal source	P_{param}	$P_{data\ base}$	calcium source	$Q_{data\ base}$
Pitt seam 8 washed	0.48	1.05	BCR1337 dolom	0.98
Pitt seam 8 unwashed	0.83	1.00	BCR1360 limes	1.00
Peabody Coal Co.	0.80	1.01	BCR1359 limes	0.99
Pitt seam unspecified	0.92	1.00	BCR1359H hyd lm	1.01
Commonw. Edison supplied	0.97	0.98	Tymochtee dolom	1.04
Park Hill coal	1.04	1.00	US limestone 18	1.02
Illinois coal unspec.	1.13	1.00	UK limestone	1.02
Welbeck coal unspec.	1.28	0.975	Stow-on-Kent	0.92

Actual SO₂ Removal (percent) axis:
75.0   80.0   85.0   90.0   95.0   100.0   105.0

Predicted SO₂ Removal (percent) axis:
105.0
100.0
95.0
90.0
85.0
80.0
75.0
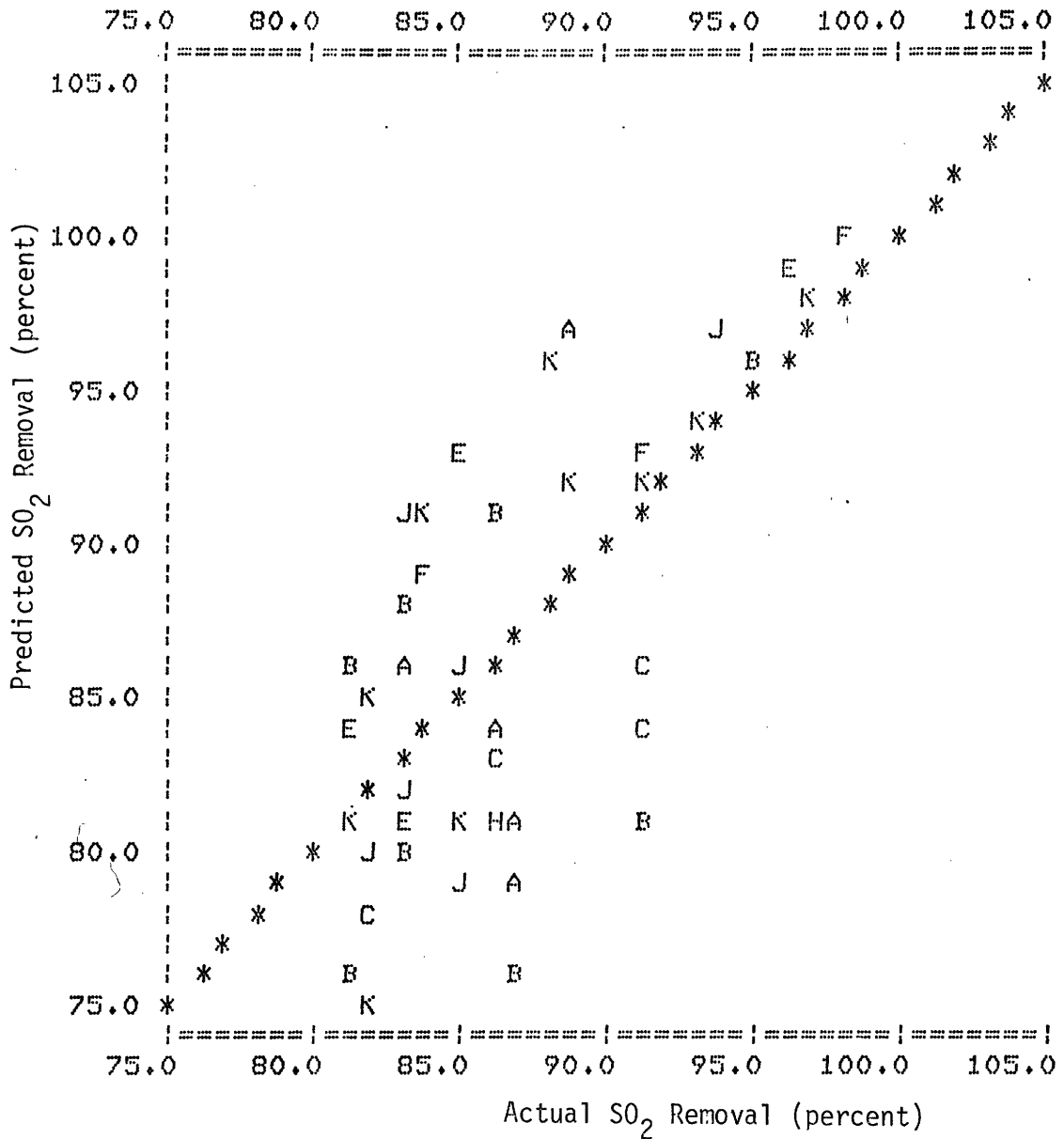
Figure 6.1-1 Scatterplot of SO_2 removal as observed versus predictions
based on a data base developed model, the least squares
fit approach concentrates on containing outlyers and has
relative indifference to values within general band around
the identity line (symbols refer to positions in data base)

Table 6.1-2 Empirical model of sulfur removal for observations exceeding 80% removal, eliminating calcium mean as a variable to increase the number of applicable experiments where all variables were reported

Symbols:

- S = sulfur removed, as percent
- C = calcium to sulfur mole ratio
- V = fluidizing velocity, in m/sec
- T = bed temperature, in degrees centigrade
- A = bed area, in square meters
- F = 1 or 0 for fines recycled or not
- G = MgO/CaO in sorbent
- P = parameter for effect of coal sources
- Q = parameter for effect of calcium sources

Model on data base where all experiments with these variables reported and with S over 80%:

$$S = (101.- \frac{30.97}{C+.3097})(1.208+\frac{.40}{V+.762}+.0151V^2)(1.+.0095A)[1.-.00022(1.8T+32.)]$$

$$*(1.-.0012F)(1.-.00476G)[1.-.00000246(1.8T+1500.)^2] \, 1.002 \, P \, Q$$

[fit=arithmetic stand dev 3.976%]
on 62 experiments

Table of parameters:

coal source	$P_{data\ base}$	calcium source	$Q_{data\ base}$
Pitt seam 8 washed	0.993	BCR1337 dolom	0.995
Pitt seam 8 unwashed	1.000	BCR1360 limes	0.991
Peabody Coal Co.	1.015	BCR1359 limes	0.992
Pitt seam unspecified	1.002	BCR1359H hyd lm	1.023
Commonw. Edison supplied	0.998	Tymochtee dolom	1.031
Park Hill coal	0.974	US limestone 18	1.007
Illinois coal unspec.	0.991	UK limestone	1.014
Welbeck coal unspec.	0.959	Stow-on-Kent	0.925

65.   75.   85.   95.   105.
|==============|==============|==============|==============|
105. | E   *
     | *
     | *
     | K   F *
     | A   B   * G
     | H H 2 H *G
95. | 2 *
     | K   H   K K   K
     | 2   H 2 H 2* H
     | J   H   *
     | CF H   *K   3
     | AK 2 *
85. | *
     | EJ B
     | *   A A
     | * HC   JC
     | *   K   3 KB   G
     | B
75. | *
     | *
     | *
     | *
     |*
65. *
|==============|==============|==============|==============|
65.   75.   85.   95.   105.
              Actual SO$_2$ Removal (percent)

Table 6.1-3 Empirical model of sulfur removal for all observations for which values of all parameters were available

Symbols:

S = sulfur removed, as percent
C = calcium to sulfur mole ratio
V = fluidizing velocity, in m/sec
A = bed area, in square meters
M = calcium particle mean size, in microns
L = sulfur content of coal, in percent
X = excess air, in percent
P = parameter for effect of coal source
Q = parameter for effect of sorbent source

Model on data base for all experiments:

$$S = (100. - \frac{209.58}{C+20.96})(1.-.0912V)(1.+.0108A)(1.-.00011M)*$$

$$(1.-.0000117L)(1.-.000516X)\ P\ Q$$

[fit=arithmetic stand dev 13.2%]
on 296 experiments

Tables of parameters:

coal source	$P_{data\ base}$	calcium source	$Q_{data\ base}$
Pitt seam 8 washed	1.789	BCR1337 dolom	1.033
Pitt seam 8 unwashed	1.688	BCR1360 limes	1.160
Peabody Coal Co.	1.615	BCR1359 limes	0.971
Pitt seam unspecified	1.721	BCR1359H hyd lm	1.228
Commonwealth Edison	1.404	Tymochtee dolom	1.230
Park Hill coal	1.712	US limestone 18	1.065
Illinois coal unspec.	1.535	UK limestone	1.000
Welbeck coal unspec.	1.693	Stow-on-Kent	1.000

Figure 6.1-3. The disappointment with this amount of deviation is apparent

when one considers the risk involved in designing a combustor to meet a

specific emissions threshold, see Figure 6.1-4.

The most important initial term in the modeling of NO_x emissions is

the fluidizing velocity, with FIF = 1.315. To show the conflicting nature

of the NO_x data, which is evident from the reports in the literature, the

next most important term for modeling is the reference number! Again,

once a first term is removed, arithmetic, that is additive, terms showed

very little potential, FIF = 1.148, and multiplicative terms were explored.

Some important cross-terms were identified and to facilitate the modeling

of these, the multiplicative terms were expanded into polynomial form, see

Table 6.1-4. Figure 6.1-5 shows the disappointing fit, which can be

significantly improved by separate modeling of the results from the

different experiments.

An example of this type of separate modeling is shown in the CO

predictions. Reference number shows the best initial FIF = 1.509, but is

suppressed in favor of bed area. After the other modeling is completed

there are a number of terms that are off by a factor of 10 to 15, see

Figure 6.1-6. These two clusters were finally isolated using excess air, see Figure
6.1-7, but nothing
was useful in modeling this effect. The best model without this step

function is shown in Table 6.1-5 with scatterplot in Figure 6.1-8.

Hydrocarbon modeling was shown to be impossible without accounting for

enormous differences between different experimenters, FIF = 1.479, with

coal top size next at FIF = 1.299 but this fit was essentially motivated by

correlation with experimenters. The best final model is shown in Table

6.1-6 with scatterplot in Figure 6.1-9.
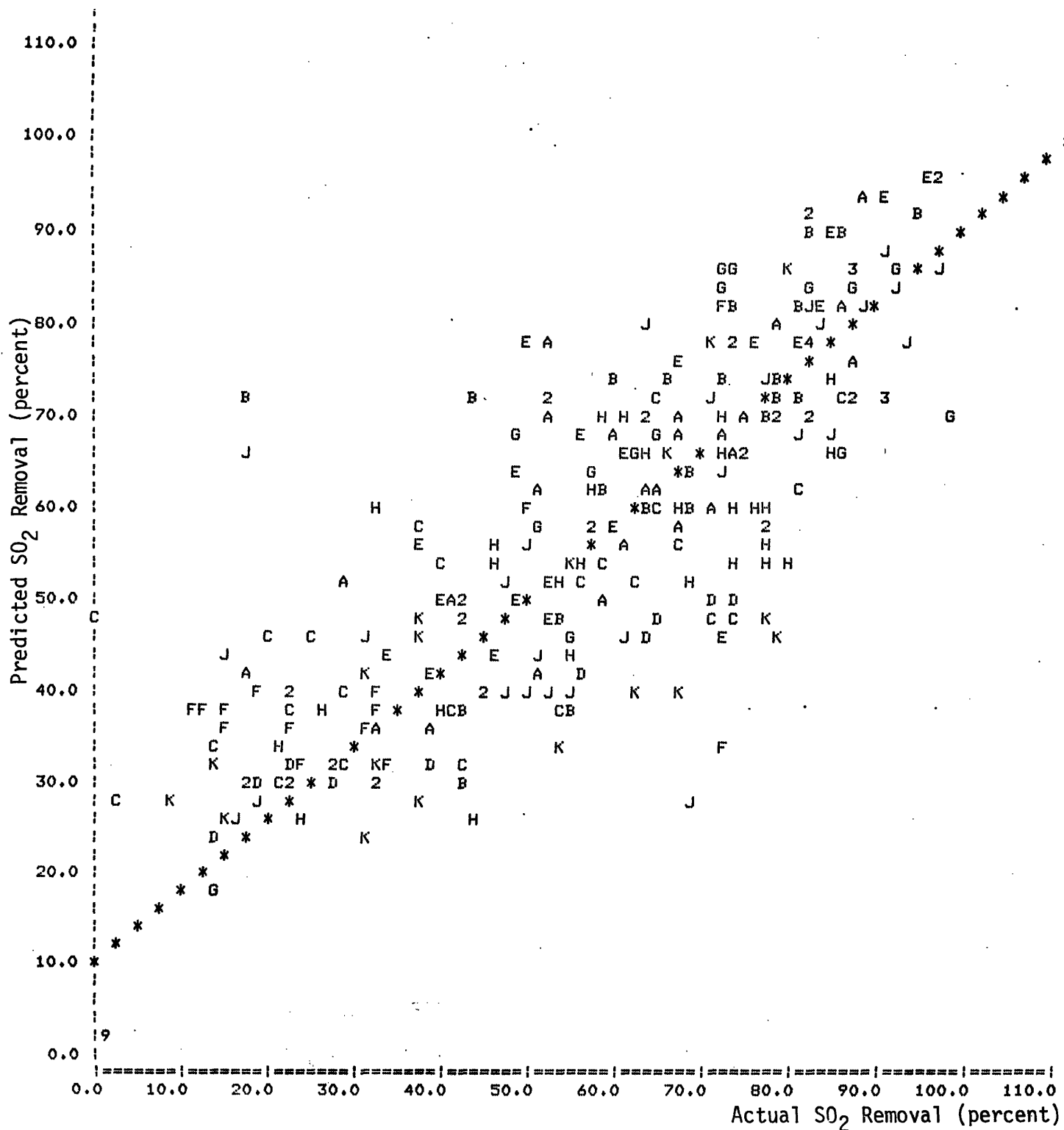
Figure 6.1-3 Scatterplot of SO_2 removal as observed versus predictions
based on all experiments, numbers refer to multiple
occurrences with the 9 in the lower left representing
about 40 experiments
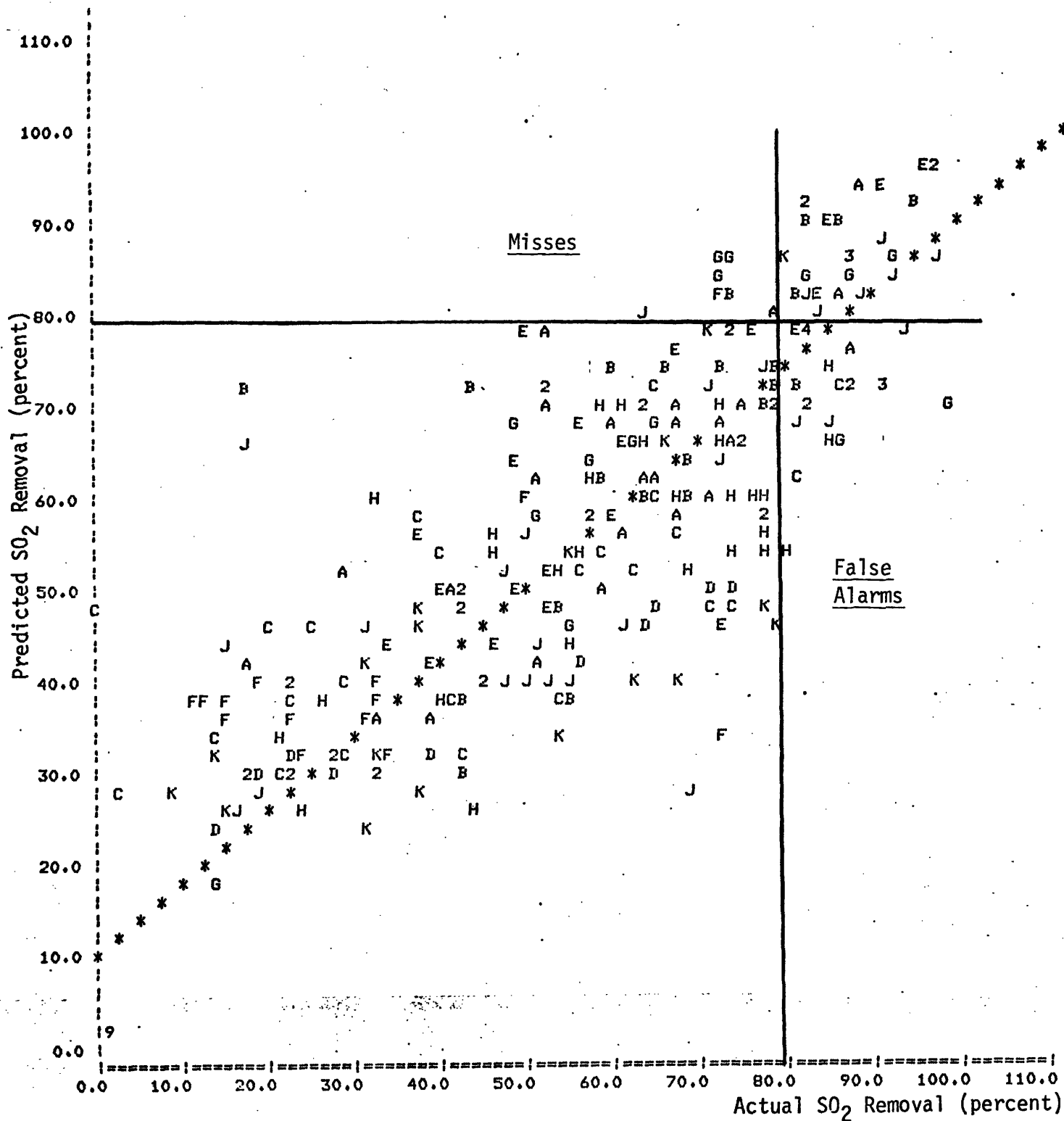
Figure 6.1-4 Using this model to design a combustor to meet the SO$_2$ standard
by removing 80% or more of the SO$_2$ would result in more than
1/5 of the designs being improper (misses) and would exclude
more appropriate designs (false alarms) than it would include,
one might conclude that there is currently too much uncertainty
to make it appropriate to design an environmentally acceptable
combustor.

Table 6.1-4   Empirical model of NO_x at outlet based on data base
             experiments

Symbols:

    N = NO_2 equivalent of NO_x at outlet, in ppm
    C = calcium to sulfur mole ratio
    V = fluidizing velocity, in m/sec
    T = bed temperature, in degrees centigrade
    X = excess air, in percent
    Q = parameter for effect of sorbent source

Model on all data base experiments for which the above variables
  were available:

  N=(-5818+82.41C+ 6.43CV + 362.V + 6.9(1.8T+32.) - .267V(1.8T+32.)

      -.065C(1.8T+32.) - .00182(1.8T+32.)^2) (1.106-.0074X) 1.053 Q

                                    [fit=arithmetic stand dev 91ppm]
                                          on 198 experiments

Table of parameters:

calcium source	$Q_{data\ base}$
BCR1337 dolom	0.90
BCR1360 limes	0.85
BCR1359 limes	1.00
BCR1359H hyd lm	0.70
Tymochtee dolom	0.85
US limestone 18	1.00
UK limestone	1.00
Stow-on-Kent	0.85
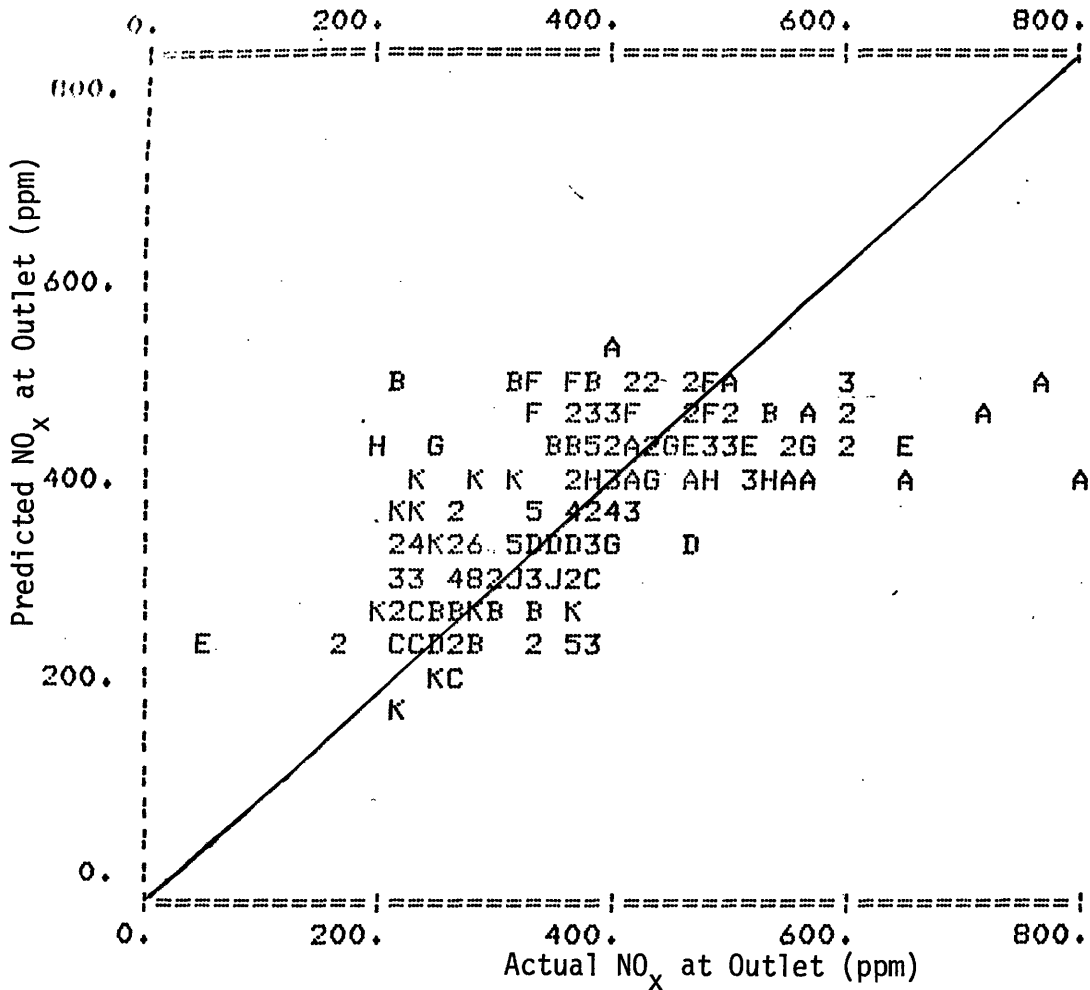
Figure 6.1-5 Scatterplot of NOx predictions versus observations,
showing general lack of good fit to data due
possibly to unmeasured variables or differences in
position or techniques of measurement

```
        1.00      2.00      3.00      4.00      5.00      6.00      7.00
       |=========|=========|=========|=========|=========|=========|
  15.0 |                                        4
       |                                        2
       |                                        4
       |                                        2
       |                                        3
  12.5 |                                        2
       |                                        H
       |
       |                                        H
  10.0 |
       |
       |
       |                            \                         J
   7.5 |
       |
       |                                        H
       |
       |                                        H
   5.0 |
       |
       |                                        H
       |
       |              A               K
   2.5 |              C               3         2
       |              6     K         2         4         J
       |              6     K         2         3         3
       |              7     3                   H         8
       |              9     2                   2         7
   0.0 A 2                  D                   E         +
       |=========|=========|=========|=========|=========|=========|
        1.00      2.00      3.00      4.00      5.00      6.00      7.00
                    Different Reference Documents
```

Error Factor (Actual CO/Predicted CO)

Figure 6.1-6   Scatterplot demonstrating the cluster of CO values that
               persistently stayed at about 10 times the predicted
               values, thus initiating cluster analysis with separate
               modeling.

```
         -20.0    -15.0    -10.0     -5.0     0.0      5.0     10.0     15.0     20.0
         |=======|=======|=======|=======|=======|=======|=======|=======|
   17.5 | .
        |
        |
        |
        |
   15.0 |              J  JH              H
        |*
        | *         J      H    HH
        |  *               H
        |   *      H      H     G
   12.5 |    *            H  G
        |     *         H  J
        |      *        H
        |       *
        |        *           H
        |         *
   10.0 |          *
        |           *
        |            *
        |         J   *
        |              *
    7.5 |               *
        |                *
        |              *
        |               * H
        |                *
        |              H *
    5.0 |                 *
        |                  *
        |              H_     *
        |                  *
        |               A *      K
    2.5 |              2      *     2K       C
        |           J  H2K   H    *  K4 JC  C
        |           KK J HJH        K*KAB 2 3
        |        J     JJJ2JKJJ  K    *   2H  A2  2
        |        J2K              22J  A  *K  C23 2 A        +
    0.0 |              2  2DJJ   233 2H2   2    A      A
        |=======|=======|=======|=======|=======|=======|=======|=======|
        -20.0    -15.0    -10.0     -5.0     0.0      5.0     10.0     15.0     20.0
                              Excess Air (percent)
```
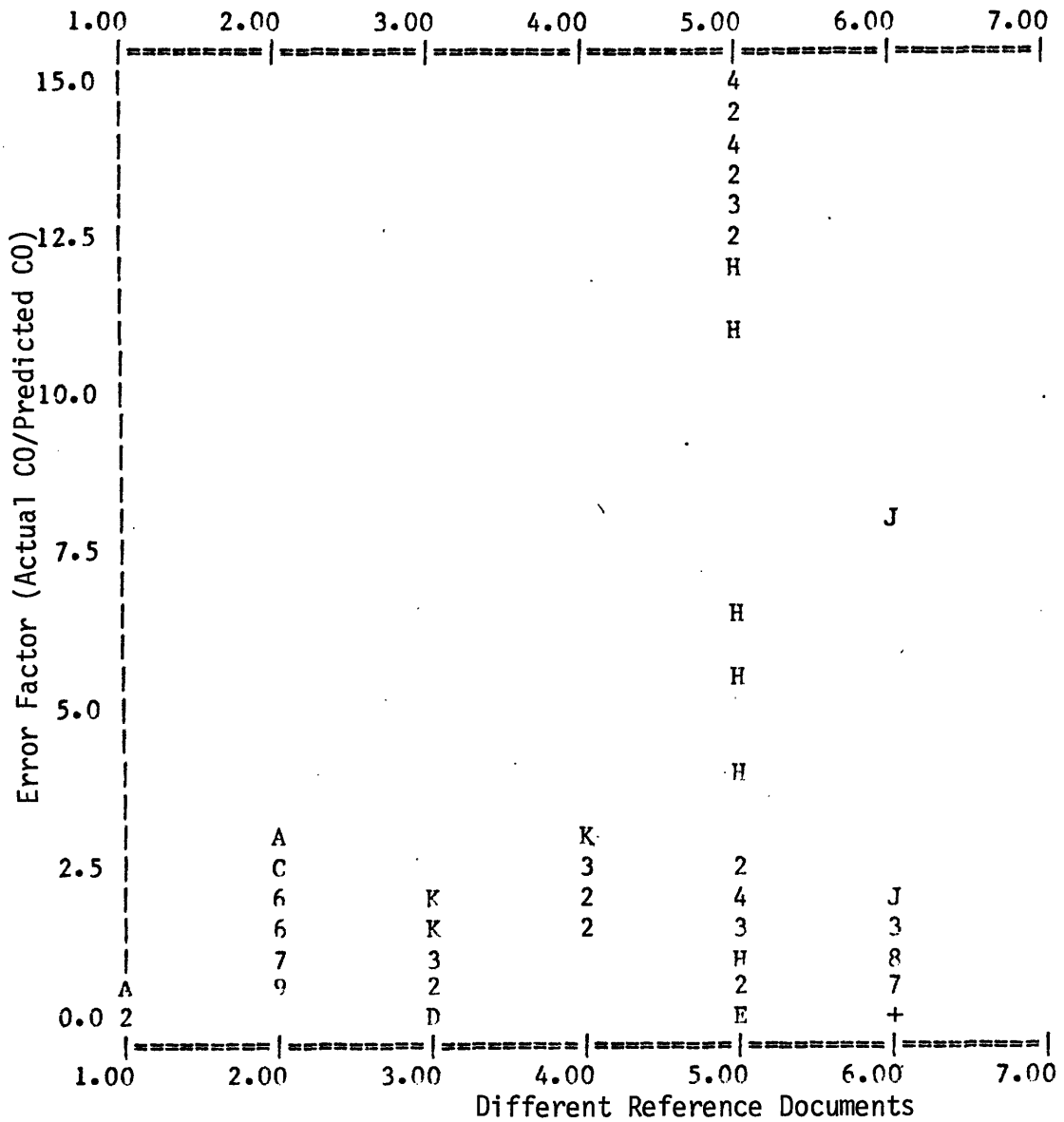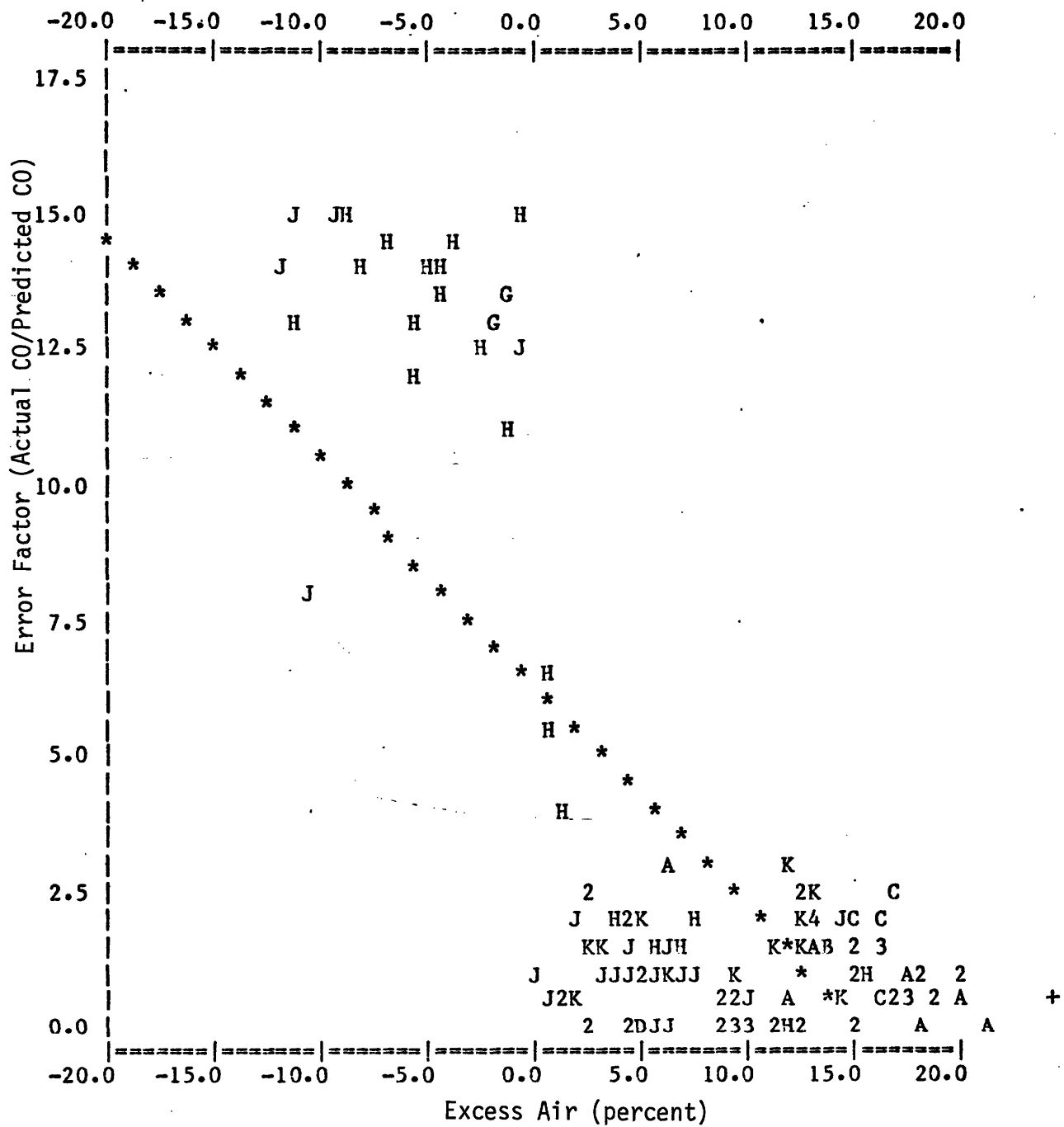
Figure 6.1-7 Scatterplot demonstrating the separation of the clusters
               in the excess air variable space, use of various
               functions over excess air did not improve the fit,
               except step-functions which essentially is identical to
               separate modeling of the clusters

Table 6.1-5 Empirical model of CO concentrations at outlet

Symbols:

        O = CO concentration at outlet, in ppm
        A = bed area, in square meters
        Z = coal top particle diameter, in microns
        X = excess air, in percent
        F = 1 or 0 variable if fines recycled or not
        L = sulfur content of coal, in percent
        V = fluidizing velocity, in m/sec

Model based on least squares best fit to <u>all</u> available data:

$$O = max[\ 50., \ (-303+149.A^{-1}-2.6A^{-2})(1.-.00008Z)(-22.276+1.199[X+15]$$

$$+98.12[X+15]^{-1})(1.+1.087F)(1.-.04116L)(1.-.0377V)\ ]$$
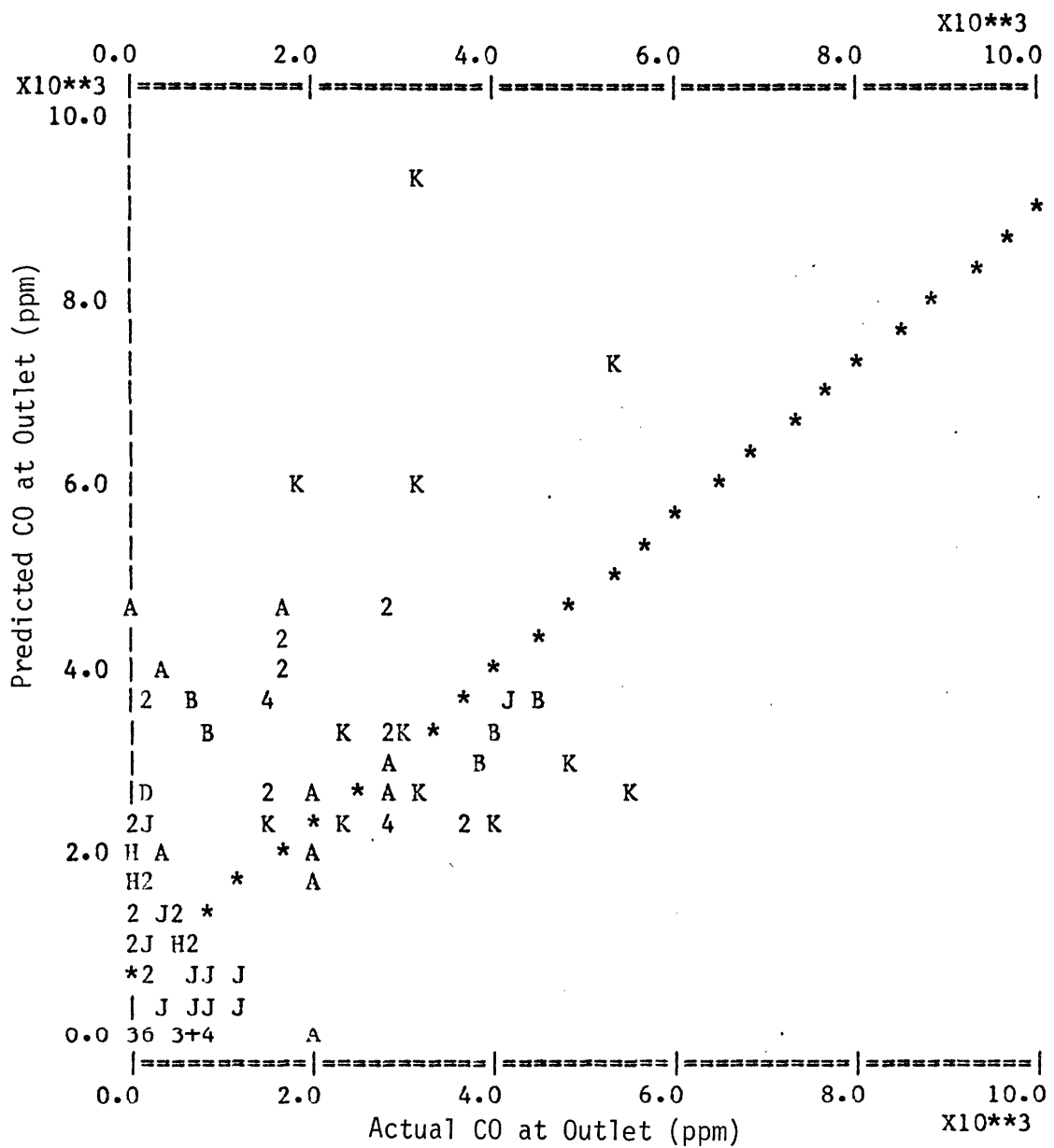
[fit=arithmetic stand dev 1010 ppm]
on 124 experiments

```
                                                              X10**3
          0.0          2.0          4.0          6.0          8.0         10.0
X10**3  |===========|===========|===========|===========|===========|
  10.0  |
        |
        |                        K
        |
        |                                                                    *
        |                                                                *
        |                                                            *
  8.0   |                                                        *
        |                                                    *
        |                        K                        *
        |                                             *
        |                                         *
        |                                     *
  6.0   |        K            K                   *
        |                                  *
        |                               *
        |                            *
        |                         *
  A     |        A           2          *
        |                    2        *
  4.0   |  A                 2      *
        | 2   B        4          *     J  B
        |     B          K   2K  *      B
        |                    A       B      K
        | D           2   A  *  A K              K
        | 2J          K   *  K  4      2  K
  2.0   |  A             *   A
        | 2      *       A
        | 2 J2 *
        | 2J  H2
        | *2   JJ  J
        |  J  JJ  J
  0.0   | 36  3+4        A
        |===========|===========|===========|===========|===========|
          0.0          2.0          4.0          6.0          8.0         10.0
                         Actual CO at Outlet (ppm)              X10**3
```

Figure 6.1-8   Scatterplot showing the relatively poor fit to the
               wildly varying CO data, this is for all the data
               together and is very similar to the superposition
               of scatterplots from the separate cluster modeling
               efforts

Table 6.1-6   Empirical model of CH_4 equivalent of the hydrocarbon
at the outlet

Symbols:

H = CH_4 equivalent of hydrocarbon concentrations at
     outlet, in ppm
A = bed area, in square meter
N = number of the reference from which the data was
     collected
C = calcium to sulfur ratio in bed, molar ratio
Z = coal top particle diameter in microns
X = excess air, in percent

Model based on all data in base:

H = 146.(1.+10.65A)(0 if N greater than 4.5 else 1.)(1.-.2249C)*

(1.-.000125Z)(1. if X less than 0. else 11.98/(X+.15))

[fit=arithmetic stand dev 11.1]
over 196 experiments
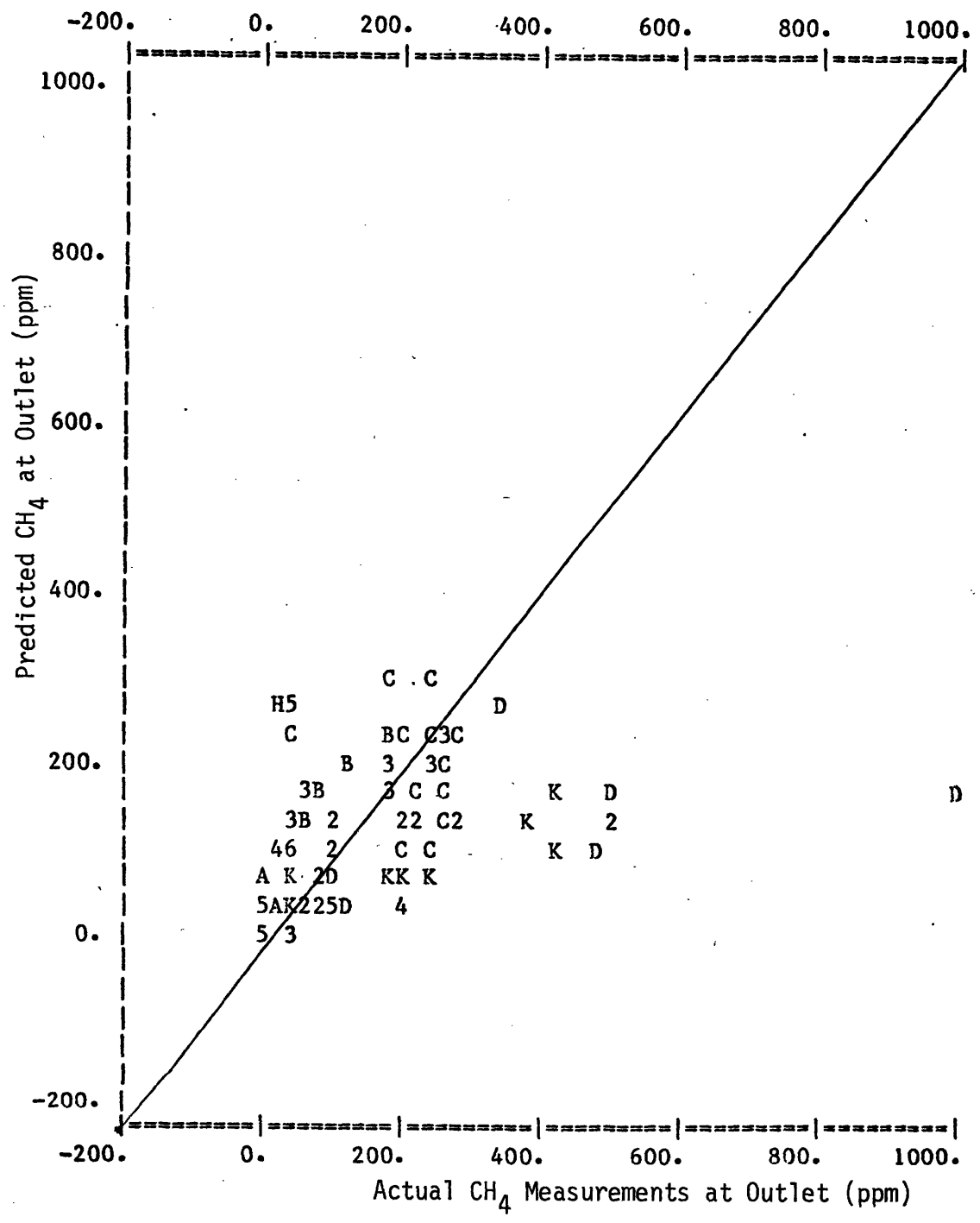
Figure 6.1-9 Scatterplot of predicted versus observed CH_4 concentrations at outlet

Dust loading at outlet showed FIF = 2.666 for fluidizing velocity and FIF = 2.551 for bed area. This shows that there might be an opportunity for a really close model for dust loading, and Table 6.1-7 and Figure 6.1-10 show this is the case.

6.2 Empirical Model of Efficiency

The best initial term for the modeling of combustion efficiency is the coal source with FIF = 1.427, see Table B-13 in Appendix B. Reference number is also high on the list. It was felt, and perhaps after later reflection improperly, that coal source should be considered a last-resort correction factor so it was avoided. It is probably legitimate to model these correction factors and then go after, for example, properties or constituent levels in the coals that correlate with the correction factors and could reasonably be responsible for these effects, such as BTU content. With fluidizing velocity removed the next best arithmetic FIF was 1.151 so multiplicative factors were explored. The final model is shown in Table 6.2-1 with scatterplot in Figure 6.2-1.

Table 6.1-7   Empirical model of outlet dust loading for all observations for which values of all variables were published

Symbols:

U = outlet dust loading, in tonnes/day
V = fluidizing velocity, in m./sec
A = bed area, in square meters
X = excess air, in percent
C = calcium to sulfur mole ratio
Z = coal top particle diameter, in microns
F = 1 or 0 indicator of fines recycled or not
G = MgO to CaO ratio in sorbent
T = bed temperature in degrees centigrade
P = parameter for effect of coal source
Q = parameter for effect of calcium source

Model based on data base for all experiments:

$$U = .1576(1.+min[0.,\{-16.774+24.56V-3.54V^2\}])(1.+3.53A)(1.-.0055X)*$$

$$(1.+.2124C+.00627C^2)(1.+.000097Z)(1.+.2165F)(1.-.014G)*$$

$$(1.+.0000122[1.8T-1555.]^2) P Q$$

[fit=arithmetic stand dev **2.94**]
on 118 experiments

Tables of parameters:

coal source	$P_{data\ base}$	calcium source	$Q_{data\ base}$
Pitt seam 8 washed	0.783	BCR1337 dolom	3.090
Pitt seam 8 unwashed	0.927	BCR1360 limes	1.720
Peabody Coal Co.	0.067	BCR1359 limes	3.713
Pitt seam unspecified	1.000	BCR1359H hyd lm	2.674
Commonwealth Edison	0.117	Tymochtee dolom	3.707
Park Hill coal	0.328	US limestone 18	0.143
Illinois coal unspec.	0.063	UK limestone	2.785
Welbeck coal unspec.	2.794	Stow-on-Kent	1.000

```
          0.        50.        100.       150.       200.       250.       300.       350.       400.
        |=========|=========|=========|=========|=========|=========|=========|=========|
 400. |
      |
      |
      |                                                                                    *
      |
350.  |                                                                                 *
      |                                                                              *
      |                                                                           *
      |                                                                         *
300.  |                                                                      *
      |                                                                    *
      |                                                                 *
      |                                                               *
250.  |                                                            *
      |                                                         *
      |                                                       *       C
      |                                                     *
      |                                                  *  C
200.  |                              C           CCC  *
      |                                          C  *  C
      |                                   H        *
      |                          D     CH      *
      |                            G  2  G  *   G
      |                          H           *
      |                  D         H  C  2C
150.  |                       HHH  *  H        H
      |                        *  HH
      |                       H        H        C
      |                     *CH
      |              GE  F  FHHH
      |             E   DD
100.  |      E        GFF*
      |        G E    *      H     H
      |              *
      |           *E      E
      |          *
      |     E       G
 50.  |          *    G
      |     E 2FE G
      |     EEGE     G
      |     * E
      |   *  E E
      6*
  0.  7G     G
        |=========|=========|=========|=========|=========|=========|=========|=========|
        0.        50.        100.       150.       200.       250.       300.       350.       400.
```

Predicted Dust Loading (lb/hr)

Actual Dust Loading (lb/hr)
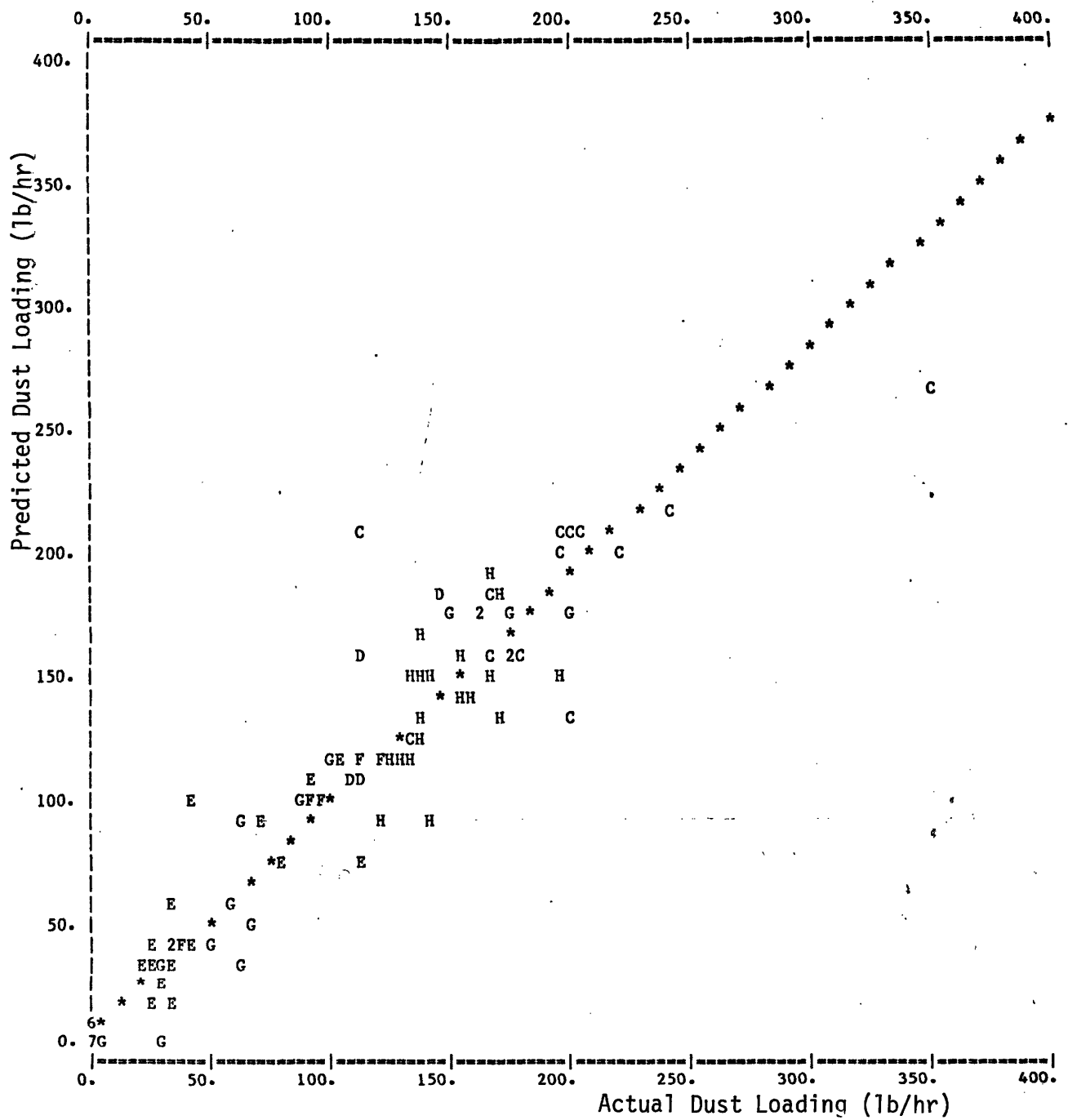
Figure 6.1-10 Scatterplot of outlet dust loading in lb/hr for observations
versus predictions, the 118 points are fit with an
arithmetic standard deviation of 6.48, lb/hr*.4536=tonne/day

Table 6.2-1 Empirical models of combustion efficiencies for which values
of all modeling parameters were available

Symbols:

    E = combustion efficiency, in percent
    V = superficial velocity, in m/sec
    D = static bed depth, in cm
    X = excess air, in percent
    A = bed area, in square meters
    C = calcium to sulfur mole ratio
    F = 1 or 0 indicator of fines recycled or not
    M = calcium particle mean size, in microns

Model based upon parametric experiments and examination of trends:

$$E=96.[1.-.036V][1.+.011max(-4.,[4.-.787D])](.972+.0035D)*$$

$$(.97+.008abs[min(7.,X)]\ )$$

Model based on data base for all experiments and least squares, rather than
main trend, modeling:

$$E = 77.79 + 14.11(1.- .1434V)(1.+.02433max[-4.,(4.-.787D)])(1.+.0746*$$

$$abs[min(7.,X)])(1.-.215A)(1.-.038C)(1.+.231F)(1.-.00016M)$$

[fit=arithmetic stand dev 2.95%]
on 144 experiments
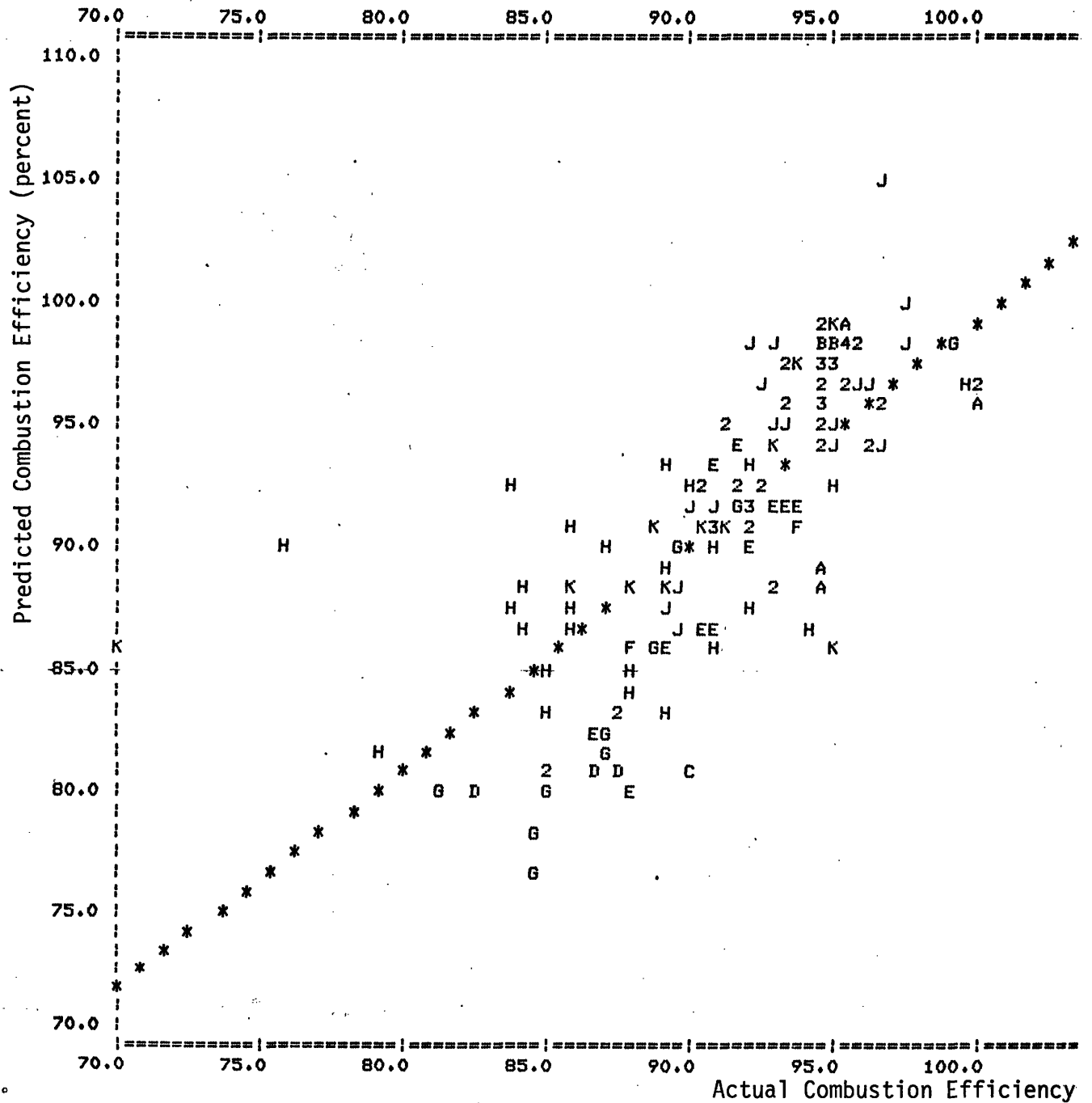
Figure 6.2-1 Scatterplot of combustion efficiency as observed versus predictions based on a data-based developed model, the least squares fit criterion concentrates on reducing outlyers, robust estimation techniques could be used to cluster the mass of points while ignoring outliers

VII. MODEL USES

Despite the fact that this is the final chapter of this report, the consideration of _model uses_ should really be the first task of any modeling effort. Referring to the list of objectives set down in Section 1.1, the _uses_ for the empirical models of an emerging, advanced energy technology can broadly be categorized as for:

1. _guidance_: toward the next best experiment or the next best design for analytic modeling; or for the setting of emission standards, and

2. _assessment_: of the expected performance and uncertainty of that performance measure; of the uncertainty or risk associated with the above guidelines; or of the gap between theoretical and empirical understandings.

Beginning this discussion with the category of guidance toward the next best experiment, there are several separate issues of importance. First, "best" must be uniquely defined. Surveys of the R & D literature, (Chen, Kirkwood, Lathrop, Pollock, 1977) and (Gruhl, _et al._, 1976), show that there are seven broad categories of objectives that might be used to define "best":

1. _economics_, in terms of cost of unit energy output, investment and operating costs;

2. _timeliness_, availability for commercial use and fit into energy-economic context;

3. _resource consumption_, including unpolluted air and water, materials, fuels, manpower, and capital;

4. _environmental_, safety, and health characteristics;

5. _basic research_, meaning those contributions that will also apply to other processes;

6. _institutional_ factors such as public image and government-industry interference and cooperation; and

7. _national security_, primarily in aiming at replacing or avoiding cartel-vulnerable products, such as oil or imported metals, and avoiding disruptions that could affect the survival of the establishment.

A quick look at this list points out the strict limitation of most empirical models, which can at most be used to tune in on best designs with respect to cost, efficiency, and emission performance measures. There is, however, no reason why some of those other performance measures could not be modeled and thus considered in the design of experiments. It must also be noted that empirical models that focus only upon expected values will be very much limited in applicability because decision makers are generally very risk averse with regard to many of the important performance indexes. Measures of uncertainty are required for these instances where a lower risk region such as a broadly level hill of high performance is much preferred compared to the absolute optimum performance that may be closely surrounded by disastrous chasms. The standard gradient search procedures for identifying optimum designs can easily be modified for consideration of risk aversion by using average or minimum performance over a range of uncertainty.

Aside from experiments aimed at best performance there are also experiments that are aimed at making the greatest reduction in the uncertainties in our knowledge about a technology. It is highly unlikely that the experiment that will provide the best performance will also reduce the most important uncertainties, and thus this is a classic dual control problem where the designer must divide, by relative weightings, his interest between performance and information. As a technology matures and becomes better understood, that weighting will slowly shift from entirely informational to entirely beneficial.

Aside from the easy cases where measurement errors are known to be responsible, the designing of experiments to most reduce the key uncertainties requires a very difficult procedure that involves:

1. prespecified priorities or weightings on the relative importance of the different types of uncertainties, a weighting that will usually vary over the range of the variables, that is, have regions where the uncertainty is more critical;

2. measures of the difficulties (including absolute constraints), such as costs, involved in changing the various variables (for example it may be less costly to change temperature than design), both for the specific experiment and for any subsequent final design, and

3. obtaining all of the available validation information about the empirical model, as the validation procedure is very similar to the process of reducing uncertainty.

With these pieces of information, some simple directions toward key data can be made, such as: (1) resolving discrepancies by repeating experiments, (2) performing tests on the experiments with the largest residuals, (3) interpolating between experiments, (4) aiming at reducing measurement errors with parametric investigations, and (5) developing correction factors that can be traced to new variables. Reducing widespread, persistent uncertainty is, however, still a formidable problem, and a well-constructed empirical model will generally, by definition, not be able to offer clues for reducing that type of uncertainty, because the modeler should previously have followed through on, and factored out, all these clues. Such widespread uncertainty may point toward missing, unmeasured variables which, of course, can only be identified by validation procedures and not by any other hints from the model. Aside from validation approaches, perhaps the only avenue left for exploring persistent uncertainty is through the highlighting of the most sensitive variables. One possible method is to trace the errors back through the model to find the minimum weighted distance change in inputs that could account for the error. Indications of which input variables may be responsible for the errors might then come from examinations of the correlations on input variables that seem to be most persistently accountable for the errors.

This procedure involves the study of the minimum compensating change in $\underline{x}_n$, called $\Delta\underline{x}_n$, using weights of the certainty with which its effects are felt to have been modeled $\underline{W}$ (a diagonal matrix of weights that may be composed of robust, measurement, or other confidence indicators), where $\Delta\underline{x}_n$ is such that

$$y_n = F(\underline{x}_n + \Delta\underline{x}_n, \underline{p}) \qquad (7.1)$$

with $\Delta\underline{x}'_n\underline{W}\Delta\underline{x}_n$ minimized. Although this is not a panacea for the problem of persistent uncertainties, it will show in some sense where the responsibility for the uncertainties can be most easily relegated.

Going back to the list of potential model uses, the design of the next best facility is completely analogous to the design of the next best experiment. For the next best experiment, the design parameters are generally fixed, and the optimization takes place in the operating variable space; for the next best facility the optimization takes place over design and operating variable space; the situations are otherwise identical.

The comparison of empirical models with analytic results has been extensively discussed. If the empirical model is simple enough, then direct structural comparisons will be possible. If either the analytic or the empirical model is complex, then side by side parametric investigations may be necessary. Such comparison of models in many respects is quite similar to some of the comparative validation procedures, and Table 5-2 contains a list of possible comparison methodologies.

The assessment categories of model uses are the most easily accomplished. In these cases the situation is identical to the guidance uses that were initially described except that there is no interest here in the uncertainty reduction aspects of that use. Models built for assessment, and the assessments themselves, should be more diligently

validated and scrutinized as it is quite common for these products to show biases that are favorable to the technology. These biases are unconscious, but to be expected, whenever modelers and investigators, being experts in an emerging technology, naturally have a stake in favorable outcomes.

VIII. REFERENCES AND BIBLIOGRAPHY

Afifi, A.A. and S.P. Azen, 1972. _Statistical Analysis: A Computer Oriented Approach_, Academic Press, New York.

Aitkin, M.A., 1974. "Simultaneous Inference and the Choice of Variable Subsets in Multiple Regression", _Technometrics_, 16, pp. 221-229.

Allen, D.M., 1971. "Mean Square Error of Prediction as a Criterion for Selecting Variables", _Technometrics_, 13(3), pp. 469-476.

Allen, D.M., 1974. "The Relationship Between Variable _Selection_ and Data Augmentation and a Method for Prediction", _Technometrics_, 16, pp. 125-127.

Andrews, D.F., 1974. "A Robust Method for Multiple Linear Regression", _Technometrics_, 16, pp. 125-127.

Andrews, D.F., _et. al._, 1972. _Robust Estimates of Location_, Princeton University Press.

Ariatti, F., _et. al._, 1972. "Methods for Electric Power System State Estimation", _Proc. 4th PSCC_, Grenoble, France.

Arnstein, S.R., 1975. "Public Participation in Technology Assessment - Working Model for Public Participation", _Public Administration Review_, 35 (1), pp. 70-73.

Arnstein, S.R., 1977. "Technology Assessment: Opportunities and Obstacles", _IEEE Transactions on Systems, Man, and Cybernetics_, SMC7(8), pp. 571-582, August.

Ashton, R.H., 1976. "Predictive Ability Criterion and User Prediction Models", _Accounting Review_, 51(3), pp. 680-682 and 49(4), pp. 719-732.

Azzi, C.F. and J.C. Cox, 1974. "Shadow Prices in Public Program Evaluation Models", _Quarterly Journal of Economics_, 88(1), pp. 158-165.

Baker, N.R., 1974. "R and D Project Selection Models - Assessment", _IEEE Transactions on Engineering Management_, 21(4), pp. 165-171.

Baker, N.R., 1975. "R and D Project Selection Models - Assessment", _R & D Management_, 5, pp. 105-111.

Beale, E.M., M.G. Kendell, D.W. Mann, 1967, "The Discarding of Variables in Multivariate Analysis", _Biometrika_, 54, pp. 357-366.

Beaton, A.E. and J.W. Tukey, 1974. "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", _Technometrics_, 16, pp. 147-185.

Bibliography Continued

Bendel, R.B. and A.A. Afifi, 1976. "A Criterion for Stepwise Regression", The American Statistician, 30(2), pp. 85-87, May.

Bendel, R.B. and A.A. Afifi, 1977. "Comparison of Stopping Rules in Forward 'Stepwise' Regression", J. Am. Stat. Assoc., 72(357), p. 46, March.

Benedetti, J.K., 1977. "On the Nonparametric Estimation of Regression Functions", J. R. Statist. Soc. B, 2, pp. 248-253.

Berndt, E.R. and N.E. Savin, 1977. "Conflict Among Criteria for Testing Hypotheses in Multivariate Linear-Regression Model", Econometrica, 45(5), pp. 1263-1277.

Bettman, J.R., 1973. "Perceived Risk and its Components - Model and Empirical Test", Journal of Marketing Research, 10(2), pp. 184-190.

Bhansali, R.J. and D.Y. Downham, 1977. "Some Properties of Order of an Autoregressive Model Selected by a Generalization of Akaikes EPF Criterion", Biometrika, 64(3), pp. 547-551.

Bitz, M., 1975. "Models for Information Evaluation", Zeitschrift fur Betriebswirtschaft, 45(9), pp. 521-546.

Blair, J.P. and S.M. Maser, 1977. "Axiomatic Versus Empirical - Models in Policy Studies", Policy Studies Journal, 5(3), pp. 282-289.

Boshier, J.F., F.C. Schweppe, and J. Gruhl, 1978. "Methodologies for Assessing the Validity of Complex Models", MIT Energy Lab Report, Cambridge, Mass.

Boshier, J., F.C. Schweppe, and J. Gruhl, forthcoming. "Validity and Uncertainty of Predictive Energy Models", Proceedings Sixth Power Systems Computation Conference, Darmstadt Germany, 8 pp., August, 1978.

Box, G.E.P., M.S. Phadke and G.C. Tiao, 1976. "Empirical-Mechanistic Modelling of Air Pollution", Proceedings of the Fourth Symposium on Statistics and the Envirionment.

Bunke, H. and O. Bunke, 1974. "Empirical Decision Problem and Choice of Regression Models", Biometrische Zeitschrift, 16(3), pp. 167-184.

Chen, K., C.W. Kirkwood, J.W. Lathrop, S.M. Pollock, 1977. "Energy R&D Modeling for Budgetary Decision", IEEE Transcations on System, Man, and Cybernetics, SMC7(4), pp. 222-227, April.

Christensen, L.R., 1973. "Simultaneous Statistical Inference in the Normal Multiple Linear Regression Model", J. Amer. Statist. Ass., 68, pp. 457-461.

Bibliography Continued

Clark, R.M., 1977. "Non-Parametric Estimation of a Smooth Regression Function", J.R. Statist. Soc. B, pp. 107-113.

Coates, J.F., 1976. "Role of Formal Models in Technology Assessment", Technological Forecasting And Social Change, 9(1-2), pp. 139-190.

Couch, G.H., A.C. Sullivan and J.A. Dembecki, 1974. "A State Estimator Oriented to a Semi-Isolated 5 GW Power Station", IEEE 1974 Summer Power Meeting, paper number C 74 346-3.

Daniel, C. and F.S. Wood, 1971. Fitting Equations to Data, John Wiley and Sons, Inc., New York.

Das Gupta, S. and M.D. Perlman, 1974. "Power of the Noncentral F-Test: Effect of Additional Variables on Hottelings T^2-Test", J. Amer. Statist. Ass., 69, pp. 174-180.

Debs, A.S., R.E. Larson and L.S. Hajdu, 1972. "On-Line Sequential State Estimation", Proc. 4th PSCC, Grenoble, France.

Dhrymes, P.J., et. al., 1972. "Criteria for Evaluation of Econometric Models", Annals of Economic and Social Measurement, I, pp. 291-324, Summer.

Dixon, W.J. and J.W. Kuzma, 1974. "Data Reduction in Large Clinical Trials", Comm. in Statist., 3, pp. 301-324.

Doksum, K., 1974. "Empirical Probability Plots and Statistical Inference for Nonlinear Models in 2-Sample Cases", Annals of Statistics, 2(2), pp. 267-277.

Dopazo, J.F., O.A. Klitin and L.S. VanSlyck, 1972. "State Calculation of Power Systems from Line Flow Measurements, Part II", IEEE Transactions, PAS-91, pp. 145-151.

Dopazo, J.F., O.A. Klitin and A.M. Sasson, 1973. "State Estimation for Power Systems: Detection and Identification of Gross Measurement Errors", Paper X-2, PICA Conference, Minneapolis.

Draper, N.R. and H. Smith, 1966. Applied Regression Analysis, John Wiley & Sons, Inc., New York.

Draper, N.R. et al., 1971. "The Distribution of Certain Regression Statistics", Biometrika, 58, pp. 295-298.

Efroymson, M.A., 1960. "Multiple Regression Analysis", in A. Ralston and H.S. Wilf (eds), Mathematical Methods for Digital Computers, John Wiley & Sons, New York.

Ezekiel, M. and K.A. Fox, 1959. Methods of Correlation and Regression Analysis, Third Edition, John Wiley & Sons, New York, p. 342.

Bibliography Continued

Farris, D.R. and A.P. Sage, 1975. "Use of Interpretive Structural Modeling to Obtain Models for Worth Assessment", Battelle Monographs, pp. 153-159, June.

Finney, D.J., 1974. "Problems, Data and Inference-Presidential Address", J.R. Statist. Soc., A, 137, pp. 1-18.

Fromm, G., W.L. Hamilton, and D.E. Hamilton, 1975. "Federally Supported Mathematical Models - Survey and Analysis", U.S.G.P.O. #038-000-00221-0, Washington, D.C.

Furnival, G.M. and R.W. Wilson, 1974. "Regression by Leaps and Bounds", Technometrics, 16, pp. 499-511.

Gallant, A.R., 1975. "Nonlinear Regression", The American Statistician, 29(2), pp. 73-81, May.

Gallant, A.R. and W.A. Fuller, 1973. "Fitting Segmented Polynomial Models Whose Join Points Have to Be Estimated", J. Am. Stat. Assoc., 68, pp. 144-147.

Gass, S.I., 1977a. "Evaluation of Complex Models", Computers and Operations Research, 4(1), March.

Gass, S.I., 1977b. "A Procedure for the Evaluation of Complex Models", Proceedings First International Conference on Mathematical Modeling, St. Louis, September.

Gnanadesikan, R., 1977. Methods for Statistical Data Analysis of Multivariate Observations, Wiley, New York.

Greenberger, M., 1977. "Closing the Circuit between Modelers and Decision Makers", EPRI Journal, pp. 6-12, October.

Greville, T.N.E., 1969. "Introduction to Spline Functions", in Theory and Applications of Spline Functions, T.N.E. Greville (ed.), Academic Press, New York, pp. 1-36.

Gruhl, J., et al., 1976. "Systematic Methodology for Comparison of Environmental Coal-Fired Electric Generation", MIT Energy Lab Technical Report MIT-EL 76-012, for Argonne National Labs, 220 pp., November.

Gruhl, J. and J.D. Teare, 1978. "Empirical Models of Emissions and Energy Efficiencies of Coal-Fired Fluidized Bed Power Plants", MIT Energy Lab Report (forthcoming).

Gungor, R.G., N.F. Tsang, and B. Webb, 1971. "A Technique for Optimizing Real and Reactive Power Schedules", IEEE Transactions, PAS-90, pp. 1781-1790.

Bibliography Continued

Geisser, S., 1975. "The Predictive Sample Reuse Method with Applications", J. Amer. Statist. Assoc., 70, pp. 320-328.

Hampel, F.R., 1973. "Robust Estimation: A Condensed Partial Survey", Zeitschrift Wahrscheinlichkeitstheorie, 27, pp. 87-104.

Handschin, E. (editor), 1972. Real-Time Control of Electric Power Systems, Chap. 2, Elsevier, Amsterdam.

Handschin, E., H. Glavitsch and J. Kohlas, 1973. "Estimation und Entdeckung Schlechter Messdaten," EDV in der Elektrizitatswirtschaft, Seminar des Oesterreichischen Produktivitatszentrum,, Vienna, April 11-12.

Handschin, E. and J. Kohlas, 1972. "Statistical Data Processing for Power System Operation", Proc. 4th PSCC, Grenoble, France.

Handschin, E., F.C. Schweppe, J. Kohlas and A. Fiechter, 1973. "Bad Data Analysis in Electric Power Systems", Brown Boveri Research Report.

Handschin, E., F.C. Schweppe, J. Kohlas and A. Fiechter, 1975. "Bad Data Analysis for Power System State Estimation", IEEE Transactions, PAS-94, pp. 329-337.

Hansen, J.V. and L.E. Heitger, 1977. "Models for Evaluating Information Updating Intervals", Information Sciences, 13(2), pp. 123-135.

Hartley, H.O., 1961. "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares", Technometrics, 3, pp. 269-280.

Harvey, A.C., 1977. "A Comparison of Preliminary Estimators for Robust Regression", J. Am. Stat. Assoc., 72(360), pp. 910-913, December.

Hawkins, D.M., 1976. "The Subset Problem in Multivariate Analysis of Variance", J. R. Statist. Soc. B, 2, pp. 132-139.

Helms, R.W., 1974. "Average Estimated Variance Criterion for Selection-of-Variables Problem in General Linear Models", Technometrics, 16(2), pp. 261-273.

Heppe, R.J. and V. Selman, 1975. "Game Theory Technology Assessment Model for Aiding National Policy in Energy Research Alternatives", Operations Resarch, 23(2), pB430.

Hill, R.W. and P.W. Holland, 1977. "Two Robust Alternatives to Least-Squares Regression", J. Am. Stat. Assoc., 72(360), pp. 828-833, December.

Hill, W.J. and R.A. Wiles, 1975. "Plant Experimentation", J. Quality Technology, 7, pp. 115-122.

Bibliography Continued

Hinich, M.J. and P.P. Talwar, 1975. "A Single Method for Robust Regression", J. Am. Stat. Assoc., 70, pp. 113-119.

Hoerl, A.E. and R.W. Kennard, 1970a. "Ridge Regression: Biased Estimation for Nonorthogonal Problems", Technometrics, 12, pp. 55-68.

Hoerl, A.E. and R.W. Kennard, 1970b. "Ridge Regression: Applications to Nonorthogonal Problems", Technometrics, 12, pp. 69-82.

Hoffman, K.C. and D.W. Jorgenson, 1977. "Economic and Technological Models for Evaluation of Energy-Policy", Bell Journal of Economics, 8(2), pp. 444-466.

Hogg, R.V., 1974. "Adaptive Robust Procedure: A Partial Review and Some Suggestions for Future Applications and Theory", J. Am. Stat. Assoc., 69, pp. 909-923, December.

Huber, P.J., 1972. "Robust Statistics: A Review", Annals of Mathematical Statistics, 43, pp. 1041-1067.

Huber, P.J., 1973. "Robust Regression: Asymptotics, Conjectures and Monte Carlo", Annals of Statistics, 1, pp. 799-821.

Ivakhnenko, A.G., et. al., 1972, "GMDH Algorithm with Successive Determination of Polynomial Trends on the Basis of Most Essential Variables", Avtomatyka (Soviet Automatic Control), No. 2.

Ivakhnenko, A.G., M.M. Todua and A.A. Forichev, 1974. "Heuristic Self-Organization", Avtomatyka (Soviet Automatic Control), pp. 94-103.

Jennrich, R.I., 1969. "Asymptotic Properties of Nonlinear Least Squares Estimators,", The Annals of Mathematical Statistics, 40, pp. 633-643.

Jonke, A.A., 1970. "Reduction of Atmospheric Pollution by the Application of Fluidized Bed Combustion", ANL/ES-CEN-1002, Argonne National Lab, Argonne Ill.

Jonke, A.A., 1971. "Reduction of Atmospheric Pollution by the Application of Fluidized Bed Combustion", Argonne National Lab Annual Report, July 1970-June 1971.

Jonke, A.A., 1975. "Fluidized Bed Combustion: A Status Report, Coal Processing Technology", CEP Tech. Manual, Vol. 2.

Judge, G.G., M.E. Bock and T.A. Yancey, 1974. "Post Data Model Evaluation", Review of Economics and Statistics, 56(2), pp. 245-253.

Kennard, R.W. and L. Stone, 1969. "Computer Aided Design of Experiments", Technometrics, 11, pp. 137-148.

Kennedy, W.J. and T.A. Bancroft, 1971. "Model Building for Prediction in Regression Based Upon Repeated Significance Tests", Annals of Mathematical Statistics, 42(4), pp. 1273-1284.

Bibliography Continued

Khilnani, A., 1974. "Parameter Sensitivity in Non-Linear Models", MIT
    Electrical Engineering Thesis, Cambridge, Mass., June.

Kohlas, J., 1972. "On Bad Data Suppression in Estimation", IEEE
    Transactions, AC-17, pp. 827-828.

Kissler, G.R., 1974. "Evaluation of Computer-Based Laboratory Simulation
    Models to Teach Scientific Research Strategies", Behavior Research
    Methods and Instrumentation, 6(2), pp. 124-126.

Larkey, P.D., 1977. "Process Models of Government Resource-Allocation and
    Program-Evaluation", Policy Sciences, 8(3), pp.269-301.

Lathrop, J.W. and K. Chen, 1976. "Comprehensive Evaluation of Long-Range
    Research and Development Strategies", IEEE Transaction on Systems,
    Man, and Cybernetics, SMC6(1), pp. 7-17, January.

Levine, C.H., 1976. "Leadership - Problems, Prospects, and Implications of
    Research Programs Arrived at Linking Empirical and Normative
    Modeling", Policy Studies Journal, 5(1), pp. 34-41.

Limaye, D.R., 1976. "Energy Policy Evaluation Modeling and Simulation
    Approaches", J. of the Am. Inst. of Plan., 42(3) p. 352.

Louis, J. and S.E. Tung, 1977. "Quarterly Report on Fluidized Bed Data
    Base Contract", M.I.T. Energy Lab, Cambridge, Mass.

Maciarello, J.A., 1975. "Dynamic Benefit Cost-Analysis - Policy Evaluation
    in a Dynamic Urban Simulation Model", Socio-Economic Planning
    Sciences, 9 (3-4), pp. 147-168.

Mallows, C.L., 1973. "Some Comments on C_p", Technometrics, 15(4),
    pp. 317-326.

Mandel, J., 1956. "Non-Additivity in Two-Way Analysis of Variance",
    J. Amer. Statist. Assoc., 56, pp. 878-888.

Mandel, J., 1964. The Statistical Analysis of Experimental Data, Wiley
    Interscience Publ., New York, pp. 261-266.

Mandel, J., 1969. "A Method for Fitting Empirical Surfaces to Physical or
    Chemical Data", Technometrics, 11, pp. 411-429.

Mandel, J. and F.L. McCrackin, 1963. "Analysis of Families of Curves",
    Journal of Research - National Bureau of Standards, Section A.
    Physics and Chemistry, 67A(3), pp. 259-267.

Manne, A.S., 1976. "ETA-Model for Energy Technology-Assessment", Bell
    Journal of Economics, 7(2), pp. 379-406.

Marquardt, D.W., 1963. "An Algorithm for Least Squares Estimation of
    Nonlinear Parameters", Journal of the Society for Industrial and
    Applied Mathematics, 11, pp. 431-441.

Bibliography Continued

Marquardt, D.W., 1970. "Generalized Inverses, Ridge Regression, Biased
Linear Estimation, and Nonlinear Estimation", Technometrics, 12,
pp. 591-612.

Marquardt, D.W., and R.D. Snee, 1975. "Ridge Regression in Practice",
American Statistician, 29, pp. 3-20.

Masiello, R.D. and F.C. Schweppe, 1971. "Power System Tracking State
Estimation", IEEE Transactions, PAS-90, pp. 1025-1033.

Maxim, L.D. and C.L. Brazie, 1973. "Multistage Input-Output Model for
Evaluation of Environmental Impact of Energy Systems", IEEE
Transaction on Systems, Man and Cybernetics, VSMC3(6), pp. 583-587.

McFarlan, D.D., 1974. "On Relationship Between Mathematical Models and
Empirical Data...," editorial in American Sociologist Review, 39(6),
pp. 883-885.

McCarthy, P.J., 1976. "The Use of Balanced Half-Sample Replication in
Cross-Validation Studies", J. Amer. Statist. Assoc., 71, pp. 596-604.

McNeil, D.R., 1977. Interactive Data Analysis, Wiley, New York.

Merrill, H.M. and F.C. Schweppe, 1971. "Bad Data Suppression in Power
System Static-State Estimation", IEEE Transactions, PAS-90,
pp. 2718-2725.

Merrill, H.M. and F.C. Schweppe, 1973. "On-Line System Model Error
Correction", IEEE Winter Power Meeting 1973, Paper C 73 106-2.

M.I.T. Model Assessment Laboratory, 1978. "Policy Model Assessment: Two
Case Studies", M.I.T. Energy Lab report for EPRI, (forthcoming), May.

Morgan, B.S., 1966. "Sensitivity Analysis and Synthesis of Multivariable
Systems," IEEE Transactions, AC-11, pp. 506-512.

Mosier, C.I., 1951. "Symposium: The Need and Means of Cross-Validation",
Education and Psychological Measurement, 11, pp. 5-11.

Mosteller, F., and J.W. Tukey, 1977. Data Analysis and Regression,
Addison-Wesley, Reading, Mass.

National Bureau of Economic Research, 1974. "TROLL Experimental Programs:
Interactive Data Analysis, Econometrics and Numerical Methods", NBER,
Computer Research Center for Economics and Management Science,
Cambridge, Mass.

National Coal Board, 1971a. "Reduction of Atmospheric Pollution", Vol. 2,
NCB, September.

National Coal Board, 1971b. "Reduction of Atmospheric Pollution", Vol. 3,
NCB, September.

Bibliography Continued

Novick, M.R., P.H. Jackson, D.T. Thayer and N.S. Cole, 1972. "Estimating Multiple Regression in m Groups: A Cross-Validation Study", Br. J. Math. Statist. Psychol., 25, pp. 33-50.

Park, C.N. and A.L. Dudycha, 1974. "Cross-Validation Approach to Sample Size Determination for Regression Models", Journal of the American Statistical Association, 69(345), pp. 214-218.

Peters, B.G., 1975. "Non-Additive Models for Policy Research-Longitudinal Evaluation", Western Political Quarterly, 28(3), pp. 542-547.

Peterson, D.W. and F.C. Schweppe, 1974. "Code for a General Purpose System Identifier and Evaluator (GPSIE)", IEEE Transactions, AC-19(6), December.

Poretta, B. and R.S. Dhillon, 1973. "Performance Evaluation of State Estimation from Line Flow Measurements on Ontario Hydro Power System", IEEE Winter Power Meeting 1973, Paper T 73 086-6.

Priestly, M.B. and M.T. Chao, 1972. "Non-Parametric Function Fitting", J. R. Statist. Soc. B, 34, pp. 385-392.

Rao, C.R., 1965. Linear Statistical Inference and its Applications, Wiley & Sons, New York, N.Y.

Reggiani, M.G. and F.E. Marchett, 1975. "On Assessing Model Adequacy", IEEE Transactions on Systems, Man and Cybernetics, SMC5(3), pp. 322-330.

Reinsch, C.H., 1971. "Smoothing by Spline Functions, II", Numer. Math., 16, pp. 451-454.

Ribbert, B., 1977. "Decision-Oriented Evaluation Models and Enterprise Evaluation by a Neutral Expert", Zeitschrift fur Betriebswirtschaft, 47(9), pp. 599-603.

Robison, E.B., 1970. "Characterization and Control of Gaseous Emissions from Coal-Fired Fluidized Bed Boilers", Pope, Evans, and Robbins, NTIS #PB198 412, October.

Rosenblatt, M., 1956. "Remarks on Some Nonparametric Estimates of a Density Function", Annals Mathematical Statistics, 27, pp. 832-837.

Rosenblatt, M.. 1971. "Curve Estimates", Ann. Math. Statist., 42, pp. 1815-1842.

Sahal, D., 1977. "Structural Models of Technology-Assessment", IEEE Transactions on Systems, Man, and Cybernetics, 7(8), pp. 582-589.

Savage, L.J., 1962. The Foundations of Statistical Inference, Methuen, London, p. 34.

Schlossmacher, E.J., 1973. "An Iterative Technique for Absolute Deviations Curve Fitting", _Journal of the American Statistical Association_, 68, pp. 857-865.

Schweppe, F.C., 1970. "Power System Static-State Estimation - Part III: Implementation", _IEEE Transactions_, PAS-89, pp. 130-135.

Schweppe, F.C., 1973. _Uncertain Dynamic Systems_, Prentice Hall, Englewood Cliffs, N.J.

Schweppe, F.C. and D.B. Rom, 1970. "Power System Static-State Estimation - Part II: Appropriate Model", _IEEE Transactions_, PAS-89, pp. 125-130.

Schweppe, F.C. and J. Wildes, 1970. "Power System Static-State Estimation - Part I: Exact Model", _IEEE Transactions_, PAS-89, pp. 120-125.

Stagg, G.W., J.F. Dopazo, O.A. Klitin and L.S. VanSlyck, 1970. "Techniques for Real-Time Monitoring of Power System Operation", _IEEE Transacations_, PAS-89, pp. 545-555.

Scott, J.E. and P. Wright, 1976. "Modeling an Organizational Buyers Product Evaluation Strategy -- and Procedural Considerations", _Journal of Marketing Research_, 13(3), pp. 221-224.

Scott, J.R., 1975. "Regression Assessment Statistic", _Journal of the Royal Statistical Society, Series C, Applied Statistics_, 24(1) pp. 42-45.

Sharif, N. and P. Adulbhan, 1975. "Mathematical Model for Industrial Project Evaluation", _Omega International Journal of Management Science_, 3(5), pp. 595-604.

Shibata, R., 1976. "Selection of Order of an Autoregressive Model by Akaikes Information Criterion", _Biometrika_, 63(1), pp. 117-126.

Shubik, M. and G.B. Brewer, 1972. "Models, Simulations and Games - A Survey", Rand Corp. Report R-1060-ARPA/RC, Santa Monica Calif., May.

Simon, H.A., 1953. "Prediction and Hindsight as Confirmatory Evidence", _Phil. Sci._, 22, pp. 227-230.

Smith, A.W., 1973. "Information Model and Evaluation Tool for R&D Projects and Contractor Selections - PRECEPT", _Industrial Marketing Management_, 2(2), pp. 163-175.

Snee, R.D., 1973. "Some Aspects of Nonorthogonal Data Analysis, Part I. Developing Prediction Equations", _J. Quality Technology_, 5, pp. 67-79.

Snee, R.D., 1977. "Validation of Regression Models - Methods and Examples", _Technometrics_, 19(4), pp. 415-428.

Sohnquist, J.A., 1970. Multivariate Model Building - Validation of a Search Strategy" _Am. Polit. Sci. Rev._, 68(4), 1974.

Bibliography Continued

Stone, M., 1974. Cross-Validating Choice and Assessment of Statistical Predictions", *J. Roy. Statist. Soc. B*, 36, pp. 111-147.

Strom, S.S., et. al., 1976. "Summary Evaluation of Atmospheric Pressure Fluidized Bed Combustion Applied to Electric Utility Large Steam Generators", Vol. III, EPRI Report FP-308, Palo Alto, CA.

Stuart, T.A. and C.J. Herget, 1973. "A Sensitivity Analysis of Weighted Least-Squares State Estimation for Power Systems", IEEE Winter Power Meeting, Paper T 73 085-8.

Subrahmanyan, K., 1971. "Discrimination in the Presence of Co-Variables", S.A. Statist. J., 5, pp. 5-14.

Tukey, J.W., 1965. "The Future of Processes of Data Analysis", *Proceedings of the Tenth Conference on the Design of Experiment in Army Research Development and Testing*, ARO-D Report 65-3, U.S. Army Res. Off., Durham.

Tukey, J.W., 1977. *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass.

Vandeven, A.H. and R. Koenig, 1976. "Process Model for Program Planning and Evaluation", *Journal of Economics and Business*, 28(3), pp. 161-170.

Wahba, G. and S. Wold, 1975. "A Completely Automatic French Curve: Fitting Spline Functions by Cross-Validation", *Comm. Statist.*, 4, pp. 1-17.

Welsch, R.E., 1974. "Validation and the Jackknife", M.I.T. Sloan School of Management, course notes, Cambridge, Mass., September 24.

Welsch, R.E., 1976. "Confidence Regions for Robust Regression", *Proceedings of the American Statistical Association*, Statistical Computing Section.

Wilkinson, G.N., 1970. "A General Recursive Procedure for Analysis of Variance", *Biometrika*, 57, pp. 19-46.

Wood, D.O., 1978. "Notes on the M.I.T. Experiment in Establishing a Model Assessment Laboratory", M.I.T. Energy Lab, Cambridge, Mass., 4 pp., March.

Woodford, C.H., 1970. "An Algorithm for Data Smoothing Using Spline Functions", *BIT*, 10, pp. 501-510.

Appendixes, pages 93 to 177, available upon special request.