MODELLING FACTORS AFFECTING RELATIVE PERFORMANCE OF TWO DUST SAMPLERS IN LABRADOR MINES

MOHAMMAD NURUL AZAM

ISBN 0-315-50457-9

MODELLING FACTORS AFFECTING RELATIVE PERFORMANCE OF TWO DUST SAMPLERS IN LABRADOR MINES

By

©MOHAMMAD NURUL AZAM

Submitted to the School of Graduate Studies in
partial fulfillment of the requirements
for the degree of Master of
Applied Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland
November, 1988

St.John's Newfoundland Canada

ABSTRACT

This study was undertaken for modelling the effect of temperature and humidity on relative performance of two dust samplers in two mining complexes in Labrador, Canada. Data were collected from five major locations. A statistical analysis is given to identify the distributional pattern of dust concentration. Then the multiple regression based on least squares as well as M-estimation of regression techniques were applied for modelling.

ACKNOWLEDGEMENTS

I express my deep sense of gratitude to Dr. B. C. Sutradhar for providing me the opportunity to work in this project and giving me his wise counsel, encouragement and support on countless occasions. Dr. Sutradhar has been exceedingly generous with his time, and he answered my questions and arguments with patience, courtesy and unfailing good humor. Dr. Sutradhar's advice and critical comments formed a very useful part of my training. Thanks also to Dr. E. Moore for being available for discussion whenever required. I can never forget his readiness at all times to discuss problems and provide encouraging comments. I owe a very special intellectual debt to him. Thanks also Dr. T. Mak for his value-able suggestions and Dr. P.P. Narayanaswami for helping me with computer programming.

Needless to say my participation in this work would not have been possible without the financial support which helped me to go through the graduate program. Fellowship and Bursary support was generously provided by the School of Graduate Studies of Memorial University of Newfoundland, Canada.

TABLE OF CONTENTS

ABSTRACT .... (i)

ACKNOWLEDGEMENTS .... (ii)

TABLE OF CONTENTS .... (iii)

LIST OF TABLES .... (v)

LIST OF FIGURES .... (vii)

CHAPTER 1. INTRODUCTION .... (1)

1.1 Background .... (1)

1.2 Plan of the study .... (6)

CHAPTER 2 . DISTRIBUTIONAL ASPECTS OF DATA .... (8)

2.1 Exploratory data analysis .... (9)

2.1.1 The Box and whisker plot of the original data .... (9)

2.1.2 Box and whisker plot excluding outliers .... (12)

2.1.3 Histograms excluding outliers .... (14)

2.1.4 Letter value display excluding outliers .... (20)

2.2 Description of the variable Y for different locations .... (24)

2.3 Confirmatory test for normality .... (26)

2.3.1 Maximum likelihood method for λ .... (26)

CHAPTER 3. FITTING MODELS TO THE DATA .... (36)

3.1 Selection of the best model .... (36)

3.2 Selection criteria .... (38)

3.3 Best selected model for different locations .... (52)

3.4 Performance of the regression model .... (53)

3.5 Partial effect of humidity or temperature on concentration .... (54)

3.6 Graphical check for goodness of fit .... (55)

CHAPTER 4. M-ESTIMATE OF REGRESSION .... (61)

4.1 Distribution of Regression estimator $\hat{\gamma}$ .... (65)

4.2 Testing the regression function .... (66)

4.3 Selected models for different locations .... (82)

4.4 Performance of the selected models .... (83)

4.5 Partial effect of humidity or temperature on concentration .... (84)

4.6 M-estimates : A comparison with multiple regression .... (85)

CHAPTER 5 . SUMMARY AND CONCLUSIONS .... (92)

REFERENCES .... (95)

LIST OF TABLES

2.1 Number of observations in this study .... (12)

2.2 Letter Value Display for Different Locations .... (21)

2.3 Values of λ and $L_{max}\lambda$ for different locations .... (35)

3.1 Values and P-values of coefficients for Location 1251 .... (39)

3.2 Values and P-values of coefficients for Location 1312 .... (41)

3.3 Values and P-values of coefficients for Location 1341 .... (43)

3.4 Values and P-values of coefficients for Location 2231 .... (45)

3.5 Values and P-values of coefficients for Location 2233 .... (47)

3.6 Values of R^2 for 31 different models for different locations .... (49)

3.7 Values of Adjusted R^2 for different locations .... (50)

3.8 Values of sum of squares of error for different locations .... (51)

3.9 Estimated Z and Y values for different locations .... (54)

3.10 Partial change of concentration for different locations .... (55)

4.1 Values and P-values of coefficients for Location 1251 .... (69)

4.2 Values and P-values of coefficients for Location 1312 .... (71)

4.3 Values and P-values of coefficients for Location 1341 .... (73)

4.4 Values and P-values of coefficients for Location 2231 .... (75)

4.5 Values and P-values of coefficients for Location 2233 .... (77)

4.6 P-Values of Chi-squre for different locations .... (79)

4.7 Values of the intercept for different locations .... (79)

4.8 Values of sum of squares of error (SSE) for different locations .... (81)

4.9 Estimated Y values for different locations .... (84)

4.10 Partial change of concentration for different locations .... (85)

LIST OF FIGURES

2.1 Box plot for Different Locations .... (10)

2.2 Box plot after deletion of outliers for different locations .... (13)

2.3 Histogram for different locations .... (15)

2.4 Plot of $L_{\max}(\lambda)$ versus λ for different locations .... (30)

3.1 Multiple plot of Z and $\hat{Z}$ for different locations .... (56)

4.1 Multiple plot of Y, $\hat{Y}_{ro}$ and $\hat{Y}_{ls}$ for different locations .... (86)

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

It is well-known that the respiratory health condition of the workers in a mine is affected by concentration of dust mainly deposited due to production activities [cf. Fairman et al.(1977); Musk et al. (1977); Gregory (1971)]. In planning for the prevention of high dust concentrations, it is important to measure the magnitude of dust produced in a mine. One requires a scientific instrument for such measurements. Common scientific instruments used in mining to measure dust concentration are:

1. Nylon Cyclone (NY) - designed to approximate the American Conference of Governmental Industrial Hygienists (ACGIH) criteria.

2. H & H Custom Metal Cyclone (ME)- designed to approximate the British Medical Research Council (BMRC) criteria.

Nylon cyclones (NY) are considerably smaller in size and weight than some other size-selection devices, and are not seriously affected by orientation. These characteristics of the nylon cyclone are particularly useful in personal sampling. For many years, the nylon cyclone, designed to conform to the ACGIH criterion, has been routinely used by the U.S. Public Health Service as a size-selective presampler to remove the non-respirable fraction of the dust. A metal cyclone of comparable size and weight has been developed in Britain. The most commonly used metal cyclone, designed by Higgins (1967), conforming to the BMRC criterion, is manufactured by Casella, England.

We note here that there are some other instruments available to measure dust concentration. For example, beginning in 1962, the methods employed to sample dust conditions centered on the use of the midget impinger. This device, developed prior to the 1950's, is a short period dust sampling instrument; it usually samples a cubic foot volume of air over a 10 minute period. It is apparent that the determination of the average dust concentration over an 8 hour shift, by means of the midget impinger, would involve the collection of a considerable number of samples and the time -consuming task of microscopy work. Even if this approach were acceptable, the accuracy of the end result would be subject to the limitations of the equipment and accompanying errors. While it should be appreciated that these instruments were not designed to assess average dust concentrations, they have satisfied the need as an engineering tool. Their main advantage includes the ability to measure peak concentrations and provide the results in a matter of hours. As a result , they serve the useful function of locating individual sources of dust, determining the effectiveness of dust control systems, and measuring dust collector efficiencies, etc.

As ME and NY (unlike the instruments of 50's and 60's) are to assess the magnitude of dust concentration in a production area , they were used in two mining complexes in Labrador in 1980 to measure dust concentration. Areas were selected throughout the mines to give a good representation of the general atmosphere under all working conditions. Once the areas were chosen, the equipment was generally fastened to a beam or a pipe by a piece of wire and hung approximately 5 feet of the floor. The instrument was then turned on and the following data other than production were recorded:

1. Experiment Type
2. Experiment Number
3. Date Weighted
4. Filter number (marked on masking tape)
5. Tare Weight (MG)
6. Gross Weight (MG)
7. Net Weight (Gross weight - Tare weight)
8. Sample Type (ME or NY)
9. Pump number (marked on back case)
10. Head Number
11. Orifice Number

12. Flow rate
13. Shift
14. Exposure Date
15. Location
16. Wet bulb temperature (WT)
17. Dry bulb temperature F (T)
18. The Percentage relative Humidity (H) were calculated using standard comfort chart
19. Time on
20. Time off
21. Total time

The five locations under this study are as follows:

1251 : IOCC Concentrator Process 1

1312 : IOCC Pellet Plant regrind mills

1341 : IOCC Pellet Plant induration Machines 1-4

2231 : Scully Process Spirals

2233 : Scully Process High Tension

The primary objective of this project is to model the peformance of dust concentration measured by ME and NY in terms of appropriate covariates. The influential covariates are production (P), temperature (T) and humidity (H). Unfortunately, in the present study the production (P) data was not available. As a result, we will confine our study to examine the effect of the covariates H and T on ratio of dust concentration .

The following formulae were used to measure the concentration of dust by two instruments ME and NY :

$$U = \frac{\text{Net weight}}{\text{Total time} \times 1.90} \times 1000 \quad , \tag{1.1}$$

$$V = \frac{\text{Net weight}}{\text{Total time} \times 1.70} \times 1000 \quad , \tag{1.2}$$

where U and V denotes the dust concentration measured by the two instruments ME and NY respectively. The dust concentration of U in (1.1) and V in (1.2) can be expressed as follows :

$$U = g_1(P, H, T) \quad , \tag{1.3}$$

$$V = g_2(P, H, T) \quad , \tag{1.4}$$

where g_1 and g_2 are usually non-linear.

Next assuming that the production has same effect on dust concentration, we eliminate the production effect by considering the ratio variable Y defined as :

$$Y = \frac{V}{U} \quad , \tag{1.5}$$

and try to determine the effect of temperature and humidity on ratio of dust concentration (Y) . The functional relationship of Y, with H and T may be defined as :

$$Y = g_3(H, T) \quad , \tag{1.6}$$

where g_3 may be linear or non-linear.

1.2 PLAN OF THE STUDY

The main objective of this study is to fit an appropriate model of type (1.6) for the different locations. In order to do this,

1. we study the distributional aspects of the data namely Y in chapter 2. Both Exploratory and Confirmatory analysis will be given in this chapter.
2. Multiple Regression by Least-Squares method is to be applied to the transformed data to fit a suitable polynomial model. Thirty two (32) different models both linear and quadratic have been exploited to choose the best fitted model using minimum error sum of squares criterion and then the effect of partial change of temperature and humidity on ratio of concentration of dust based on the best fitted model will be discussed in chapter 3.
3. In chapter 4, we use the whole data set without deletion or transformation and apply M and L_1 method of estimations as a robust procedure to select a

suitable model on the basis of minimum error sum of squares. We also study the effect of partial change of temperature and humidity on ratio of concentration of dust based on the best fitted model .

4. The summary and conclusion of the study has been presented in the last chapter (chapter 5).

CHAPTER 2

DISTRIBUTIONAL ASPECTS OF DATA

As it was mentioned in chapter 1, we are interested in modeling the ratio of concentration of dust Y, as a function of H and T. For such modeling it is important to diagnose the distributional pattern of the dependent variable Y , while the independent variables H and T are assumed to be fixed. This is because, if the distribution of Y is not Gaussian the statistical inference based on normality will be inappropriate. For example, the classical F - statistics constructed to test the goodness of fits of a regression model may not have the F distribution if the distribution of the dependent variable is skewed. From this point of view, it is important to study the behavior of the dependent variable before modeling its relation with other variables . Hence we analyze our data (Y) in section 2.1 through exploratory technique (cf.Tukey 1977; Mosteller 1977 and Hartwig 1979) which is traditionally known as Exploratory Data Analysis (EDA). Further we discuss a confirmatory analysis (Tukey 1980) on the distributional pattern of Y in section 2.3 .

2.1 EXPLORATORY DATA ANALYSIS

In the following subsection, we examine the symmetry of Y through Box-and-Whisker plotting (McGill et al., 1978).

2.1.1 THE BOX-AND-WHISKER PLOT OF THE ORIGINAL DATA

Box-and-Whisker plot (Box-plot) is the appropriate graphical tool for detecting the symmetry of the data. Box-plots also help to identify the extreme observations in the data set. More specifically it shows the median of a data batch by a '+' sign. If the '+' sign is located in the middle of the given box, the data is assumed to be symmetrical provided there is no indication of heavy or thin tails. If data batches contain outliers, they naturally stand out from the rest of the data. Values between the inner and outer fence (Velleman and Hoaglin, 1981) are possible outliers, and in Box-plot they are indicated by '*' . Values beyond the outer fence are probable outliers and in Box-plot they are indicated by '0' . The inner and outer fence are defined as follows:

H-spread = (upper hinge) - (lower hinge)

Lower limit of inner fence = (lower hinge) - (1.5 x (H-spread))

Upper limit of inner fence = (upper hinge) + (1.5 x (H-spread))

Lower limit of outer fence = (lower hinge) - (3.0 x (H-spread))

Upper limit of outer fence = (upper hinge) + (3.0 x (H-spread))

The Box-plot for five different locations are shown in Figure 2.1 .

Figure 2.1 Box-plots for Different Locations.

Location 1251 (N = 35)

```
                        -------
      O          *--------I + I-----              O
                        -------
  +---------+---------+---------+---------+---------+------Y
0.20      0.40      0.60      0.80      1.00      1.20
```

Location 1312 (N = 28)

```
                                   -----------
    O       *                 -----I      + I-------
                                   -----------
  +---------+---------+---------+---------+---------+------Y
 0.24      0.36      0.48      0.60      0.72      0.84
```

Location 1341 (N = 44)

```
                ----------------
  ---------------I    +        I-------------------
                ----------------

----+---------+---------+---------+---------+---------+--Y
  0.20      0.40      0.60      0.80      1.00      1.20
```

Location 2231 (N = 24)

```
                            -------------
     *                ---------I  +      I-----------
                            -------------

+---------+---------+---------+---------+---------+------Y
0.630   0.700     0.770     0.840     0.910     0.980
```

Location 2233 (N = 35)

```
                 --------------
   *---------------I    +     I----------          *
                 --------------

--+---------+---------+---------+---------+---------+----Y
 0.48     0.60      0.72      0.84      0.96      1.08
```

It is clear from the Figure 2.1 that there are 3, 2, and 1 possible and probable outliers for location 1251, 1312, and 2231 respectively. Location 2233 has also 4 possible outliers which is not clear from the Figure 2.1, since 3 values were approximately same and they were shown in a single '*' sign . The possible and probable outliers for location 1251 are 0.27318, 0.51795 and 1.25317; for 1312 are 0.26080 and 0.36982 ; for location 2231 is 0.66032; and for location 2233 are 0.49069, 1.11736, 1.11765, 1.11769. Location 1341 does not have any outliers.

Temporarily, the above mentioned outliers will be excluded from the analysis for fitting suitable regression model as in section 3.1.2 of chapter 3 . However, in chapter 4 we will attempt to fit suitable models based on all information including the outliers.

TABLE 2.1

Number of Observations in this Study

Location	Original	Excluding outliers
1251	35	32
1312	28	26
1341	44	44
2231	24	23
2233	35	31

2.1.2 BOX-AND-WHISKER PLOT EXCLUDING POSSIBLE AND/OR PROBABLE OUTLIERS

In this subsection , we draw the Box-plot for the data excluding outliers as mentioned above. We do not give the Box-plot for location 1341, as there is no outliers in this location.

Figure 2.2. Box-plot after Deletion of Outliers for Different Locations.

Location 1251 (N = 32)

```
                   ----------------
-------------------I       +       I---------------
                   ----------------

----+---------+---------+---------+---------+---------+--Y
  0.560     0.640     0.720     0.800     0.880     0.960
```

Location 1312 (N = 26)

```
                 ---------------------
  ---------------I          +         I----------------
                 ---------------------

------+---------+---------+---------+---------+---------+Y
    0.650     0.700     0.750     0.800     0.850     0.900
```

Location 2231 (N = 23)

```
                 -----------------
     ------------I    +          I----------------
                 -----------------

+---------+---------+---------+---------+---------+------Y
0.750     0.800     0.850     0.900     0.950     1.000
```

Location 2233 (N = 31)

```
                  ----------------
    --------------I    +         I--------------------
                  ----------------

----+---------+---------+---------+---------+---------+--Y
  0.560     0.640     0.720     0.800     0.880     0.960
```

2.1.3 HISTOGRAMS EXCLUDING POSSIBLE AND/OR PROBABLE OUTLIERS

The histograms for each location are displayed in this sub section. For the construction of histogram equal class intervals were considered for each location. Y is plotted on horizontal axis and the counts for each class plotted on a vertical axis. Note that the range of different locations are different by nature. Consequently, the class intervals do not mean the same interval for different locations. Histograms for different locations are shown in the following figure .

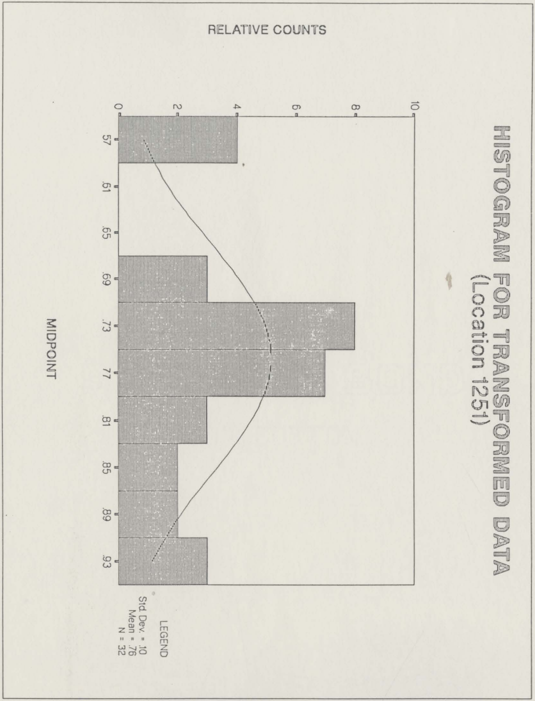


Figure 2.3 Histogram for different locations.

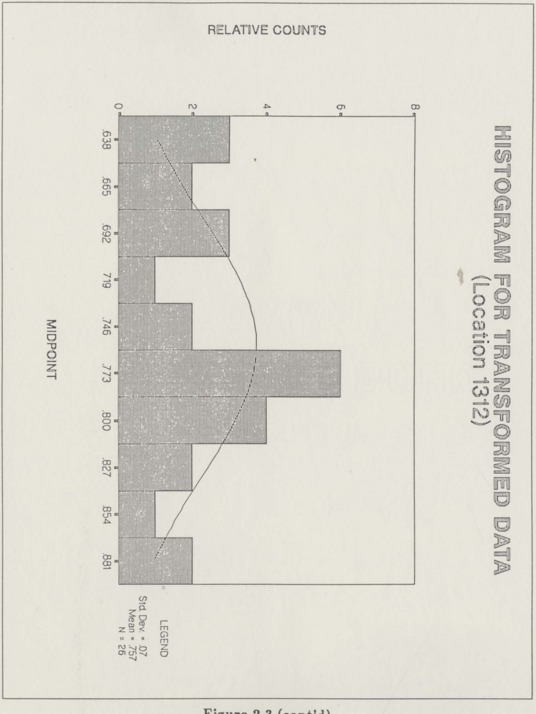


Figure 2.3 (cont'd)

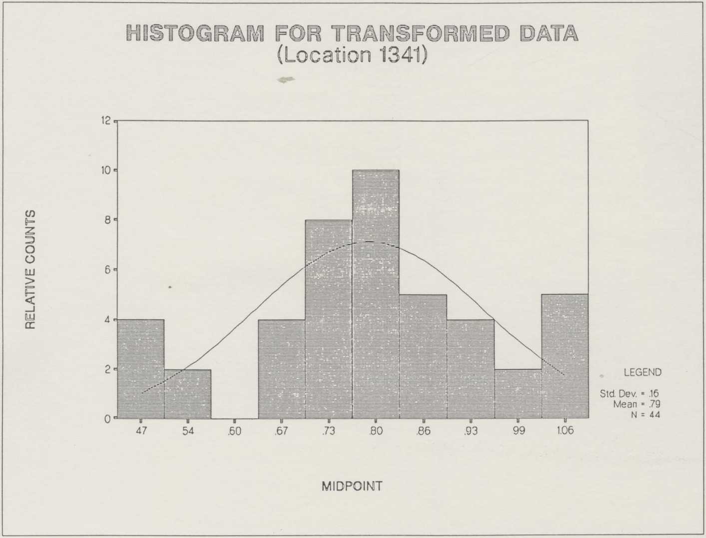


Figure 2.3 (cont'd)

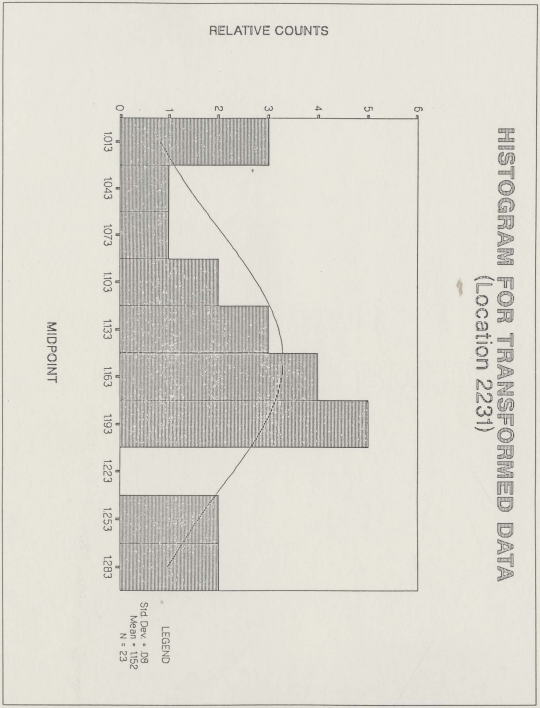


Figure 2.3 (cont'd)

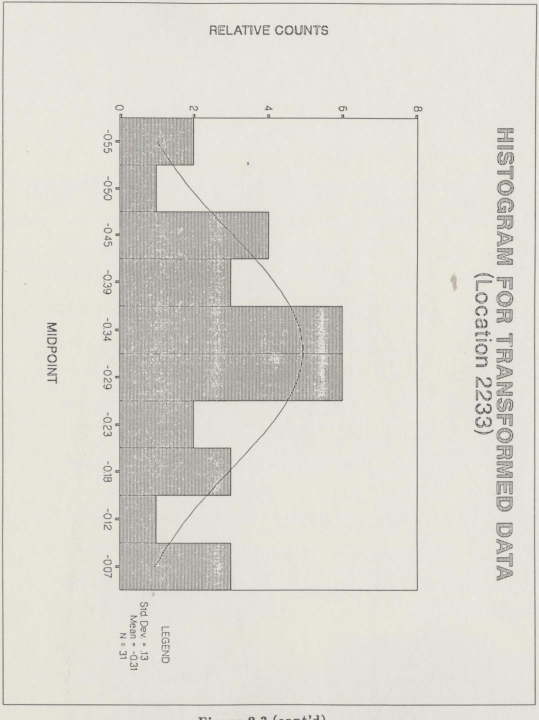


Figure 2.3 (cont'd)

2.1.4 LETTER-VALUE DISPLAY EXCLUDING POSSIBLE AND/OR PROBABLE OUTLIERS

So far we have presented graphical summaries of the data by using Box-plots and histograms. We now investigate the data batches by using numerical summaries. For each location, we observe the letter-value spreads H, E, D, and C [Velleman and Hoaglin, (1981)]. The median M splits an ordered data batch in half. The letter H denotes the hinges which are the summary values in the middle of each half of the data. They are about a quarter of the way in from each end of the ordered batch. Similarly, the letter E denotes the eighths and they are the middle values for the outer quarters of the data. These values are about an eighth of the way in from each end of the ordered batch. The difference between the lower hinge and upper hinge is known as the H-spread. Similarly, the E-spread is the difference between the lower eighth and the upper eighth, that is , the E-spread gives the range of the middle three-quarters of the data, and known as eighths. The letter D is familiar as sixteenth, and so on.

For examining the shape of the data, the data spreads as defined above will be compared to the spreads for the Gaussian distribution. The standard Gaussian spreads are: H-spread = 1.349, E-spread = 2.301, D-spread = 3.068,and C-spread = 3.726. We can compare the spreads of the data with the Gaussian spreads by dividing the spread values of the data by the Gaussian spread values to obtain quotients. If the data resemble a sample from a Normal distribution, then all of these quotients will be nearly the same. A clear trend in the quotients derived from the spreads provides an indication of how the data depart from Normal shape. If the quotients grow, the tails of the batch are heavier than the tails of the Normal shape. If the quotients shrink, the tails of the data are lighter. The median, Mids, spreads, and Quotients for different spreads for different locations are summarized in the following table 2.2 .

TABLE 2.2

Letter Value Display for Different Locations

Location 1251 (N = 32)

Depth	Lower	Upper	Mid	Spread	Quotient
M			0.755		
H	0.707	0.820	0.764	0.113	0.084
E	0.632	0.880	0.756	0.248	0.108
D	0.567	0.913	0.740	0.346	0.113
C	0.552	0.933	0.742	0.381	0.102

Location 1312 (N = 26)

Depth	Lower	Upper	Mid	Spread	Quotient
M			0.782		
H	0.702	0.802	0.752	0.099	0.073
E	0.655	0.823	0.739	0.167	0.073
D	0.630	0.867	0.749	0.237	0.077
C	0.625	0.890	0.757	0.265	0.071

Location 1341 (N = 44)

Depth	Lower	Upper	Mid	Spread	Quotient
M			0.630		
H	0.500	0.798	0.649	0.297	0.220
E	0.309	0.989	0.649	0.680	0.296
D	0.233	1.078	0.656	0.845	0.275
C	0.214	1.108	0.661	0.895	0.240

Location 2231 (N = 23)

Depth	Lower	Upper	Mid	Spread	Quotient
M			0.855		
H	0.832	0.914	0.873	0.083	0.062
E	0.795	0.966	0.880	0.171	0.074
D	0.785	0.994	0.889	0.208	0.068

Location 2233 (N = 31)

Depth	Lower	Upper	Mid	Spread	Quotient
M			0.724		
H	0.668	0.794	0.731	0.125	0.093
E	0.630	0.858	0.744	0.228	0.099
D	0.581	0.925	0.753	0.344	0.112
C	0.565	0.945	0.755	0.380	0.102

The results of the above tables along with the results of Box-plots from Figure 2.2 and Histograms from Figure 2.3 are discussed in the following section.

2.2 DESCRIPTION OF THE VARIABLE Y FOR DIFFERENT LOCATION

Location 1251

A first look at the histogram shows a concentration of data towards the center of the distribution. Box-plot shows no indication of skewness and this is verified by results of the table 2.1, where the midsummaries, which do not show a significant increasing or decreasing trend. The quotients of the spreads are close to constant. Thus, the distribution of Y for this location may considered as normal distribution.

Location 1312

Histogram for this location gives the impression that data is skewed and it has longer left tail. Box-plot shows a slight skewness to the left . This pattern of the data is also verified by midsummaries . Therefore, the data for this location appears to be negatively skewed.

Location 1341

The histogram for this location shows a concentration towards the center of the distribution. The Box-plot in Figure 2.1 shows a slight skewness to the right which is verified by the slightly increasing midsummary values. Hence we may consider the data as positively skewed.

Location 2231

The Box-plot indicates that data are skewed to the right. Increasing values of the midsummaries supports that data for this location is Skewed. Therefore, the normality of the distribution of Y in this location is questionable.

Location 2233

The Box-plot for this location shows log tail towards right. Histogram shows a concentration towards the center of the distribution. Midsummaries shows an increasing trend, indicating a long right tailed distribution. Thus the normality of the distribution of Y in this location is also questionable.

2.3 CONFIRMATORY TEST FOR NORMALITY THROUGH POWER TRANSFORMATIONS

In the previous section we have seen that in location 1251 and 1312 the distribution of Y is close to normal. But the normality for the location 1341, 2231 and 2233 are questionable. in order to apply the classical results based on normality need some transformation for these locations. Although we feel that no transformation is needed for location 1251 and 1312, for the sake of completeness we study the necessity of the transformation for all locations. We use the well-known Box and Cox (1964) power transformations which is widely used in statistical literature. The transformation is given by :

$$Z = \begin{cases} (\frac{Y^{\lambda}-1}{\lambda}), & \text{for } \lambda \neq 0, \\ \log Y, & \text{for } \lambda = 0, \end{cases} \tag{2.1}$$

where Y is the old variable , λ is the parameter to be determined .

2.3.1 MAXIMUM LIKELIHOOD METHOD FOR λ

We assumed that the observations $Y_1, Y_2, Y_3, \ldots, Y_n$ are independently normally distributed with constant variance and with expectations specified by a model linear in a set of parameters β. We restrict our attention to transformations indexed by unknown parameter λ, and then estimate λ and the other parameters of model by standard methods of inferences. It is assumed that for each λ, Z is a monotonic function of Y over the admissible range. We have $Y = [\ Y_1, Y_2, Y_3, \ldots, Y_n\ ]$, as an $n \times 1$ vector of observations , and we assume that the appropriate linear model for the problem is :

$$E(Z) = A\ \beta, \tag{2.2}$$

where Z is the common vector of transformed observations, A is a known matrix of order $n \times k$ (say) and β is a $k \times 1$ vector of unknown parameters associated with the transformed observations.

It is assumed that for some unknown λ, the transformed observations Z_i (i $=$ 1, 2, . . .,n) satisfy the full normal theory assumptions, i.e., are Z's independently normally distributed with constant variance σ^2, and expectations as given in (2.2). The probability density for the untransformed observations, and hence the likelihood in relation to these original observations is obtained by multiplying the normal density by the Jacobian of the transformation.

The likelihood in relation to the original observations Y is thus

$$\frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left[-\frac{(Z - A\ \beta)'(Z - A\ \beta)}{2\ \sigma^2}\right] J(\lambda; Y), \tag{2.3}$$

where

$$J(\lambda, Y) = \prod_1^n \frac{\partial Z_i}{\partial Y_i} = \prod_1^n Y_i^{\lambda-1}, \quad \text{for } all\ \ \lambda.$$

We follow Box and Cox (1964) and use large-sample maximum-likelihood theory to make inference about the parameters λ, in (2.3) which leads directly to the point estimates of the parameters. For given λ, (2.3) is, except for a constant factor, the likelihood for a standard Least-squares problem, i.e., the maximum-likelihood estimates of the β's are the least-square estimates for the dependent variable Z and the estimate of σ^2 , denoted for fixed $\hat{\sigma}^2(\lambda)$, is

$$\hat{\sigma}^2 (\lambda) = \frac{Z' A_r Z}{n} = \frac{S(\lambda)}{n} \tag{2.4}$$

where, when A is of full rank,

$$A_r = I - A (A'A)^{-1} A',$$

and S (λ) is the residual sum of squares in the analysis of variance of Z.

Thus for fixed λ, the maximized log likelihood , except for constant is

$$L_{\max}(\lambda) = -\frac{1}{2} n \log \hat{\sigma}^2(\lambda) + \log J(\lambda; Y), \tag{2.5}$$

so that $\log J(\lambda, Y) = (\lambda - 1) \sum_{i=1}^{n} \log Y_i$.

The $L_{\max}(\lambda)$ in (2.5) will be plotted against trial series of values of λ. We will choose that value of λ which gives $\max\{L_{\max}(\lambda)\}$. If $L_{\max}(\lambda)$ occurs at -1, 0, and .5 then the required transformations should be Reciprocal, Log, Square-root respectively. If $L_{\max}(\lambda)$ occurs at $\lambda = 1$, we do not need any transformation.

We plot $L_{\max}(\lambda)$ against possible value of λ in Figure 2.4 and we also summarize the values of λ for different locations in Table 2.3 .

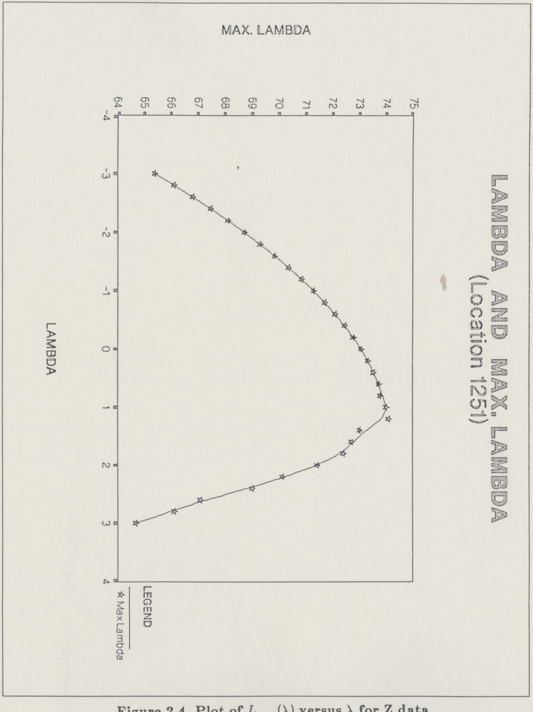


Figure 2.4 Plot of $L_{\max}(\lambda)$ versus λ for Z data

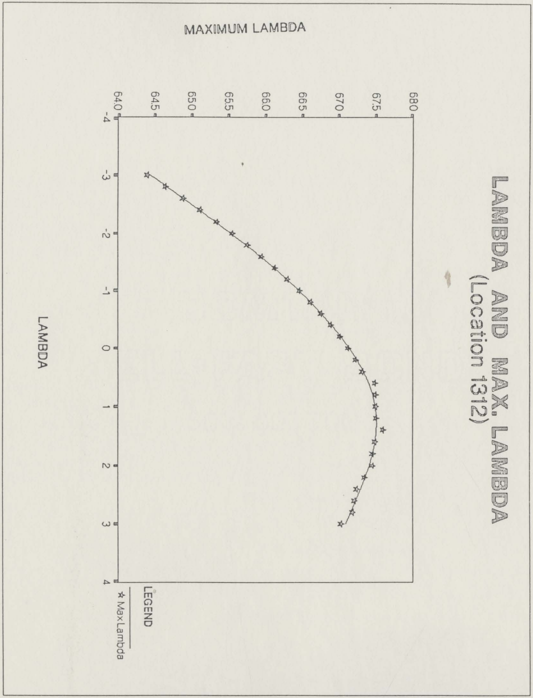


Figure 2.4 (cont'd)

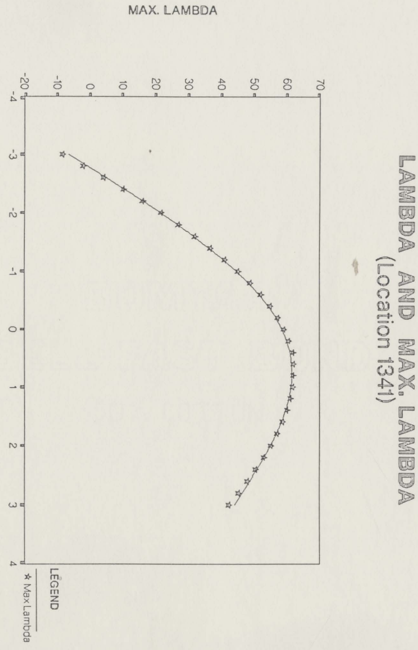


Figure 2.4 (cont'd)

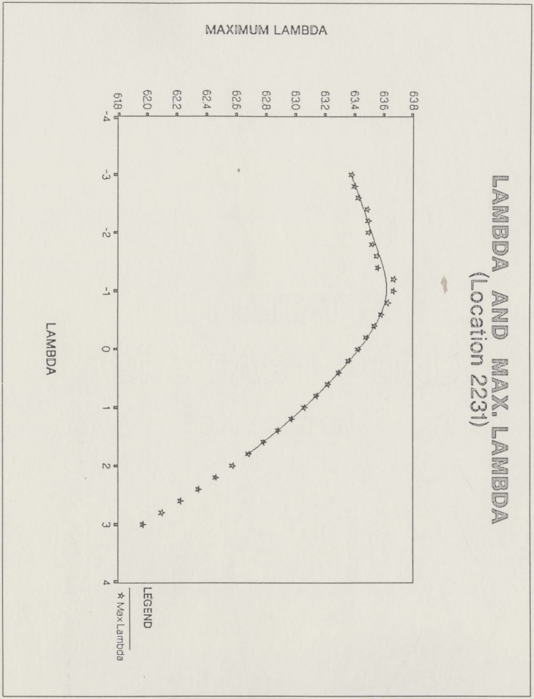


Figure 2.4 (cont'd)

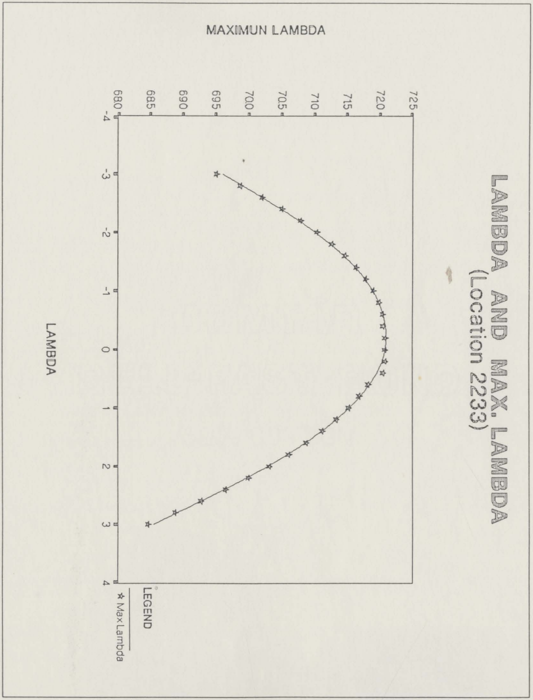


Figure 2.4 (cont'd)

TABLE **2.3**

Values of λ and $L_{max}(\lambda)$ for Different Locations

Location	L_{max}	λ
1251	74.097	1.2
1312	67.595	1.4
1341	62.038	0.6
2231	63.670	-1.2
2233	72.083	-0.2

From table 2.3 it is apparent that location 1251 and 1312 do not need any transformation, locations 1341, 2231 and 2233 needs square-root, reciprocal and log transformation respectively .

CHAPTER 3

FITTING MODELS TO THE TRANSFORMED DATA

Consider the transformed data Z obtained in the last chapter . The Z data refers to original Y data for location 1251 and 1312 after deletion of outliers. The Z data refers to the square root of the Y data for location 1341 and inverse and log transformation of the Y data after deletion of outliers in locations 2231 and 2233 respectively. In this chapter our objective is to fit appropriate models for Z in terms of H and T. All possible linear and quadratic regression (cf. Belsley 1980; Daniel 1980; Draper 1981; Seber 1977; Wetherill 1986; Wishart 1953) functions will be examined in order to choose the best fitted model.

3.1 SELECTION OF THE BEST MODEL

Consider a general regression model

$$Z = X\beta + \epsilon \ , \tag{3.1}$$

where Z is a $n \times 1$ response variable (transformed or/and deleted), X is a known design matrix of order $n \times k$, β is a $k \times 1$ vector of unknown parameters and ϵ is a $n \times 1$ error variable. As the transformed data is normal, without any further clarification, we assume that $\epsilon \sim N(0, \sigma^2 I_n)$. For the case where first or second degree equation is of interest, the appropriate design matrix X will be of the form:

$$X = \begin{bmatrix} 1 & t_1 & h_1 & h_1 t_1 & t_1^2 & h_1^2 \\ 1 & t_2 & h_2 & h_2 t_2 & t_2^2 & h_2^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & t_n & h_n & h_n t_n & t_n^2 & h_n^2 \end{bmatrix} \tag{3.2}$$

For a given model, linear or quadratic $\hat{\beta} = (X'X)^{-1}X'Z$ is the Best Linear Unbiased Estimates (BLUE) of β as ϵ has $N(0, \sigma^2 I_n)$. In order to choose the best model we test the linear hypothesis

$$\mathrm{H}_0 : \beta = 0 \text{ versus } \mathrm{H}_1 : \text{the negation of } \mathrm{H}_0 \quad , \tag{3.3}$$

by using the classical F-statistic

$$W = \frac{\dfrac{Z'X\,(X'X)^{-1}X'Z}{(k-1)}}{\dfrac{Z'[I - X\,(X'X)^{-1}X']Z}{(n-k)}} \quad , \tag{3.4}$$

which follows usual non-central F-distribution (F') distribution with (k - 1) and (n - k) degrees of freedom (d.f), and the noncentrality parameter given by $\delta = \dfrac{\beta'(X'X)\beta}{2\sigma^2}$. Under the H_0, W has the central F distribution with (k - 1) and (n - k) d.f. The higher values of W would lead the rejection of the null hypothesis.

The following criteria will be examined to choose the best models.

3.2 SELECTION CRITERIA

1. The overall model fits at less than 5% level of significance.
2. Individual coefficients of the related variable should be significant at less than 5% level of significance.
3. Error sum of squares of the fitted model will be minimum.

For clearity, for every location we produce the p-values $P(W \leq W^0) = 1 - p$, where W^0 is the calculated value W . Also we produce sum of squares error (SSE), the p - values for each coefficients .

TABLE 3.1

Values and P-values of Coefficients for Location 1251 .

Model	Cons	T	H	TH	T^2	H^2
02	1.35	-.0089	-	-	-	-
(.091)	(.000)	(.091)	-	-	-	-
03	.865	-	-.0017	-	-	-
(.293)	(.000)	-	(.293)	-	-	-
04	.952	-	-	-.00005	-	-
(.084)	(.000)	-	-	(.084)	-	-
05	1.05	-	-	-	-.00007	-
(.095)	(.000)	-	-	-	(.095)	-
06	.807	-	-	-	-	-.00001
(.327)	(.000)	-	-	-	-	(.327)
07	1.95	-.0143	-.0037	-	-	-
(.024)	(.000)	(.012)	(.032)	-	-	-
08	1.73	-.0109	-	-.00006	-	-
(.023)	(.000)	(.033)	-	(.030)	-	-
09	5.56	-.134	-	-	.00092	-
(.204)	(.425)	(.516)	-	-	(.544)	-
10	1.84	-.0144	-	-	-	-.00003
(.026)	(.000)	(.013)	-	-	-	(.036)
11**	1.00	-	.0116	-.00023	-	-
(.020)	(.000)	-	(.029)	(.010)	-	-
12	1.47	-	-.0036	-	.00011	-
(.025)	(.000)	-	(.033)	-	(.013)	-
13	1.23	-	-.0132	-	-	.00009
(.460)	(.029)	-	(.439)	-	-	(.497)
14	1.36	-	-	-.00006	-.00008	-
(.024)	(.000)	-	-	(.031)	(.034)	-
15	1.30	-	-	-.00020	-	-.00007
(.024)	(.000)	-	-	(.001)	-	(.034)
16	1.35	-	-	-	-.00011	-.00003
(.028)	(.000)	-	-	-	(.013)	(.037)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.1 (cont'd)

Values and P-values of Coefficients for Location 1251 .

Model	Cons	T	H	TH	T^2	H^2
17	.70	.0044	.0163	.00030	-	-
(.054)	(.757)	(.896)	(.648)	(.576)	-	-
18	6.29	-.143	-.0037	-	.00095	-
(.050)	(.337)	(.459)	(.034)	-	(.505)	-
19	2.07	-.0140	-.0080	-	-	.00003
(.060)	(.002)	(.017)	(.614)	-	-	(.784)
20	5.83	-	-.133	-.000055	.00090	-
(.049)	(.372)	-	(.491)	(.032)	(.527)	-
21	1.56	-.0063	-	-.00012	-	.00005
(.056)	(.013)	(.646)	-	(.522)	-	(.722)
22	6.12	-.141	-	-	.00094	-.00003
(.055)	(.352)	(.467)	-	-	(.513)	(.038)
23	.78	-	-.0184	-.00033	.00005	-
(.053)	(.460)	-	(.579)	(.507)	(.836)	-
24	1.057	-	.0097	-.00023	-	.000014
(.054)	(.041)	-	(.591)	(.015)	-	(.908)
25	1.59	-	-.0080	-	.00010	.00003
(.063)	(.004)	-	(.615)	-	(.018)	(.783)
26	1.34	-	-	-.00013	-.00004	.00004
(.057)	(.000)	-	-	(.492)	(.678)	(.688)
27	4.62	-.106	.0096	-.0002	.00077	-
(.100)	(.574)	(.637)	(.802)	(.730)	(.619)	-
28	.80	.0035	.0139	-.00028	-	.00001
(.111)	(.764)	(.923)	(.772)	(.629)	-	(.940)
29	6.51	-.145	-.0084	-	.00097	.00004
(.101)	(.332)	(.460)	(.598)	-	(.504)	(.764)
30	5.52	-.124	-	-.00011	.00086	.00003
(.099)	(.411)	(.533)	-	(.563)	(.553)	(.766)
31	.83	-	.0169	-.00032	.00004	.00001
(.11)	(.548)	-	(.705)	(.551)	(.859)	(.957)
32	5.13	-.116	.0036	-.00016	.00082	.00002
(.176)	(.563)	(.623)	(.945)	(.810)	(.608)	(.863)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.2

Values and P-values of Coefficients for Location 1312

Model	Cons	T	H	TH	T^2	H^2
02	1.19	-.0071	-	-	-	-
(.018)	(.00)	(.018)	-	-	-	-
03	.820	-	-.0009	-	-	-
(.500)	(.00)	-	(.500)	-	-	-
04	.914	-	-	-.00004	-	-
(.075)	(.00)	-	-	(.075)	-	-
05	.971	-	-	-	-.00006	-
(.021)	(.00)	-	-	-	(.021)-	
06	.782	-	-	-	-	-.00001
(.608)	(.00)	-	-	-	-	(.608)
07	1.31	-.0075	-.0014	-	-	-
(.034)	(.00)	(.013)	(.260)	-	-	-
08	1.23	-.0059	-	-.00003	-	-
(.030)	(.00)	(.050)	-	(.219)	-	-
09	3.26	-.0746	-	-	.00055	-
(.035)	(.09)	(.231)-	-	(.276)	-	
10	1.26	-.0075	-	-	-	-.00001
(.039)	(.00)	(.014)	-	-	-	(.322)
11	0.86	-	.00624	-.00013	-	-
(.019)	(.00)	-	(.030)	(.007)	-	-
12	1.08	-	-.00145	-	-.00006	-
(.038)	(.00)	-	(.254)-	(.015)-		
13	1.58	-	-.0249	-	-	.00019
(.190)	(.00)	-	(.083)	-	-	(.093)
14	1.04	-	-	-.00003	-.00005	-
(.032)	(.00)	-	-	(.205)	(.055)	-
15**	1.06	-	-	-.00013	-	.00005
(.010)	(.00)	-	-	(.003)	-	(.015)
16	1.03	-	-	-	-.00006	-.00001
(.044)	(.00)	-	-	-	(.015)	(.313)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.2 (cont'd)

Values and P-values of Coefficients for Location 1312

Model	Cons	T	H	TH	T^2	H^2
17	-.88	.0288	.0334	-.00058	-	-
(.021)	(.48)	(.168)	(.097)	(.083)	-	-
18	2.92	-.0609	-.00115	-	.00044	-
(.062)	(.138)	(.343)	(.379)	-	(.404)	-
19	1.83	-.00673	-.0192	-	-	.00014
(.036)	(.00)	(.024)	(.145)	-	-	(.174)
20	2.74	-.0560	-	-.00002	.00041	-
(.058)	(.17)	(.388)	-	(.340)	(.439)	-
21	.33	.0224	-	-.00046	-	.00020
(.006)	(.41)	(.069)	-	(.016)	-	(.021)
22	2.92	-.0621	-	-	.00045	-.00001
(.070)	(.15)	(.339)	-	-	(.399)	(.474)
23	.21	-	.0266	-.00046	.00018	-
(.024)	(.68)	-	(.104)	(.087)	(.202)	-
24	1.43	-	-.0124	-.00012	-	.00014
(.020)	(.001)	-	(.350)	(.012)	-	(.157)
25	1.61	-	-.0190	-	-.00005	.00014
(.041)	(.001)	-	(.153)	-	(.029)	(.184)
26	1.01	-	-	-.00046	.00018	.00020
(.005)	(.00)	-	-	(.012)	(.056)	(.016)
27	-1.98	.0597	.0385	-.00066	-.00021	-
(.047)	(.59)	(.548)	(.141)	(.130)	(.749)	-
28	-.59	.0353	.0169	-.00067	-	-.00017
(.012)	(.62)	(.084)	(.415)	(.041)	-	(.079)
29	4.58	-.0928	-.0241	-	.00070	.00018
(.037)	(.036)	(.153)	(.077)	-	(.183)	(.090)
30	1.47	-.0152	-	-.00044	.00030	.00020
(.014)	(.433)	(.803)	(.022)	(.531)	(.027)	
31	.72	-	.0095	-.00056	.00023	.00018
(.013)	(.204)	-	(.591)	(.036)	(.087)	(.070)
32	-.17	.0236	.0147	-.00063	.00008	.00017
(.029)	(.964)	(.807)	(.600)	(.130)	(.901)	(.092)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.3

Values and P-values of Coefficients for Location 1341

Model	Cons	T	H	TH	T^2	H^2
02	.87	-.00127	-	-	-	-
(.443)	(.00)	(.443)	-	-	-	-
03	.83	-	-.00151	-	-	-
(.396)	(.00)	-	(.396)	-	-	-
04	.85	-	-	-.00003	-	-
(.405)	(.00)	-	-	(.405)	-	-
05	0.86	-	-	-	-.00002	-
(.216)	(.00)	-	-	-	(.216)	-
06	0.82	-	-	-	-	-.00003
(.256)	(.00)	-	-	-	-	(.256)
07	1.22	-.0045	-.0049	-	-	-
(.103)	(.00)	(.051)	(.048)	-	-	-
08	1.01	-.0021	-	-.00005	-	-
(.357)	(.00)	(.224)	-	(.227)	-	-
09	-.434	.0400	-	-	-.00031	-
(.001)	(.202)	(.000)	-	-	(.000)	-
10	1.10	-.0038	-	-	-	-.00006
(.092)	(.00)	(.063)	-	-	-	(.042)
11	.843	-	-.0009	-.00002	-	-
(.692)	(.00)	-	(.821)	(.870)	-	-
12	1.15	-	-.0058	-	-.00004	-
(.023)	(.00)	-	(.014)	-	(.009)	-
13	.690	-	-.0084	-	-	-.00014
(.285)	(.00)	-	(.270)	-	-	(.183)
14	.998	-	-	-.00006	-.00002	-
(.161)	(.00)	-	-	(.146)	(.087)	-
15	.786	-	-	-.00003	-	-.00004
(.500)	(.00)	-	-	(.740)	-	(.406)
16	1.02	-	-	-	-.00004	-.00007
(.027)	(.00)	-	-	-	(.015)	(.018)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.3 (cont'd)

Values and P-values of Coefficients for Location 1341

Model	Cons	T	H	TH	T^2	H^2
17	1.67	-.0122	-.0254	.00037	-	-
(.011)	(.00)	(.001)	(.003)	(.011)	-	-
18	-.081	.0360	-.0045	-	-.0003	-
(.000)	(.821)	(.001)	(.033)	-	(.000)	-
19	1.11	-.0039	-.0006	-	-	-.00005
(.194)	(.001)	(.140)	(.950)	-	-	(.646)
20**	-.336	.0427	-	-.00008	-.00034	-
(.000)	(.302)	(.000)	-	(.028)	(.000)	-
21	1.05	-.0049	-	-.00011	-	-.00013
(.102)	(.000)	(.030)	-	(.226)	-	(.044)
22	-.18	.0358	-	-	-.00029	-.00005
(.000)	(.606)	(.001)	-	-	(.000)	(.051)
23	1.26	-	-.0214	.00030	-.00008	-
(.003)	(.00)	-	(.001)	(.011)	(.000)	-
24	.696	-	.0088	-.00001	-	-.00014
(.477)	(.00)	-	(.296)	(.917)	-	(.190)
25	1.14	-	-.0056	-	-.00004	-.000003
(.058)	(.00)	-	(.559)	-	(.027)	(.980)
26	.931	-	-	.00010	-.00004	-.00013
(.036)	(.00)	-	-	(.235)	(.008)	(.028)
27	-.506	.0470	.0029	-.00013	-.00004	-
(.001)	(.517)	(.026)	(.810)	(.538)	(.005)	-
28	2.52	-.0204	-.0612	.00062	-	.00028
(.005)	(.00)	(.001)	(.004)	(.002)	-	(.061)
29	-.029	.0366	-.00793	-	-.00031	.00004
(.001)	(.94)	(.001)	(.345)	-	(.000)	(.669)
30	-.372	.0440	-	-.0001	-.00035	.00001
(.001)	(.349)	(.001)	-	(.304)	(.000)	(.871)
31	1.64	-	-.0441	.0004	-.00012	.00021
(.002)	(.000)	-	(.005)	(.003)	(.000)	(.097)
32	-.81	.0530	.0107	-.0002	-.00039	-.00004
(.002)	(.626)	(.141)	(.786)	(.629)	(.042)	(.835)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.4

Values and P-values of Coefficients for Location 2231

Model	Cons	T	H	TH	T^2	H^2
02	1.11	.0007	-	-	-	-
(.884)	(.001)	(.884)	-	-	-	-
03	1.13	-	.0003	-	-	-
(.835)	(.000)	-	(.835)	-	-	-
04	1.10	-	-	.00001	-	-
(.674)	(.000)	-	-	(.674)	-	-
05	1.13	-	-	-	.00001	-
(.889)	(.000)	-	-	-	(.889)	-
06	1.15	-	-	-	-	.0000
(.975)	(.000)	-	-	-	-	(.975)
07	.781	.0046	.0014	-	-	-
(.843)	(.242)	(.589)	(.576)	-	-	-
08	.832	.0035	-	.00003	-	-
(.787)	(.111)	(.585)	-	(.504)	-	-
09	-.374	.050	-	-	-.0004	-
(.957)	(.948)	(.795)	-	-	(.798)	-
10	.976	.0025	-	-	-	.00001
(.957)	(.113)	(.770)	-	-	-	(.797)
11	1.03	-	-.0046	.00011	-	-
(.671)	(.00)	-	(.435)	(.391)	-	-
12	.922	-	.0014	-	.00004	-
(.847)	(.032)	-	(.580)	-	(.595)	-
13**	.375	-	.0259	-	-	-.00021
(.185)	(.368)	-	(.069)	-	-	(.071)
14	.939	-	-	.00003	.00003	-
(.789)	(.008)	-	-	(.506)	(.558)	-
15	.835	-	-	.00014	-	-.0005
(.409)	(.002)	-	-	(.187)	-	(.210)
16	1.06	-	-	-	.00002	.00001
(.960)	(.005)	-	-	-	(.778)	(.803)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.4 (cont'd)

Values and P-values of Coefficients for Location 2231

Model	Cons	T	H	TH	T^2	H^2
17	3.51	-.0390	-.0394	.00066	-	-
(.536)	(.109)	(.248)	(.202)	(.185)	-	-
18	-.22	.038	.0014	-	-.0003	-
(.948)	(.970)	(.846)	(.608)	-	(.864)	-
19	.157	.0031	.0261	-	-	-.0002
(.328)	(.826)	(.705)	(.074)	-	-	(.086)
20	.13	.027	-	.00003	-.0002	-
(.923)	(.982)	(.891)	-	(.540)	(.906)	-
21	1.56	-0154	-	.0003	-	-.00013
(.364)	(.026)	(.242)	-	(.086)	-	(.108)
22	-.37	.047	-	-	-.0004	.00001
(.986)	(.950)	(.810)	-	-	(.820)	(.819)
23	2.06	-	-.0318	.0005	-.0003	-
(.599)	(.051)	-	(.247)	(.225)	(.307)	-
24	.382	-	.022	.00004	-	-.0002
(.332)	(.371)	-	(.207)	(.729)	-	(.113)
25	.326	-	.0262	-	.00003	-.0002
(.326)	(.656)	-	(.073)	-	(.692)	(.084)
26	1.08	-	-	.00027	-.00012	-.00013
(.381)	(.003)	-	-	(.092)	(.261)	(.115)
27	14.8	-.376	-.0734	.0012	.0025	-
(.483)	(.157)	(.218)	(.096)	(.090)	(.264)	-
28	-.69	.0145	.0414	-.00018	-	-.00024
(.494)	(.875)	(.805)	(.602)	(.844)	-	(.278)
29	2.53	-.078	.0283	-	.00067	-.00022
(.473)	(.664)	(.692)	(.077)	-	(.681)	(.087)
30	6.14	-.167	-	.00035	.00122	-.00016
(.459)	(.354)	(.443)	-	(.073)	(.485)	(.088)
31	-.24	-	.0411	-.00017	.00012	-.00024
(.489)	(.909)	-	(.526)	(.813)	(.763)	(.224)
32	3.24	.469	.0211	-.00012	.00096	-.00004
(.489)	(.809)	(.567)	(.782)	(.819)	(.879)	(.475)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.5

Values and P-values of Coefficients for Location 2233

Model	Cons	T	H	TH	T^2	H^2
02	-.616	.0047	-	-	-	-
(.024)	(.00)	(.024)	-	-	-	-
03	-.315	-	.00001	-	-	-
(.995)	(.001)	-	(.995)	-	-	-
04	-.460	-	-	.00006	-	-
(.102)	(.000)	-	-	(.102)	-	-
05	-.454	-	-	-	.00003	-
(.041)	(.000)	-	-	-	(.041)	-
06	-.319	-	-	-	-	.000003
(.919)	(.000)	-	-	-	-	(.919)
07	-.792	.0058	.0026	-	-	-
(.040)	(.000)	(.012)	(.234)	-	-	-
08	-.728	.0044	-	.00005	-	-
(.024)	(.000)	(.029)	-	(.120)	-	-
09	-1.41	.0298	-	-	-.00019	-
(.022)	(.008)	(.062)	-	-	(.109)	-
10	-.736	.0057	-	-	-	.00003
(.039)	(.000)	(.012)	-	-	-	(.230)
11**	-.439	-	-.0098	.00021	-	-
(.006)	(.000)	-	(.006)	(.002)	-	-
12	-.589	-	.0026	-	.00004	-
(.068)	(.000)	-	(.261)	-	(.022)	-
13	-.168	-	-.0079	-	-	.00010
(.837)	(.525)	-	(.561)	-	-	(.555)
14	-.586	-	-	.00006	.00003	-
(.034)	(.000)	-	-	(.103)	(.043)	-
15	-.609	-	-	.00020	-	-.00011
(.012)	(.000)	-	-	(.003)	-	(.013)
16	-.539	-	-	-	.00004	.00003
(.067)	(.000)	-	-	-	(.021)	(.254)

Figures in the parenthesis indicate P-value
** indicates best selected model

TABLE 3.5 (cont'd)

Values and P-values of Coefficients for Location 2233

Model	Cons	T	H	TH	T^2	H^2
17	.373	-.0125	-.0320	.00056	-	-
(.004)	(.424)	(.085)	(.019)	(.011)	-	-
18	-1.45	.0277	.0020	-	-.00017	-
(.040)	(.007)	(.086)	(.370)	-	(.167)	-
19	-.748	.0058	.0005	-	-	.00003
(.095)	(.031)	(.016)	(.968)	-	-	(.868)
20	-1.32	.0239	-	.00004	-.00015	-
(.032)	(.014)	(.149)	-	(.256)	(.232)	-
21	-.597	-.00037	-	.00020	-	-.00012
(.033)	(.003)	(.934)	-	(.130)	-	(.244)
22	-1.41	.0277	-	-	-.00017	.00002
(.039)	(.009)	(.086)	-	-	(.166)	(.358)
23	-.149	-	-.0243	.00043	-.00006	-
(.006)	(.457)	-	(.018)	(.008)	(.122)	-
24	-.366	-	-.0136	.00021	-	.00005
(.018)	(.123)	-	(.253)	(.002)	-	(.738)
25	-.538	-	-.00001	-	.00004	.00003
(.151)	(.075)	-	(.000)	-	(.029)	(.843)
26	-.605	-	-	.00022	-.00001	-.00013
(.032)	(.000)	-	-	(.066)	(.793)	(.150)
27	-.80	-.0237	-.0370	.00064	.00007	-
(.011)	(.471)	(.386)	(.042)	(.032)	(.669)	-
28	.705	-.0145	-.0459	.00061	-	.00013
(.008)	(.237)	(.058)	(.028)	(.008)	-	(.360)
29	-1.40	.0277	-.0006	-	-.00017	.00003
(.084)	(.020)	(.092)	(.963)	-	(.174)	(.837)
30	-1.06	.0152	-	.00015	-.00011	-.00008
(.055)	(.090)	(.447)	-	(.318)	(.425)	(.451)
31	.042	-	-.0341	.00045	-.00007	.00010
(.012)	(.898)	-	(.048)	(.007)	(.099)	(.466)
32	1.41	-.0319	-.0556	.00074	.0001	.00015
(.017)	(.267)	(.265)	(.035)	(.020)	(.525)	(.309)

Figures in the parenthesis indicate P-value
** indicates best selected model

Table 3.6

Values of the R^2 for 31 Different Models in Each Location

Model	1251	1312	1341	2231	2233
02	0.092	0.211	0.014	0.001	0.164
03	0.037	0.019	0.017	0.002	0.000
04	0.096	0.126	0.017	0.090	0.090
05	0.090	0.203	0.036	0.001	0.136
06	0.032	0.011	0.031	0.000	0.000
07	0.227	0.255	0.105	0.017	0.206
08	0.230	0.263	0.049	0.024	0.235
09	0.104	0.253	0.296	0.004	0.238
10	0.222	0.245	0.110	0.004	0.207
11	0.235*	0.292	0.018	0.039	0.305*
12	0.224	0.248	0.168	0.016	0.174
13	0.052	0.135	0.059	0.155*	0.013
14	0.228	0.258	0.085	0.023	0.215
15	0.228	0.328*	0.033	0.085	0.273
16	0.219	0.238	0.161	0.004	0.254
17	0.236	0.352	0.241	0.106	0.378
18	0.239	0.278	0.373	0.018	0.261
19	0.229	0.316	0.110	0.162	0.207
20	0.241	0.283	0.378*	0.024	0.275
21	0.233	0.424	0.142	0.151	0.273
22	0.234	0.270	0.361	0.007	0.262
23	0.236	0.343	0.294	0.092	0.364
24	0.236	0.355	0.060	0.161	0.308
25	0.226	0.307	0.168	0.163	0.176
26	0.232	0.433	0.191	0.149	0.274
27	0.243	0.355	0.379	0.167	0.383
28	0.236	0.442	0.307	0.164	0.399
29	0.242	0.373	0.376	0.170	0.262
30	0.244	0.434	0.378	0.174	0.291
31	0.237	0.441	0.342	0.165	0.378
32	0.244	0.442	0.379	0.165	0.408

* indicates R^2 value for the selected model

Table 3.7

Values of the adjusted R^2 for 31 Different Models in Each Location

Model	1251	1312	1341	2231	2233
02	0.062	0.179	0.000	0.000	0.135
03	0.005	0.000	0.000	0.000	0.000
04	0.066	0.090	0.000	0.000	0.058
05	0.060	0.169	0.013	0.000	0.106
06	0.000	0.000	0.008	0.000	0.000
07	0.174	0.190	0.061	0.000	0.149
08	0.177	0.199	0.003	0.000	0.180
09	0.042	0.187	0.262	0.000	0.184
10	0.168	0.179	0.066	0.000	0.150
11	0.182*	0.230	0.000	0.000	0.255*
12	0.171	0.182	0.128	0.000	0.115
13	0.000	0.059	0.013	0.071*	0.000
14	0.175	0.193	0.041	0.789	0.159
15	0.174	0.270*	0.000	0.000	0.221
16	0.165	0.172	0.120	0.000	0.117
17	0.154	0.263	0.184	0.000	0.309
18	0.158	0.180	0.325	0.000	0.179
19	0.147	0.223	0.043	0.000	0.119
20	0.160	0.185	0.331*	0.000	0.194
21	0.151	0.345	0.078	0.017	0.192
22	0.152	0.170	0.313	0.000	0.180
23	0.155	0.254	0.241	0.000	0.294
24	0.154	0.267	0.000	0.028	0.231
25	0.143	0.212	0.106	0.030	0.084
26	0.150	0.355	0.130	0.011	0.194
27	0.131	0.232	0.315	0.000	0.288
28	0.123	0.336	0.236	0.000	0.306
29	0.130	0.253	0.311	0.000	0.149
30	0.131	0.327	0.314	0.000	0.182
31	0.123	0.334	0.275	0.000	0.282
32	0.098	0.303	0.298	0.000	0.290

* indicates Adjusted R^2 value for the selected model

Table 3.8

Values of the Sum of Squares of Error for 31 Different Models in Each Location

Model	1251	1312	1341	2231	2233
02	0.2847	0.1138	1.095	0.1578	0.4656
03	0.3020	0.1415	1.091	0.1576	0.5571
04	0.2883	0.1261	1.092	0.1566	0.5072
05	0.2852	0.1150	1.070	0.1578	0.4815
06	0.3035	0.1427	1.077	0.1580	0.5569
07	0.2424	0.1075	0.994	0.1553	0.4423
08	0.2415	0.1064	1.056	0.1542	0.4264
09	0.2810	0.1079	0.781	0.1573	0.4242
10	0.2440	0.1089	0.989	0.1573	0.4418
11	0.2398*	0.1022	1.091	0.1518	0.3874*
12	0.2433	0.1056	0.924	0.1554	0.4599
13	0.2972	0.1249	1.045	0.1334*	0.5500
14	0.2421	0.1071	1.016	0.1543	0.4371
15	0.2422	0.0970*	1.074	0.1445	0.4053
16	0.2429	0.1100	0.932	0.1573	0.4593
17	0.2396	0.0936	0.843	0.1413	0.3462
18	0.2385	0.1041	0.696	0.1551	0.4116
19	0.2417	0.0987	0.988	0.1324	0.4418
20	0.2380	0.1035	0.691*	0.1541	0.4040
21	0.2404	0.0832	0.952	0.1342	0.4051
22	0.2402	0.1034	0.709	0.1568	0.4109
23	0.2394	0.0948	0.784	0.1435	0.3540
24	0.2397	0.0931	1.044	0.1326	0.3857
25	0.2426	0.1000	0.923	0.1323	0.4593
26	0.2407	0.0819	0.898	0.1349	0.4041
27	0.2374	0.0931	0.670	0.1316	0.3438
28	0.2396	0.0805	0.769	0.1321	0.3351
29	0.2377	0.0905	0.693	0.1311	0.4109
30	0.2372	0.0816	0.690	0.1305	0.3951
31	0.2394	0.0807	0.730	0.1319	0.3467
32	0.2371	0.0805	0.689	0.1301	0.3296

* indicates Sum of Squares of Error of the selected model

3.3 BEST SELECTED MODEL FOR DIFFERENT LOCATION

According to the criteria given above, we now summarize the best model for each location from Table 3.1 through Table 3.5.

Location 1251

$$Z = 1.0005 + 0.0116\ H - 0.0002\ TH$$

Location 1312

$$Z = 1.0626 - 0.0001\ TH + 0.0001\ H^2$$

Location 1341

$$Z = -0.3365 + 0.0427\ T - 0.0001\ TH - 0.0003\ T^2$$

Location 2231

$$Z = 0.3750 + 0.0259\ H - 0.0002\ H^2$$

Location 2233 :

$$Z = -0.4385 - 0.0098\ H + 0.0002\ TH$$

REMARKS:

The fitted equation for different locations shows no linear but quadratic relationship of ratio of concentration of dust with the covariates temperature and humidity. Humidity is more influential covariate than temperature because it is present in all best selected model in all location. It also reveals from the selected model that interaction between humidity and temperature is found in all locations except 2231. Model for the location 2231 shows that ratio of the concentration of dust is not influenced by temperature.

3.4 PERFORMANCE OF THE REGRESSION MODELS

Now we evaluate the performance of the above regression equations at average temperature and humidity levels i.e., for $H=\overline{H}$ and $T=\overline{T}$. Z and Y values for each of the five locations are as follows :

TABLE 3.9

Estimated Z and Y values for Different locations at Average Level of T and H

Variable	1251	1312	1341	2231	2233
$\overline{H}$	64.220	66.380	29.500	66.040	37.740
$\overline{T}$	67.344	61.538	64.000	60.174	64.840
Z	0.880	1.095	0.979	1.213	-0.319
Y	0.880	1.095	0.958	0.824	0.727

It reveals from Table 3.9 that the performance of the two instruments is not the same. Note that Y is the ratio of NY and ME. It is clear that the performance of NY over ME is 88, 110, 96, 82 and 73 percent for Locations 1251, 1312, 1341, 2231 and 2233 respectively, which shows that in all locations except 1312 ratio of concentration of dust measured by ME is higher than NY at average level of temperature and humidity.

3.5 PARTIAL EFFECT OF HUMIDITY OR TEMPERATURE ON CONCENTRATION

We calculated the partial derivatives of ratio of concentration of dust at the mean change of temperature and/or mean humidity for the selected models in each location . Results are shown in Table 3.10

TABLE 3.10

Partial Change of Concentration for Different Locations

Locations	$\frac{\partial Z}{\partial H}$	$\frac{\partial Z}{\partial T}$	$\frac{\partial^2 Z}{\partial H \partial T}$	$\frac{\partial^2 Z}{\partial H^2}$	$\frac{\partial^2 Z}{\partial T^2}$
1251	-0.00187	-0.01284	-0.0002	0.0000	0.0000
1312	*	-0.00664	-0.0001	0.0002	0.0000
1341	-0.00640	*	-0.0001	0.0000	-0.0006
2231	*	0.00000	0.0000	-0.0004	0.0000
2233	0.00317	0.00755	0.0002	0.0000	0.0000

* indicates that variable still present

Table 3.10 shows that partial change of the ratio of concentration of dust for unit change of humidity at average level of temperature is maximum for location 2233 and minimum in location 1341. It also reveals that partial change of ratio of concentration of dust for the unit change of temperature at average level of humidity is maximum in location 2233 nd minimum in location 1251.

3.6 GRAPHICAL CHECK FOR GOODNESS OF FIT

In this section we check the goodness of the fit of the model by displaying the multiple line plot of observed Z and fitted Z i.e., $\hat{Z}$ against sequence of the observations.

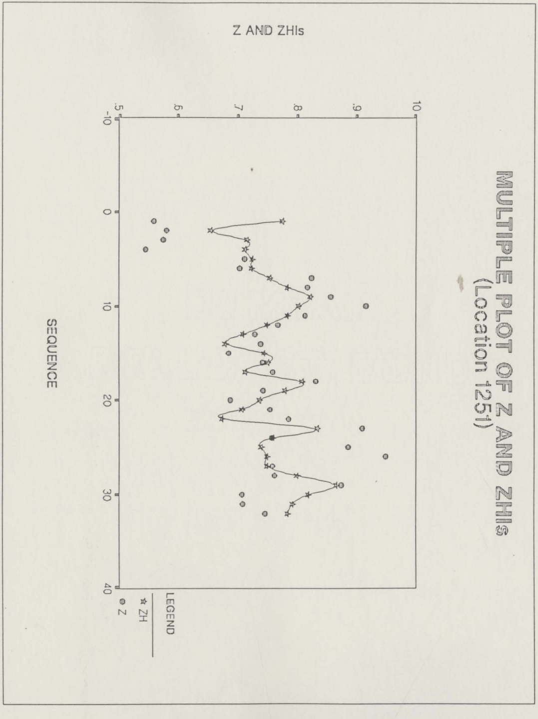


Figure 3.1 Multiple Plot of Z and $\hat{Z}$

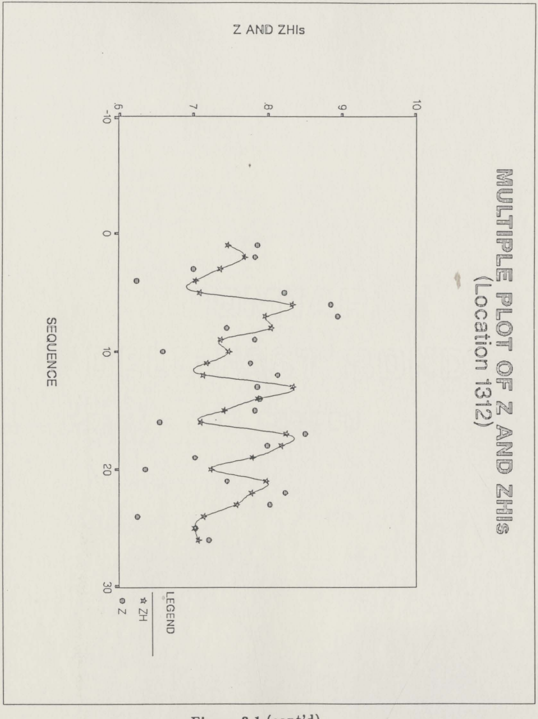


Figure 3.1 (cont'd)

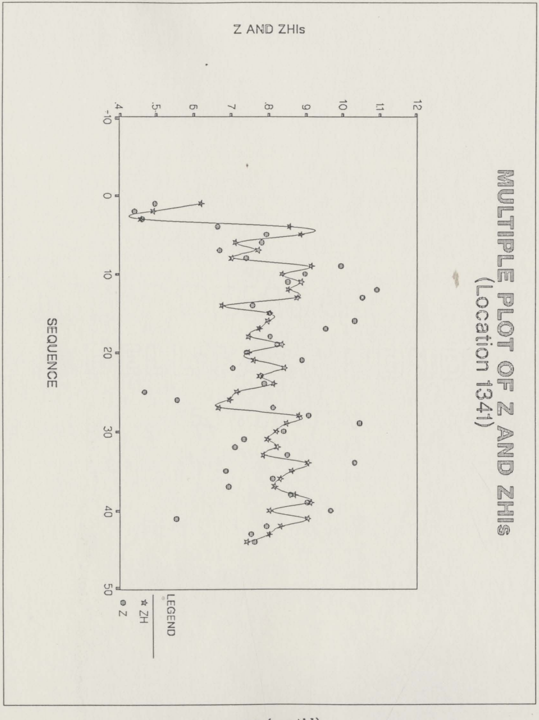


Figure 3.1 (cont'd)

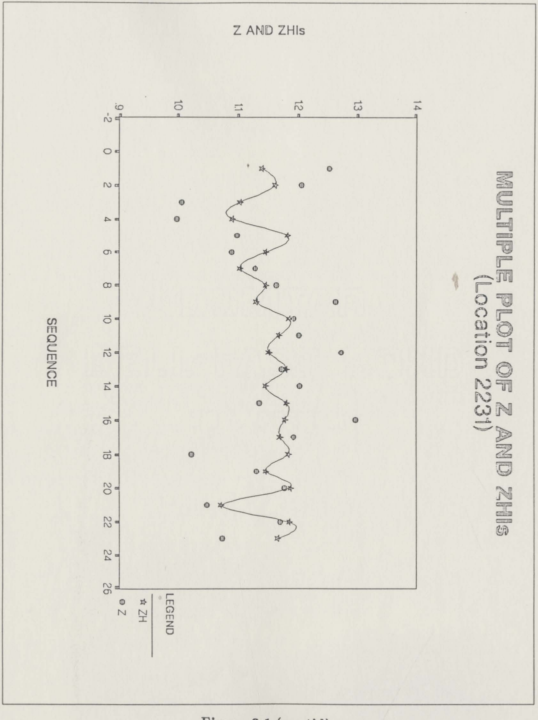


Figure 3.1 (cont'd)

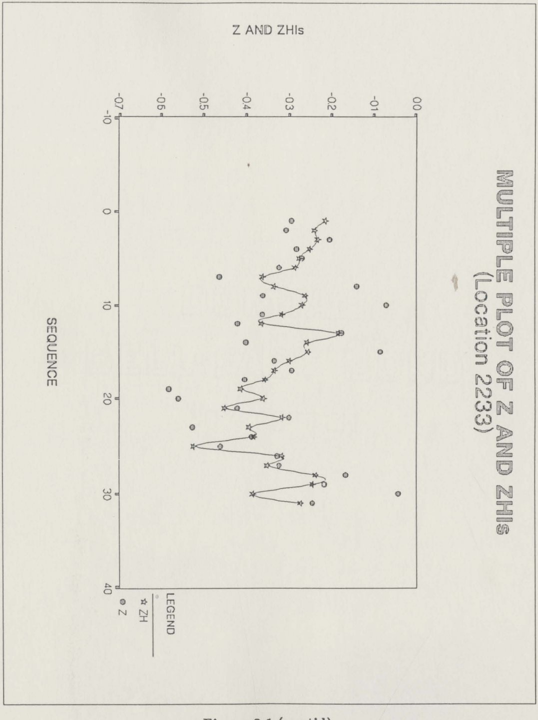


Figure 3.1 (cont'd)

CHAPTER 4

M - ESTIMATE OF REGRESSION

Consider the model

$$Y = X\,\Gamma + \epsilon^{*} \quad , \tag{4.1}$$

where Y is a $n \times 1$ vector which represents the ratio of concentration of dust as described in chapter 1.

X is a known design matrix of order $n \times k$, of the form :

$$X = \begin{bmatrix} t_1 & h_1 & h_1 t_1 & t_1^2 & h_1^2 \\ t_2 & h_2 & h_2 t_2 & t_2^2 & h_2^2 \\ . & . & . & . & . \\ . & . & . & . & . \\ t_n & h_n & h_n t_n & t_n^2 & h_n^2 \end{bmatrix} ,$$

and $\Gamma = [\gamma_0 , \gamma_1 , \ldots, \gamma_{k-1}]'$, a vector of unknown parameters.

In (4.1) ϵ^* is a $n \times 1$ error variable which is generally asymmetric.

Let $\Gamma = (\gamma_0, \gamma)$, where $\gamma = \gamma_1 , \gamma_2 \cdots, \gamma_{k-1}{}'$ is the $(k-1) \times 1$ vector of regression coefficients, γ_0 is the intercept.

For the case when ϵ^* in (4.1) has asymmetric distribution, we will estimate γ by M-estimation technique. It is well known that for the case when

i. the distribution of ϵ^* symmetrical but contains some outliers,

ii. the distribution of ϵ^* is either positively or negatively skewed,

one can not use the multiple regression technique to estimate Γ . In such case, the M-estimates of regression is highly appropriate (Andrews 1974; Andrews et al. 1972; Bickel 1976; Carroll 1978, 1979; Hinich 1975; Hogg 1967, 1974; Huber 1964,1972,1973; Jackel 1972; Jureckova 1971; Moberg et al. 1980; Stigler 1977; Yohai 1974) for estimation of γ . However, in Carroll(1979, JASA) it has been shown that $var(\gamma_0)$ is inflated where $\hat{\gamma}_0$ is the M-estimate of γ_0 . As we are primarily concerned about γ , we use M-estimation technique to estimate this parameter.

As the change of origin does not effect γ_i ($i = 1, 2, \ldots, k-1$) , we will consider

X as a centered design matrix. For example $\sum_{i=1}^{n} t_i = 0$, and $\sum_{i=1}^{n} h_i = 0$.

We adopt the notation of Hogg (1979) in explaining the distributional properties of the M-estimate of γ . The M-estimation of γ_i $(i = 1, 2, \ldots, k-1)$ requires to minimize

$$\sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i\, \gamma}{S}\right) , \tag{4.2}$$

where

$$\rho'(X) = \psi(X) = \begin{cases} \sin(X/a) , & |X| <= a\ \pi, \\ 0 , & |X| > a\ \pi, \end{cases} \tag{4.3}$$

and

$$S = \text{Median of the nonzero deviations } \frac{|Y_i - X_i\, \tilde{\gamma}|}{.6745} , \tag{4.4}$$

$\tilde{\gamma}$ being the preliminary estimate of γ obtained by using the least absolute value technique and $a = 2.1$. In calculation of S the nonzero deviations have been considered because as too many zero deviations would produce small S , while the actual measure of dispersion should not be too small .

To minimize the function (4.7), we take its first derivatives with respect to γ and equate those expression to zero . The equations are:

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - X_i \gamma}{S}\right) X_{ij} = 0, \quad j = 1, 2, \ldots, (k-1) \quad , \tag{4.5}$$

Solving these equation (4.10) is equivalent to solve

$$\sum_{i=1}^{n} w_i \; X_{ij} \; (Y_i - X_i \; \gamma) = 0, \qquad j = 1,2,\ldots,k \quad , \tag{4.6}$$

where

$$w_i = \frac{\psi[(Y_i - X_i \tilde{\gamma}) / S]}{(Y_i - X_i \tilde{\gamma}) / S}, \quad i = 1,2,\ldots,n \quad , \tag{4.7}$$

$$= \frac{\sin\left[\dfrac{\text{Residual (i) / S}}{2.1}\right]}{\text{Residual (i) / S}}$$

The solution to these approximate equations (4.11) is :

$$\hat{\gamma} = (X'WX)^{-1} X'WY \quad , \tag{4.8}$$

where W is the diagonal matrix as shown below :

$$W = \begin{bmatrix} w_1 & 0 & . & . & 0 \\ 0 & w_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & w_k \end{bmatrix} \qquad (4.9)$$

The $\hat{\gamma}$ in (4.8) provides us with a new start which we use in (4.4) and (4.6) to recompute w_i. The recomputed w_i is then used in (4.13) to compute $\hat{\gamma}$. This cycle is continued until $\hat{\gamma}$ is stable.

4.1 DISTRIBUTION OF REGRESSION ESTIMATOR $\hat{\gamma}$ (γ_i ($i = 1, 2, \ldots, k-1$))

The distribution of $\hat{\gamma}$ for small sample case is not studied in the literature. The asymptotic theory suggests that (Hill and Holland (1977)) $\sqrt{n}\ (\hat{\gamma} - \gamma)$ has an approximate k-variate normal distribution with mean vector 0 and a suitable covariance matrix estimated by

$$S^* = \hat{\mathrm{Cov}}(\hat{\gamma}) = \left(\frac{n}{n-p}\right) S^2 \frac{\frac{1}{n}\sum \psi^2[(Y_i - X_i\gamma) / S]}{\left[\frac{1}{n}\sum \psi'[(Y_i - X_i\gamma) / S]\right]^2} (X'X)^{-1} , \qquad (4.10)$$

which reduces to

$$S^{*} = \hat{Cov}(\hat{\gamma}) = Cov(\hat{\gamma}) = (\frac{n}{n-p})S^{2}\frac{\frac{1}{n}\sum \sin\psi^{2}[(Y_{i} - X_{i}\gamma)/S]}{\frac{1}{n}\sum \cos\psi[(Y_{i} - X_{i}\gamma)/S]^{2}}(X'X)^{-1}. \tag{4.11}$$

4.2 TESTING REGRESSION FUNCTION ($H_0 : \gamma = 0$)

In order to choose the best model we test the linear hypothesis

$$H_0 : \gamma = 0 \text{ versus } H_1 : \text{the negation of } H_0 , \tag{4.12}$$

by using the χ^2 - statistics

$$S_1 = \gamma' S^{*} \gamma \tag{4.13}$$

where S^{*} is given in (4.16) . The test statistics S_1 in (4.13) follows $\chi'^{2}(k,\lambda)$, the non-central χ^2 - distribution with n degrees of freedom and non-centrality parameter $\lambda = \frac{1}{2}\gamma' \, cov(\hat{\gamma}) \, \gamma$. Under the H_0, χ^2 reduces to the central χ^2, with n degrees of freedom.

To test individual coefficients we use normal statistic

$$Z = \frac{\sqrt{n}\ \hat{\gamma}_i}{\sqrt{S^*_{ii}}} , \tag{4.14}$$

where S^*_{ii} is the i th diagonal elements of S^* and Z distributed as standard normal variate with mean 0 and variance 1.

Tables (4.1) through (4.5) shows values of the individual coefficients of the model and their respective P-values for 31 different model in each location. Table (4.6) P-value of the 31 different models for each location.

We estimate γ_0 by

$$\hat{\gamma}_0 = \overline{Y} - \sum_{i=1}^{k-1} \hat{\gamma}_i\ \overline{X}$$

We remark that the M-estimatimator of γ_0 is not quite suitable as its sampling error is generally inflated for the skewed data. It would have been more appropriate to estimate γ_0 by Jacknife technique but was not chosen in the present study as we are primarily interested in slop parameters.

Table (4.7) through (4.11) shows the estimated value of $\hat{\gamma}_0$ and Sum of Squares of Errors (SSE).

SELECTION CRITERIA

1. The overall model fits at less than 5% level of significance.
2. Individual coefficients of the related variable should be significant at less than 5% level of significance.
3. Error sum of squares of the fitted model will be minimum.

TABLE 4.1

Values and P-values of Coefficients for Location 1251 .

Model	T	H	TH	T^2	H^2
02	-.0082	-	-	-	-
	(.0000)	-	-	-	-
03	-	-.0047	-	-	-
	-	(.0696)	-	-	-
04	-	-	-.00002	-	-
	-	-	(.0000)	-	-
05	-	-	-	-.00006	-
	-	-	-	(.0000)	-
06	-	-	-	-	-.000004
	-	-	-	-	(.0702)
07	-.0149	-.0041	-	-	-
	(.0000)	(.0000)	-	-	-
08	-.0109	-	-.00006	-	-
	(.0000)	-	(.0000)	-	-
09	-.201	-	-	.00142	-
	(.0000)	-	-	(.0000)	-
10	-.0143	-	-	-	-.00003
	(.0000)	-	-	-	(.0000)
11	-	.0120	-.00024	-	-
	-	(.0000)	(.0000)	-	-
12	-	-.0042	-	.00011	-
	-	(.0000)	-	(.0000)	-
13	-	-.0070	-	-	.00005
	-	(.0041)	-	-	(.0041)
14	-	-	-.00006	-.00008	-
	-	-	(.0000)	(.0000)	-
15	-	-	-.00024	-	-.00009
	-	-	(.0000)	-	(.0000)
16	-	-	-	-.00011	-.00003
	-	-	-	(.0000)	(.0000)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.1 (cont'd)

Values and P-values of Coefficients for Location 1251

Model	T	H	TH	T^2	H^2
17	.028	.0042	-.00067	-	-
	(.0000)	(.0000)	(.0000)	-	-
18	-.218	-.0026	-	.0015	-
	(.0000)	(.0000)	-	(.0000)	-
19	-.013	-.026	-	-	.00017
	(.0000)	(.0000)	-	-	(.0000)
20	-	-.228	-.00004	.0016	-
	-	(.0000)	(.0000)	(.0000)	-
21	.0029	-	-.00029	-	.00012
	(.1346)	-	(.0000)	-	(.0000)
22	-.2270	-	-	.0016	-.00002
	(.0000)	-	-	(.0000)	(.0000)
23**	-	.0431	-.00069	.00022	-
	-	(.0000)	(.0000)	(.0000)	-
24	-	-.0076	-.00023	-	.00014
	-	(.0117)	(.0000)	-	(.0000)
25	-	-.0264	-	-.00010	.00002
	-	(.0000)	-	(.0000)	(.0000)
26	-	-	-.0003	.00003	.00012
	-	-	(.0000)	(.0801)	(.0000)
27	-.167	.026	-.0004	.00134	-
	(.0003)	(.0005)	(.0002)	(.0000)	-
28	.0057	.0106	-.00032	-	.00005
	(.2017)	(.1112)	(.0017)	-	(.0102)
29	-.214	-.0119	-	.00148	.00007
	(.0000)	(.0001)	-	(.0000)	(.0025)
30	-.201	-	-.00013	.0015	.00004
	(.0000)	-	(.0010)	(.0000)	(.0180)
31	-	.0152	-.00038	.00007	.00005
	-	(.0314)	(.0001)	(.0672)	(.0154)
32	-.100	.0576	-.00076	.00010	.00007
	(.0271)	(.0000)	(.0000)	(.0029)	(.0044)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.2

Values and P-values of Coefficients for Location 1312

Model	T	H	TH	T^2	H^2
02	-.0108	-	-	-	-
	(.0000)	-	-	-	-
03	-	-.00013	-	-	-
	-	(.3347)	--	-	
04	-	-	-.00003	-	-
	-	-	(.0000)	-	-
05	-	-	-	-.00009	-
	-	-	-	(.0000)	-
06	-	-	-	-	-.000002
	-	-	-	-	(.2070)
07	-.0109	-.00040	-	-	-
	(.0000)	(.1118)	-	-	-
08	-.0106	-	-.00006	-	-
	(.0000)	-	(.1161)	-	-
09	-.0273	-	-	.00013	-
	(.0458)	-	-	(.1542)	-
10	-.0108	-	-	-	-.000002
	(.0000)	-	-	-	(.2253)
11	-	.0103	-.00018	-	-
	-	(.0000)	(.0000)	-	-
12	-	-.00040	-	-.00009	-
	-	(.1084)	-	(.0000)	-
13	-	-.0262	-	-	.00020
	-	(.0000)	-	-	(.0000)
14	-	-	-.000008	-.00008	-
	-	-	(.0705)	(.0000)	-
15	-	-	-.00015	-	-.00007
	-	-	(.0000)	-	(.0000)
16	-	-	-	-.000088	-.000002
	-	-	-	(.0000)	(.2428)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.2 (cont'd)

Values and P-values of Coefficients for Location 1312

Model	T	H	TH	T^2	H^2
17	.029	.0364	-.00062	-	-
	(.0000)	(.0000)	(.0000)	-	-
18	-.281	-.00024	-	.00014	-
	(.0464)	(.2416)	-	(.1511)	-
19	-.010	-.0107	-	-	.00008
	(.0000)	(.0009)	-	-	(.0012)
20	-	-.264	-.000005	.00013	-
	-	(.0591)	(.1657)	(.1753)	-
21	.0188	-	-.00044	-	.00020
	(.0000)	-	(.0000)	-	(.0000)
22	-.0285	-	-	.00014	-.000001
	(.0449)	-	-	(.1466)	(.3349)
23**	-	.0231	-.00039	.00011	-
	-	(.0000)	(.0000)	(.0021)	-
24	-	-.00031	-.00017	-	.000074
	-	(.4664)	(.0000)	-	(.0030)
25	-	-.0102	-	-.00008	.00008
	-	(.0015)	-	(.0000)	(.0022)
26	-	-	-.0005	.00017	.00021
	-	-	(.0000)	(.0000)	(.0000)
27	-.215	.075	-.00128	-.00116	-
	(.0000)	(.0000)	(.0000)	(.0000)	-
28	.0411	.0246	-.00080	-	.00017
	(.0000)	(.0000)	(.0000)	-	(.0000)
29	-.055	-.0131	-	.00036	.00010
	(.0012)	(.0003)	-	(.0064)	(.0004)
30	.0233	-	-.00048	-.00001	.00021
	(.0451)	-	(.0000)	(.4541)	(.0000)
31	-	.0118	-.00058	.00022	.00017
	-	(.0059)	(.0000)	(.0000)	(.0000)
32	.218	.0580	-.00145	-.00109	.00021
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.3

Values and P-values of Coefficients for Location 1341

Model	T	H	TH	T^2	H^2
02	-.00084	-	-	-	-
	(.0164)	-	-	-	-
03	-	-.0033	-	-	-
	-	(.0000)	-	-	-
04	-	-	-.00006	-	-
	-	-	(.0000)	-	-
05	-	-	-	-.00001	-
	-	-	-	(.0000)	-
06	-	-	-	-	-.00005
	-	-	-	-	(.0000)
07	-.00519	-.0070	-	-	-
	(.0000)	(.0000)	-	-	-
08	-.00183	-	-.00008	-	-
	(.0000)	-	(.0000)	-	-
09	.0565	-	-	-.00044	-
	(.0000)	-	-	(.0000)	-
10	-.0042	-	-	-	-.00008
	(.0000)	-	-	-	(.0000)
11	-	-.0027	-.00001	-	-
	-	(.0019)	(.3417)	-	-
12	-	-.0086	-	.00006	-
	-	(.0000)	-	(.0000)	-
13	-	-.0058	-	-	-.00013
	-	(.0007)	-	-	(.0000)
14	-	-	-.00009	-.00002	-
	-	-	(.0000)	(.0000)	-
15	-	-	-.00003	-	-.00006
	-	-	(.0770)	-	(.0000)
16	-	-	-	-.00004	-.00099
	-	-	-	(.0000)	(.0000)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.3 (cont'd)

Values and P-values of Coefficients for Location 1251 .

Model	T	H	TH	T^2	H^2
17	-.0166	-.0355	-.00051	-	-
	(.0000)	(.0000)	(.0000)	-	-
18	.0513	-.0069	-	.0004	-
	(.0000)	(.0000)	-	(.0000)	-
19	-.0054	-.0115	-	-	.00006
	(.0000)	(.0000)	-	-	(.0144)
20**	.0513	-	-.00013	-.00049	-
	(.0000)	-	(.0000)	(.0000)	-
21	-.0058	-	-.00013	-	-.00017
	(.0000)	-	(.0000)	-	(.0000)
22	.0516	-	-	-.00042	-.00007
	(.0000)	-	-	(.0000)	(.0000)
23	-	-.0310	.00042	-.00012	-
	-	(.0000)	(.0000)	(.0000)	-
24	-	.0060	-.00001	-	-.00013
	-	(.0014)	(.4043)	-	(.0000)
25	-	-.0144	-	-.00006	.00007
	-	(.0000)	-	(.0000)	(.0000)
26	-	-	.00013	-.00005	-.00018
	-	-	(.0000)	(.0000)	(.0000)
27	.0774	.0132	-.00037	-.00057	-
	(.0000)	(.0000)	(.0000)	(.0000)	-
28	-.0318	-.1000	.00097	-	.0005
	(.0000)	(.0000)	(.0000)	-	(.0000)
29	.0522	-.0173	-	-.00045	.00014
	(.0000)	(.0000)	-	(.0000)	(.0000)
30	.0679	-	-.00020	-.00053	.00006
	(.0000)	-	(.0000)	(.0000)	(.0001)
31	-	-.0717	.00062	-.00017	.00039
	-	(.0000)	(.0000)	(.0000)	(.0000)
32	.0717	.0037	-.00027	-.00055	.00006
	(.0000)	(.3494)	(.0045)	(.0000)	(.1376)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.4

Values and P-values of Coefficients for Location 2231

Model	T	H	TH	T^2	H^2
02	.00178	-	-	-	-
	(.0185)	-	-	-	-
03	-	-.00073	-	-	-
	-	(.0029)	-	-	-
04	-	-	-.00002	-	-
	-	-	(.0020)	-	-
05	-	-	-	.00002	-
	-	-	-	(.0090)	-
06	-	-	-	-	-.000004
	-	-	-	-	(.0180)
07	-.00002	-.00069	-	-	-
	(.4950)	(.0717)	-	-	-
08	-.00027	-	-.00001	-	-
	(.4083)	-	(.0260)	-	-
09	.0648	-	-	-.00052	-
	(.0293)	-	-	(.0328)	-
10	.0017	-	-	-	-.000001
	(.1377)	-	-	-	(.3799)
11	-	-.00020	-.00002	-	-
	-	(.4231)	(.1671)	-	-
12	-	-.00077	-	.000001	-
	-	(.0523)	-	(.4760)	-
13	-	-.0210	-	-	.00016
	-	(.0000)	-	-	(.0000)
14	-	-	-.00002	.000002	-
	-	-	(.0227)	(.4320)	-
15	-	-	-.00006	-	-.00002
	-	-	(.0008)	-	(.0101)
16	-	-	-	.00001	-.000001
	-	-	-	(.1786)	(.3288)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.4 (cont'd)

Values and P-values of Coefficients for Location 1251 .

Model	T	H	TH	T^2	H^2
17	.0428	.0402	-.00067	-	-
	(.0000)	(.0000)	(.0000)	-	-
18	.0796	-.0009	-	-.0007	-
	(.0127)	(.0277)	-	(.0125)	-
19	.0015	-.0204	-	-	.00016
	(.1499)	(.0000)	-	-	(.0000)
20	.0900	-	-.00002	-.00074	-
	(.0065)	-	(.0064)	(.0067)	-
21	.0151	-	-.00021	-	.00010
	(.0000)	-	(.0000)	-	(.0000)
22	.0681	-	-	-.00055	-.000002
	(.0261)	-	-	(.0288)	(.2936)
23	-	.0335	-.00056	.00029	-
	-	(.0000)	(.0000)	(.0000)	-
24	-	-.0240	.00003	-	.00017
	-	(.0000)	(.0660)	-	(.0000)
25	-	-.0208	-	-.00001	.00016
	-	(.0000)	-	(.2423)	(.0000)
26	-	-	-.00022	.00012	.00010
	-	-	(.0000)	(.0000)	(.0000)
27	.2802	.0637	-.0011	-.0018	-
	(.0000)	(.0000)	(.0000)	(.0000)	-
28	-.0002	-.0226	-.000004	-	.0002
	(.4937)	(.0441)	(.4897)	-	(.0000)
29**	.1419	-.0246	-	-.0012	.00019
	(.0000)	(.0000)	-	(.0000)	(.0000)
30	.1777	-	-.00030	-.0013	.00014
	(.0000)	-	(.0000)	(.0000)	(.0000)
31	-	-.0284	.00006	-.00004	.00019
	-	(.0047)	(.3001)	(.2881)	(.0000)
32	.2107	.0197	-.00054	-.0015	.0001
	(.0111)	(.2367)	(.0539)	(.0076)	(.0466)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.5

Values and P-values of Coefficients for Location 2233

Model	T	H	TH	T^2	H^2
02	.0042	-	-	-	-
	(.0000)	-	-	-	-
03	-	-.0016	-	-	-
	-	(.0000)	-	-	-
04	-	-	.00003	-	-
	-	-	(.0000)	-	-
05	-	-	-	.00003	-
	-	-	-	(.0000)	-
06	-	-	-	-	-.00002
	-	-	-	-	(.0000)
07	.0049	.0025	-	-	-
	(.0000)	(.0000)	-	-	-
08	.0035	-	.00005	-	-
	(.0000)	-	(.0000)	-	-
09	.0222	-	-	-.00014	-
	(.0000)	-	-	(.0000)	-
10	.0046	-	-	-	.00003
	(.0000)	-	-	-	(.0000)
11	-	-.0080	.00017	-	-
	-	(.0000)	(.0000)	-	-
12	-	.0020	-	.00004	-
	-	(.0000)	-	(.0000)	-
13	-	-.0106	-	-	.00011
	-	(.0000)	-	-	(.0001)
14	-	-	.00005	.00003	-
	-	-	(.0000)	(.0000)	-
15	-	-	.00016	-	-.00008
	-	-	(.0000)	-	(.0000)
16	-	-	-	.00003	.000023
	-	-	-	(.0000)	(.0000)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.5 (cont'd)

Values and P-values of Coefficients for Location 2233

Model	T	H	TH	T^2	H^2
17	-.0066	-.0196	.00035	-	-
	(.0000)	(.0000)	(.0000)	-	-
18	.0234	.00036	-	-.00015	-
	(.0000)	(.1758)	-	(.0000)	-
19	.0048	.0010	-	-	.000015
	(.0000)	(.2980)	-	-	(.2573)
20	.0132	-	.00004	-.000071	-
	(.0000)	-	(.0000)	(.0001)	-
21	.0004	-	.00015	-	-.00007
	(.0000)	-	(.2322)	-	(.0000)
22	.0228	-	-	-.00014	.000004
	(.0000)	-	-	(.0000)	(.1880)
23	-	-.0148	.00027	-.00003	-
	-	(.0000)	(.0000)	(.0000)	-
24	-	-.0092	.00016	-	.00002
	-	(.0014)	(.4043)	-	(.1411)
25	-	-.00058	-	.00004	.00003
	-	(.3866)	-	(.0000)	(.1155)
26	-	-	.00015	.000002	-.00008
	-	-	(.0000)	(.3197)	(.0000)
27	-.1587	-.0238	.00042	.000053	-
	(.0001)	(.0000)	(.0000)	(.0114)	-
28	-.0092	-.0330	.00041	-	.0001
	(.0000)	(.0000)	(.0000)	-	(.0000)
29	.0191	-.0005	-	-.00011	.00002
	(.0000)	(.4121)	-	(.0000)	(.1787)
30	.0051	-	.00013	-.000032	-.00006
	(.0375)	-	(.0000)	(.0491)	(.0000)
31**	-	-.0277	.00033	-.00049	.00010
	-	(.0000)	(.0000)	(.0000)	(.0000)
32	-.0176	-.0365	.00047	.00005	.00011
	(.0001)	(.0000)	(.0000)	(.0298)	(.0000)

Figures in the parenthesis indicate P-values
** indicates the best selected model

TABLE 4.6

P-values of Chi-square for Different Location

Model	1251	1312	1341	2231	2233
02	0.00006	0.00000	1.00000	1.00000	0.00000
03	1.00000	1.00000	0.08825	1.00000	0.99985
04	0.94713	0.66336	0.19321	0.99166	0.99992
05	0.00443	0.00000	0.98669	1.00000	0.00000
06	1.00000	1.00000	0.00349	1.00000	1.00000
07	0.00000	0.00000	0.00000	0.99664	0.00000
08	0.00000	0.00000	0.13483	0.98814	0.00000
09	0.00000	0.00000	0.00000	1.00000	0.00000
10	0.00000	0.00000	0.00000	1.00000	0.00000
11	0.00000	0.00000	0.09277	0.98906	0.00000
12	0.00000	0.00000	0.00000	0.99612	0.00000
13	1.00000	0.23504	0.00109	0.00000	0.74560
14	0.00000	0.00000	0.00000	0.98996	0.00000
15	0.00000	0.00000	0.00475	0.94356	0.00000
16	0.00000	0.00000	0.00000	1.00000	0.00000
17	0.00000	0.00000	0.00000	0.00000	0.00000
18	0.00000	0.00000	0.00000	0.99466	0.00000
19	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.00000	0.00000	0.00000*	0.99233	0.00000
21	0.00000	0.00000	0.00000	0.00000	0.00000
22	0.00000	0.00000	0.00000	0.99968	0.00000
23	0.00000*	0.00000*	0.00000	0.00014	0.00000
24	0.00000	0.00000	0.00100	0.00000	0.00000
25	0.00000	0.00000	0.00000	0.00000	0.00000
26	0.00000	0.00000	0.00000	0.00000	0.00000
27	0.00000	0.00000	0.00000	0.00000	0.00000
28	0.00000	0.00000	0.00000	0.00000	0.00000
29	0.00000	0.00000	0.00000	0.00000*	0.00000
30	0.00000	0.00000	0.00000	0.00000	0.00000
31	0.00000	0.00000	0.00000	0.00000	0.00000*
32	0.00000	0.00000	0.00000	0.00000	0.00000

* indicates P-value of the selected model

TABLE 4.7

Values of the Intercept for Different Location

Model	1251	1312	1341	2231	2233
02	1.300	1.3900	0.702	0.757	0.488
03	0.779	0.7340	0.745	0.912	0.822
04	0.841	0.8524	0.761	0.928	0.693
05	1.020	1.0700	0.709	0.804	0.632
06	0.764	0.7330	0.700	0.884	0.785
07	2.010	1.4300	1.187	0.911	0.350
08	1.730	1.4100	0.905	0.906	0.415
09	7.830	1.9000	-1.090	-1.149	-0.075
10	1.830	1.4100	1.010	0.768	0.414
11	1.000	0.7670	0.744	0.935	0.670
12	1.520	1.0800	1.140	0.918	0.529
13	0.990	1.5600	0.610	1.520	0.989
14	1.370	1.0800	0.910	0.918	0.529
15	1.390	1.0560	0.660	1.022	0.520
16	1.360	1.0700	0.950	0.829	0.576
17	-0.950	-0.9480	1.860	-1.739	1.101
18	8.710	1.9400	-0.580	-1.457	-0.132
19	2.610	1.7000	1.270	1.393	0.391
20	8.930	1.8900	-0.958*	-1.781	0.129
21	1.290	0.4680	0.974	0.330	0.496
22	8.890	1.9400	-0.763	-1.226	-0.113
23	0.039*	0.3470*	1.329	-0.197	0.806
24	1.600	1.0700	0.614	1.532	0.687
25	2.180	1.3700	1.268	1.474	0.587
26	1.380	1.0200	0.845	0.810	0.509
27	6.060	-7.7700	-1.585	-9.679	1.460
28	0.830	-0.9600	3.413	1.583	1.457
29	8.860	3.1300	-0.398	-2.680*	-0.020
30	8.050	0.3350	-1.125	-4.544	0.367
31	0.880	0.6890	2.023	1.764	1.084*
32	2.850	-7.3800	-1.267	-6.169	1.775

* indicates Intercept of the selected model

TABLE 4.8

Values of the Sum of Squares of Error(SSE) for Different Location

Model	1251	1312	1341	2231	2233
02	0.8213	0.3957	2.642	0.1365	0.6896
03	0.8560	0.5139	2.584	0.1357	0.7758
04	0.8644	0.5028	2.592	0.1357	0.7841
05	0.8224	0.3953	2.607	0.1365	0.7032
06	0.3035	0.5151	2.553	0.1369	0.7644
07	0.9064	0.3966	2.416	0.1358	0.7184
08	0.2415	0.3960	2.548	0.1356	0.7204
09	0.8050	0.3963	1.990	0.1349	0.6602
10	0.8852	0.3964	2.417	0.1363	0.7228
11	0.8748	0.3896	2.582	0.1357	0.6776
12	0.9116	0.3973	2.278	0.1357	0.7239
13	0.8703	0.5082	2.523	0.1222	0.7470
14	0.8952	0.3966	2.480	0.1356	0.7313
15	0.9190	0.3953	2.549	0.1347	0.7120
16	0.8875	0.3963	2.309	0.1364	0.7207
17	0.8088	0.3954	2.127	0.1365	0.6584
18	0.8436	0.3971	1.796	0.1337	0.6620
19	0.9846	0.3980	2.437	0.1217	0.7126
20	0.8379	0.3970	1.705*	0.1329	0.6903
21	0.9272	0.4083	2.361	0.1252	0.7198
22	0.8325	0.3969	1.834	0.1347	0.6609
23	0.2394*	0.2394*	2.002	0.1393	0.6623
24	0.9532	0.3891	2.523	0.1213	0.6849
25	0.9922	0.3985	2.273	0.1219	0.7180
26	0.9288	0.4175	2.257	0.1269	0.7208
27	0.8073	0.4260	1.784	0.1224	0.6619
28	0.8784	0.4091	1.941	0.1228	0.6328
29	0.8788	0.4001	1.781	0.1068*	0.6738
30	0.8487	0.4138	1.775	0.1157	0.7012
31	0.8704	0.4063	1.856	0.1215	0.5903*
32	0.7632	0.4750	1.784	0.1171	0.6368

* indicates Sum of Squares of Error of the selected model

4.3 SELECTED MODELS FOR DIFFERENT LOCATIONS

According to the criteria given above, we summarize the best model for each location from Table 4.1 through Table 4.8 .

Location 1251

$$Y = -0.0395 + 0.0431\ H - 0.00069\ TH + 0.00022\ T^2$$

Location 1312

$$Y = 0.368 + 0.0231\ H - 0.00039\ TH + 0.00011\ T^2$$

Location 1341

$$Y = -0.9585 + 0.0613\ T - 0.00013\ TH - 0.00049\ T^2$$

Location 2231

$$Y = -2.6795 + 0.1419\ T - 0.0246\ H - 0.0012\ T^2 + 0.00019\ H^2$$

Location 2233

$$Y = 1.0836 - 0.0277\ H + 0.00033\ TH - 0.00009\ T^2 + 0.0001\ H^2$$

REMARKS:

The best model selected by M-estimation technique for different locations shows that there is no simple linear relationship of the ratio of concentration of dust with temperature and humidity . The selected best model for each locations are all quadratic. It reveals from the selected models that the interaction effect are present in all location except location 2231.

4.4 PERFORMANCE OF THE SELECTED MODELS

Now we evaluate the performance of the above regression equations at average temperature and humidity levels i.e., for $H=\overline{H}$ and $T=\overline{T}$. Y values for each of the five locations are as shown in Table 4.9.

TABLE 4.9

Estimated Y values for Different locations at Average Level of T and H

Variable	1251	1312	1341	2231	2233
$\bar{H}$	63.710	65.710	29.500	66.500	37.000
$\bar{T}$	67.057	61.929	64.000	59.958	65.200
Y	0.748	0.721	0.712	0.719	0.609

It reveals from Table 4.9 that performance of NY over ME for location 1251, 1312, 1341, 2231 and 2233 are 75, 72, 71, 72 and 61 percent respectively which confirms the result in Knight and Moore (1984). However, the result indicate that performance of ME is better than NY.

4.5 PARTIAL EFFECT OF HUMIDITY OR TEMPERATURE ON CONCENTRATION

We evaluate the partial deravites of ratio of concentration of dust at the mean temperature and/or mean humidity for the selected models in each location . Results are shown in Table 4.10 .

TABLE 4.10

Partial Change of Concentration for Different Locations

Locations	$\frac{\partial Y}{\partial H}$	$\frac{\partial Y}{\partial T}$	$\frac{\partial^2 Y}{\partial H \partial T}$	$\frac{\partial^2 Y}{\partial H^2}$	$\frac{\partial^2 Y}{\partial T^2}$
1251	-0.00317	-0.01445	-0.00069	0.00044	0.00000
1312	-0.00105	-	-0.00039	0.00039	0.00022
1341	-0.00832	-	-0.00013	-0.00000	-0.00098
2231	-	-	0.00000	0.00038	-0.00240
2233	-	-	0.00033	0.00010	-0.00018

- indicates that variable still present

Table 4.10 shows that partial change of the ratio of concentration of dust for the unit change of humidity at average level of temperature is maximum for location 1312 and minimum for location 1341.

4.6 M-ESTIMATES : A COMPARISON WITH MULTIPLE REGRESSION TECHNIQUE

In this section we compare the Y with the predicted value of Y by M-method (Y_{ro}) and Multiple regression method (Y_{ls}) . we also show the sum of squares of error obtained by these methods .

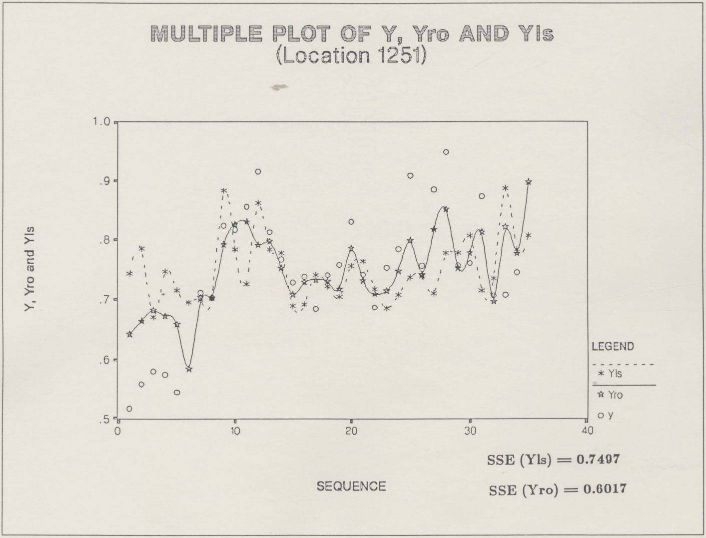
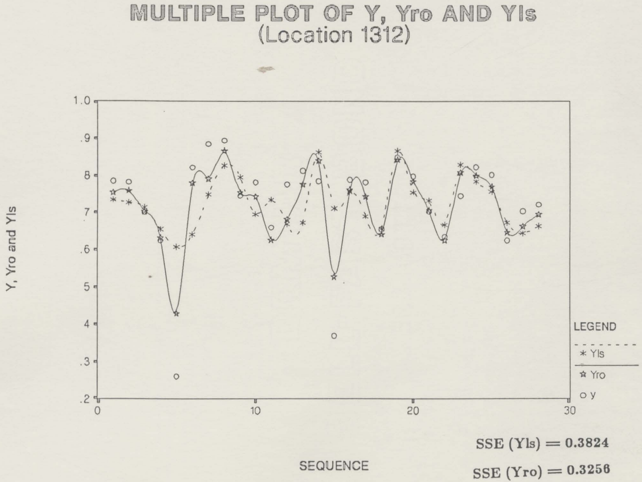


Figure 4.1 Multiple plot of Y Yro Yls



SSE (Yls) = 0.3824

SSE (Yro) = 0.3256

Figure 4.1 (cont'd)

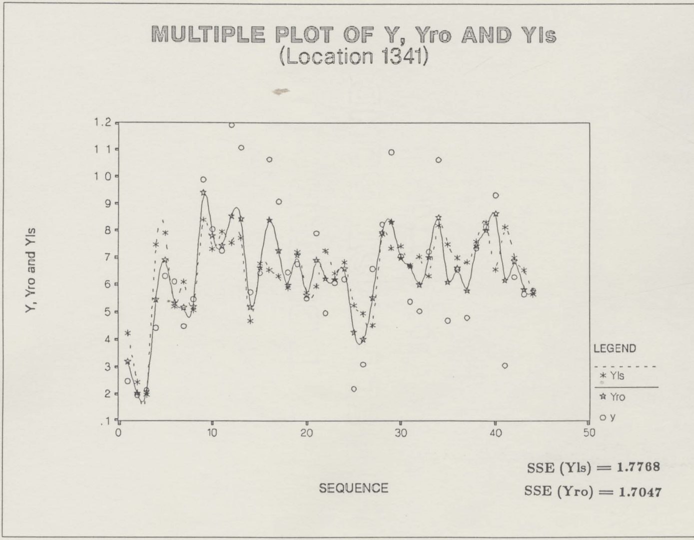


Figure 4.1 (cont'd)

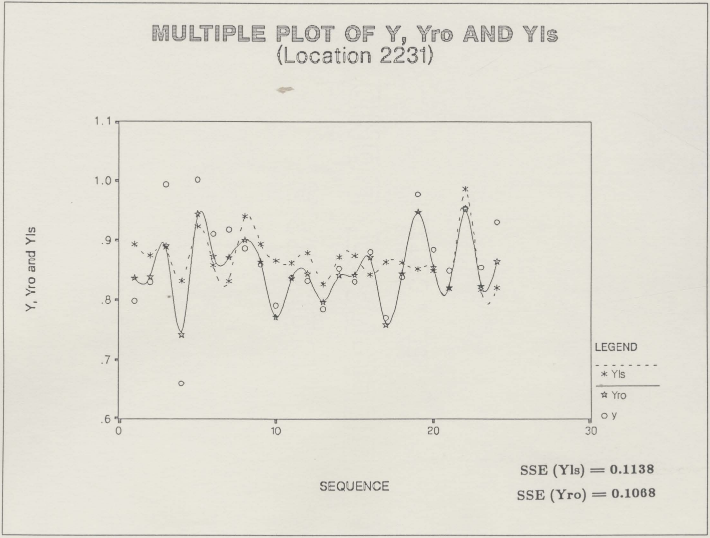


Figure 4.1 (cont'd)

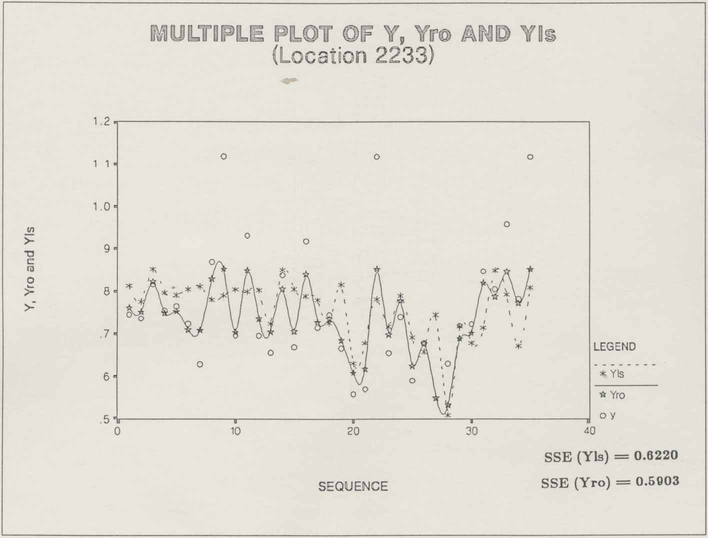


Figure 4.1 (cont'd)

It is clear from Figure 4.1 that the M - estimation technique fits the observed data better than Multiple regression by least square method. This is not surprising because the data were generally asymmetric including some outliers. Thus we conclude that M-estimation technique always perform better than Multiple regression by least square method in this particular situation.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The data on ratio of dust concentration measured by two machines ME and NY , were collected for five major locations in two mines in Labrador, Canada. The dust concentration is usually affected by production, humidity and temperature. There was no data available on production. Consequently, the ratio of dust concentration measured by two machines were considered and its relationship with temperature and humidity was examined. As a basic tool of statistical analysis the distribution of the ratio of concentration was discussed in chapter 2 . It was found that the distribution of the ratio variable is approximately normal for location 1251 and 1312. The data for other three locations were highly skewed. Consequently for fitting the model by multiple regression based on least squares method the data for these three locations were transformed by using Box and Cox power transformation function. The square root transformation were applied in location 1341, inverse transformation in location 2231 and log transformation in location 2233.

Multiple regression based on least squares method using the transformed data shows that there is no linear relationship of ratio of concentration of dust with temperature and humidity. Interaction between temperature and humidity were found to be significant in all locations except location 2231. It also shows that in location 2231, ratio of concentration of dust is not influenced by temperature. The coefficient of determination (R^2) was found to be highest (37.8 %) for location 1341 and lowest (15.5 %) for location 2231. The transformed ratio of dust concentration level was found to be higher for ME machine in comparison to NY machine except for location 2231. Partial change of the transformed ratio of concetration of dust for unit change of humidity at average level of temperature was maximum for location 2233 and minimum for location 1341. Similarly, the partial change of the transformed ratio of concentration of dust for unit change of temperarure at average humidity level was maximum for location 2233 and minimum for location 1251.

As the distribution of the ratio of concentration for most of the location were found asymmetric the M-estimation of regression was studied in chapter 3. It was found that there was no linear relationship of ratio of concentration of dust with humidity and temperature. The interaction effect was found to be significant for all location except 2231. The ratio of dust concentration was found to be higher for ME machine than NY machine. The partial change of the ratio of concentration of dust for the unit change of humidity at average level of temperature is maximum for location 1312 and minimum for location 1341.

It has been found that the best selected model based on M-estimation technique is not adequate when fitted by multiple regression based on least squares principle. This is not surprising because the data were asymmetric, whereas multiple regression principle requires symmetric data for testing the significance of the model. For all locations the sum squares of error by M-estimation technique were less than the sum of squares of error based on least square principle , this gives the impression that M-estimation technique always performs better than estimation by least squares method.

REFERENCES

Andrews, D.F (1974), " A robust method for multiple linear regression," Technometrics, vol. 16, 523-531.

Andrews, D.F., Bickel, P.J., Hample, F.R., Huber, P.J., Rogers, W.H. and Tukey J.W. (1972), " Robust estimations of location : survey and advances," Princeton, NJ: Princeton University Press.

Beisley, D. A., Kuh, E. and Welsch, R. E. (1980), " Regression diagnostics: Identifying influential data and sources of collinearity, " Wiley, New York.

Bickel, P.J. (1976) , " Another look at robustness," Scandinavian Journal of Statistics, vol. 3 , 145-168.

Box, G.E.P. and D.R. Cox. (1964), " An Analysis of Transformations," Journal of the Royal Statistical Society, vol. B-26, 211-243.

Carrol, R.J. (1978), " An investigation into the effects of asymmetry on robust estimates of regression," University of North Carolina Department of Statistics Mimeo Series, No. 1172 .

Carroll, R.J. (1979), " On estimating variances of robust estimators when the errors are asymmetric," Journal of the American Statistical Association, Vol. 74, 674-679 .

Crow, E.L. and Siddiqui, M.M. (1967), " Robust estimates of location," Journal of the American Statistical Association, Vol. 62 , 353-389.

Cody, R.P. and Smith, J.K. (1985), " Applied statistics and the SAS progaamming language," Elsevier Science Publishing, New York.

Daniel, C. and Wood, F.S. (1980)," Fitting Equations to Data," Wiley, New York.

Dixon, W.J. (1988), " BMDP Statistical Software Manual" , Barkeley, California.

Draper, N.R. and Smith, H. (1981) , " Applied regression analysis," Wiley, New York,

Fairman, R.P., O'Brain, R.J., Swecker, S., Amandos, H.E., and Shoub, E.P.(1977), " Respiratory status of surface coal miners in the United States," Archives of Environmental Health, September/october, 211-215.

Knight, G. and Moore, E. (1984) " Comparison of respirable dust samplers for use in hard rock mines," American Industrial hygiene conference, Detriot.

Gregory, J. (1971) , " A study of 340 cases of chronic bronchitis," Archives of Environmental Health, vol. 22 , 428-439.

Hartwig, F. and Dearing, B.N. (1979), " Exploratory data analysis ," Sage University Paper series on Quantitative Applications in the Social Sciences, 07-001. Beverly Hills and London: Sage Pubns.

Higgins, R. and Dewell, P. (1967), " A gravimetric Size-Selecting Personal Dust Sampler, Inhaled Particles and Vapours II," Editor C .N. Davies , 575

Pergmon, London.

Hill, R.W. and Holland, P.W. (1977), " Two robust alternatives to Least-Squares regression," Journal of the American Statistical Association, Vol.72,1041-1067.

Hinich, M.J. and Cohen, M.L. (1975) " A simple method for robust regression," Journal of the American Statistical Association, Vol. 70 , 113-119.

Hogg, R.V.(1967), " Some observations on robust estimation," Journal of the American Statistical Association, Vol. 62 , 1179-1186.

Hogg, R.V.(1974), " Aduptive robust Procedures : a partial review and some suggestions for future applications and theory," Journal of the American Statistical Association, Vol. 69 , 909-927.

Hogg, R.V.(1979), " An introduction to robust estimation," In Robustness in statistics (Launer, R.L. and Wilkinson, G.N., ed.), Academic press, New York, 1-18.

IMSL, Inc. (1987) , User's Manual STAT/Library Fortran Subroutines for Statistical Analysis .

Jaeckel, L.A. (1971), " Some flexible estimates of location," Annals of Mathematical Statistics, vol. 42, 1540-1552.

Jaeckel, L.A. (1972), " Estimating regression coefficients by minimizing the dispersion of the residuals," Annals of Mathematical Statistics, vol. 43, 1449-1458.

Jureckova, J. (1971), " Nonparametric estimate of regression coefficients," Annals of Mathematical Statistics, vol. 42, 1328-1338.

Huber, P.J. (1964), " Robust estimation of a location parameter," Annals of Mathematical Statistics, vol. 35, 73-101.

Huber, P.J. (1972), " Robust statistics : a review, " Annals of Mathematical Statistics, vol. 43, 1041-1067.

Huber, P.J. (1973), " Robust regression : asymptotics, conjectures and Monte Carlo, " Annals of Mathematical Statistics, vol. 1, 799-821.

McGill, R., Tukey, J.W. and Larsen, W.A. (1978), " Variations of box plots," Journal of the American Statistical Association, Vol. 32 , 12-16.

Minitab Inc. (1988), Minitab Reference Manual, Release 6.1

Moberg, T.E., Ramberg, J.S., and Randels, R. H. (1980), " An adaptive multiple regression procedure based on M-estimators," Technometrics, vol. 22, 213-224.

Mosteller, F. and Tukey, J.W.(1977), " Data analysis and regression," Reading, MA: Addison-Wesley.

Musk, A.W., Peters, J.W., Wegmon, D.H., and Fine, L.J. (1977), " Pulmonary function in granite dust exposure: a four year follow up," American review of respiratory disease, vol. 115, 769-776.

Ryan, B.F., Joiner, B.L. and Ryan, T.A. (1985), " Minitab Handbook," Second edition, Duxbury Press.

Seber, G.A.F. (1977), " Linear Regression Analysis," Wiley, New York.

SPSS Inc. (1988), SPSS-X User's Guide, New York.

SPSS Inc. (1986), SPSS Graphics , New York.

SPSS Inc. (1985), SPSS-X Advanced Statistics Guide, New York.

Stigler, S.M. (1977), " Do robust estimators work with real data?" Annals of Statistics, vol. 5, 1055-1098.

Tukey, J.W. (1977), " Exploratory data analysis," Reading, MA: Addison-Wesley,

Tukey, J.W. (1975), " Instead of Gauss-Markov least squares, what ?,"In Applied statistics (R.P. Gupta, ed.), North-Holland, Amsterdam, 351-372.

Tukey, J.W. (1980) , " We need both exploratory and confirmatory, "Journal of the American Statistical Association, Vol. 34, 23-25.

Velleman, P. F. and Hoaglin, D. C. (1981), " Applications, Basics and Computing of exploratory data analysis," Duxbury press, Boston .

Wetherill, G.B. (1986), " Regression analysis with applications," Chapman and hall, London.

Wishart, J. and Metakides, T. (1953) , " Orthogonal polynomial fitting," Biometrika, vol. 40, 361-369.

Yohai, V.J. (1974), " Robust estimation in the linear model," Annals of Statistics, vol. 2, 562-567.