TREND ANALYSIS OF MONTHLY ACID RAIN DATA – '80 - '86

By

SHIYING WU

B. S. (Mathematics), Peking University, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1988

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date Oct., 17, 1988

DE-6 (2/88)

ABSTRACT

Three-way median polish is used to model the monthly concentrations of three kinds of ions in precipitation, namely sulphate, nitrate and hydrogen ions. In contrast to previous findings that the wet acid deposition had decreased from late 70's to early 80's, the results suggest that there is a V-shaped trend for wet acid deposition during the period of 1980 - 1986 with the change point around 1983. Strong seasonality is also discovered by the analysis. Nonparametric monotone trend tests are performed on the data collected from 1980 to 1986 and on the data collected from 1983 to 1986 separately. The results are consistent with the findings from the median polish approach. A nonparametric slope estimate of the trend is obtained for each monitoring station. Based on these estimates, the slope estimate is obtained by Kriging interpolation for each integer degree grid point of longitude and latitude across the 48 conterminous states in the United States.

Also, a geographical pattern in the data is suggested by hierarchical clustering and by median polishing.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

ACKNOWLEDGEMENT

I would like to thank Professor J. V. Zidek for his guidance and assistance in producing this thesis. I would like to express my gratitude to Doctor Jian Liu, who kindly gave me helpful advice and carefully read this thesis.

The work reported here relies heavily on the data provided by the Battelle's Pacific Northwest Laboratories; special thanks go to Doctor A. R. Olsen and Judy Sweeney. I also would like to express my appreciation to Doctor H. Joe and P. Schumacher for their assistance with the computing issues.

This work was partially supported by the Environmental Protection Agency (U.S.) through a co-operative research agreement with Societal Institute for Mathematics in Society.

Chapter 1 INTRODUCTION

The Acid Deposition System (ADS) is an integrated, centralized repository for data from monitoring networks in North America. The purposes of ADS are (1) to facilitate access to deposition data collected by different organizations, (2) to provide annual statistical summaries of available data, and (3) to maintain the data for assessment of long-term trends. A complete description of ADS is available in a system design and user's code manual by Olsen and Slavich (1986).

The data set used in this study is obtained from ADS Data Summary (Olsen and Slavich ,1986) and it contains only the data from the NADP network, the largest and most important network among the networks represented in the ADS Data Base.

Three important ions of the data set have been singled out for study in this report: SO_4, NO_3 and H^+ (derived from pH). Separate analyses of each have been done in this study.

1.1 Purpose and Scope

The primary purpose of this study is to detect and estimate the possible temporal trends in different levels of chemical constituents of acid deposition at different locations. The analysis is divided into 2 sections, one for the data collected from 1980 to 1986 (the "historical data"), and the other for the data collected from 1983 to 1986 (the "recent data"). In this way we have tried to differentiate between what happened in the past seven years from what happened in the last four years.

A secondary purpose of this study is the classification of the stations into groups such that patterns in the deposition chemistry of the stations in each group are similar. The ultimate aim is to reduce the number of stations in the network to reduce the cost of maintaining the network without loosing too much information. But this issue will not be addressed in this report.

Spatial patterns and seasonal patterns of the level of chemical concentration in wet deposition are examined as well.

Over time, the NADP network came to include more than 200 monitoring stations. Many of them either started to operate after 1980 or have a lot of missing data. For convenience and with little apparent loss of information, we consider only the stations with 5 or fewer missing observations. The resulting sets of stations are different for various chemicals under consideration. Roughly speaking, there are just over 80 stations providing the recent data and just over 30 stations providing the historical data. The stations are located throughout the United States and there are more stations in the East than in the West.

1.2 A Review of Previous Works

A lot of assessments of acid precipitation have been done during the past years. Among those are Schertz and Hirsch (1985), Eynon and Switzer (1983), Finklestein (1984), Cape and Fowler (1984), Vong *et al* (1985), Bilonick (1985, 1987), Le and Petkau (1986), Dana and Easter (1986), Lettenmaier (1986), and Altwicker and Johannes (1987).

Finklestein (1984) estimates the variograms of H^+, SO_4, NO_3 and NH_4. These variogram estimates are computed from the observations collected at 31 NAPD stations during the period from July 1979 to June 1980, and are shown to be distance dependent in all cases, increasing with the distance between stations.

Bilonick (1985) estimated the space-time semi-variograms of sulfate deposition using the data collected from USGS network and did a 3-dimensional Kriging. He (1987) used indicator Kriging and pointed out that "it is a nonparametric technique" and "is useful for highly skewed data distribution and is resistant to outliers", while "the entire probability distribution function at a given point is estimated".

Dana and Easter (1986) applied concentration-time regression analysis to the data from MAP3S network, collected from 1977 to 1983, to estimate temporal trends. Their conclusions are that the levels of sulfur, H^+, NH_4, and NO_3 declined during that period.

Schertz and Hirsch's study cited above contains three parts: (1) modeling variations in concentration; (2) trend analysis; and (3) multiple station analysis. In the first part, they developed a model:

$$\log(C)=a+b^*\log(P)+c*(2\pi T)+d*(2\pi T),$$

and used regression to estimate the parameters. In the second part, they used the Seasonal Kendall Test to detect trends in the concentrations of constituents with the conclusion that most of the stations showed no trends or down trends during the period from 1978 to 1983. In the third part, they analyzed the relationship between the correlation coefficients and the distance of pairs of stations. Our study is more closely related to the second part of their work.

1.3 Data Base Description

ADS requires that networks provide documentation on their network operation and that a minimum level of information accompany each sample result. Networks not able to provide the required information are unable to transfer data to ADS. In preparing concentration and deposition summaries, four related steps occur. First, network protocols and data screening procedures are determined and an algorithm to translate this information along with the sample results to the ADS data base is constructed. Second, valid sample criteria for the data summary are determined. They are: (a) all sampling periods for which it is known that no precipitation occurred are considered valid sample periods; (b) the wet deposition sample must be a wet-only sample; (c) Wet deposition samples that have insufficient precipitation to complete a chemical analysis for a specified ion are invalid for that specific ion species; (d) an individual ion species concentration accompanied by a comment code designating the measurement to be "suspect" or "invalid" is declared as an invalid sample; and (e) The actual sampling period for a wet deposition sample must be close to the network's protocol sampling period. Third, data completeness measures for each summary are computed. Fourth, criteria based on data completeness measures and site representativeness are selected for reporting a specific data summary. For more details, see Olsen and Slavich (1986).

One of the networks contained in ADS is NADP/NTN. The National Atmospheric Deposition Program (NADP) was established in 1978 to monitor trends in the exposure of various ecosystems to acidic deposition in the United States. The NADP was created by the Association of State Agricultural Experiment Stations (originally as North Central Regional Project NC-141, now Interregional Project IR-7) to conduct research on atmospheric deposition and its effects, in cooperation with federal, state, and private research agencies. A major program objective is to discover and characterize biologically important geographical and temporal trends in the chemical climate of North America through the continued development and maintenance of a deposition monitoring network.

Since its inception the network has grown from 22 operational sites during 1978 to 222 sites in 1986.

The Deposition Monitoring Task Group of the Interagency Task Force on Acid Precipitation was charged in the National Acid Precipitation Assessment Plan (Interagency Task Force on Acid Precipitation, 1982) with developing a National Trends Network (NTN). The objective of the 150-station National Trends Network is to provide long-term monitoring (10-years minimum) at sites across the United States that represent broad regional characteristics of the chemistry of wet deposition. Robertson and Wilson (1985) describe the design of NTN. Many existing NADP sites were selected as NTN sites. Because of the operating modes of NADP and NTN, the two networks are considered to be a single network and the acronym NADP/NTN is used to refer to sites which are either NADP or both NADP and NTN. In the following, however, we refer to it simply as NADP.

The NADP monitoring protocol is based on a weekly (Tuesday to Tuesday) sampling protocol with wet-only sample collection. The NADP program has developed and adheres to strict requirements regarding sample collection and analysis. The requirements assure uniformity in siting criteria, sampling protocol, analytical chemistry techniques, data handling, and overall network operation. All NADP precipitation chemistry samples are analyzed by the Central Analytical Laboratory at the Illinois State Water Survey. For specific details the reader may consult existing publications on siting criteria (NADP 1984a), site operation and collection protocol (NADP 1982), overall quality assurance plans (NADP 1984b), and analytical procedures(NADP 1980).

Chapter 2 METHODOLOGY BACKGROUND

The statistical procedures involved in this paper are clustering, median polishing, nonparametric trend testing and slope estimation, and Kriging.

2.1 Clustering

In the single sample problem, $(x_{i1}, \cdots, x_{in})$ is the ith observed n-dimensional sample, x_i, $i = 1, ..., m$, which may well be heterogeneous. The aim of cluster analysis is to group these samples into g homogeneous classes where g is unknown, $g \leq m$. The multi-sample variant involves x_{ik}, $k = 1, \cdots, n_i$, the observations of the ith sample x_i, $i=1, \cdots, m$ and again the aim is to group the m samples into homogeneous classes (Mardia, Kent, Bibby, 1979).

Here, we consider only hierarchical methods which cluster the g groups into g+1 groups according to the distances between the samples, dividing one group into two and keeping the others the same. One way to do this is follows. Start with $\{C_i(1) = x_i, i = 1, \cdots, m\}$. Suppose $\{C_i(p), i = 1, \cdots, m - p + 1\}$ are the clusterings at step p. Define $D_{ij}(p)$ to be a measure of distance between $C_i(p)$ and $C_j(p)$. Let

$$D_{12}(p) = \mathbf{min}\{D_{ij}(p): i, j = 1, \cdots, m\text{-}p+1, i \neq j\}.$$

Then set

$$C_1(p+1) = C_1(p) \cup C_2(p)$$

and

$$C_i(p+1) = C_{i+1}(p), \qquad i = 1, \cdots, m\text{-}p+1.$$

Continue this procedure until all the inter-cluster distances are greater than D_0 where D_0 is an arbitrary threshold value.

If d_{rs} is taken to be the distance between $\mathbf{x}_r$ and $\mathbf{x}_s$ (the distance can be defined in different ways), and if

$$D_{ij}(p) = \mathbf{min}\{d_{rs}\text{: (r, s) such that } x_r \in C_i(p),\ x_s \in C_j(p)\}$$

the method is called a single linkage method. If

$$D_{ij}(p) = \mathbf{max}\{d_{rs}\text{: (r, s) such that } x_r \in C_i(p),\ x_s \in C_j(p)\}$$

then the method is called a complete linkage method.

The clusters obtained by a single linkage method are "rod" type elongated clusters without nuclei. This leads to a chaining effect. Chaining can be misleading if items at opposite ends of the chain are quite dissimilar (Johnson and Wichern, 1982).

On the other hand, the clusters obtained by the complete linkage method tend to be compact clusters without a chaining effect. Thus the within-group distances of the resulting groups are all less than the threshold value D_0.

Since the ultimate purpose of the clustering used here is to choose representatives from each cluster and we expect the inter-group distances to be small, we choose complete linkage in this study. For the same reason, the distances used in this study are defined as

$$d_{ij}^2 = \sum_{k=1}^{n}(x_{ik}-x_{jk})^2/(n-1), \qquad i, j = 1, \cdots, m,$$

where $n = n_i$.

In the single sample case, such a d_{ij} is equivalent to Euclidean distance since the method is invariant under any monotonic transformation of d_{ij}. In multi-sample problems where the samples are one dimensional, if $\mathbf{x}_i$ has mean μ_i and variance σ_i^2 $(i = 1,\cdots, m)$, and if the covariance of $\mathbf{x}_i$ and $\mathbf{x}_j$ is denoted by ρ_{ij}, d_{ij}^2 is an estimate of

$$\mathrm{Var}(x_i\text{-}x_j) = \sigma_i^2 + \sigma_j^2 - 2\,\rho_{ij} + (\mu_i - \mu_j)^2, \qquad i, j=1, \cdots, m.$$

Thus a small d_{ij} requires that both ρ_{ij} be large and that $(\mu_i - \mu_j)^2$ be small. This means that the two samples vary in a similar fashion.

Some problems may be considered both single sample problem and multi-sample problem as we will see in chapter 3.

2.2 Nonparametric Monotone Trend Test and Slope Estimator

Let $x_1,\cdots, x_n$ be a sequence of observation ordered by time. We are interested in the null hypothesis

H_0: the observations are randomly ordered, i.e., $x_1,\cdots, x_n$ are i.i.d. samples,

and the alternative hypothesis

H_1: there is a monotone trend over time , i.e., $F_{\mathbf{x}_i}(x) \geq (\text{or} \leq)\ F_{\mathbf{x}_j}(x)$

for all $i < j$ with at least one strict inequality, where $F_{x_i}(x)$ is the cumulative density function of random variable x_i.

Let

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

Mann (1945) proposed the following test statistic:

$$S = \sum_{i<j} \operatorname{sgn}(x_j - x_i).$$

Under H_0, the test statistic has mean 0 and variance

$$\sigma^2 = [n(n-1)(2n+5)/18]^2$$

and S/σ is asymptotically N(0,1).

Kendall (1975) gives the mean and variance of S under H_0 given the possibility that there may be ties in the x values:

$$E(S) = 0\ ,$$

$$\operatorname{Var}(S) = \{\ n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)\ \}/18\ ,$$

where t is the extent of any given tie (number of x's involved in a given tie) and $\sum_t$ denotes the summation over all ties.

Both Mann and Kendall derive the exact distribution of S for $n \le 10$ and show that even for $n=10$ the normal approximation is excellent, provided one uses a continuity correction, i.e., computes the standard normal variate Z by

$$Z = \begin{cases} (S-1)(\mathrm{Var}(S))^{-1/2} & S > 0 \\ 0 & S = 0 \\ (S+1)(\mathrm{Var}(S))^{-1/2} & S < 0 \end{cases} .$$

Then in a two-sided test, a positive value of Z indicates an up trend and a negative value of Z indicates a down trend. This test is commonly called Mann-Kendall test. Bradley (1968, p228) notes that when this test is used as a test of randomness against normal regression alternatives, this test has an asymptotic relative efficiency of 0.98 relative to the parametric test based on the regression slope coefficient.

Kendall's (1970) τ test for correlation considers a more general case. Suppose $(x_1,y_1),\cdots,(x_n,y_n)$ are a sequence of bivariate observations, ordered by time. The hypothesis to be tested is:

$$H_0: \tau = 0$$

versus

$$H_1: \tau \neq 0,$$

where $\tau = 2P(X_j > X_i | Y_j > Y_i) - 1$. In their procedure , a point estimator of τ is given by

$$\hat{\tau} = \sum_{i<j} \mathrm{sgn}[(x_j - x_i)(y_j - y_i)].$$

For the Mann's test, $y_i = i,\ i = 1, \cdots, n$.

Sometimes the time series data of interest show that there exist seasonal patterns in the data and thus the hypotheses mentioned above may be too restrictive. However, different procedure are given for dealing with such cases.

Let $X=(X_1, \cdots, X_p)$, where $X_g=(x_{1g}, \cdots, x_{ng})^t$ is a subsample of season g and x_{ig} is the observation obtained in the ith year and the gth season. A procedure proposed by Dietz and Killen (1981) can be used in this case: The null hypothesis under consideration is that the p vectors are randomly ordered vs. the alternative hypothesis that there is monotone trend in one or mere of the p variables. Let

$$S = (S_1, \cdots, S_p)^t,$$

and

$$\Sigma = (\sigma_{gh}),$$

where

$$S_g = \sum_{i<j} \operatorname{sgn}(x_{jg} - x_{ig}), \quad g=1,\cdots,p,$$

$$\sigma_{gg} = n(n-1)(2n+5)/18, \quad g = 1, \cdots, p,$$

and

$$\sigma_{gh} = \frac{1}{3}\left\{ \sum_{i<j} \operatorname{sgn}[(x_{jg}-x_{ig})(x_{jh}-x_{ih})] + \sum_{(i,j,k)} \operatorname{sgn}[(x_{jg}-x_{ig})(x_{jh}-x_{kh})] \right\},$$

$g \neq h$, the estimated covariances of S_g and S_h. Then $S^t\Sigma^- S$ is asymptotically χ^2_q, where Σ^- is any generalized inverse of Σ and $q \leq p$ is the rank of Σ.

Farrell (1980), following Sen (1968), proposed another test procedure in which the data are "deseasonalised" first. Let $y_{ij} = x_{ij} - x_{\cdot j}$ where $x_{\cdot j} = \frac{1}{n}\sum_i x_{ij}$, R_{ij} is the rank of y_{ij} among all np y_{lm}'s, $l = 1, \cdots, n$, $m = 1, \cdots, p$, and H_0 represents the hypothesis of no trend. Then

$$T = 12p^2\left(n(n+1)\sum_{i,j}(R_{ij} - R_{\cdot j})^2\right)^{-1}\left(\sum_i[(i - (n+1)/2)(R_{i\cdot} - (np+1)/2)]\right)$$

is asymptotically N(0,1), where $R_{\cdot j} = \frac{1}{n}\sum_i R_{ij}$, $R_{i\cdot} = \frac{1}{p}\sum_j R_{ij}$.

Hirsh *et al.* (1982) defined a multivariate extension of the Mann-Kendall test called the Seasonal Kendall Test. Let H_0 and H_1 be the hypothesis given by

H_0: $(x_{1j}, \cdots, x_{nj})$ are independent and identical sample for $j = 1, \cdots, p$ and x_{ij}'s are independent.

and

H_1: one or more $(x_{1j}, \cdots, x_{nj})$'s are not independent and identical sample.

Let

$$S_g = \sum_{i<j} \operatorname{sgn}(x_{gj} - x_{gi}), \quad g = 1, \cdots, p$$

and

$$S' = \sum_{g=1}^{p} S_g .$$

Then E(S')=0, and

$$Var(S')=\sum_{j=1}^{p} Var(S_j) .$$

Let

$$Z' = \begin{cases} (S'-1)(Var(S'))^{-1/2} & S' > 0 \\ 0 & S' = 0 \\ (S'+1)(Var(S'))^{-1/2} & S' < 0 ; \end{cases}$$

then Z' is asymptotically N(0,1) under H_0. They demonstrate that the normal approximation is quite accurate even for sample sizes as small as n=2, p=12.

The Seasonal Kendall Test is similar to a test proposed by Jonckheere (1954) as a multivariate extension of the sign test for the case when the number of observations is greater than 2. In the case that n_js are equal, the seasonal test and Jonckheere's (1954) test are equivalent.

Hirsch and Slack (1984) modified their seasonal Kendall test using Dietz and Killen's estimator of the covariances of S_g and S_h. So the variance of S' becomes

$$Var(S')=\sum_{g=1}^{p} Var(S_g) + \sum_{g \neq h} cov(S_g,S_h) .$$

The approximation is good in this test when p = 12 and n > 10. Compared with the seasonal Kendall test, this test is less powerful but more robust against serial correlation.

As van Belle and Hughes (1984) suggest in their paper, Hirsch's (1982) and Farrell's (1980) procedures may be based on the same model, i.e.

$$x_{ij} = u + a_i + b_j + e_{ij}, \qquad i = 1, \cdots, n, \; j = 1, \cdots, p,$$

where

$$\sum_i a_i = \sum_j b_j = 0,$$

a_i is the yearly component, b_j is the seasonal component and e_{ij} are i.i.d. with $E(e_{ij}) = 0$. The common hypothesis being tested is

$$H_0: a_1 = \cdots = a_n = 0$$

versus

$$H_1: a_1 \leq \cdots \leq a_n \text{ or } a_1 \geq \cdots \geq a_n$$

with at least one strict inequality. These are true for Hirsch's (1984) modified procedure too.

van Belle and Hughes (1984) conclude that Farrell's (1980) procedure is more powerful when there are no missing data since ranking all together preserves the relative ranks between seasons which are lost in the Seasonal Kendall Test, while Hirsch's (1982) is easier to compute when there are missing data.

Only the Diets and Killen's test is valid when the trends in different seasons are heterogeneous but Hirsch (1982) argues that it is probably only applicable for sets of data including at least 40 years of monthly values.

All the other tests of seasonal data mentioned above will be misleading if the trends are not homogeneous among seasons especially when there are opposing trends in different seasons. So van Belle and Hughes (1984) develop a 2-way ANOVA-like nonparametric trend test which can test for the homogeneity of trend direction at different locations and different seasons. When there is only one location, it looks like a 1-way ANOVA and the statistic they proposed is

$$\chi^2_{homog} = \chi^2_{total} - \chi^2_{trend} = \sum_{g=1}^{p} Z_j^2 - \bar{Z}^2,$$

where

$$Z_g = S_g(Var(S_g))^{-1/2} \ ,$$

$$\bar{Z} = \sum_{g=1}^{p} Z_g /p,$$

and S_g is the Mann - Kendall trend statistic for the gth season.

If the trend for each season is in the same direction then χ^2_{homog} has a chi squared distribution with (p-1) degrees of freedom χ^2_{q-1}. If χ^2_{trend} exceeds the predefined critical value, then the null hypothesis of homogeneous seasonal trends is rejected in which case the Seasonal Kendall Test does not apply. However, if that hypothesis is accepted, then χ^2_{homog} is the statistic used to test the hypothesis that the common trend direction is significantly different from 0.

van Belle and Hughes (1984) point out that the validity of these χ^2 tests requires that the S_g be independent. A procedure for testing the contrasts is proposed and possible use of Newman-Keul's procedure to group the seasons are illustrated in their paper as well.

Sen (1968b) developed a nonparametric procedure to estimate the slope of a possible existing trend with a 100(1-α)% confidence interval and it is robust against gross data errors and outliers. Let

$$Q_{ij} = (x_j - x_i)/(j - i) .$$

Suppose there are N such (i, j) pairs, the median of which is Sen's estimator of the slope.

A simple way to get a 100(1 - α)% confidence interval is by using normal approximation. Suppose $Q_{(1)}, \cdots, Q_{(N)}$ are the order statistics of the Q_{ij}'s. Then $[Q_{(m1)}, Q_{(m2+1)}]$ is a 100(1 - α)% confident interval of the slope estimator, where

$$m1 = (N - Z_{1-\alpha/2}(Var(S))^{1/2})/2 ,$$

$$m2 = (N + Z_{1-\alpha/2}(Var(S))^{1/2})/2 ,$$

$Z_{1-\alpha/2}$ is the (1 - α/2) quantile of the standard normal distribution.

A seasonal Kendall slope estimator is given by Gibert (1987). Suppose there are N_k pairs of x_{ik}, x_{jk} such that i < j. Then there are N_k Q_{ijk}-values where $Q_{ijk} = (x_{jk} - x_{ik})/(j - i)$ for the kth season. Let $N = N_1 + \cdots + N_p$. In all, there are N slope estimates Q_{ijk} for all the seasons conbined. The median of these N Q_{ijk}s is the seasonal Kendall slope estimator and if $Q_{(1)}, ..., Q_{(N)}$ are the order statistics of Q_{ijk}s, then $[Q_{(m1)}, Q_{(m2+1)}]$ is a 100(1 - α)% confidence interval of the slope estimator where

$$m_1 = (N - Z_{1-\alpha/2}(Var(S'))^{1/2})/2 ,$$

$$m2 = (N + Z_{1-\alpha/2}(Var(S'))^{1/2})/2 ,$$

and $Z_{1-\alpha/2}$ is the (1 - α/2) quantile of the standard normal distribution.

2.3 Median Polish

Tukey (1977) proposed a procedure called median polish to fit a three-way model:

$$y_{ijk} = u + a_i + b_j + c_k + ab_{ij} + ac_{ik} + bc_{jk} + e_{ijk} \qquad (1)$$

where

u is the common effect,

a_i is th ith row effect,

b_j is the jth column effect,

c_k is the kth layer effect,

ab_{ij} is the interaction of a_i and b_j,

ac_{ik} is the interaction of a_i and c_k,

bc_{jk} is the interaction of b_j and c_k,

e_{ijk} is the random error of y_{ijk}.

Roughly speaking, this procedure uses medians of the data to estimate the main effects and the interactions in the same way as means are used to estimate the effects in the case of ANOVA. Note that when we use means for fitting, the main effects and the interactions can be found in one "iteration", further iteration leaving the result unchanged. When we use medians, the first iteration may not be adequate.

Tukey (1977) developed one-way and two-way median polish as well while three-way polish is the generalization of two-way procedure. Turkey pointed out that the generalization of this three-way polish to a more-way case is straightforward. Here we describe the three-way median polish procedure only.

Suppose y_{ijk} is the observation under the ith level of factor a, the jth level of factor b, and the kth level of factor c($i = 1, \cdots, I, j = 1, \cdots, J, k = 1, \cdots, K$). Let, in general, x_M denote the median of any given set of numbers, $x_1, \cdots, x_k$; these may have additional subscripts like $x_{ij1}, x_{ij2}, \cdots, x_{ijk}$ where now x_{ijM} amounts to computing the median over the last subscript, and let

$$r_{ijk} = y_{ijk,} \text{ for all } i, j, k.$$

We fit the model (1) as follows:

(a) let $u^{(0)} = a_i^{(0)} = b_j^{(0)} = c_k^{(0)} = ab_{ij}^{(0)} = ac_{ik}^{(0)} = bc_{jk}^{(0)} = 0,\ r_{ijk}^{(0)} = y_{ijk},$ for all i, j, k;

(b) for n = 0, 1, 2, ···, let

$$ab_{ij}^{(3n+1)} = ab_{ij}^{(3n+0)} + r_{ijM}^{(3n+0)},$$

$$a_i^{(3n+1)} = a_i^{(3n+0)} + ac_{iM}^{(3n+0)},$$

$$b_j^{(3n+1)} = b_j^{(3n+0)} + bc_{jM}^{(3n+0)},$$

$$u^{(3n+1)} = u^{(3n+0)} + c_M^{(3n+0)},$$

and

$$r_{ijk}^{(3n+1)} = r_{ijk}^{(3n+0)} - r_{ijM}^{(3n+0)},$$

$$ac_{ik}^{(3n+1)} = ac_{ik}^{(3n+0)} - ac_{iM}^{(3n+0)},$$

$$bc_{jk}^{(3n+1)} = bc_{jk}^{(3n+0)} - bc_{jM}^{(3n+0)},$$

$$c_k^{(3n+1)} = c_k^{(3n+0)} - c_M^{(3n+0)};$$

(c) let

$$ac_{ik}^{(3n+2)} = ac_{ik}^{(3n+1)} + r_{iMk}^{(3n+1)},$$

$$a_i^{(3n+2)} = a_i^{(3n+1)} + ab_{iM}^{(3n+1)},$$

$$c_k^{(3n+2)} = c_k^{(3n+1)} + bc_{Mk}^{(3n+1)},$$

$$u^{(3n+2)} = u^{(3n+1)} + b_M^{(3n+1)},$$

and

$$r_{ijk}^{(3n+2)} = r_{ijk}^{(3n+1)} - r_{iMk}^{(3n+1)},$$

$$ab_{ij}^{(3n+2)} = ab_{ij}^{(3n+1)} - ab_{iM}^{(3n+1)},$$

$$bc_{jk}^{(3n+2)} = bc_{jk}^{(3n+1)} - bc_{Mk}^{(3n+1)},$$

$$b_j^{(3n+2)} = b_j^{(3n+1)} - b_M^{(3n+1)};$$

(d) let

$$bc_{ij}^{(3n+3)} = bc_{ij}^{(3n+2)} + r_{Mjk}^{(3n+2)},$$

$$b_j^{(3n+3)} = b_j^{(3n+2)} + ab_{Mj}^{(3n+2)},$$

$$c_k^{(3n+3)} = c_k^{(3n+2)} + ac_{Mk}^{(3n+2)},$$

$$u^{(3n+3)} = u^{(3n+2)} + a_M^{(3n+2)},$$

and

$$r_{ijk}^{(3n+3)} = r_{ijk}^{(3n+2)} - r_{Mjk}^{(3n+2)},$$

$$ab_{ij}^{(3n+3)} = ab_{ij}^{(3n+2)} - ab_{Mj}^{(3n+2)},$$

$$ac_{ik}^{(3n+3)} = ac_{ik}^{(3n+2)} - ac_{Mk}^{(3n+2)},$$

$$a_i^{(3n+3)} = a_i^{(3n+2)} - a_M^{(3n+2)}.$$

(e) Repeat steps (b) to (d) until the sth stage, where s is the smallest number such that either

$$r_{ijM}^{(s)} = ac_{iM}^{(s)} = bc_{jM}^{(s)} = c_M^{(s)} = 0$$

or

$$r_{iMk}^{(s)} = ab_{iM}^{(s)} = bc_{Mk}^{(s)} = b_M^{(s)} = 0$$

or

$$r_{Mjk}^{(s)} = ab_{Mj}^{(s)} = ac_{Mk}^{(s)} = a_M^{(s)} = 0$$

Tukey (1977) suggested that two cycles of (b) - (d) would be enough -- whether or not we are at the end of the process. He also suggested using the box plot to show the relative magnitude of the main effects, interactions and the residuals. Such a plot can display the significant main effects and interactions.

Comparing median polish and mean polish (using means to estimate the effects), median polish minimize the sum of absolute values of the residuals while mean polish minimize the sum of squares of the residuals.

Median polish is more robust than mean polish against outliers, namely, the median tends to leave outliers as spikes while the mean tends to reduce them and spread the deviations around. Whether or not the outliers are deleted from the data has a lesser effect on the result of median polish than on that of mean polish.

2.4 Kriging and Universal Kriging

Kriging is the name given to best linear unbiased estimation of a stochastic process by generalized least squares and its name comes from D. R. Krige, a mining engineer in South Africa. It is used to fit a surface by regression techniques.

Definition: A stochastic process Z(x) is said to be strictly stationary if the joint probability distribution function of n arbitrary points is invariant under translation (Delhomme, 1976). i.e.,

$$f(Z(x_1), \cdots, Z(x_n)) = f(Z(x_1+h), \cdots, Z(x_n+h))$$

where h is arbitrary.

Definition: A stochastic process Z(x) is said to be weakly stationary if its first two moments are invariant under translation, namely,

$$E(Z(x)) = \text{constant} \qquad\qquad \text{for all } x$$

and

$$\mathrm{Cov}(Z(x)), Z(y)) = C(x - y),$$

i.e., it depends only on x - y.

Definition: The covariance of Z(x) is said to be isotropic if

$$\mathrm{Cov}(Z(x), Z(y)) = C(|x - y|),$$

i.e., it depends only on the distance between x and y.

Definition: A stochastic process Z(x) is said to be intrinsic if the first two moments of Z(x+h) - Z(x) depend only on h.

Definition: $r(h) = \frac{1}{2}\mathrm{Var}(Z(x+h) - Z(x))$ is called semivariogram. 2r(h) is called variogram.

Observe that r(h) is essentially the negative of the covariance. If

$$E(Z(x)) = \text{constant},$$

then

$$r(h) = \frac{1}{2} E[Z(x+h) - Z(x)]^2$$

and it may be estimated by (Delfiner, 1975)

$$r(h) = \frac{1}{N_h} \sum_{i=1}^{N_h} [Z(x_i+h) - Z(x_i)]^2$$

where N_h = number of pairs of observation separated by h and $Z(x_i+h)$, $Z(x_i)$ are observations.

There are several commonly used variogram models when Z is istropic. One of those is called generalized covariance model which is used when $E(Z(x)) \neq$ constant.

Definition: $m(x) = E(Z(x))$ is called the drift of $Z(x)$.

Let

$$e(x) = Z(x) - m(x);$$

then

$$Z(x) = m(x) + e(x), \tag{1}$$

where $E(e(x)) = 0$. In (1), e(x) represents the random fluctuation and m(x) represents the slowly varying, continuous features of Z(x). Therefore, m(x) may be approximated within a restricted neighborhood by

$$m(x) = \sum_{i=1}^{L} a_p f_p(x) \tag{2}$$

where a_p are constant coefficients and f_p are arbitrary functions of the point x. In particular, if the $\{f_i(x)\}$ are spatial polynomials, Kriging is called "universal Kriging".

From (1), it can be seen that to estimate m(x) from the data we need to know the covariance of e(x), and to estimate e(x) we need to know m(x). But usually in practice, neither is known and both must be estimated from the data.

Matheron (1973) and Delfiner (1975) have developed a technique to estimate m(x) and the covariance of e(x) simultaneously; the key concept is the generalized increment.

Definition: A generalized increment of order k is a linear combination of the sample value $\sum_i ß_i Z(x_i)$ for which $\sum_i ß_i f(x_i) = 0$, where $f(x_i)$ is a polynomial of order less than or equal to k.

In the plane $(x_i = (x_{1i}, x_{2i}))$, this condition yields:

k = 0, $\qquad \sum_i ß_i = 0,$

k = 1, $\qquad \sum_i ß_i = \sum_i ß_i x_{1i} = \sum_i ß_i x_{2i} = 0,$

k = 2, $\qquad \sum_i ß_i = \sum_i ß_i x_{1i} = \sum_i ß_i x_{2i} = \sum_i ß_i x_{1i}^2 = \sum_i ß_i x_{2i}^2 = \sum_i ß_i x_{1i} x_{2i} = 0.$

An increment of order k will filter out a polynomial trend of degree k. If $\hat{Z} = \sum_i \lambda_i Z(x_i)$ is a Kriging estimate of Z(x), then $Z(x) - \hat{Z}$, the Kriging error, is a generalized increment since it can be written as $-Z(x) + \sum_i \lambda_i Z(x_i)$ and we see that $ß_0 = -1$ and $ß_i = \lambda_i$ for all i.

Definition: If there exists a function K(h) such that for any kth order generalized increment $\sum_i ß_i Z(x_i)$,

$$\mathrm{Var}\left(\sum_i ß_i Z(x_i) \right) = \sum_{i,j} ß_i ß_j K(x_i - x_j) \tag{3}$$

then K(h) is called a generalized covariance function, and Z(x) is known as an intrinsic process of order k.

Matheron (1973) shows that any isotropic($h = |h|$) generalized covariance function defines a class of functions which are all equivalent up to an addition of an arbitrary even-powered polynomial of degree less or equal to 2k. For example, for k=1, $K(h) = -h$ and $K(h) = -h + h^2$ are equivalent. For this reason, only the essential odd powers are used when writing a generalized covariance. For orders up to 2, the isotropic polynomial generalized covariance kernels are (Jernigan,1986):

$$K(h) = \begin{cases} C\delta + a_0 h, & k = 0 \\ C\delta + a_0 h + a_1 h^3, & k = 1 \\ C\delta + a_0 h + a_1 h^3 + a_2 h^5, & k = 2 \end{cases} \tag{4}$$

where $a_0, a_1, a_2 \leq 0$, and in R^2, $a_1 \geq -10\sqrt{a_0 a_2 /3}$, in R^3, $a_1 \geq -\sqrt{10 a_0 a_2}$,

$$\delta = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}.$$

The constraints insure that (3) is nonnegative. Polynomial generalized covariances with coefficients satisfying these constraints are said to be admissible.

Our objective is to estimate Z(x) by $\hat{Z} = \sum_i \lambda_i Z(x_i)$ such that

$$\begin{cases} E(\hat{Z}(x) - Z(x)) = 0 \\ Var(\hat{Z}(x) - Z(x)) = E(\hat{Z}(x) - Z(x))^2 \text{ is minimized.} \end{cases} \tag{5}$$

Consider the case of planar region($x = (x_1, x_2)$). Using the previous assumption of a polynomial drift of order less than or equal to k,

$$m(x) = \sum_p a_p f_p(x),$$

where $f_p(x)$ are given by $\{1\}$ for $k = 0$, $\{1, x_1, x_2\}$ for $k = 1$, $\{1, x_1, x_2, x_1^2, x_2^2, x_1x_2\}$ for $k = 2$, it has been shown that the λ's satisfy (5) can be found by solving the following system of equations, known as the universal Kriging system:

$$\begin{cases} \sum_j \lambda_i K(x_i - x_j) + \sum_p \mu_p f_p(x_i) = K(x_i - x) & \text{for all } i \\ \sum_j \lambda_i f_p(x_j) = f_p(x) & \text{for all } p. \end{cases} \tag{6}$$

If we know (a) the order of the drift k, and (b) the coefficients of the generalized covariance function, we can solve (6) to estimate (Kriging) Z(x). (6) has a unique solution for λ_i, provided that the drift functions f_p are algebraically linearly independent, i.e.,

$$\sum_p a_p f_p(x) = 0 \quad \text{for all } x \text{ if and only if } a_p = 0 \text{ for all } p.$$

(a) Drift order identification

Here is a procedure to determine the order of drift k (Devary and Rice, 1982): for different values of k, assuming Z(x) is an intrinsic function of order k, delete each observation $Z(x_r)$ in turn and calculate $Z^{(k)}(x_r)$, the estimate of $Z(x_r)$, from the remaining observations using the generalized covariance function

$$K(h) = - h.$$

This particular generalized covariance function is valid for any value of k. Typically k is chosen to be 0, 1, or 2.

The best choice of k depends on the residuals $\{Z(x_i) - Z^{(k)}(x_i), \text{all } i\}$, k = 0, 1, 2. Three criteria are set up to compare the residuals:

1. rank $\{|Z(x_i) - Z^{(k)}(x_i)|, \text{all } i, k = 0, 1, 2\}$. The value of k with the smallest average rank is the preferred order of drift value;

2. the value of k with the smallest MSE is preferred, where

$$MSE_k = \sum_i (Z(x_i) - Z^{(k)}(x_i)), \qquad k = 0, 1, 2;$$

3. the value of k with $|Z_k| \leq 1.96$ is preferred, where

$$Z_k = \bar{e}_k / S(\bar{e}_k),$$

$$\bar{e} = \sum_i (Z(x_i) - Z^{(k)}(x_i)),$$

$$S(\bar{e}) = [\sum_i (Z(x_i) - Z^{(k)}(x_i) - \bar{e})^2/(n-1)]^{\frac{1}{2}},$$

since under the hypothesis that the true value of k is less than or equal to k, Z_k is asymptotically N(0, 1).

So, typically, a k with minimum MSE and/or average rank is selected.

(b) Selection of the generalized covariance function.

For a chosen k, the model for a polynomial generalized covariance function of order k is

$$K(h) = C\delta + \sum_{p=0}^{k} a_p |h|^{2p+1}, \quad \text{where} \quad \delta = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}.$$

Delete the rth observation from data, then estimate $Z(x_r)$ with the initial estimates of the coefficients $\{C = 1, a_0 = -1, a_1 = 1, a_2 = -1\}$. For example, if k = 1, the initial estimates of K(h) is

$$K(h) = 1 - h + h^3.$$

Suppose

$$Z_r(\beta) = \sum_j \beta_j^{(r)} Z(x_j)$$

is the error of the Kriging estimate of $Z(x_r)$; by equation(3),

$$\mathrm{Var}(Z_r(\beta)) = E(Z_r(\beta))^2 = \sum_{i,j} \beta_i^{(r)} \beta_j^{(r)} K(x_i - x_j) \tag{6}$$

or

$$E(Z_r(\beta))^2 = C \sum_j (\beta_j^{(r)})^2 + \sum_{p=1}^{k} a_p \sum_{i,j} \beta_i^{(r)} \beta_j^{(r)} |x_i - x_j|^{2p+1} \tag{7}$$

Let

$$T_0^r = \sum_j (\beta_j^{(r)})^2,$$

$$T_p^r = \sum_{i,j} \beta_i^{(r)} \beta_j^{(r)} |x_i - x_j|^{2p+1}, \qquad p = 1, \cdots, k.$$

Equation (7) can be written as

$$E(Z_r(\beta))^2 = CT_0^r + \sum_{p=1}^{k} a_p T_p^r \tag{8}$$

which is linear in the coefficients. To determine the coefficients, Devary and Rice (1982) suggest regressing $E(Z_r(\beta))^2$ upon $T_0^r, \cdots, T_k^r$ with the corresponding constraints on the coefficients shown in (4). The set of coefficients obtained by this regression is used to reestimate the observations, yielding a new set of errors and T^r values. The regression, (8), is performed again and again until the parameters converge. The proposed iteration is to be performed on all of the generalized covariance function of order k.

Once several sets of admissible parameters have been estimated it remains to choose the best one. Since $Z_r(ß)$ is a Kriging error and $E(Z_r(ß))^2$ is a estimate of the Kriging variance, if the estimate of K(h) is correct, $E(Z_r(ß))^2$ should be close to $Var(Z_r(ß))$ defined by (6). Therefore, the quantity

$$\rho = \sum_r (Z_r(ß))^2 / \sum_r \sigma_r^2$$

should be close to 1, where $\sigma_r^2 = \sum_{i,j} ß_i^{(r)} ß_j^{(r)} K(x_i - x_j)$. In practice, a jackknife estimator of ρ is recommended by (Delfiner,1975):

$$\rho = 2R - (n_1 r_1 + n_2 r_2) / (n_1 + n_2),$$

where

$$R = \sum_r Z_r(ß)^2 / \sum_r \sigma_r^2,$$

$$r_1 = \sum_{r \in I_1} Z_r(ß)^2 / \sum_{r \in I_1} \sigma_r^2,$$

$$r_2 = \sum_{r \in I_2} Z_r(ß)^2 / \sum_{r \in I_2} \sigma_r^2,$$

n_1 = number of r's such that $r \in I_{1,}$

n_2 = number of r's such that $r \in I_2$.

It is sufficient to let $I_1 = \{1, \cdots, n/2\}$, $I_2 = \{n/2 + 1, \cdots, n\}$ where n is the total number of observations. The generalized covariance is chosen as the one with the minimum MSE and jacknife statistic $\rho \in (0.75, 1.25)$. For $n \leq 50$, the minimal MSE criterion should be used.

There have been some objections and reluctance to use this technique because the resulting covariance function is difficult to interpret. Journel and Huijbregts (1978), Devary and Doctor (1981), as well as Neuman and Jaeobson (1984) have suggested different approaches to estimate the covariance function.

Chapter 3 APPLICATIONS AND OVERVIEW

Identical statistical analyses are applied to the "historical" data collected from 1980 to 1986 and the "recent" data collected from 1983 to 1986 for three ions: SO_4, NO_3 and H^+. In general, "recent data" is just a part of the "historical data" except that more than doubled number of sites are involved in the former.

3.1 Transformation and Clustering

A preliminary exploration of the distribution of the data provides important information which enables us to analyze the data efficiently and accurately. And transforming the data so that they are approximately normally distributed also makes the analysis simpler. Three commonly used transformations, $y=x^{1/2}$, $y=x^{1/4}$ and $y=\log(x)$ are examined separately. It seems that in most of the cases, the log transformation is the best, in the sense that the resulting empirical distribution is approximately symmetric. However the empirical distribution does have relatively long tails compared with the normal. This is similar to previous conclusions about transforming environmental data (c.f. Gilbert, 1987, p152).

The hierarchical clustering analysis with complete linkage described in Section 2.1 can be applied to either the untransformed or transformed data to cluster the sites under consideration. The difference is that if we cluster the log transformed data we can expect more clusters among the sites with smaller measurements than those with larger measurements. This is because log is a concave function. Since, in general, we are more interested in the sites with larger measurements, and we want to study them more carefully, we decided to cluster the untransformed data. Furthermore, since the precipitation volume weighted mean gives the quantity of chemicals in wet deposition rather directly, we apply this method to the volume weighted mean for all the components under study.

For the three ions under study, the results of clustering analysis show that all the sites can be grouped into about three clusters (though the NO_3 data from 1980 to 1986 is actually grouped into two clusters), namely, the subregion of the United States with the highest concentration of industry, a roughly concentric area around that region, and the rest of the United States. This result agrees with common sense; industries which emit more SO_2 , NO_x and thus have more concentrated emission patterns should be more or less contiguous.

After clustering the data, the histogram of the log transformed data is drawn for each cluster. It turns out that in most of the cases, the data within clusters have histograms are approximately symmetric but with relatively long tails. Sometimes these tails are heavy and sometimes not. This suggusts some uncertainty about whether or not the data can be treated as normal. However, our analysis does not rely on normality so this issue is not of great concern.

The remainder of our analyses are based on log transformed data, for both monthly precipitation volume weighted means and the monthly medians though only the histograms of the former are examined. We refer the log transformed data simply as "data" in the remaining parts of this chapter.

3.2 Trend and Seasonality

Since the normality of the data is doubted, as many cases when people are dealing with precipitation chemistry data, using the parametric methods such as analysis of variance to estimate trend and seasonality may not be appropriate. But this does not cause any difficulties in using polish method. In addition, the median is resistant to outliers, which appear in our case quite often. Therefore we decided to use median polish to extract out the trend and the seasonality of the data.

The three-way median polish method described in Section 2.2 is applied to both the "historical" and "recent" data, as well as to both the monthly volume weighted mean and monthly median for all three ions. For each ion, either "historical data" or "recent data", and either monthly volume weighted mean or monthly median, model (2.3.1) can be rewritten as

$$C = u + M_i + Y_j + S_k + MY_{ij} + MS_{ik} + YS_{jk} + e_{ijk}, \qquad (3.2.1)$$

where u is the common effect, M_i is the ith monthly effect, i = 1, ⋯, 12 standing for Jan., Feb., ⋯, Dec. respectively, Y_j is the jth yearly effect, where for the "historical" data set, j = 1, ⋯, 7 stands for 1980,1981, ⋯, 1986 respectively, and for the "recent" data, j = 1, ⋯, 4 stands for 1983, 1984, ⋯, 1986, respectively, S_k is the kth site effect, k = 1, ⋯, K, where with "historical" data, K = 86 for H^+ and K = 81 for the other ions. With "recent" data, K = 32 for H^+ and K = 31 for the other ions, and e_{ijk} is the residual.

The estimated main effects, namely, monthly, yearly and site effects are plotted. Residuals are displayed using boxplots where each box corresponds to a site. A summary of the three-way polish is given using boxplots where each box represents a source of main effects, interactions or residuals. This summary enables us to see how big the main effects and interactions are compared with the residuals.

3.3 Nonparametric Test, Slope Estimate and Kriging

Some characteristics of precipitation chemistry data, such as nonnormality, the existence of missing data and the limited number of observations along the time, create some difficulties in using the traditional parametric statistical methods to test the trend in such cases. But these cause no difficulties in using the nonparametric trend tests described in Section 2.2. This is why nonparametric trend test is used in this study.

As mentioned in Section 2.2, for data with seasonal patterns, the test procedure which deseasonalizes the data first and then ranks them altogether can preserve the relative ranks between seasons; but these are lost if the data are ranked within each season. Consequently, Farrell's (1980) test is more powerful. Considering that the length of the "recent" data record is merely 4 years, to test the trend over time we prefer Farrell's (1980) test. However, since there are some missing values in the data set which would make the computations very complicated, we decided instead to deseasonalize the data first and then use the Mann-Kendall Test to test whether or not there is a monotone trend over time in the deseasonalized data. For each individual site, the data are deseasonalized using the one-way median polish by subtracting the median of the values of each month from the data for that month. Then the Mann-Kendall Test is applied to the deseasonalized data to obtain the test statistics and the corresponding p-values. Sen's nonparametric slope estimates are obtained with 80% confidence intervals. Since the alternative hypothesis of the test is that the trend is monotone, for some of these tests our failure to reject the null hypothesis may result from the V-shaped trend in the data. This is suggested in particular by the plots of the yearly effects of some data obtained from the three-way median polish (for example, the monthly volume weighted mean of SO_4 for 1980 - 1986). Similarly, a small slope estimate could be derived from the V-shaped trend.

Nonparametric slope estimates are plotted on the map of the United States. If the lower 80% confidence bound of the slope estimate at a point is greater than 0 then a "+" is plotted at that point. If the upper 80% confidence bound of the slope estimate at a point is less than 0 then a "-" is plotted at that point. Otherwise a "0" is plotted on the point.

If the p-value of the Mann-Kendall Test is less than 0.2 at a point, then an "x" is plotted on the map of the United States. The resulting plots are quite similar to the plots produced by the slope estimates of the trend if we convert "+" and "-" in the latter into an "x". This is not surprising since we expect that the 80% confidence interval of the slope estimate and the p-value of the Mann-Kendall Test at the $\alpha = 0.2$ level to give us about the same information. The plots of p-values for the Mann-Kendall Test are not included in the report since the plots produced by the slope estimates of the trend contain more detailed information.

Based on the slope estimates of the trend, universal Kriging estimates of the trend with estimated standard errors are obtained at each integer degree grid point of longitude and latitude across the 48 continental states in the United States. The estimates and the estimated standard errors are calculated for the nearest 8 neighbors of the point being estimated for the "historical" data and for the nearest 10 neighbors of the point being estimated for the "recent" data. Note that usually the estimates and the standard errors estimated by Kriging are based on actual observations rather than some kind of estimates treated as data. So we have to interpret them with caution since the estimates and the estimated standard errors are different from the real Kriging estimates and the standard errors.

The results of Kriging are plotted on the map of the United States. If the upper one standard error bound is less than 0 at some point then a "-" is plotted at the point. If the lower one standard error bound is greater than 0 at some point then a "+" is plotted at the point. Otherwise nothing is plotted.

Chapter 4 ANALYSES AND CONCLUSIONS FOR SULPHATE

Monthly precipitation volume weighted mean and monthly median are are referred to as "weighted mean" and "median", respectively, below.

4.1 Results for the Historical Data

"Historical data" refers to the data collected from January, 1980 to December, 1986. For SO_4, there are 31 sites with 5 or fewer missing observations during the period. The locations of these 31 sites are plotted in Figure 4.1.1. From this figure we can see that more sites are located in the East than the West. The results of this section are based on the data obtained from these 31 sites.

4.1.1 Clustering and Transformation

Before clustering the data, one outlier was deleted from the data set, namely the observation at the site with ID=040a (see Olsen and Slavich, 1986), observed in September, 1986. Its value is 71.71, much larger than 3.2, the average value for that site. Figure 4.1.2 (a) shows the hierarchical cluster structure of the weighted means. The sites are partitioned quite naturally into three clusters. The locations of the sites labelled by cluster are shown in Figure 4.1.3. The pattern is obvious: cluster 2 is located in the subregion of highest industrial concentration; cluster 1 is located around that region; and cluster 3 is spread over the rest of the United States. This indicates that the industrial areas emit more SO_2 than other areas and the pattern of emission is consistent with intuition. A cluster analysis was done without deleting outliers. The clusters are the same however, except for site 040a which is ruled out of any cluster (Figure 4.1.2 (b)).

The histograms of the original data and the transformed data under the three kinds of transformations described in Section 3.1, are shown in Figure 4.1.4. It appears that the histogram of the log transformed data fits the normal curve best. The other three are right skewed. So the log transformation is adopted in this case and the remainder of our analysis of the sulphate data is based on the transformed data. The histograms of the log transformed data for each cluster are depicted in Figure 4.1.5. The extremely long right tail in the histogram of cluster 2 results from the outlier at site 040a. In general, these three histograms are not unduly skewed but do have relatively long tails.

4.1.2 Trend, Seasonality and Spatial Patterns

The analysis described below is done for both log transformed monthly volume weighted mean and log transformed monthly median SO_4 data. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster, in the manner described in Section 3.2. Figures 4.1.6 and 4.1.7 show the plots of the yearly effects of the three clusters for weighted mean and median, respectively. They show similar patterns although the one for weighted means is more obvious. All of the three clusters show down-trends before 1983 and cluster 3 shows a down-trend after 1983 as well. But clusters 1 and 2 show up-trends after 1983. This indicates that the concentration of SO_4 in wet deposition increased after 1983 for the most industrialized subregion of the United States and that the down-trends for the rest of the United States have not changed much up to 1986. The yearly effects of the medians are more pronounced than those of the means in 1980 and gradually became smaller than the former in 1986. This indicates that in 1980, more than half of the sites had yearly effects larger than the average of their corresponding cluster, while in 1986, the situation was reversed: less than half the sites had the yearly effects larger than the average of the corresponding cluster.

Figure 4.1.8 and Figure 4.1.9 show the monthly effects for the weighted means and medians, respectively. They have almost the same pattern . The patterns for all three clusters are similar as well. The effects are high in summer and low in winter; spring and autumn fall in between and the transitions are quite smooth. The fact that the seasonal patterns are very similar for all three clusters suggests a reassuring spatial stability in this seasonal pattern. But the complexity of the acid deposition process makes it unreasonable to infer that the emission of SO_2 has the same pattern.

Figures 4.1.10 to 4.1.12 display the site effects of $\log(SO_4)$ by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. The site effect of Station 039a for weighted means in Figure 4.1.10 are quite different from that of the medians. This may be caused by the fact that concentration tends to be smaller when the volume of the precipitation is greater. For example, if $c_1 = c_2 = 10$ and $c_3 = 1$ are the concentrations, while $v_1 = v_2 = 1$, and $v_3 = 10$ are the volumes, the median of c1, c2 and c3 is 10 but the volume weighted mean is about 1.36.

Figures 4.1.13 to 4.1.15 are the boxplots of the residuals of the three-way median polish of the $\log(SO_4)$ data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order and comparable. The differences between the boxplots for the means and for the medians are even smaller. In Figure 4.1.13 ((a) and (b)), we do observe a few extreme outliers from Station 039a, and from Station 049a and Station 168a as well. For these sites, further detailed study is needed. In Figure 4.1.14 ((a), (b)), an extreme outlier is observed in the box for Station 040a. This is the one which was deleted when we did the clustering. It was not deleted here because it cannot affect the result of the median polish by its extremely large value. In Figure 4.1.15 ((a), (b)), the boxes for Stations 059a and 074a have extreme outliers and they need a more detailed study as well. But no analyses for specific site are included in this paper.

Figures 4.1.16 to 4.1.18 summarize the median polish for the three clusters, respectively where (a) is for the weighted meana and (b) is for the medians. In each figure, boxes represent the main effects, interaction and residuals respectively. We regard interactions as non-negligible if the sizes of the boxes for the interactions are comparable with those for main effects. In Figures 4.1.16 (a) and (b), the boxes for monthly effects are relatively large indicating that the variation caused by seasonality is larger than the other effects. The boxes for the interactions of yearly effects and site effects are relatively small indicating that the trends for all the site in the cluster 1 are similar. This convinces us that the pattern of trends is relatively stable. The boxes for the other interactions are similar in size to those of the main effects, which suggests that they have to be taken into account. For cluster 2, the same thing happens again, namely the monthly effects are relatively large and the interactions between yearly effects and site effects are relatively small (see Figures 4.1.17 (a) and (b)). Figures 4.1.18 (a) and (b) show the summaries for cluster 3. We observe relatively large monthly effects and relatively small year and site interactions as before. The interpretation for cluster 1 applies here as well.

4.1.3 The Results of Trend Testing, Slope Estimation and Kriging

Mann-Kendall Test and Sen's nonparametric slope estimation procedure are applied to the deseasonalized data for each site in the manner described in Section 3.3. A Mann-Kendall Test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval are obtained for each site for both the weighted means and medians. These are shown in Tables 4.1.1 and 4.1.2 respectively. In Table 4.1.1, 24 out of the 31 sites have significant trends at the p=0.2 level. In Table 4.1.2, 22 out of the 31 sites have significant trends at the p=0.2 level.

The results of the nonparametric slope estimation are plotted in Figures 4.1.19 (a) and (b) with symbols defined in Section 3.3, for the weighted means and medians respectively. If the lower 80% confidence bound of the slope estimate is greater at any site than 0 then a "+" is plotted; if the upper 80% confidence bound of the slope estimate is less than 0 a "-" is plotted. Otherwise a "0" is plotted . "+", "-" and "0" represent respectively, an up-trend, a down-trend and no-trend. The patterns of these two figures are very similar. About 2/3 of the sites show down-trends and they are mainly located in the Eastern United States. The other 1/3 of the sites show no-trend and no site had an up trend. During the period of 1980 - 1986 most of the sites have down trends. This conclusion agrees with some previous findings that the wet deposition concentration of SO_4 had decreased from late 1970's to early 1980's (Schertz and Hirsh,1985, Dana and Easter, 1987, Seilkop and Finkelstein, 1987). However, as we mentioned before, this could be a misleading conclusion in the presences of a V-shaped trend. At the very least, this result says that the concentrations of SO_4 for the latter years of the sampling period are smaller than that of early years of the same period.

The results of Kriging the slope estimates are displayed in Figures 4.1.20(a) and (b) As before, Figure 4.1.20(a) is for the weighted means and (b) is for medians. A "+" means an up-trend is estimated at the corresponding point and a "-" means a down-trend, estimated at the point. Again, these two plots are similar. Both Figures 4.1.20 (a) and (b) show that there is a down-trend for the most of the United States except that a few small areas in the Eastern United States show no trend. Considering that these results were obtained from the same set of data which produce Figures 4.1.19 (a) and (b), this is quite natural. The difference between these two figures are probably caused by (a) the differences in magnitudes of the slope estimates of weighted means and of medians, or (b) the fact that the estimated order of the variograms is one for the weighted means and zero for the medians. Note that these two figures can only be used as a guideline for the spatial distribution of the trends and should be read in conjunction with Figure 4.1.19. At any particular point where the observations were not made, interpolation may not be reliable. This is because (a) estimation is based on the estimated slopes, not on the observations directly and (b) the spatial resolution of the measurements is low and therefore, the correlations between sites are small.

4.2 Results for the Recent Data

"Recent data" refers to the data collected from Jan., 1983 to Dec., 1986. For SO_4, there are 81 sites with 5 or fewer missing observations during the period. The location of these 81 sites are plotted in Figure 4.2.1. From this figure we can see that more sites are located in the Northeast than the rest of the areas. The results of this section are based on the data observed at these 81 sites.

4.2.1 Clustering and Transformation

Before clustering the weighted mean data, two outliers were deleted from the data set, namely the observations for the Stations 035a and 040a (see Olsen and Slavich, 1986). The first one is observed on January, 1985 with value of 365.56, much larger than 6.39 which is the second largest observation at the same site; the second one is observed on September, 1986, the same one mentioned in Section 4.1.1. Figure 4.2.2 (a) shows the hierarchical cluster structure of the weighted means. The sites partition naturally into three clusters except for Stations 070a and 071a. These two sites are assigned to cluster 3, the closest one according Figure 4.2.2 (a), for technical convenience. The locations of the sites labelled by cluster are shown in Figure 4.2.3. It gives a pattern similar to the one in Figure 4.1.3: cluster 1 is located in the subregion of highest industrial concentration; cluster 2 is located around that region; and cluster 3 is located throughout the rest of the United States, corresponding to cluster 2, 1 and 3 in Figure 4.3.1 respectively. This gives us the same indications we mentioned in Section 4.1.1, namely, that the industrial areas emit more SO_2 than other areas and the pattern of emission is consistent. A cluster analysis was done without deleting outliers. The clusters are the same except for Stations 035a and 040a which are ruled out of any cluster (Figure 4.2.2 (b)).

The histograms of the original data and the transformed data under the three kinds of transformation described in Section 3.1, are shown in Figure 4.2.4. It appears that both histograms of the log and 1/4 power transformed data fit the normal curve. The other two are right skewed. We adopt the log transformation in this case since it is the best transformation in a lot of other cases, particularly in the case of "historical data". The remaining analyses of this section are based on the log transformed data. The histograms of the log transformed data for each cluster are depicted in Figure 4.2.5. Again the extremely long right tail in the histogram of cluster 1 results from the outlier at site 040a (the outlier at site 035a is not included in the plots). The histograms for the other two clusters have long left tails.

4.2.2 Trend, Seasonality and Spatial Patterns

The analysis described below is applied to both log transformed monthly volume weighted means and log transformed monthly medians. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster. Figures 4.2.6 and 4.2.7 show the plots of the yearly effects of the three clusters for weighted mean and median respectively, which show similar patterns. Clusters 1 and 2 show up-trends while the trend of cluster 3 is probably going down. This conclusion, obtained from 81 stations instead of 31, is consistent with what we got in Section 4.1.2.

That is that the concentration of SO_4 in wet deposition increased for the most industrialized subregion of the United States during the period of 1983 to 1986 though the precise change point in 1983 is not clear yet. In contrast to the situation in Figure 4.1.6 and 4.1.8, it is hard to say the yearly effects of the weighted means or those of the medians is larger.

Figures 4.2.8 and 4.2.9 show the monthly effects for the weighted means and medians, respectively. Although more than half of the data used in the present analyses were not included in the historical data, the patterns are almost the same as those for the latter displayed in Figures 4.1.8 and 4.1.9: high in summer, low in winter with spring and autumn in between. This convinces us again of spatial stability in the seasonal pattern.

Figures 4.2.10 to 4.2.12 show the site effects of $\log(SO_4)$ by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. For cluster 1, the stations located in the center of the industrial region tend to have larger effects. No unusual results are found in these figures.

Figures 4.2.13 to 4.2.15 are the boxplots of the residuals of the three-way median polish of the $\log(SO_4)$ data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order and comparable, and comparable to those in Figures 4.1.13 to 4.1.15 as well. The differences between the boxplots for the means and for the medians are even smaller. Extreme outliers are found in Stations 040a, 168a, 249a and 350a in Figures 4.2.13 (a) or (b), 039a and 049a in Figures 4.2.14 (a) or (b) and 035a, 059a, 074a, 255a, 271a and 354a in Figures 4.2.15 (a) or (b). Among these stations, 039a, 049a, 168a, 040a, 059a and 074a have extreme outliers observed in Figures 4.1.13 to 4.1.15 and those outliers may be the same ones. All these sites need further study.

Figures 4.2.16 to 4.2.18 summarize the median polish for the three clusters respectively where (a) is for weighted means and (b) is for medians. In each figure, boxes represent the main effects, interaction and residuals. The main difference between these figures and Figures 4.1.16 to 4.1.18 is that the variation of the yearly effects for 1983 to 1986 is much smaller than those for 1980 to 1986. In Figures 4.2.16 (a) and (b), the monthly effects are larger than the other effects as in Figures 4.1.16 ((a) and (b)) while the yearly effects and the interactions of yearly effects and site effects are relatively small. Cluster 2 has the same characteristics as cluster 1 (see Figures 4.2.17 (a) and (b)). Figures 4.2.18 (a) and (b) show the summaries for cluster 3. That the station effects are relatively large is perhaps due to assigning Stations 070a and 071a to this cluster. In general, the trends for 1983 to 1986 are small in magnitude and not very consistent for all sites.

4.2.3 The Results of Trend Testing, Slope Estimation and Kriging

The Mann-Kendall Test and Sen's slope estimation procedure which were applied to "historical data" are used for "recent data" as well. A Mann-Kendall test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval were obtained for each site for both the weighted means and medians. These are shown in Tables 4.2.1 and 4.2.2 respectively. In Table 4.2.1, 40 out of the 81 sites have significant trends at the p=0.2 level. In Table 4.2.2, 31 out of the 81 sites have significant trends at the p=0.2 level.

Figures 4.2.19 (a) and (b) are obtained by the same method used to produce Figures 4.1.19, for weighted means and medians respectively. Again, "+", "-" and "0" represent an up-trend, a down-trend and no-trend, respectively. In Figure 4.2.19 (a), about 1/4 of the sites show up-trends and are mainly located in the Eastern United States. Approximately 1/4 of the sites show down-trend and are located mainly in the west and middle of the U. S. The rest of the sites show no-trend and are located throughout the United States. The pattern of Figure 4.2.19 (b) is basically the same except that there are more "no-trend" stations and fewer "up-trend" stations. This probably is due to the fact that usually the estimates based on medians is more conservative than those based on means. The patterns in these two figures are consistent with the results of median polish since the site with "-"'s belong mainly to cluster 3 and those with "+"'s, to clusters 1 and 2. Comparing Figures 4.2.19 ((a) and (b)) with Figures 4.1.19 ((a) and (b)), the conclusion is that though the trends for the stations located in the Western United States continue to decline, most of the stations located in the Eastern United States no longer have down-trend and some of them have an up-trend during 1983 to 1986! This implies that some of the stations do have a V-shaped trend.

The results of Kriging the slope estimates are displayed in Figures 4.2.20(a) and (b) where (a) is for the weighted means and (b) is for medians. As before, a "+" means an up-trend and a "-" means a down-trend estimated at the corresponding point, respectively. Again, these two plots are similar except that both the "+" area and "-" areas in (b) are smaller than those in (a). Both Figures 4.2.20 (a) and (b) show that there is a down-trend for the most of the Western United States, an up-trend in certain area of the Eastern United States and no-trend in the rest of the areas. Comparing these two figures with Figures 4.1.20 ((a) and (b)), we can see that the areas with "-" shrink significantly towards the West and the areas with "+" appear in the East. The difference in the patterns of the trends is quite remarkable. For the reason given in Section 4.1.3, these two figures can only be used as indicators of the spatial distribution of the trends as the interpolation error for a specific point could be big.

The Locations Of The 31 Monitoring Stations From 1980 To 1986
(For Sulphate)

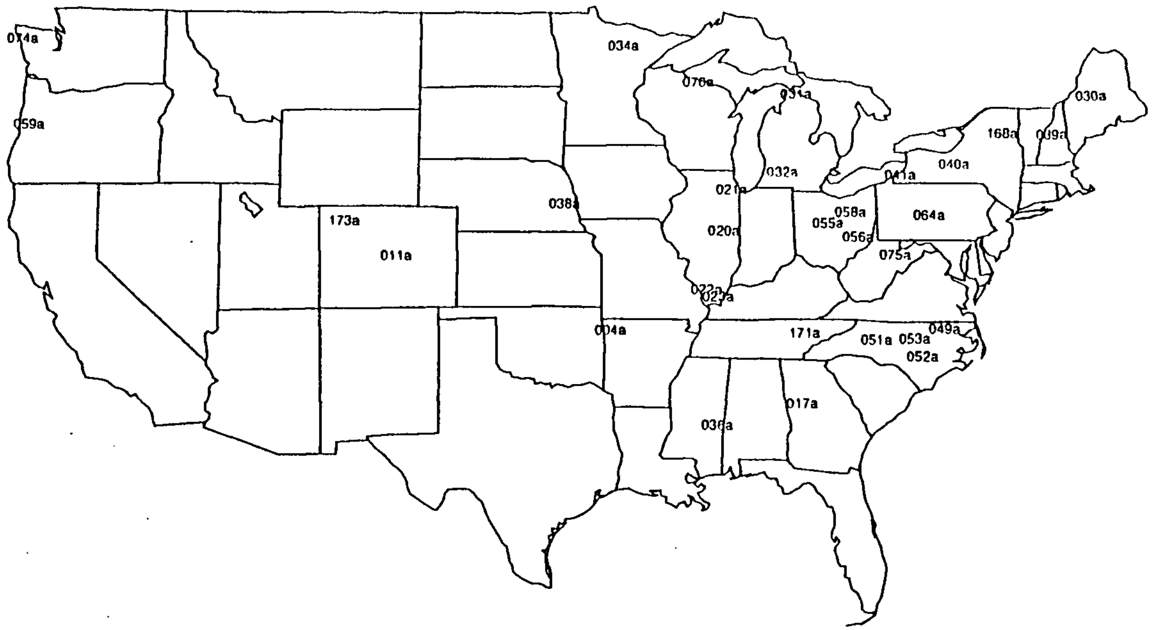


Figure 4.1.1

Clustering of SO4 monthly volume weighted mean based on sqrt(MSE)

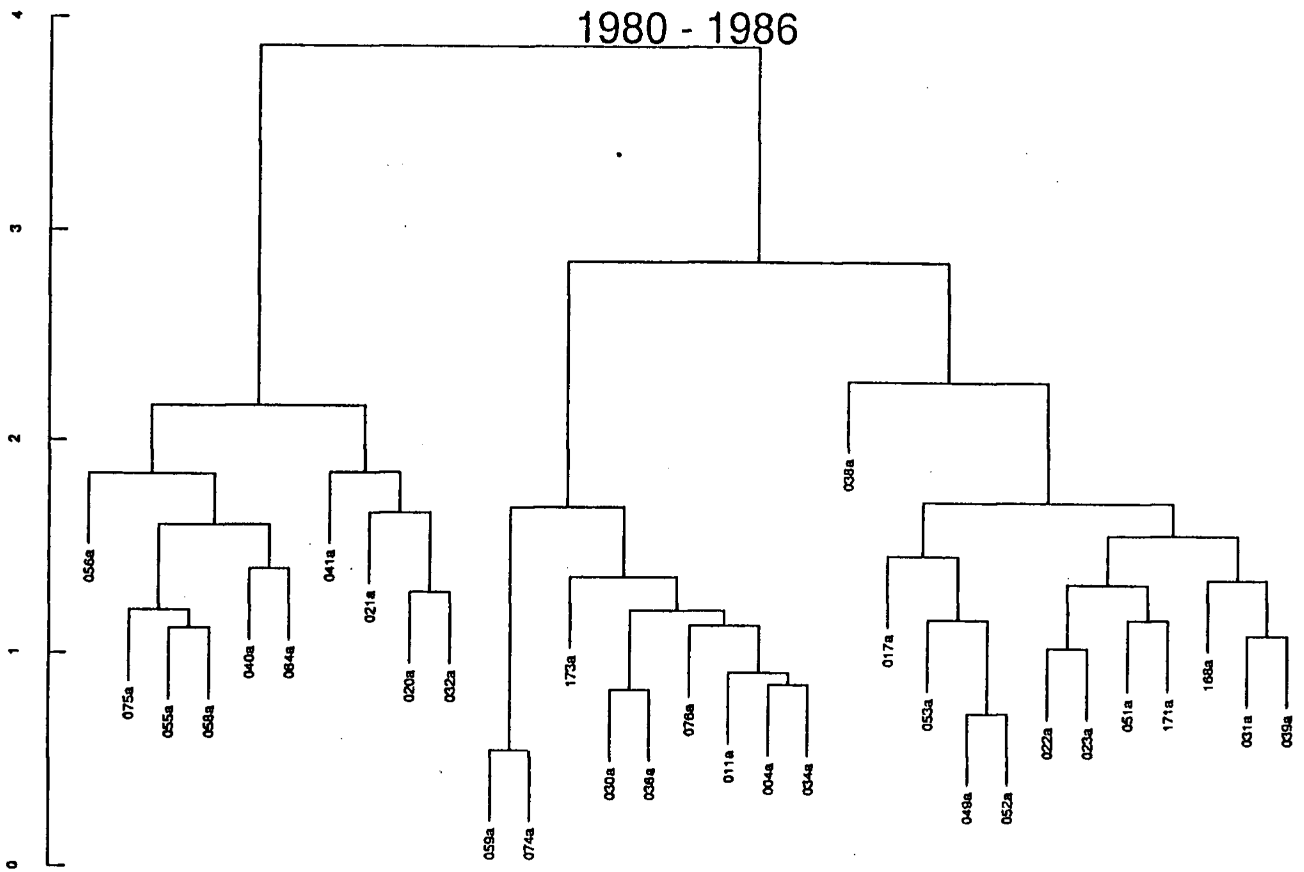


Figure 4.1.2(a)

Clustering of SO4 monthly volume weighted mean based on sqrt(MSE)

1980 - 1986 (with outliers)

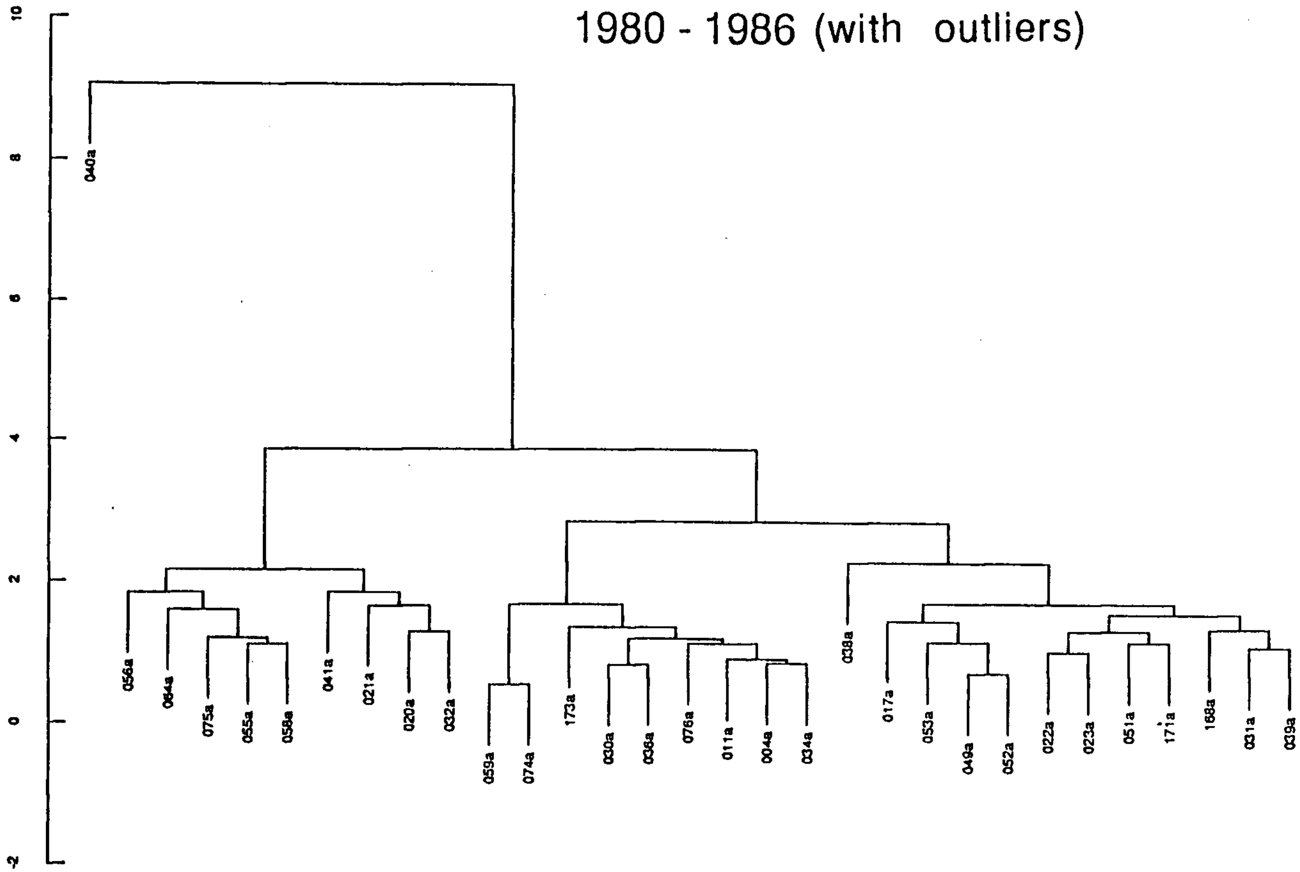


Figure 4.1.2(b)

Clusters of SO4 monthly volume weighted mean based on sqrt(MSE)

1980 - 1986 (k=3)

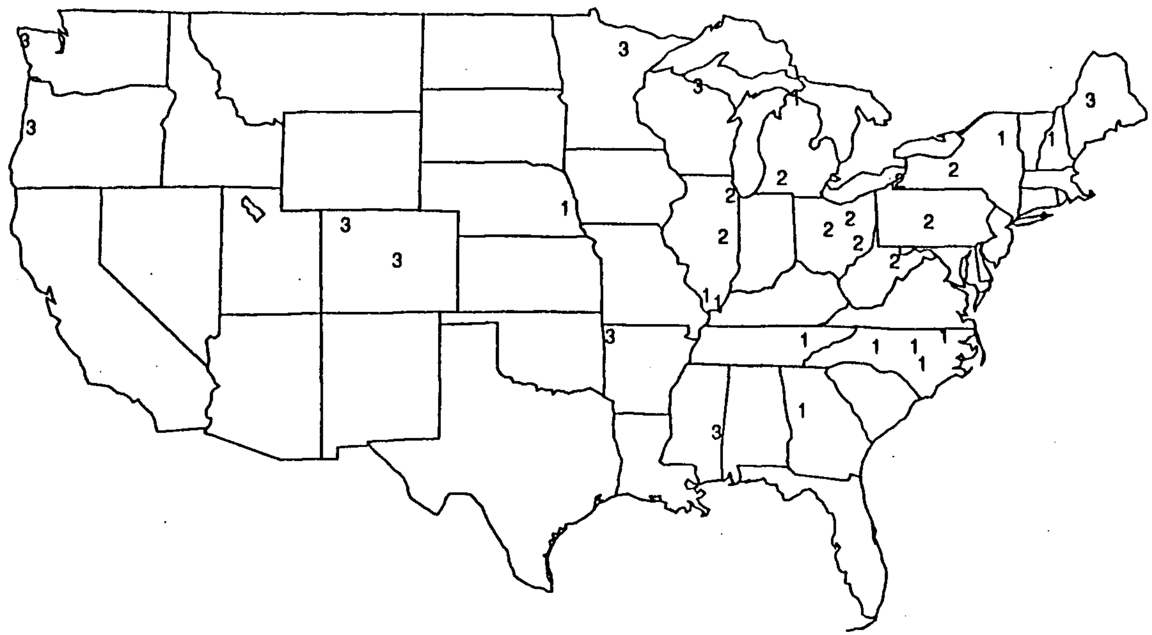


Figure 4.1.3

Histograms of Transformed SO4 (Volume Weighted Mean, 80-86)

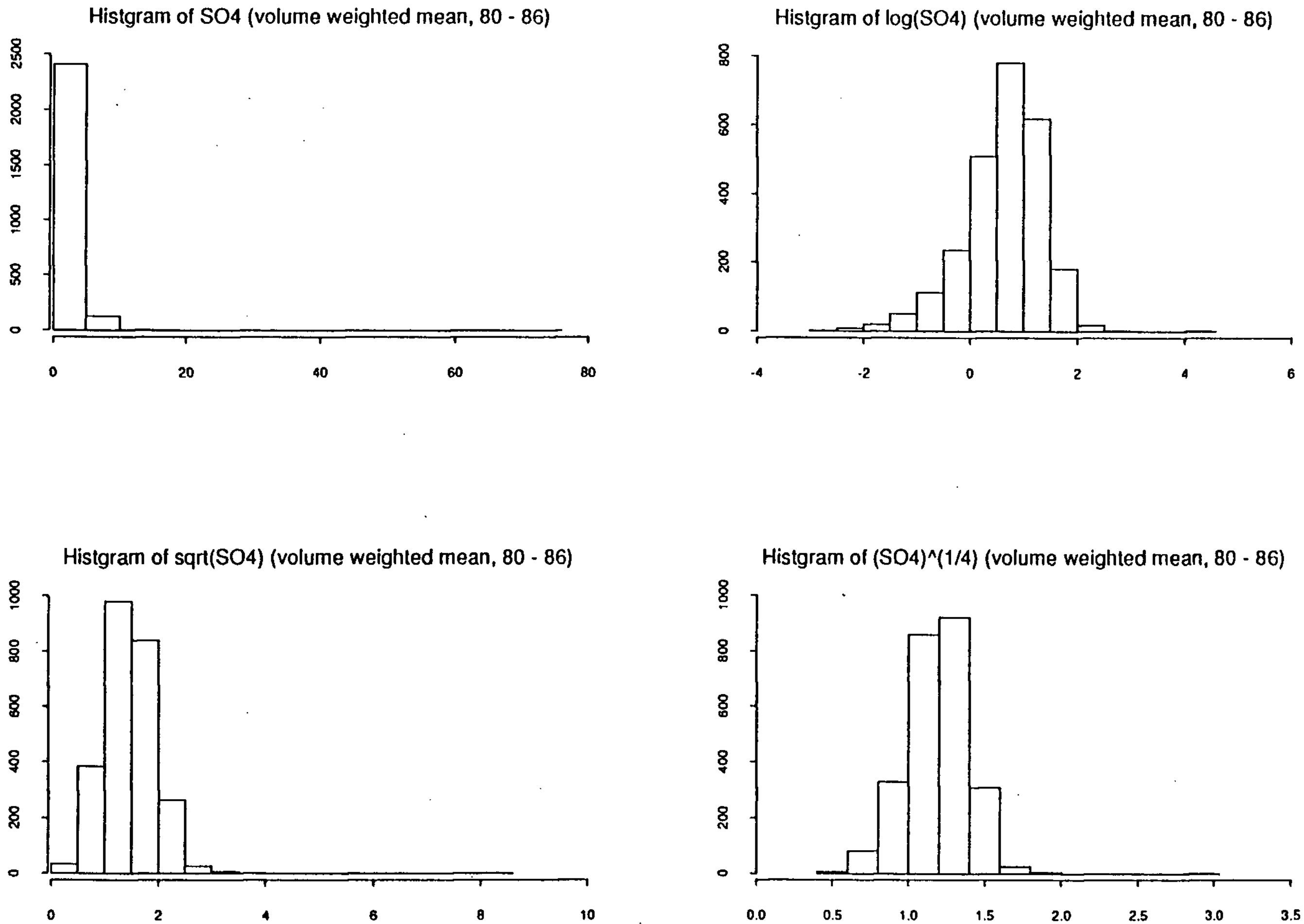


Figure 4.1.4

Histograms of log(SO4) by Clusters (Volume Weighted Mean, 80-86)

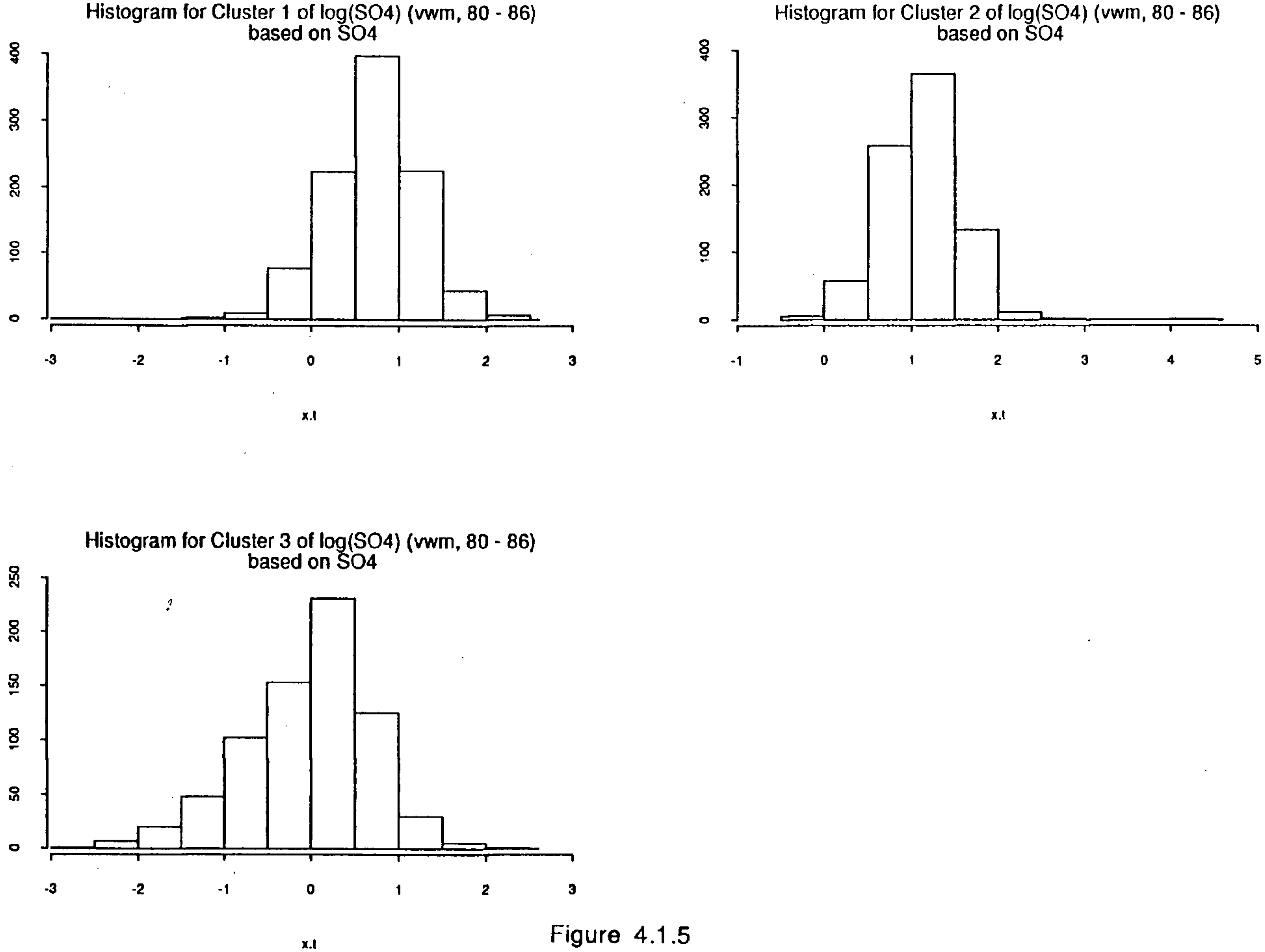


Figure 4.1.5

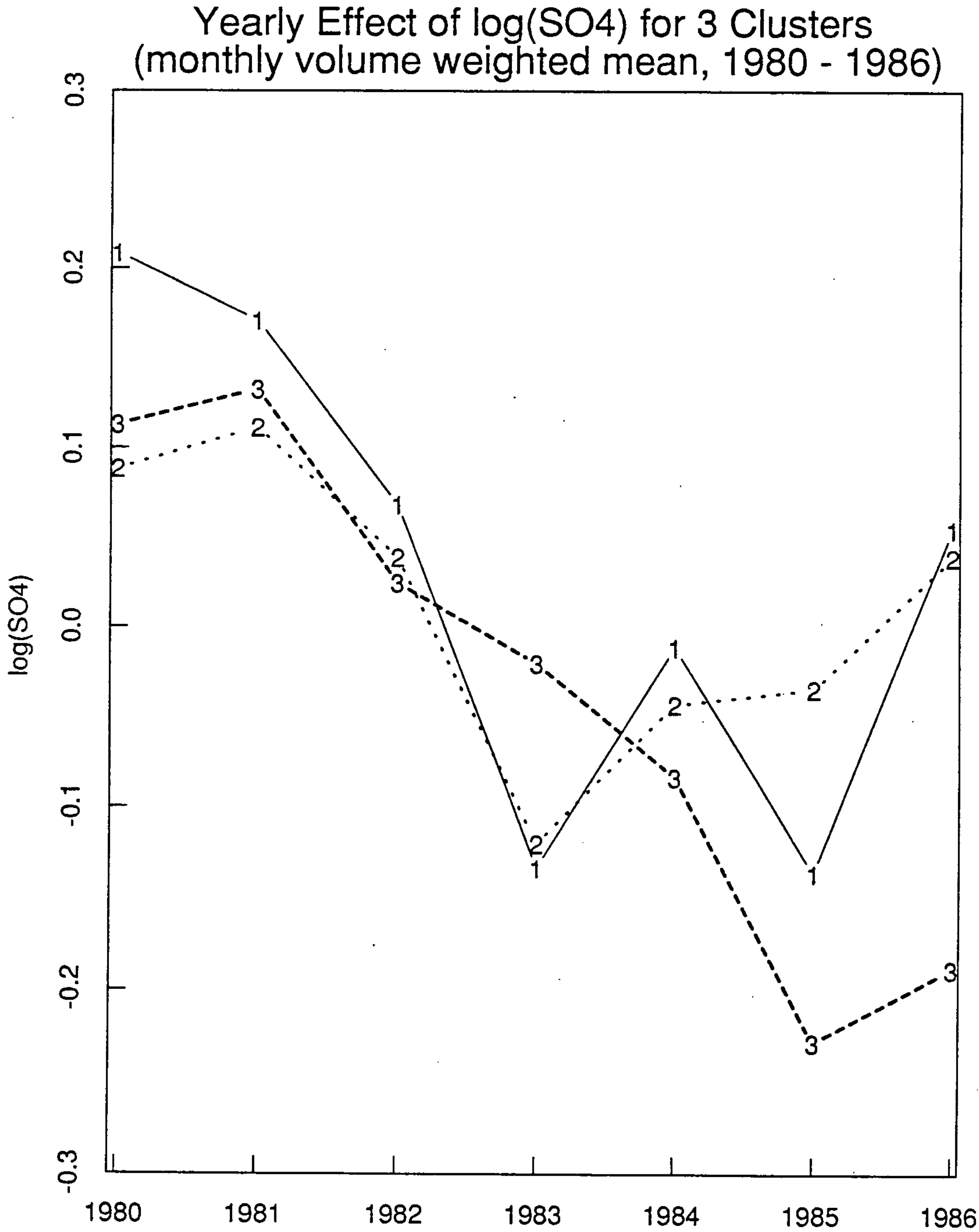

Yearly Effect of log(SO4) for 3 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(SO4)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
1980
1981
1982
1983
1984
1985
1986

Figure 4.1.6

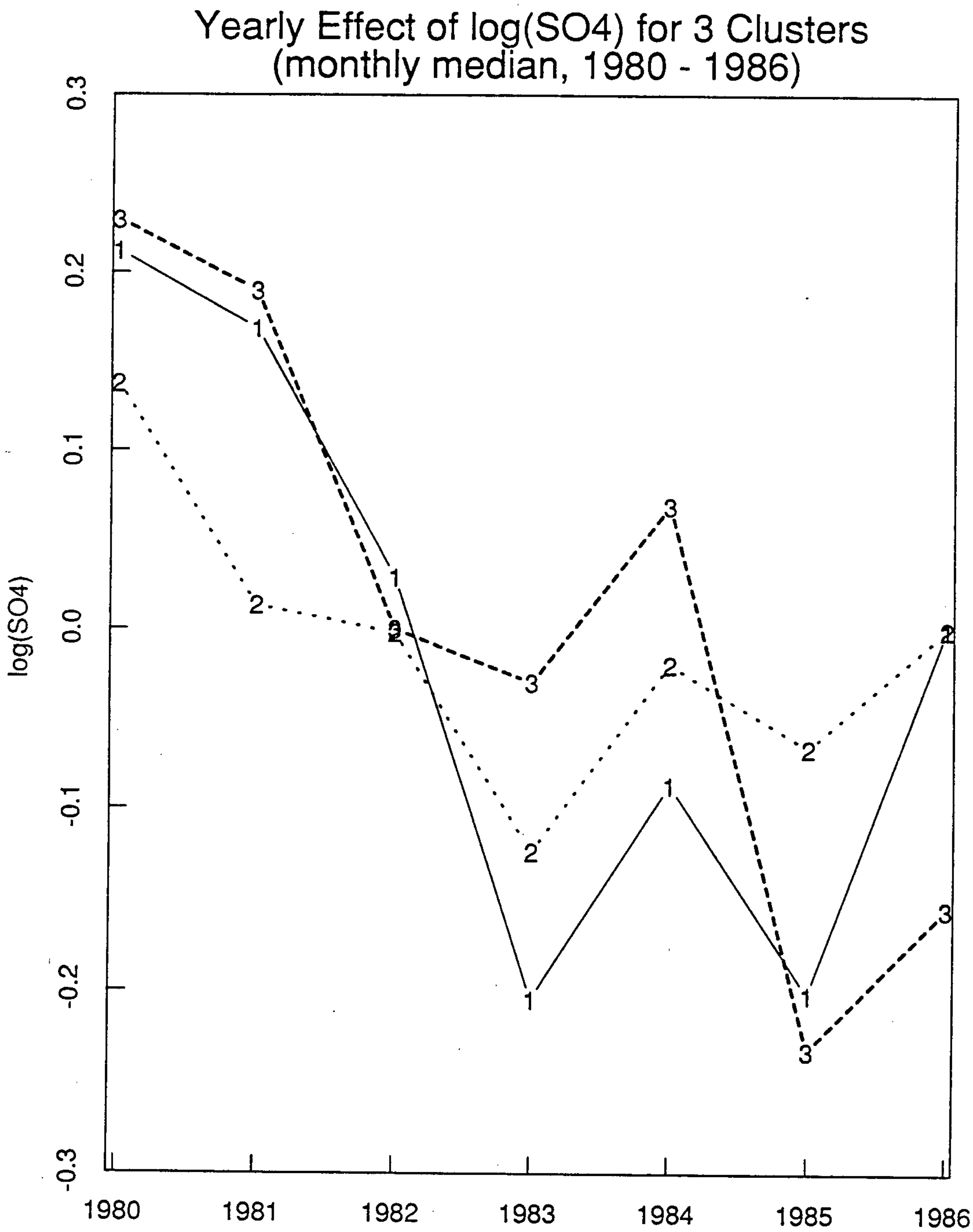
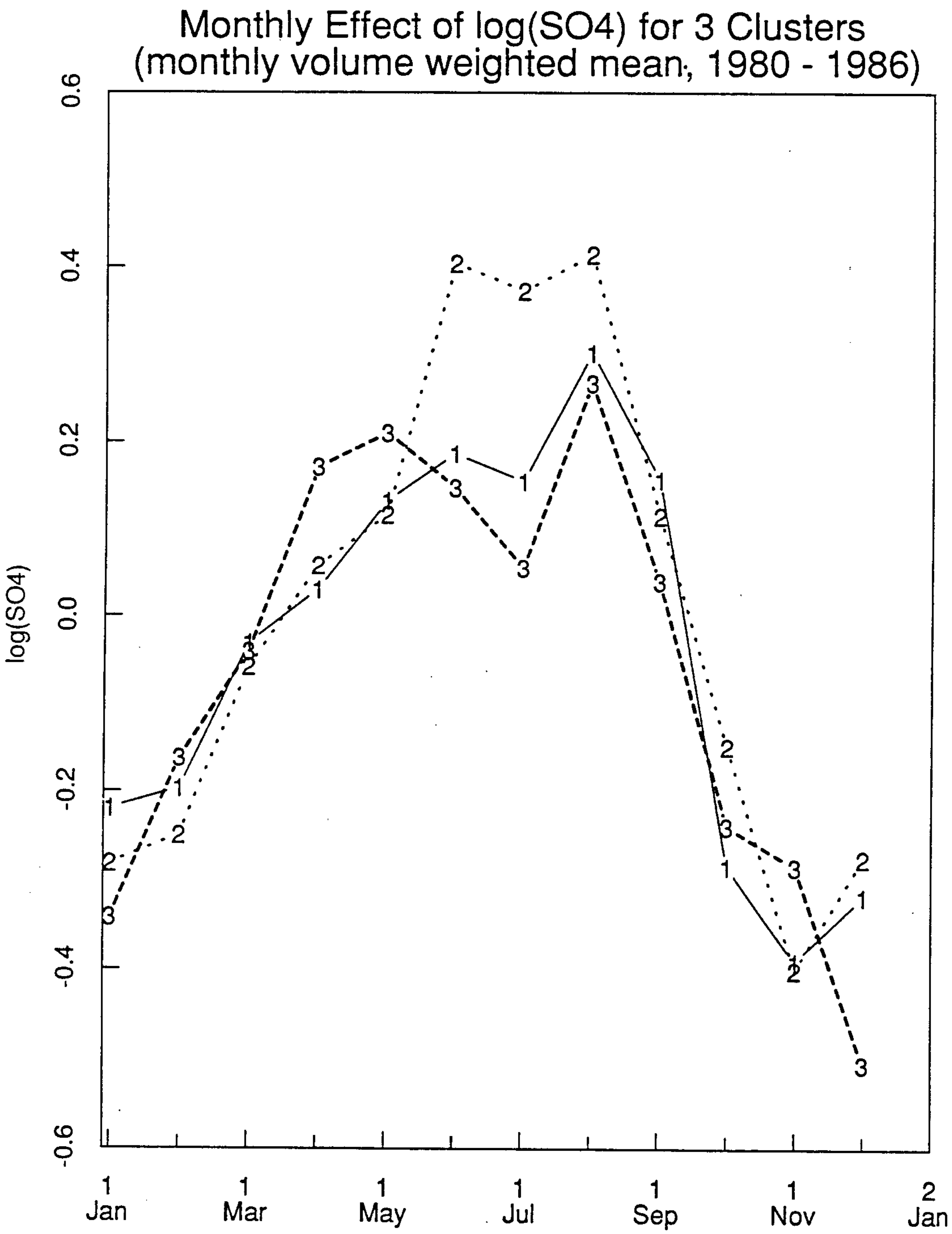
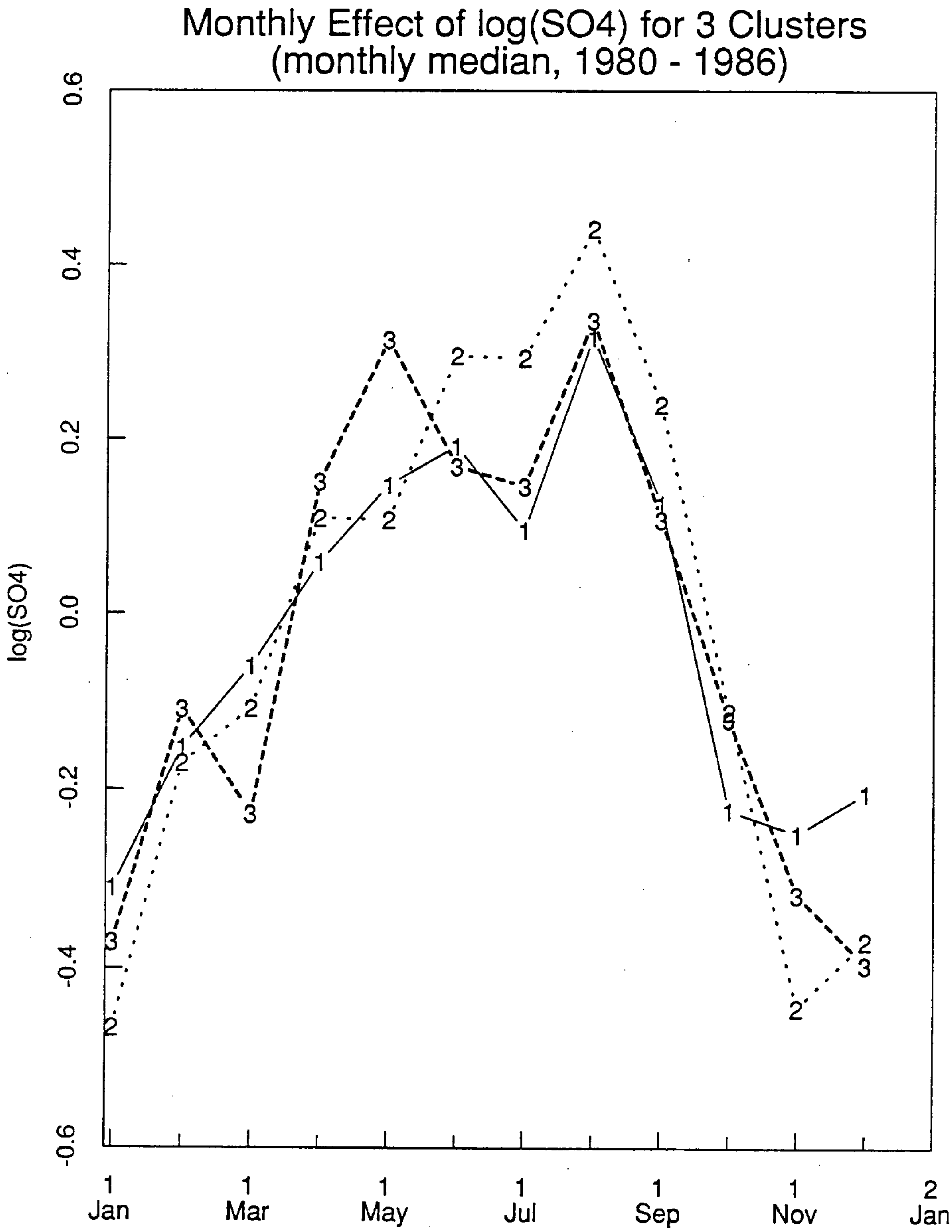

Yearly Effect of log(SO4) for 3 Clusters
(monthly median, 1980 - 1986)
log(SO4)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
1980
1981
1982
1983
1984
1985
1986

Figure 4.1.7



Monthly Effect of log(SO4) for 3 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(SO4)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 4.1.8



Monthly Effect of log(SO4) for 3 Clusters
(monthly median, 1980 - 1986)
log(SO4)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 4.1.9

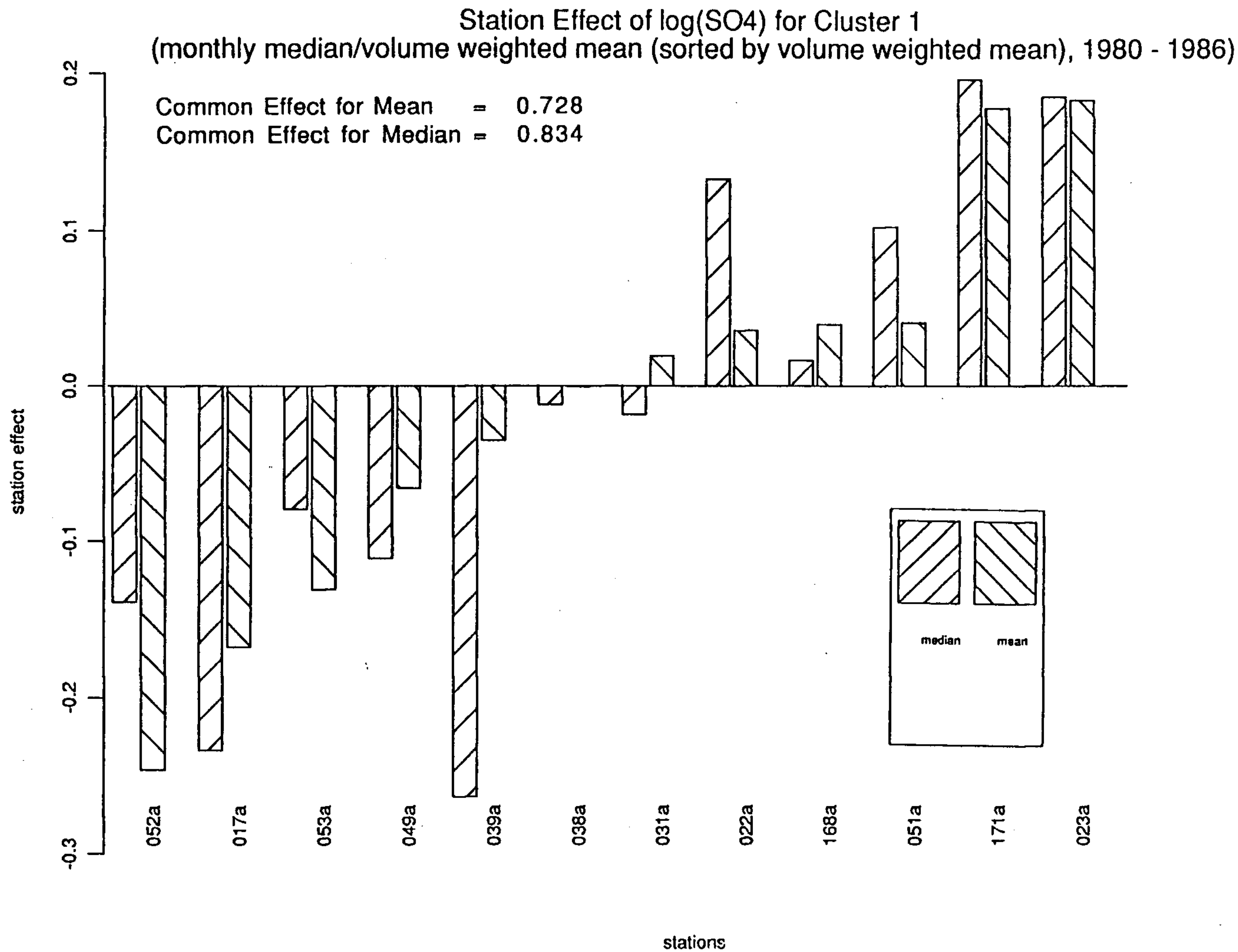

Station Effect of log(SO4) for Cluster 1
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)
Common Effect for Mean = 0.728
Common Effect for Median = 0.834
station effect
0.2
0.1
0.0
-0.1
-0.2
-0.3
052a
017a
053a
049a
039a
038a
031a
022a
168a
051a
171a
023a
median
mean
stations

Figure 4.1.10

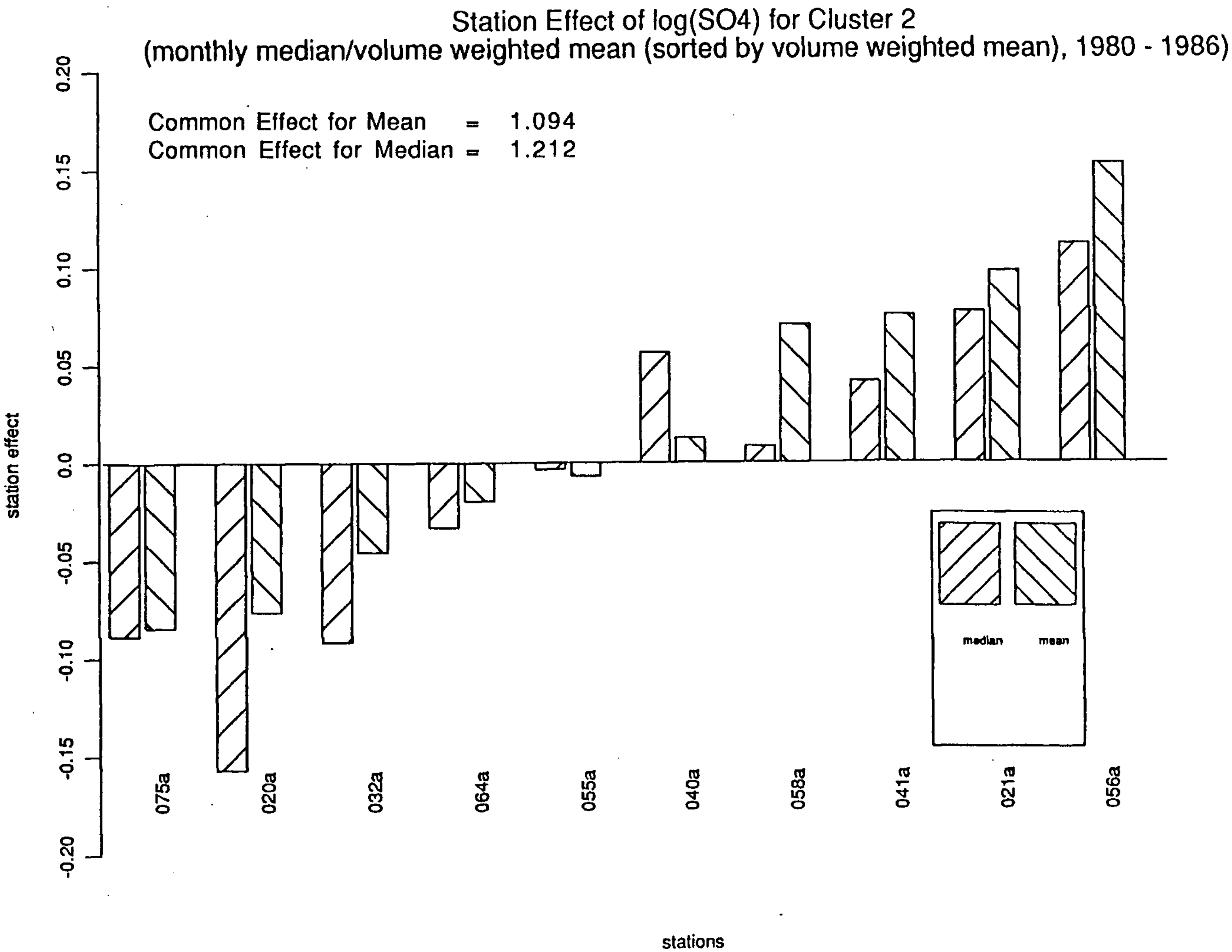
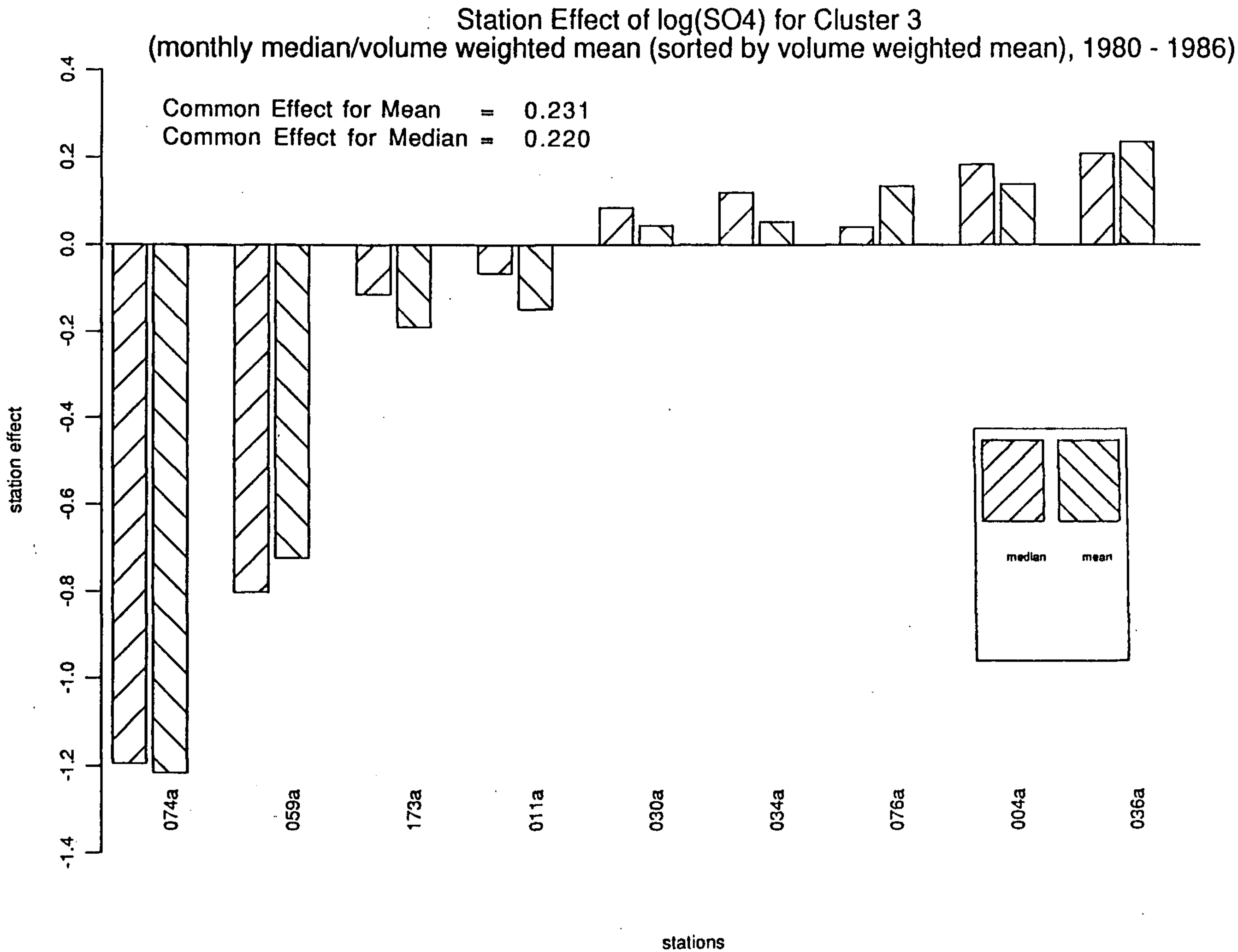

Station Effect of log(SO4) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)
Common Effect for Mean = 1.094
Common Effect for Median = 1.212
station effect
0.20
0.15
0.10
0.05
0.0
-0.05
-0.10
-0.15
-0.20
075a
020a
032a
064a
055a
040a
058a
041a
021a
056a
median
mean
stations

Figure 4.1.11



Station Effect of log(SO4) for Cluster 3
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)
Common Effect for Mean = 0.231
Common Effect for Median = 0.220
station effect
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
074a
059a
173a
011a
030a
034a
076a
004a
036a
median
mean
stations

Figure 4.1.12

Boxplot for the resid. of log(SO4)
(80-86, monthly volume weighted mean, clust 1)

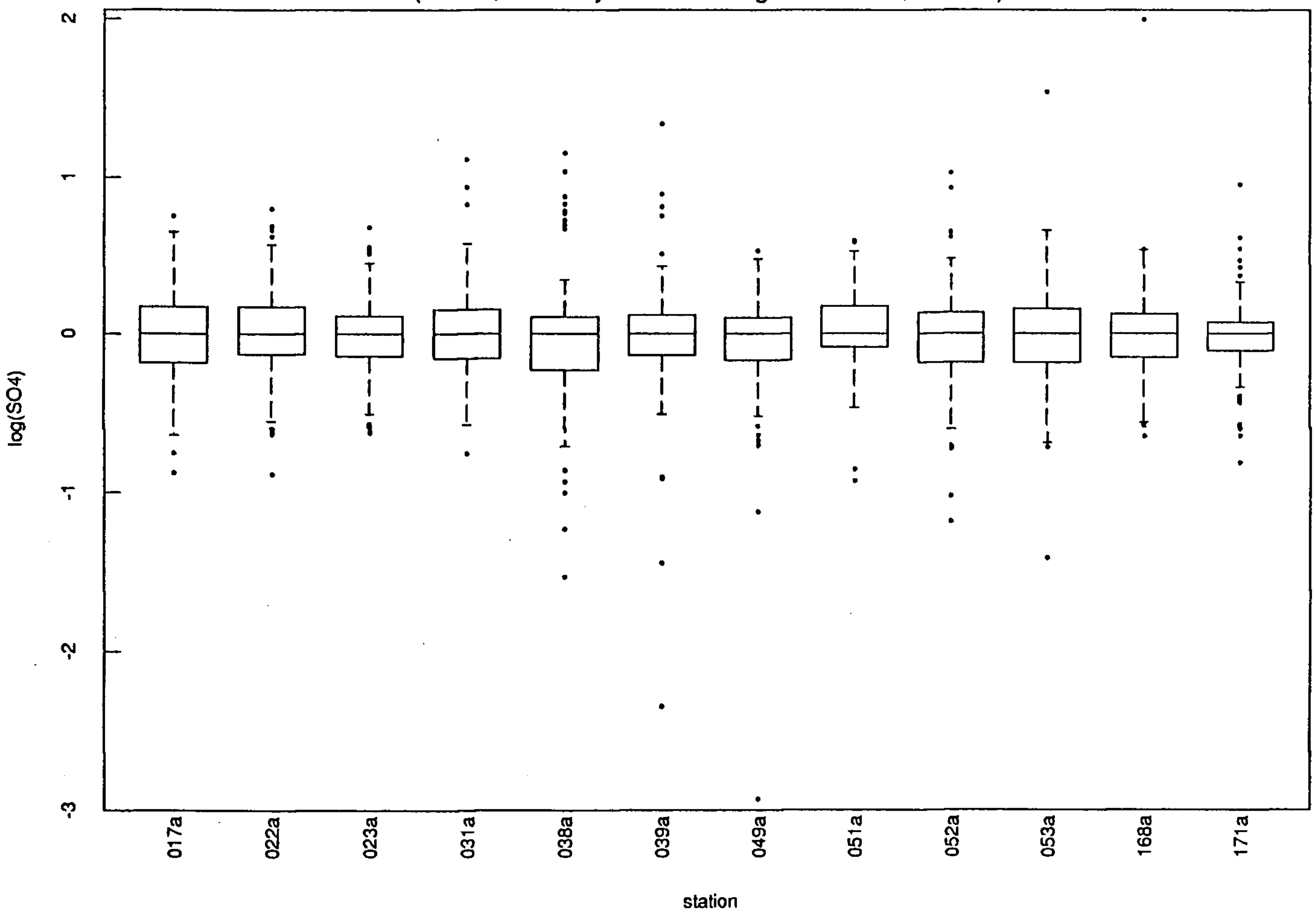


Figure 4.1.13(a)

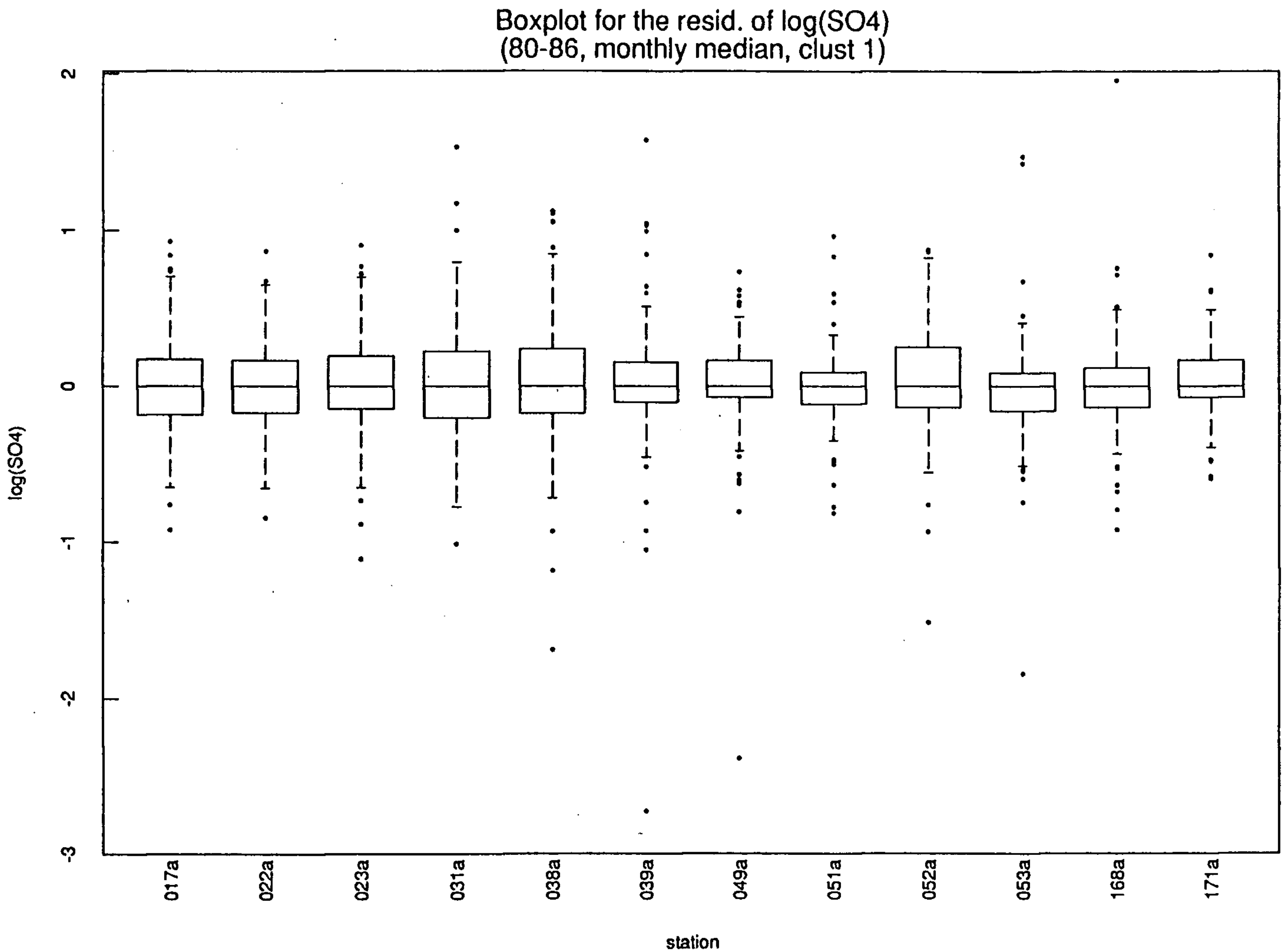


Figure 4.1.13(b)

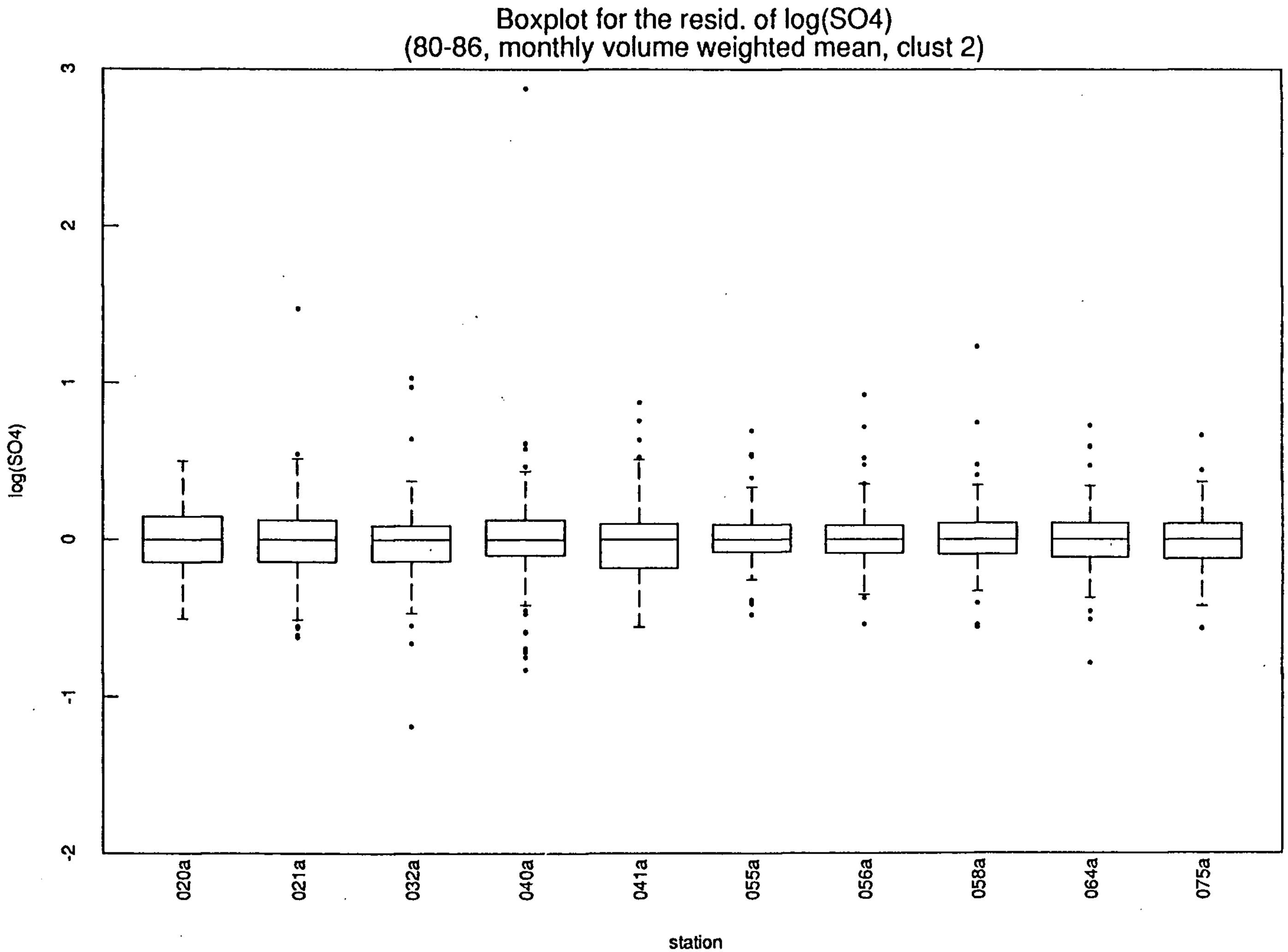


Figure 4.1.14(a)

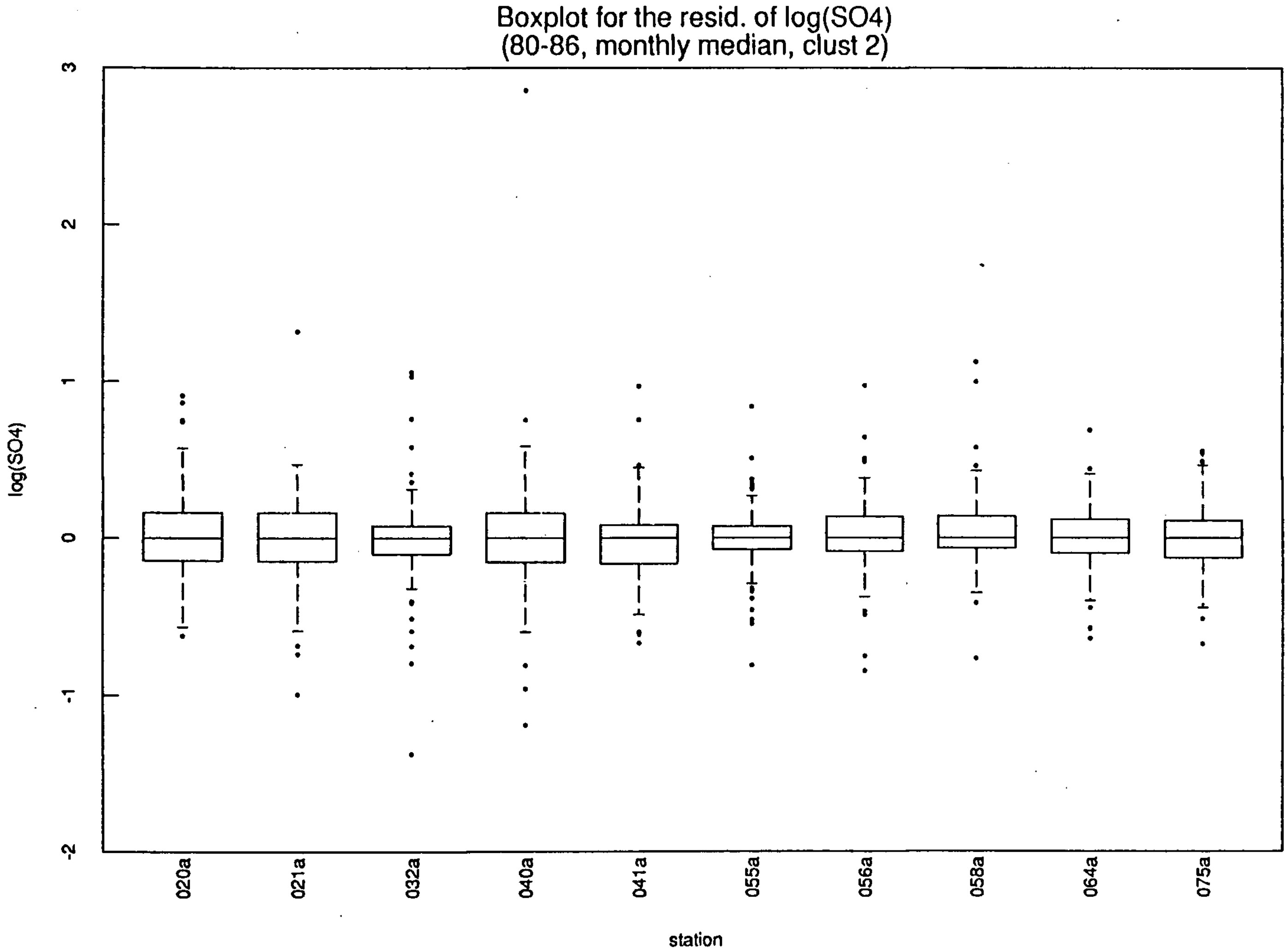


Figure 4.1.14(b)

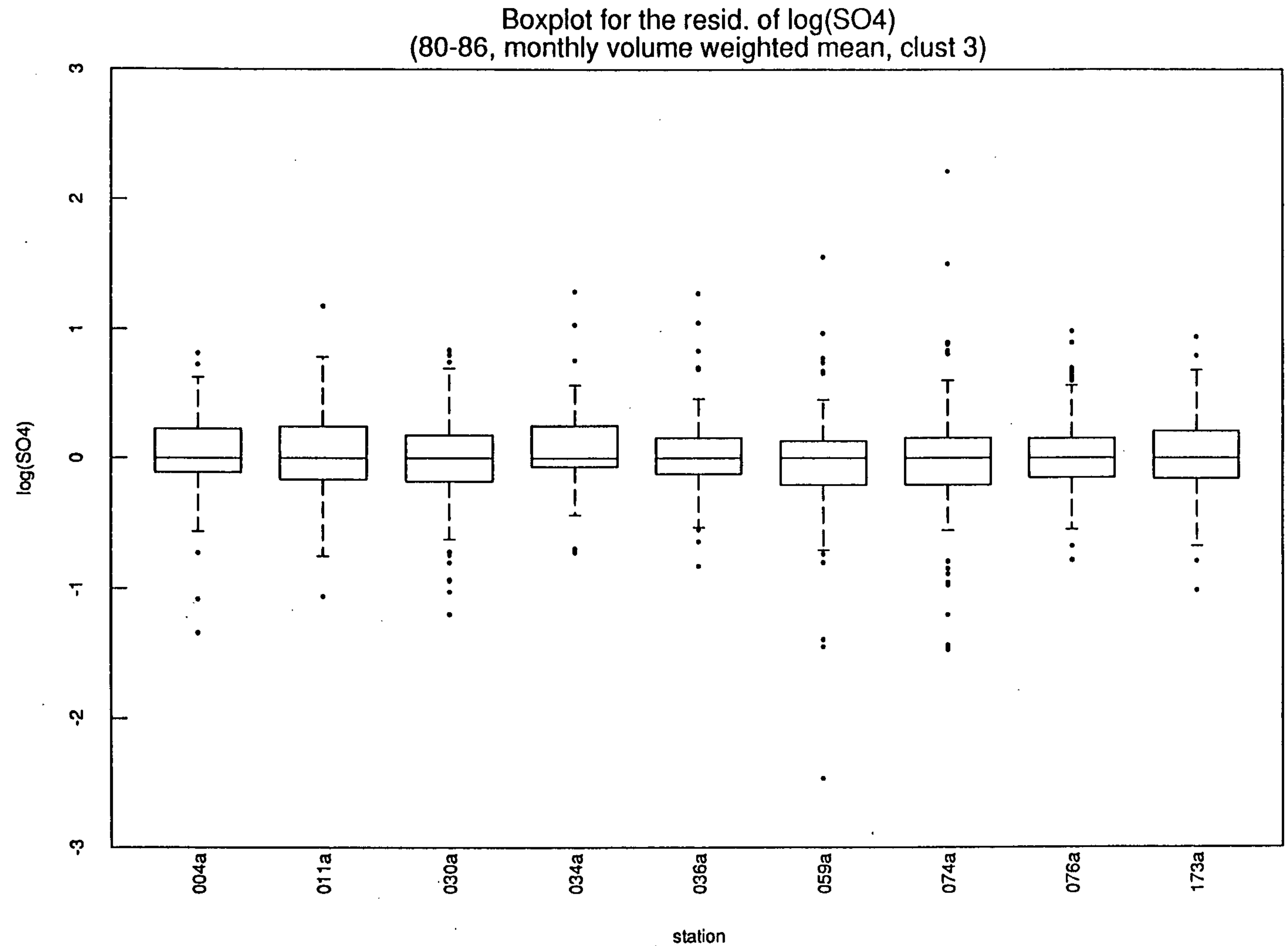


Figure 4.1.15(a)

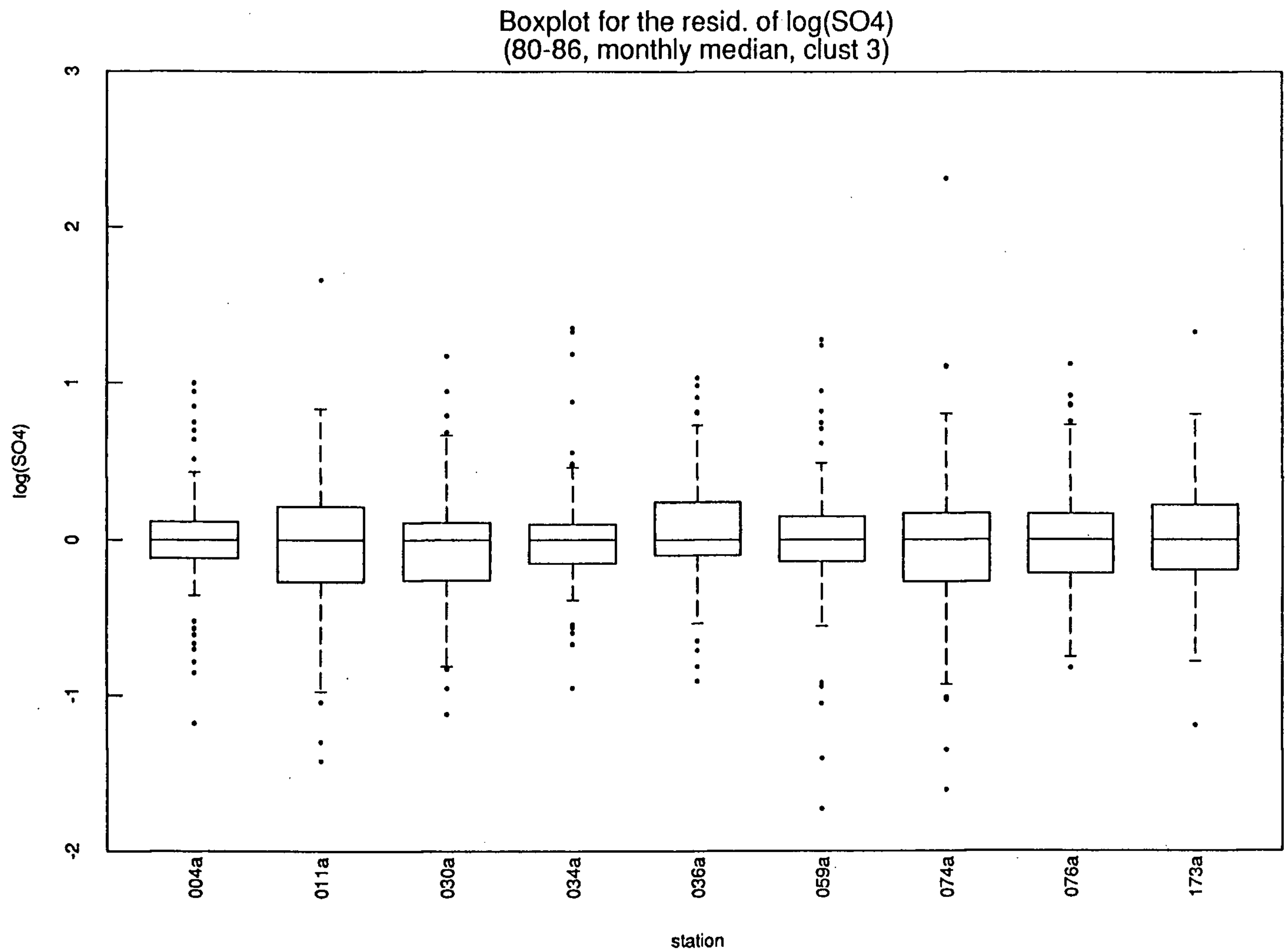


Figure 4.1.15(b)

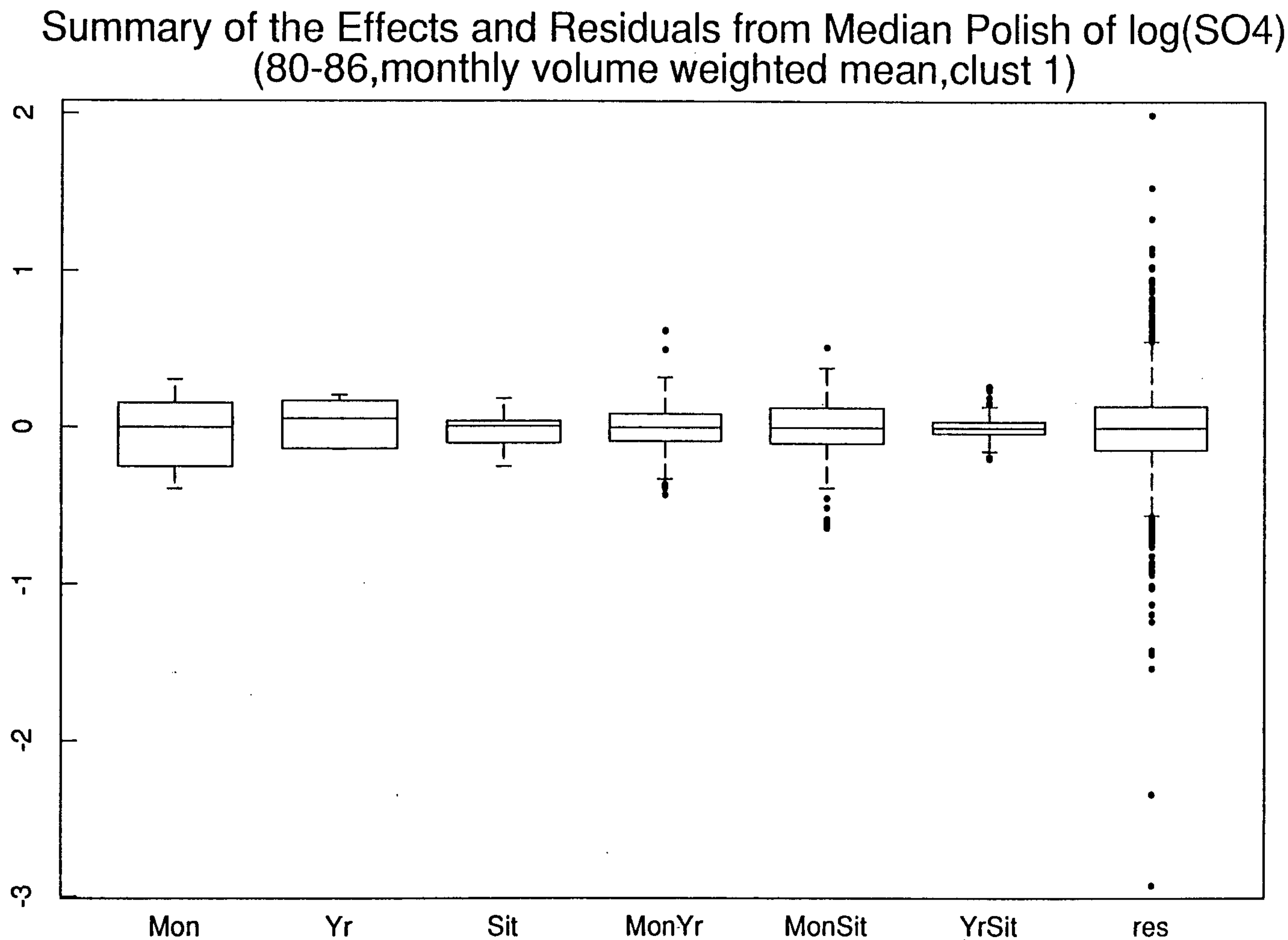


Figure 4.1.16(a)

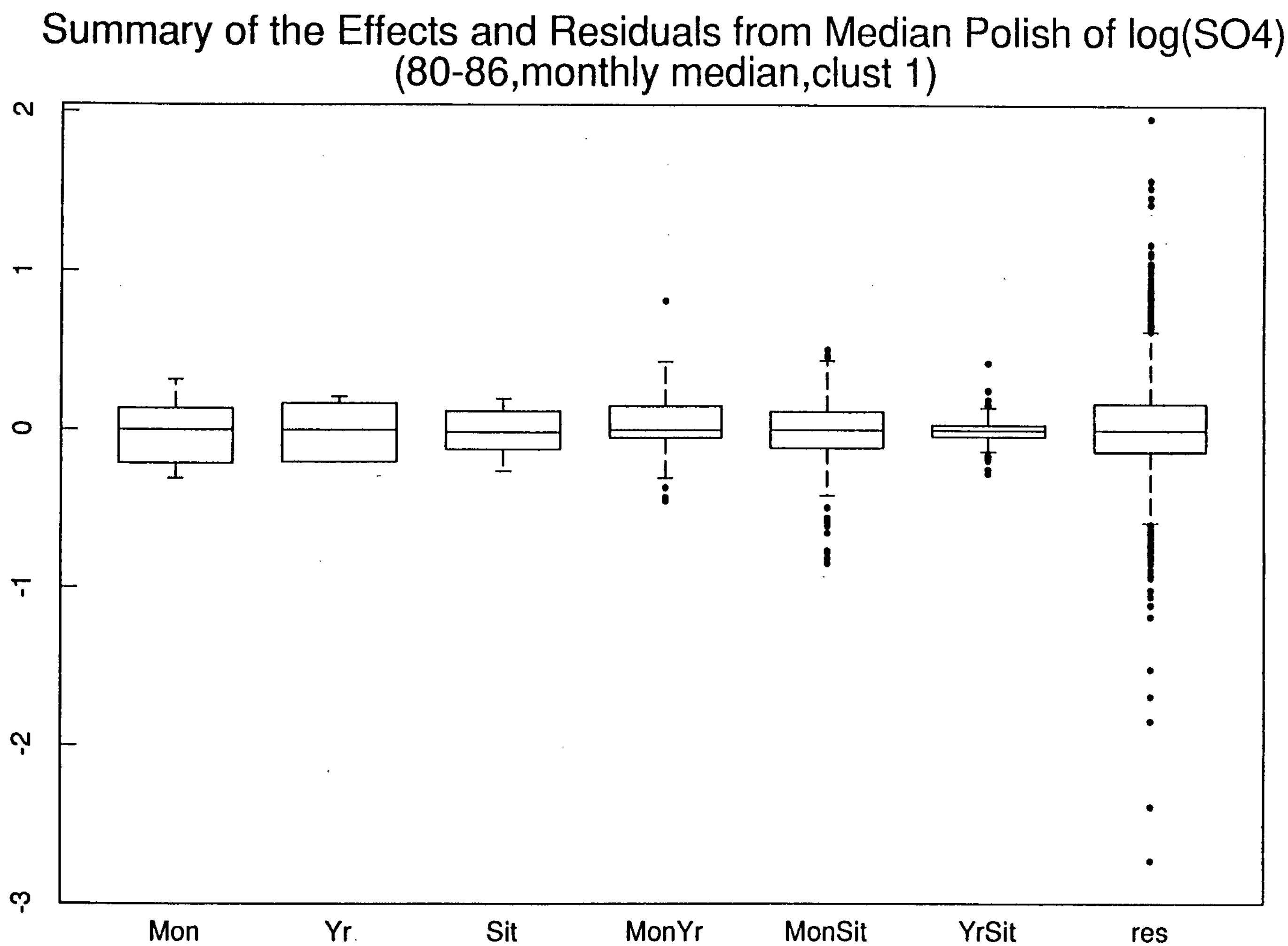


Figure 4.1.16(b)

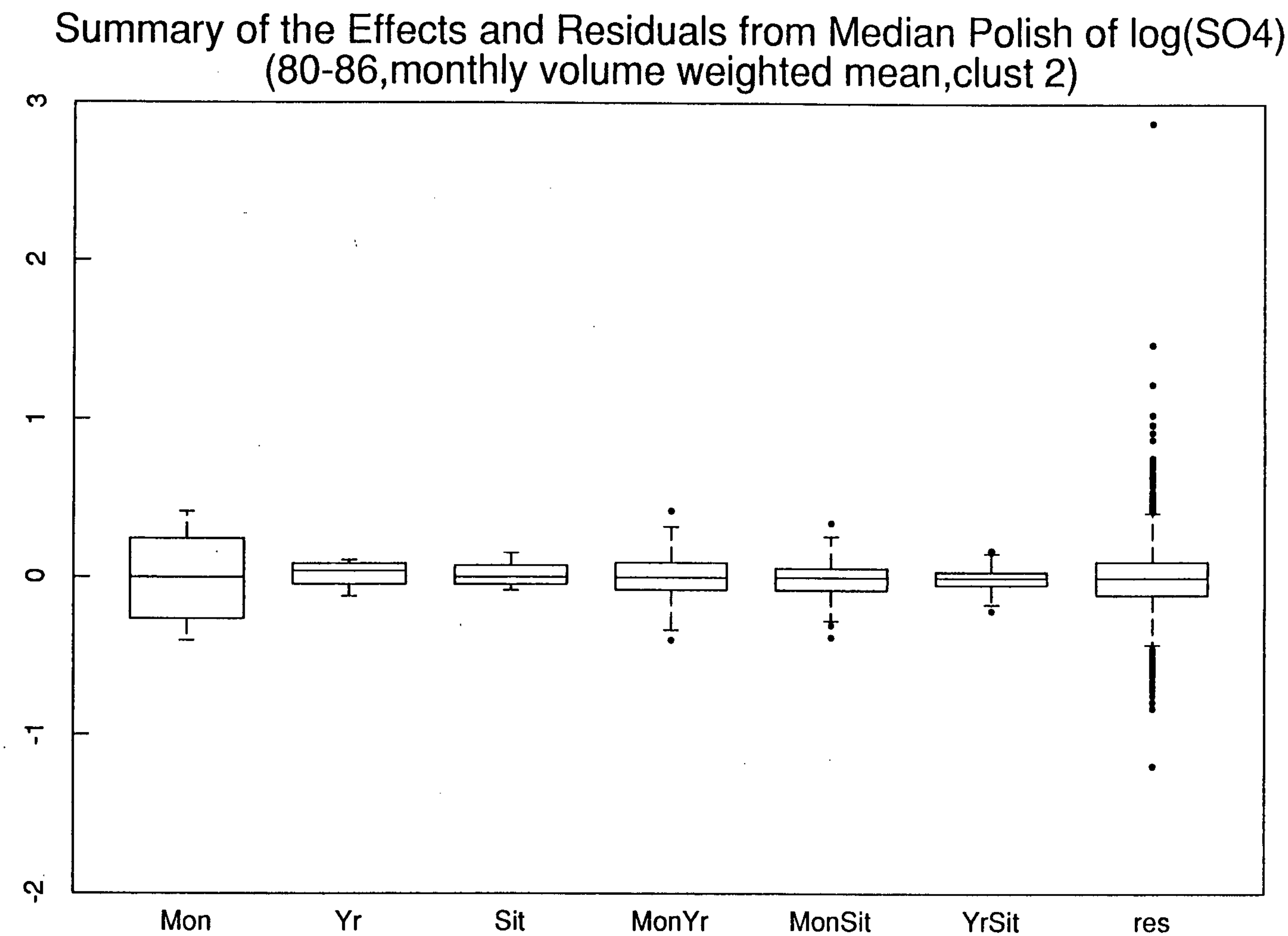


Figure 4.1.17(a)

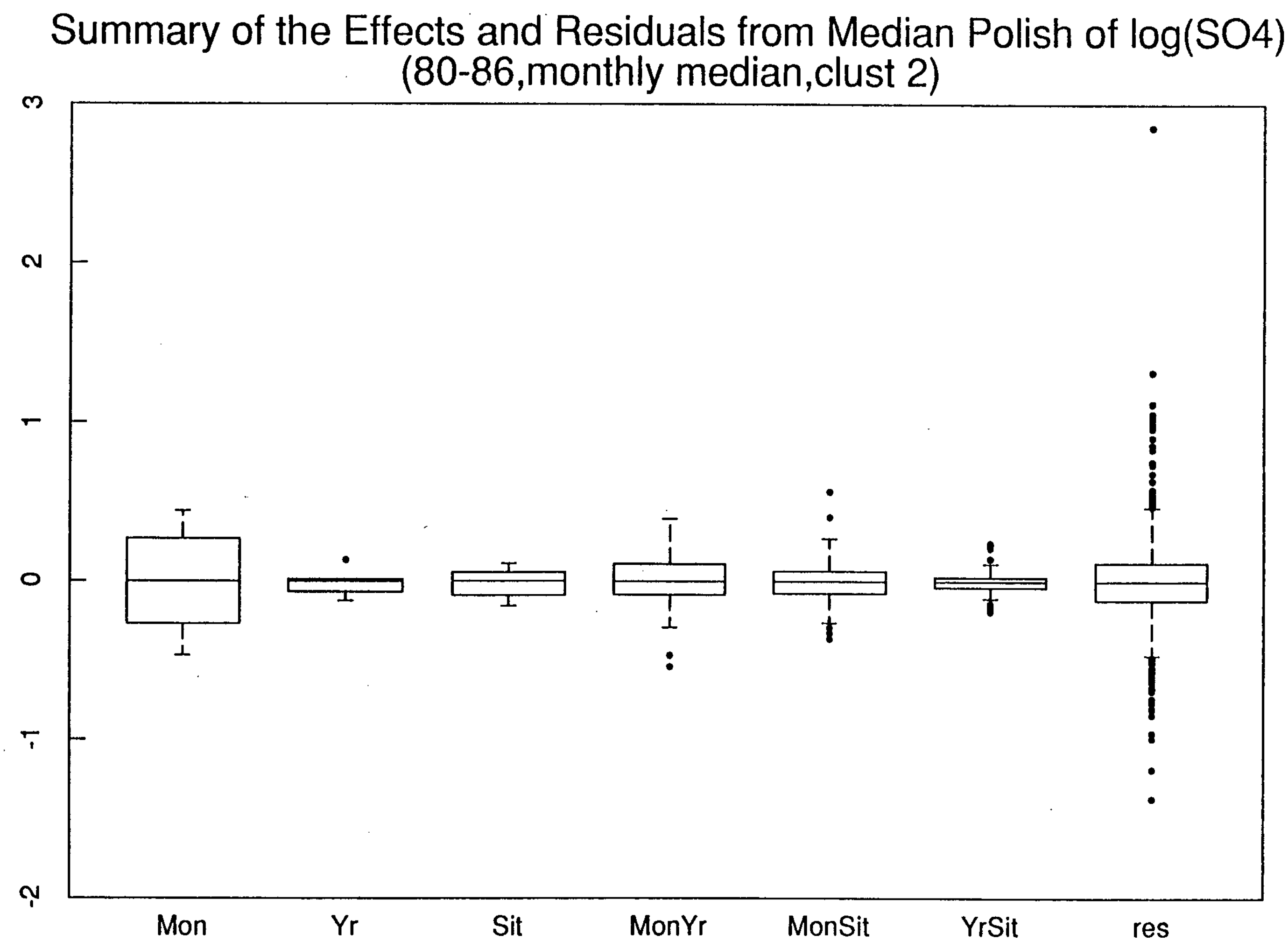


Figure 4.1.17(b)

Summary of the Effects and Residuals from Median Polish of log(SO4)
(80-86,monthly volume weighted mean,clust 3)

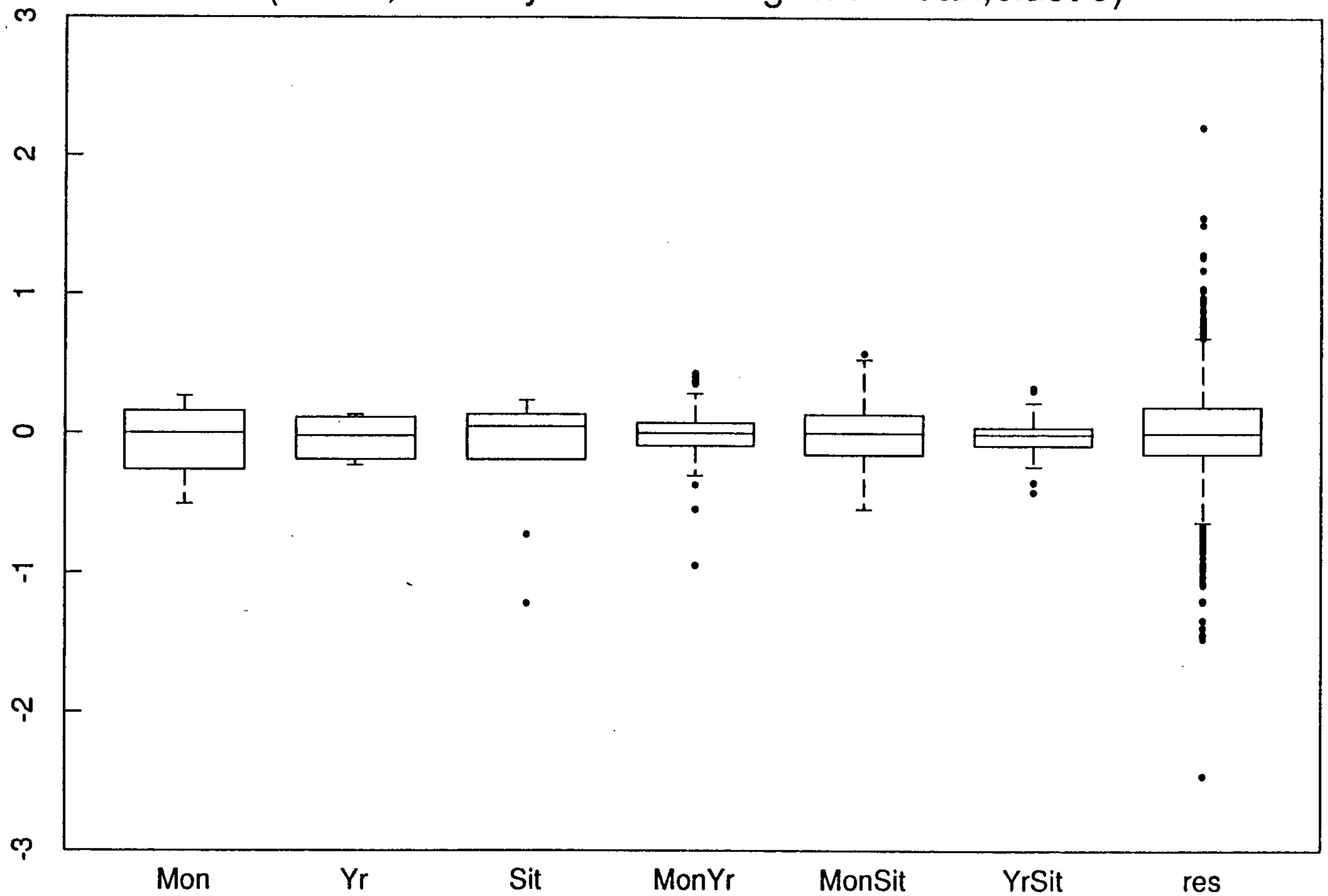


Figure 4.1.18(a)

Summary of the Effects and Residuals from Median Polish of log(SO4)
(80-86,monthly median,clust 3)

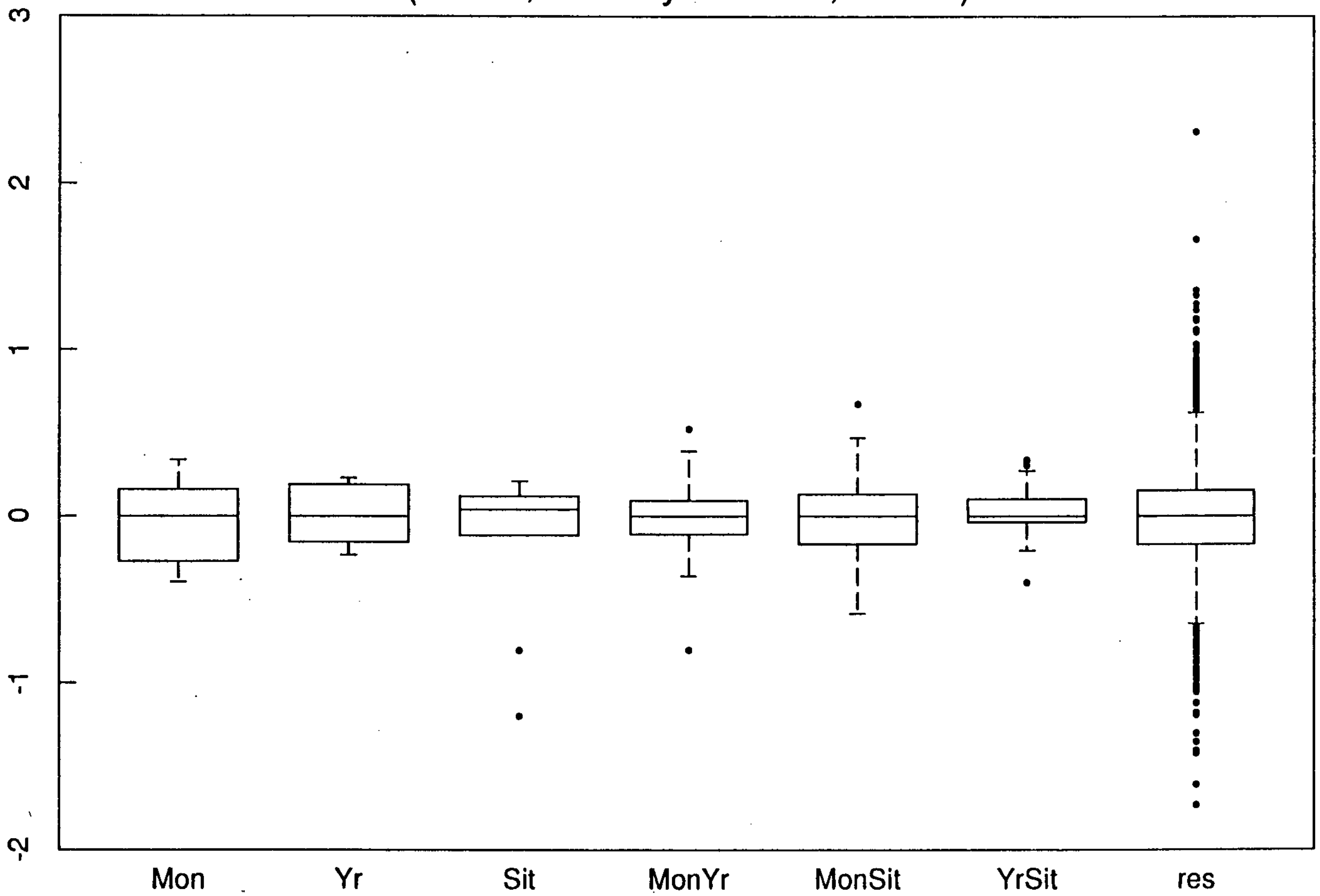


Figure 4.1.18(b)

Trend of log(SO4) from 1980 to 1986 at the 31 Stations
(calculated by monthly volume weighted mean)

Figure 4.1.19(a)

Trend of log(SO4) from 1980 to 1986 at the 31 Stations
(calculated by monthly median)

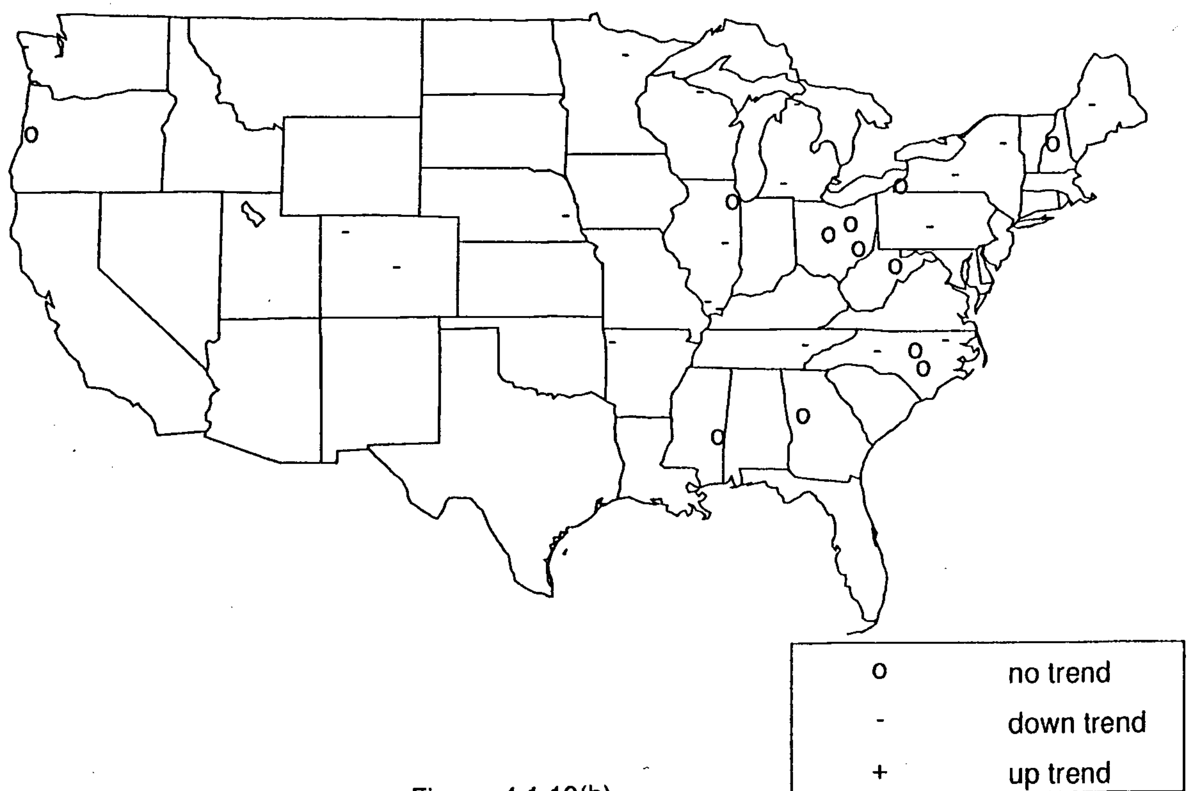


Figure 4.1.19(b)

Trend of log(SO4) from 1980 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

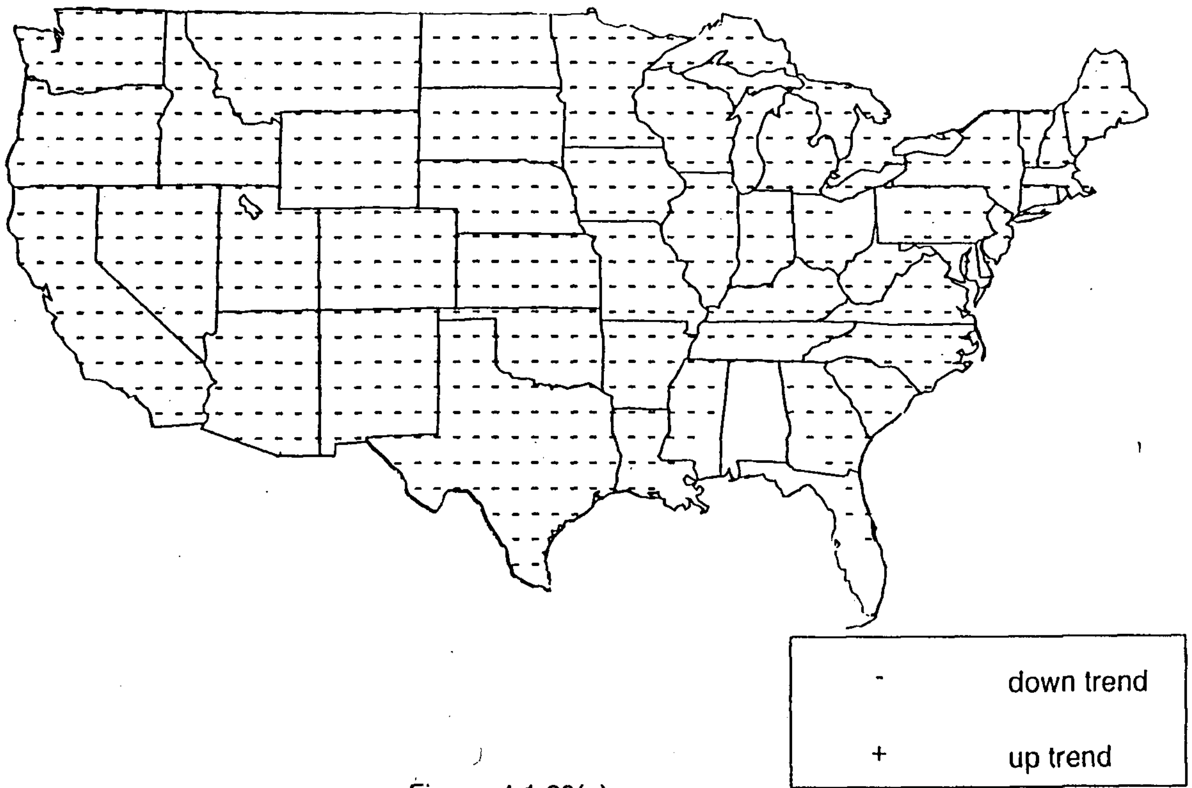


Figure 4.1.20(a)

Trend of log(SO4) from 1980 to 1986 in the USA
(calculated by Kriging from monthly median)

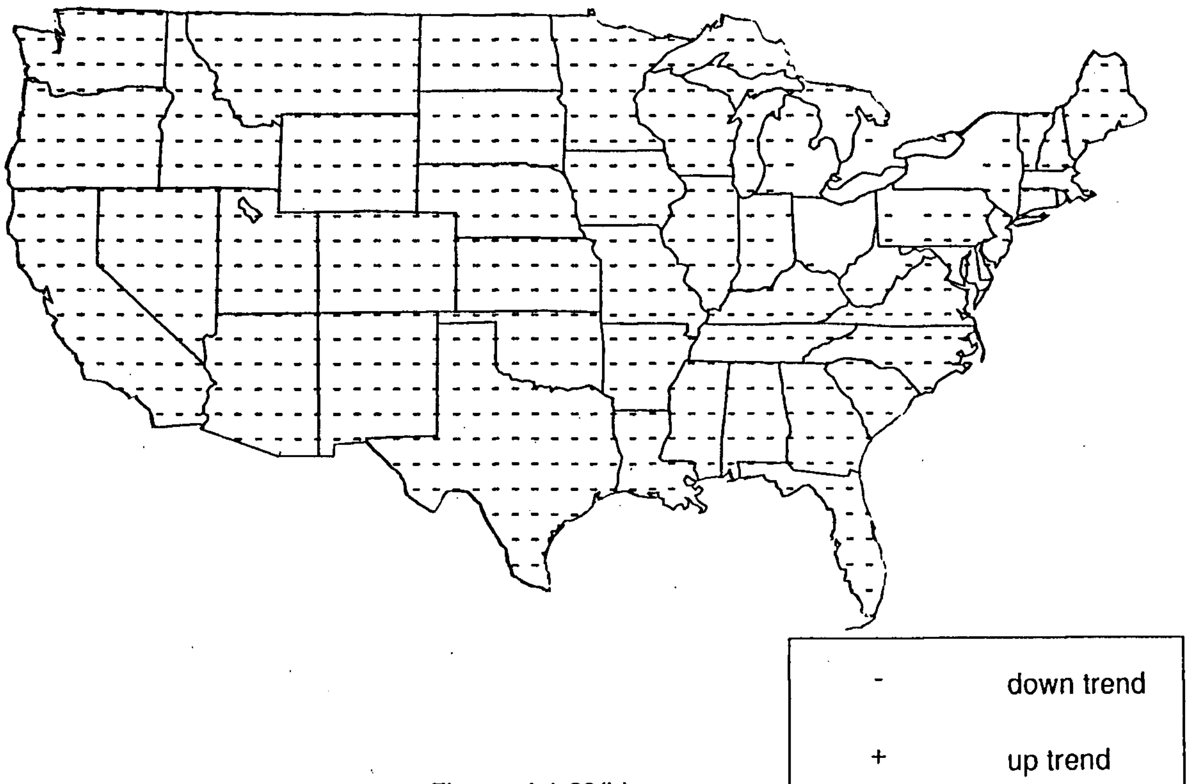


Figure 4.1.20(b)

The Locations Of The 81 Monitoring Stations From 1983 To 1986 (For Sulphate)

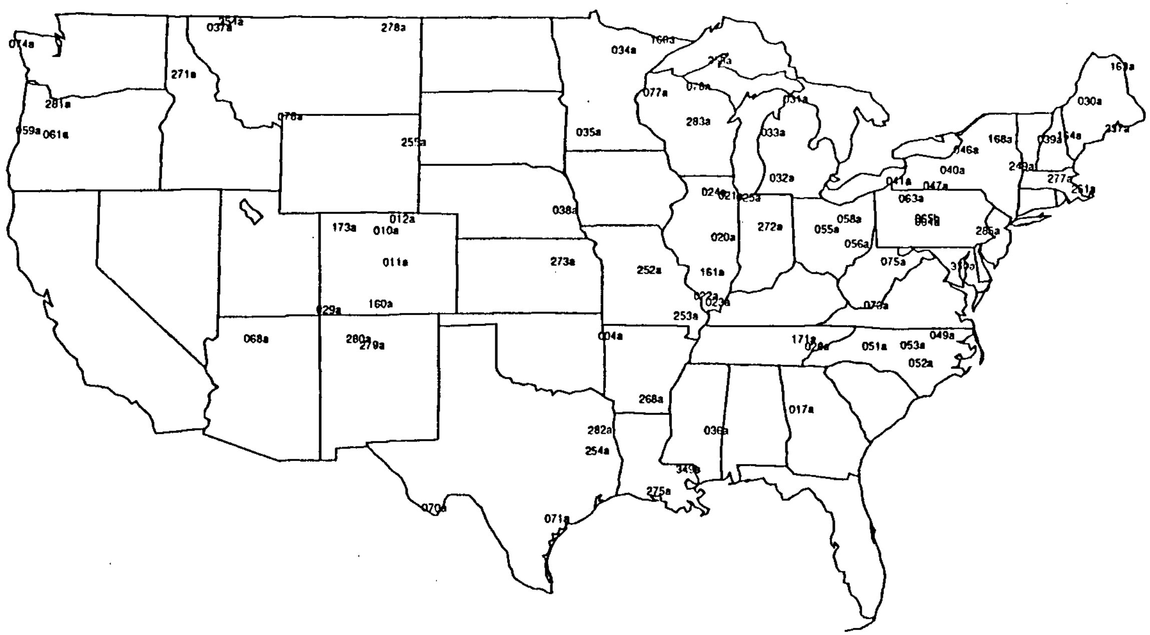


Figure 4.2.1

Clustering of SO4 monthly volume weighted mean based on sqrt(MSE)

1983 - 1986

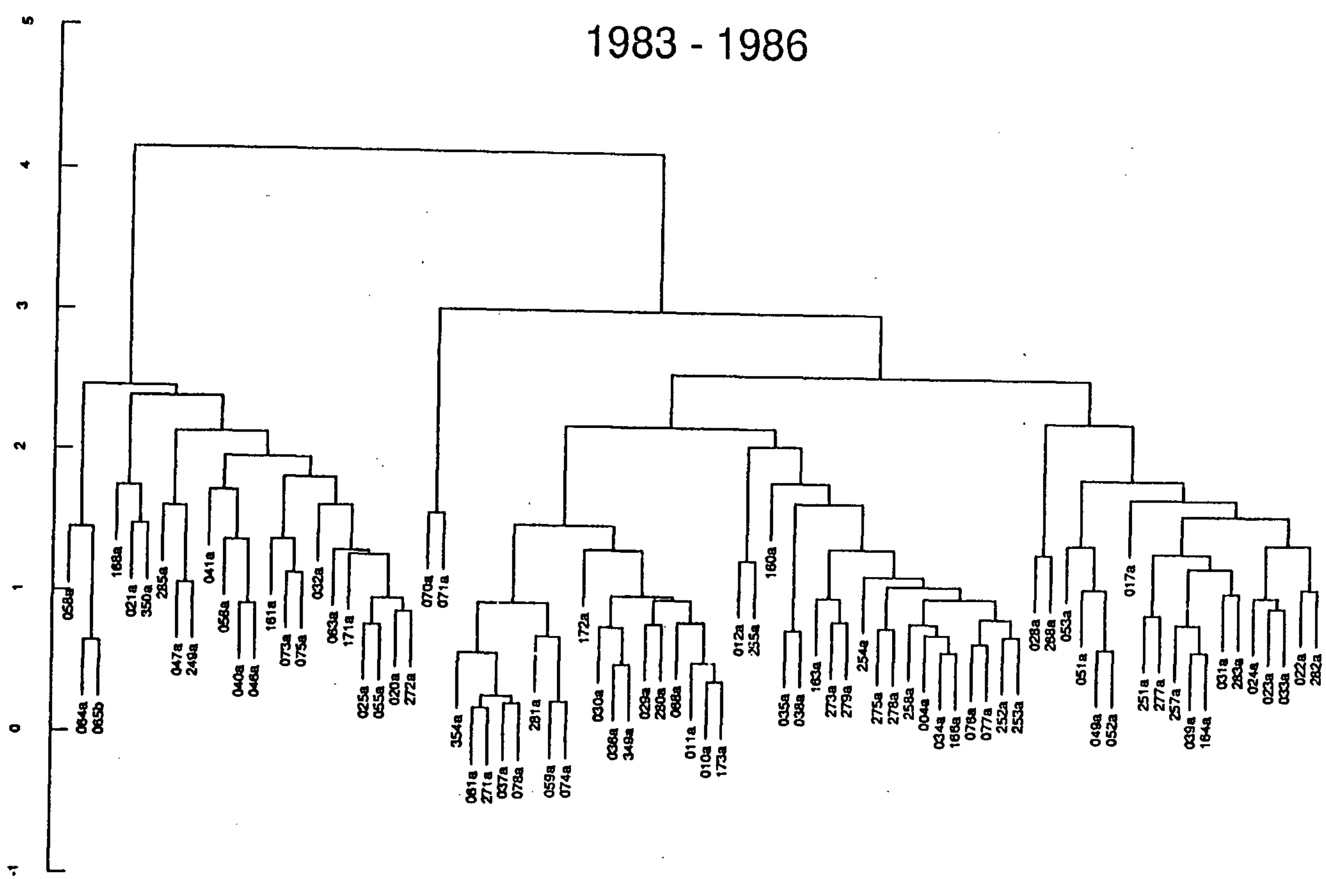


Figure 4.2.2(a)

Clustering of SO4 monthly volume weighted mean based on sqrt(MSE)

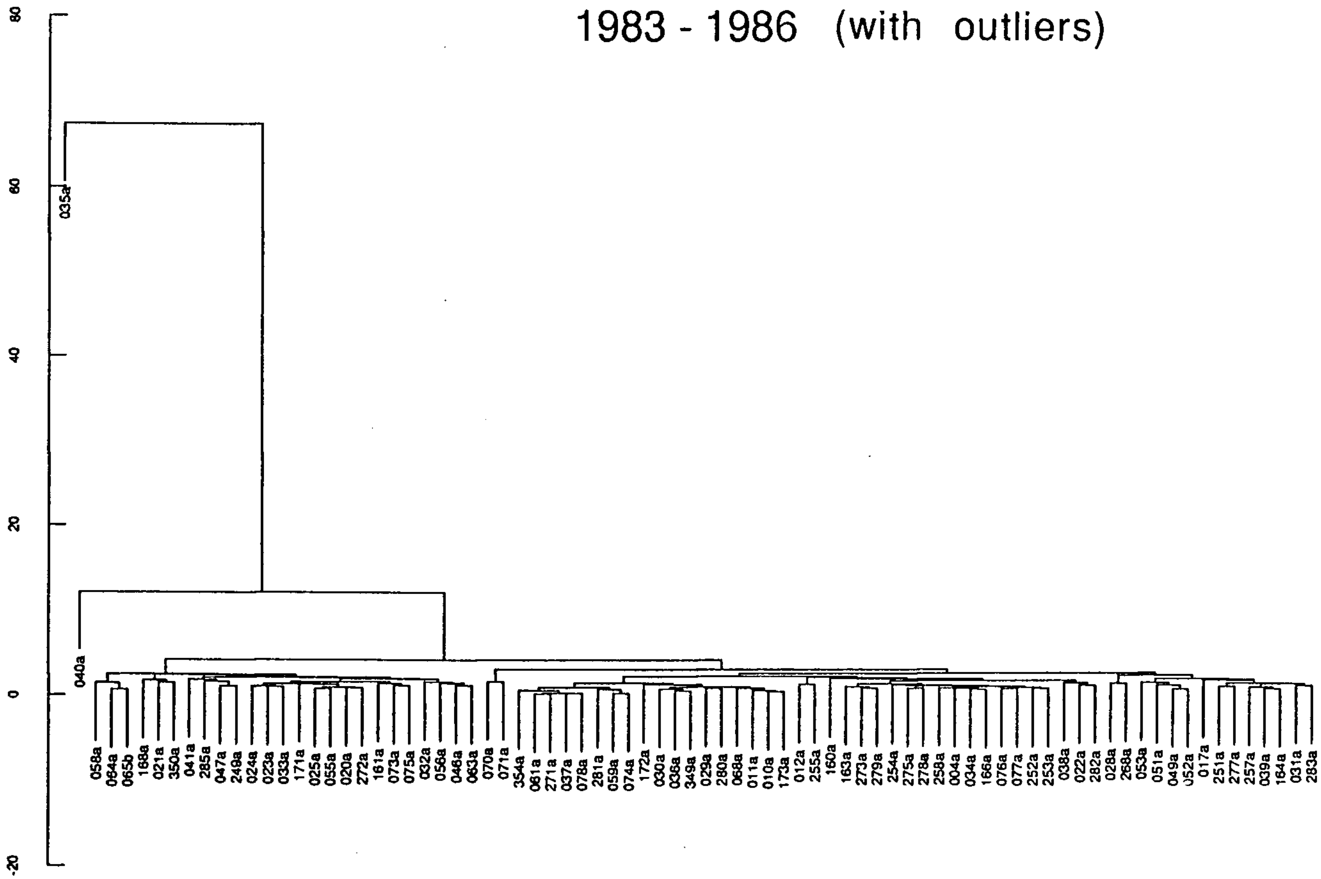


Figure 4.2.2(b)

Clusters of SO4 monthly volume weighted mean based on sqrt(MSE)

1983 - 1986 (k=3)

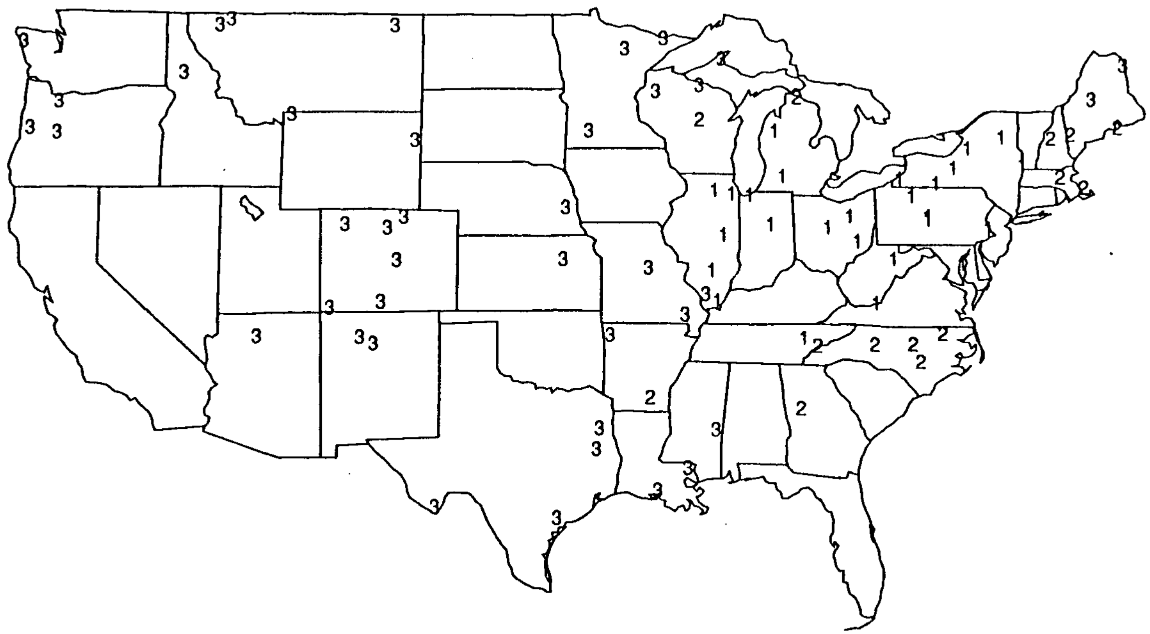


Figure 4.2.3

Histograms of Transformed SO4 (Volume Weighted Mean, 83-86)

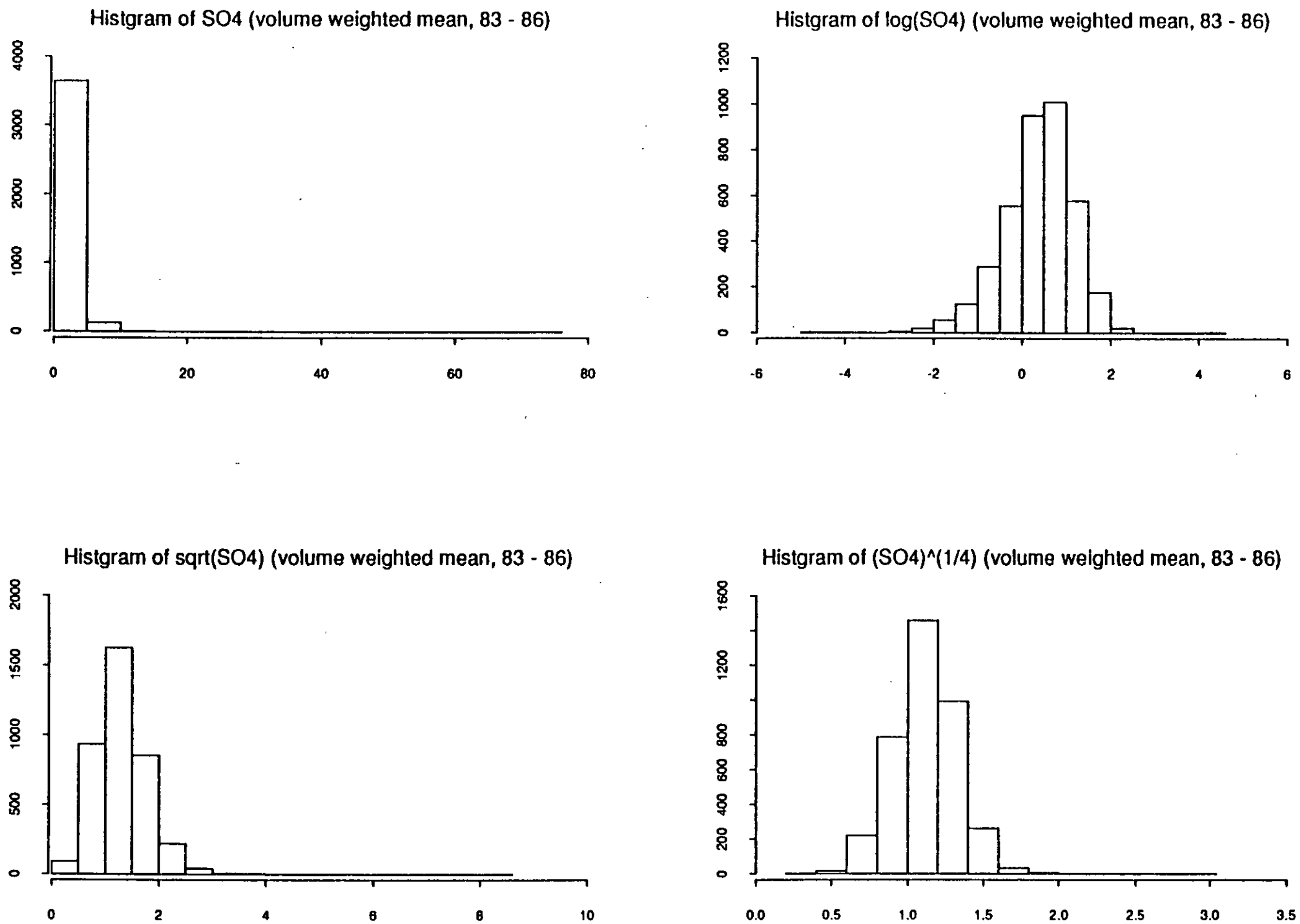


Figure 4.2.4

Histograms of log(SO4) by Clusters (Volume Weighted Mean, 83-86)

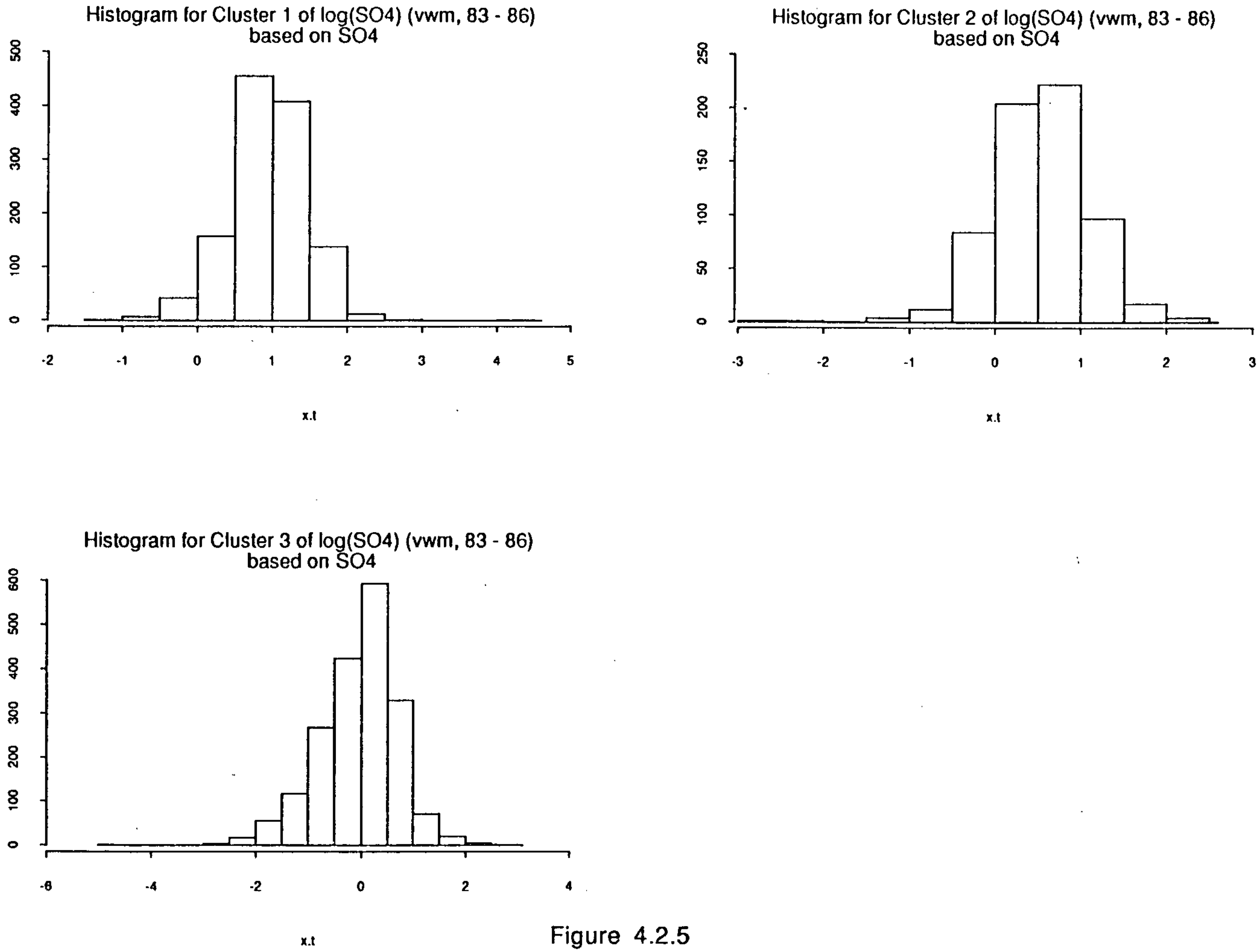
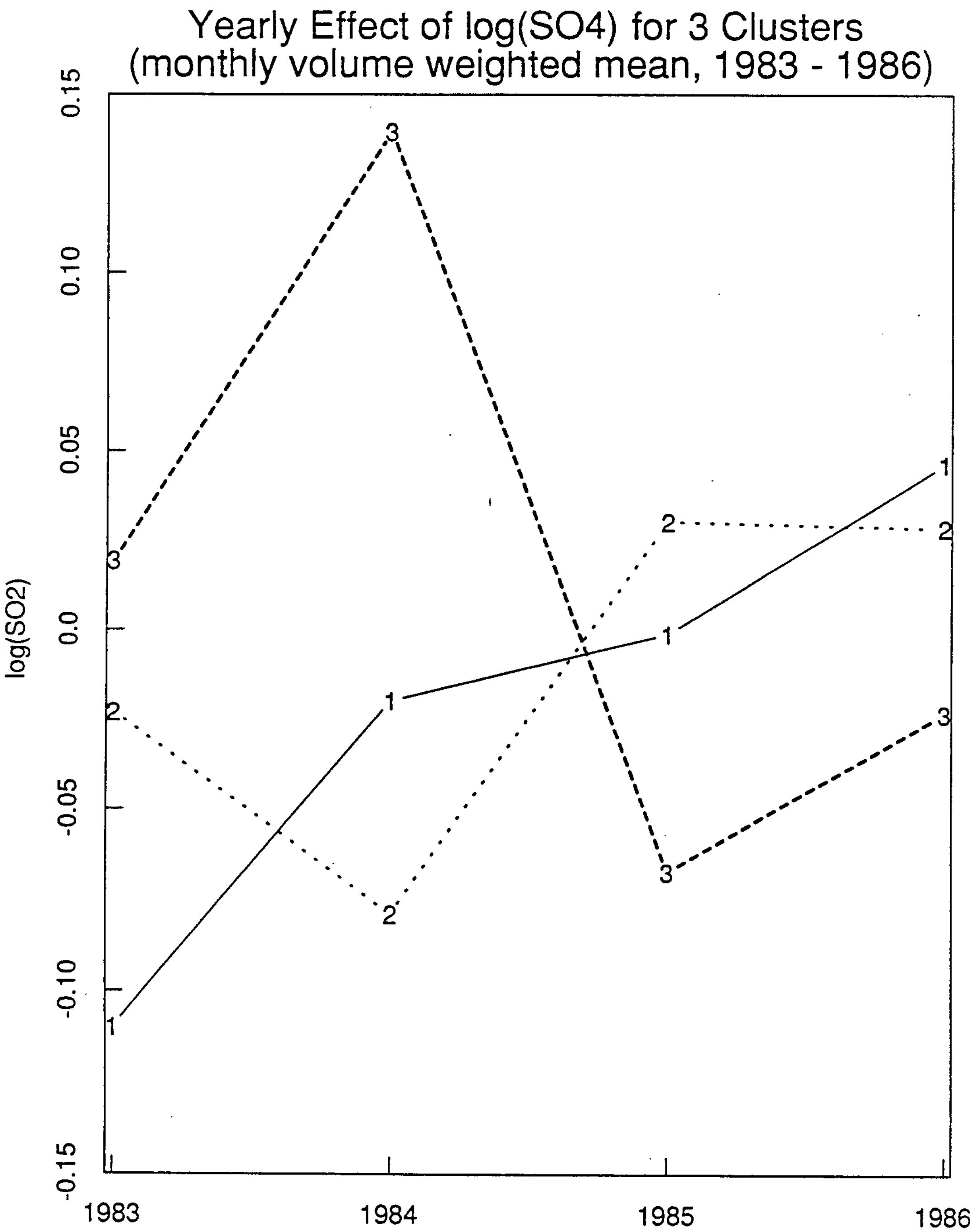
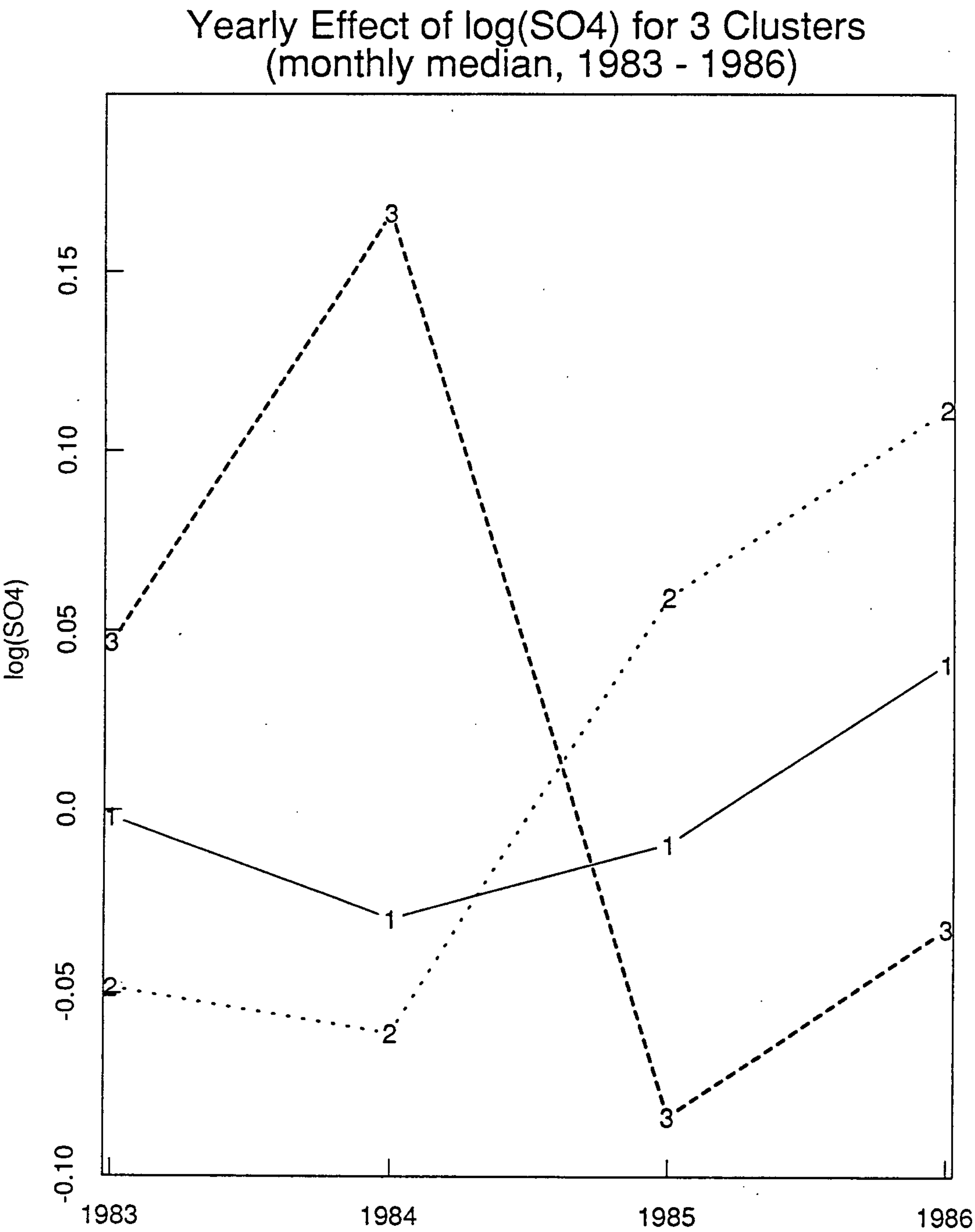


Figure 4.2.5



Yearly Effect of log(SO4) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)
log(SO2)
0.15
0.10
0.05
0.0
-0.05
-0.10
-0.15
1983
1984
1985
1986

Figure 4.2.6



Yearly Effect of log(SO4) for 3 Clusters
(monthly median, 1983 - 1986)
log(SO4)
0.15
0.10
0.05
0.0
-0.05
-0.10
1983
1984
1985
1986

Figure 4.2.7

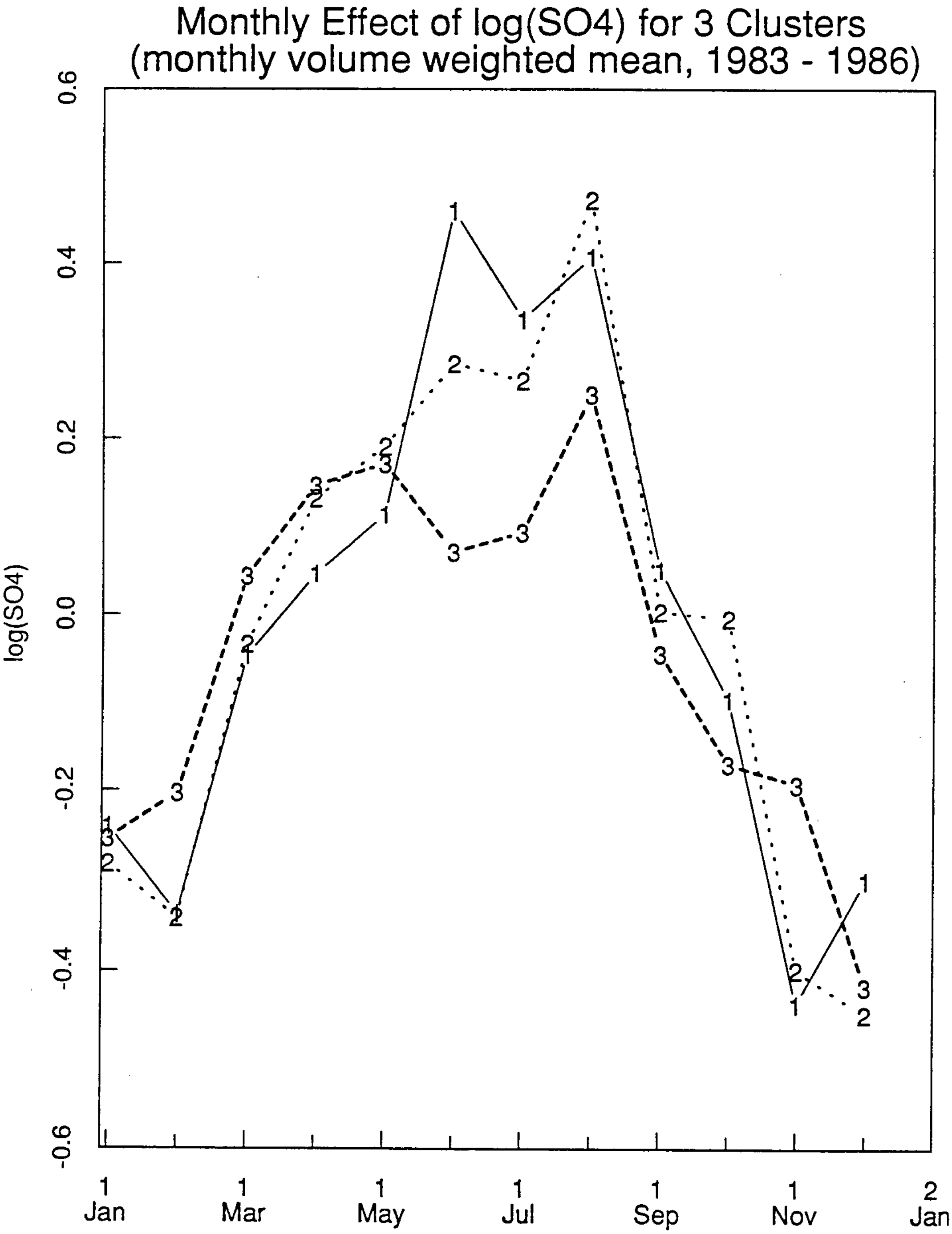
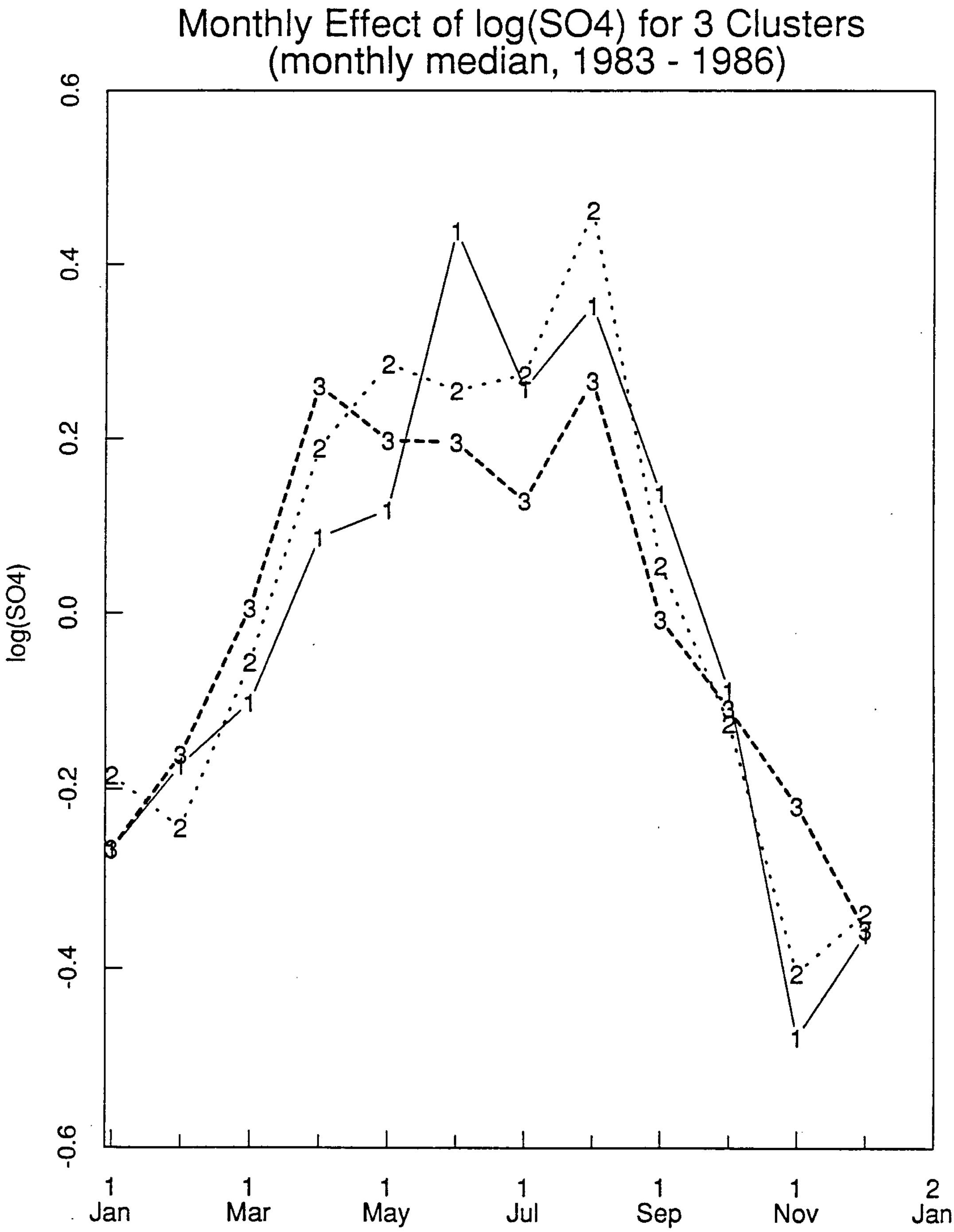

Monthly Effect of log(SO4) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)
log(SO4)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 4.2.8



Monthly Effect of log(SO4) for 3 Clusters
(monthly median, 1983 - 1986)
log(SO4)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 4.2.9

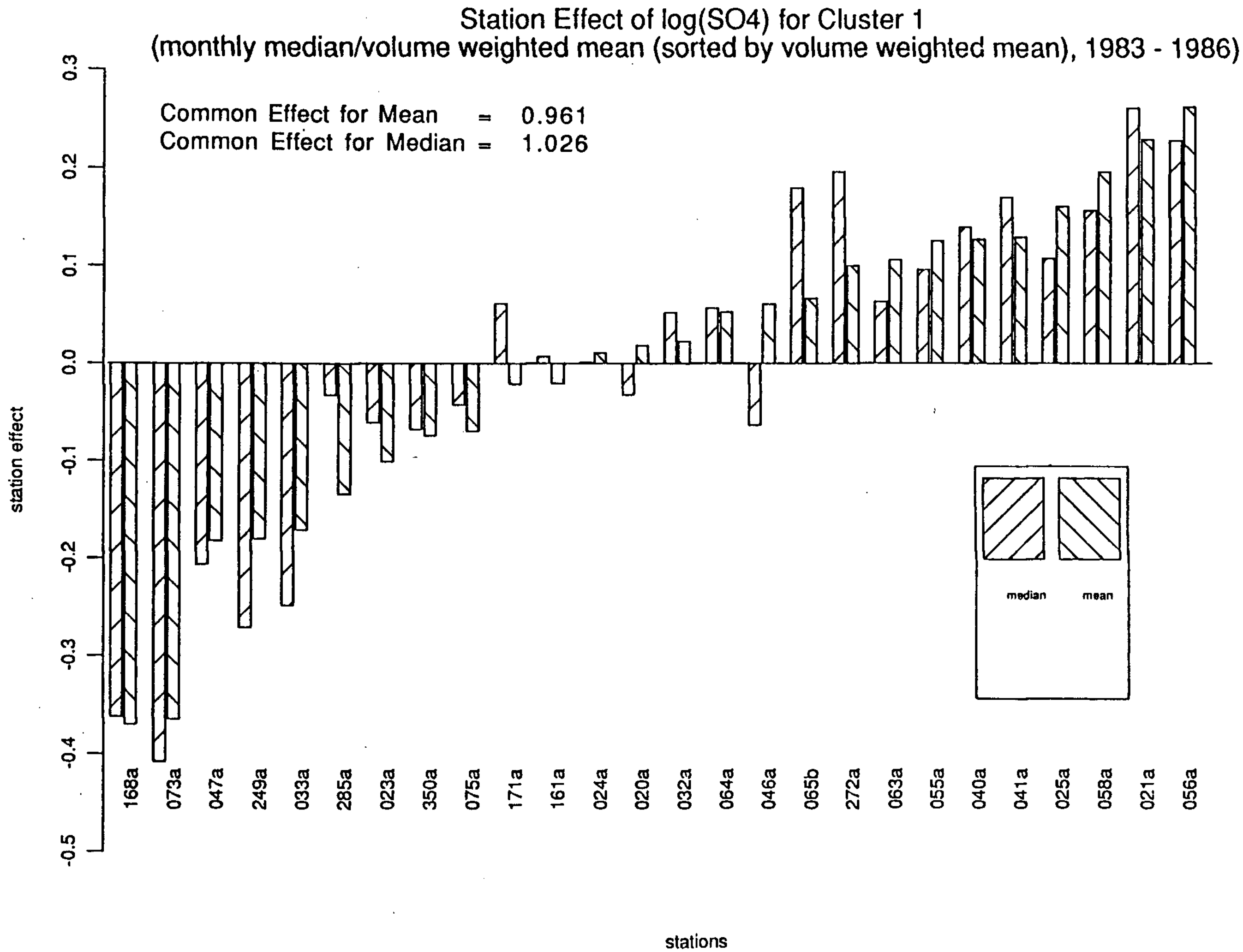

Station Effect of log(SO4) for Cluster 1
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 0.961
Common Effect for Median = 1.026
station effect
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
168a
073a
047a
249a
033a
285a
023a
350a
075a
171a
161a
024a
020a
032a
064a
046a
065b
272a
063a
055a
040a
041a
025a
058a
021a
056a
median
mean
stations

Figure 4.2.10

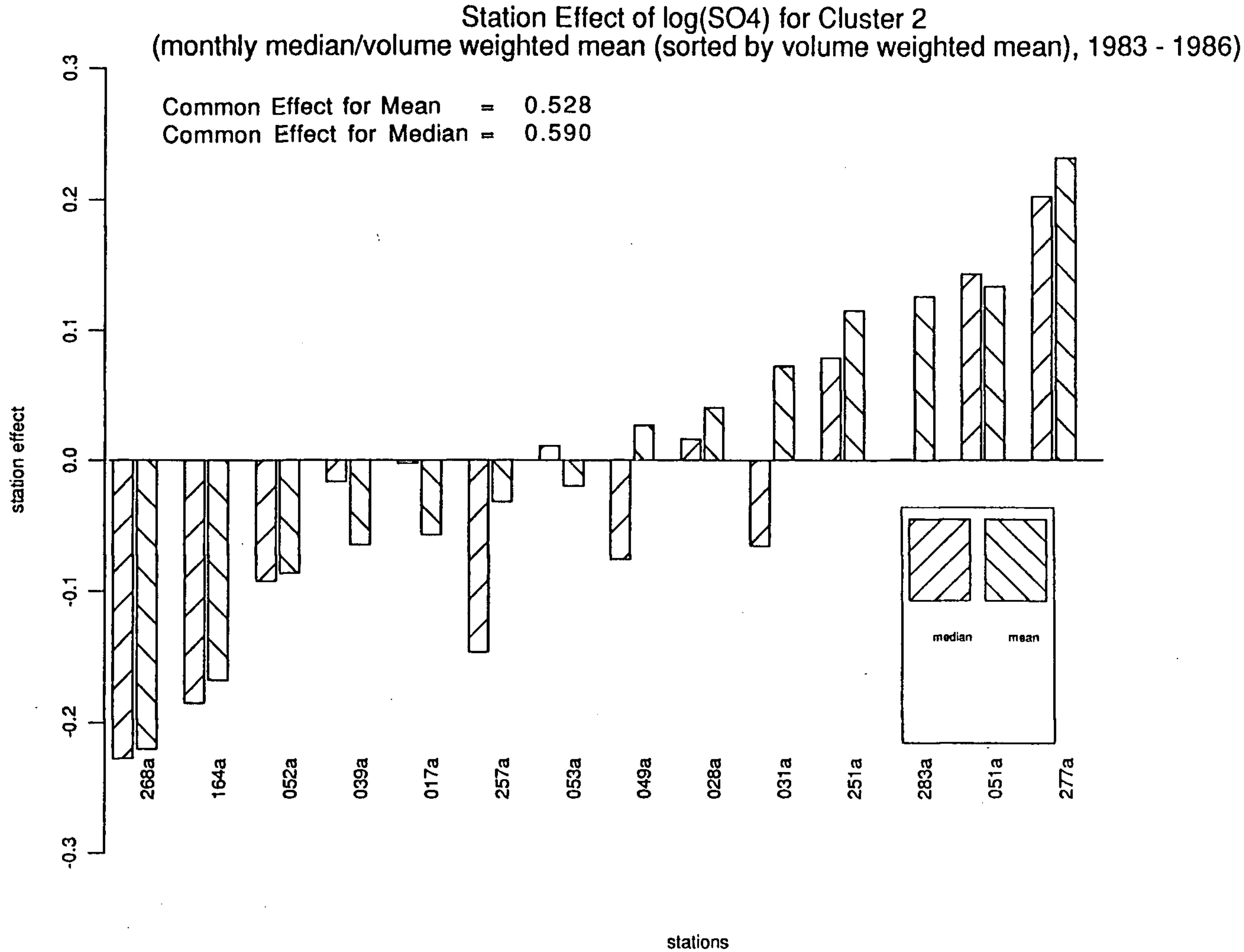

Station Effect of log(SO4) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 0.528
Common Effect for Median = 0.590
station effect
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
268a
164a
052a
039a
017a
257a
053a
049a
028a
031a
251a
283a
051a
277a
median
mean
stations

Figure 4.2.11

Station Effect of log(SO4) for Cluster 3
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)

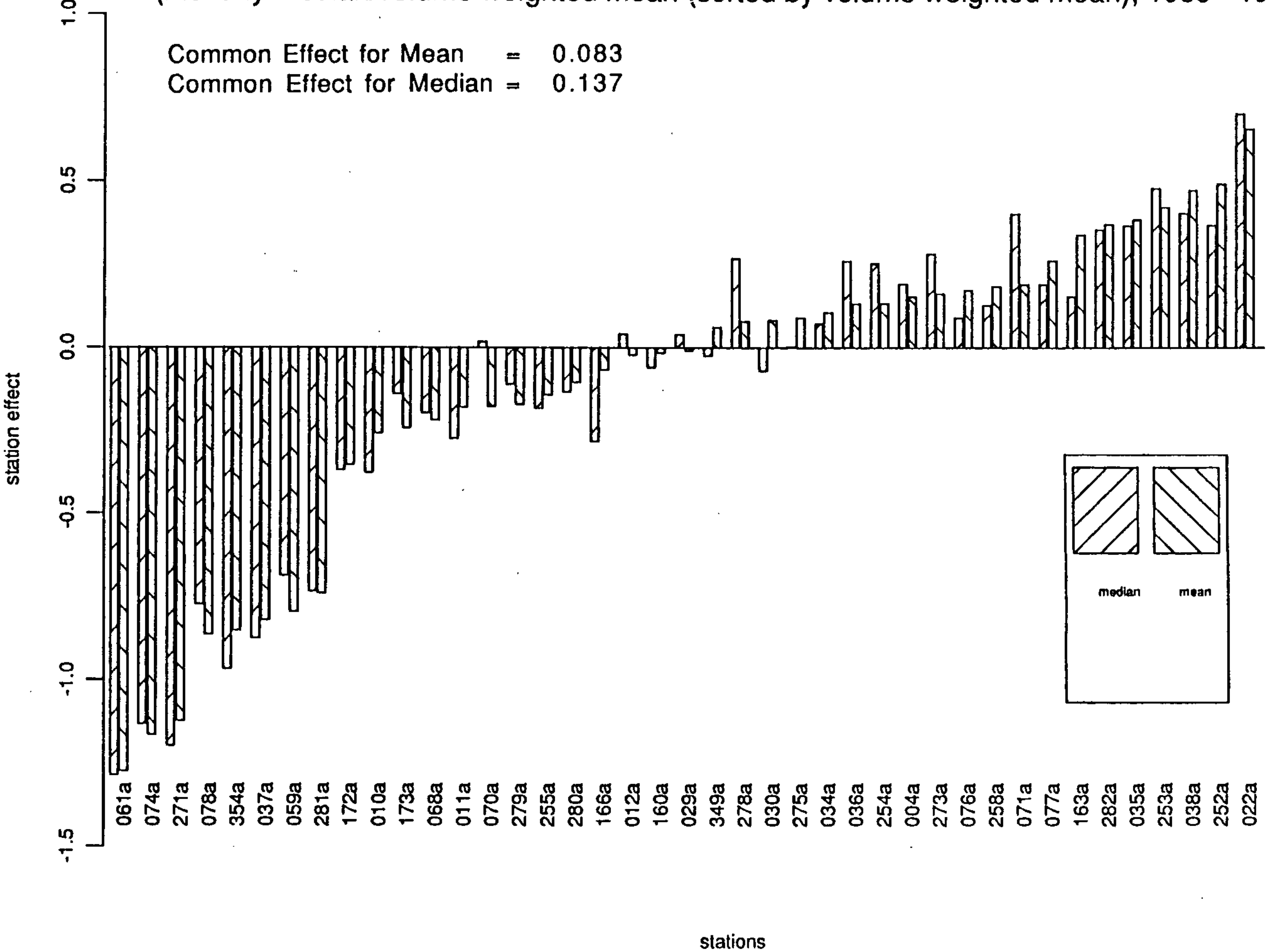


Figure 4.2.12

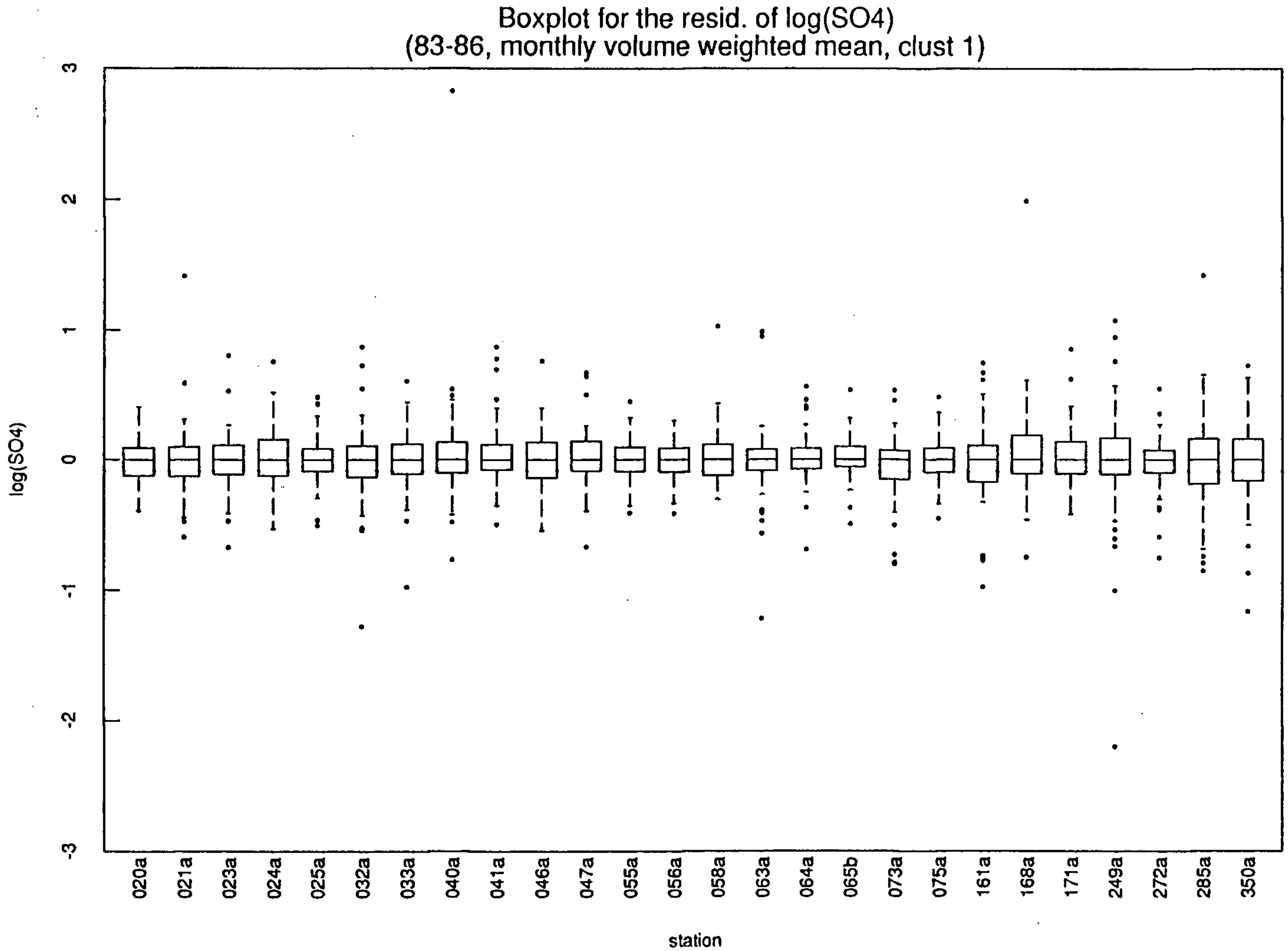


Figure 4.2.13(a)

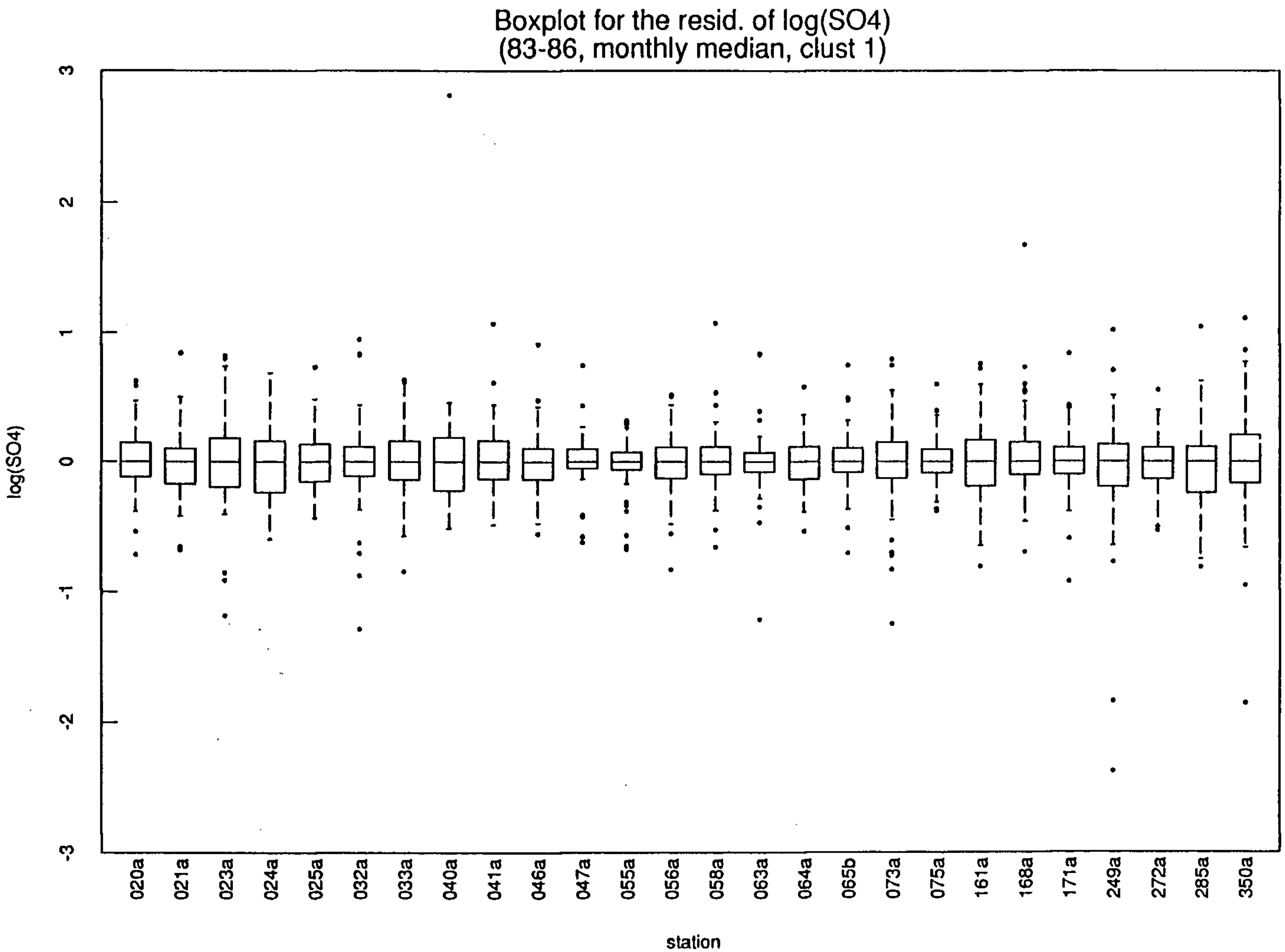


Figure 4.2.13(b)

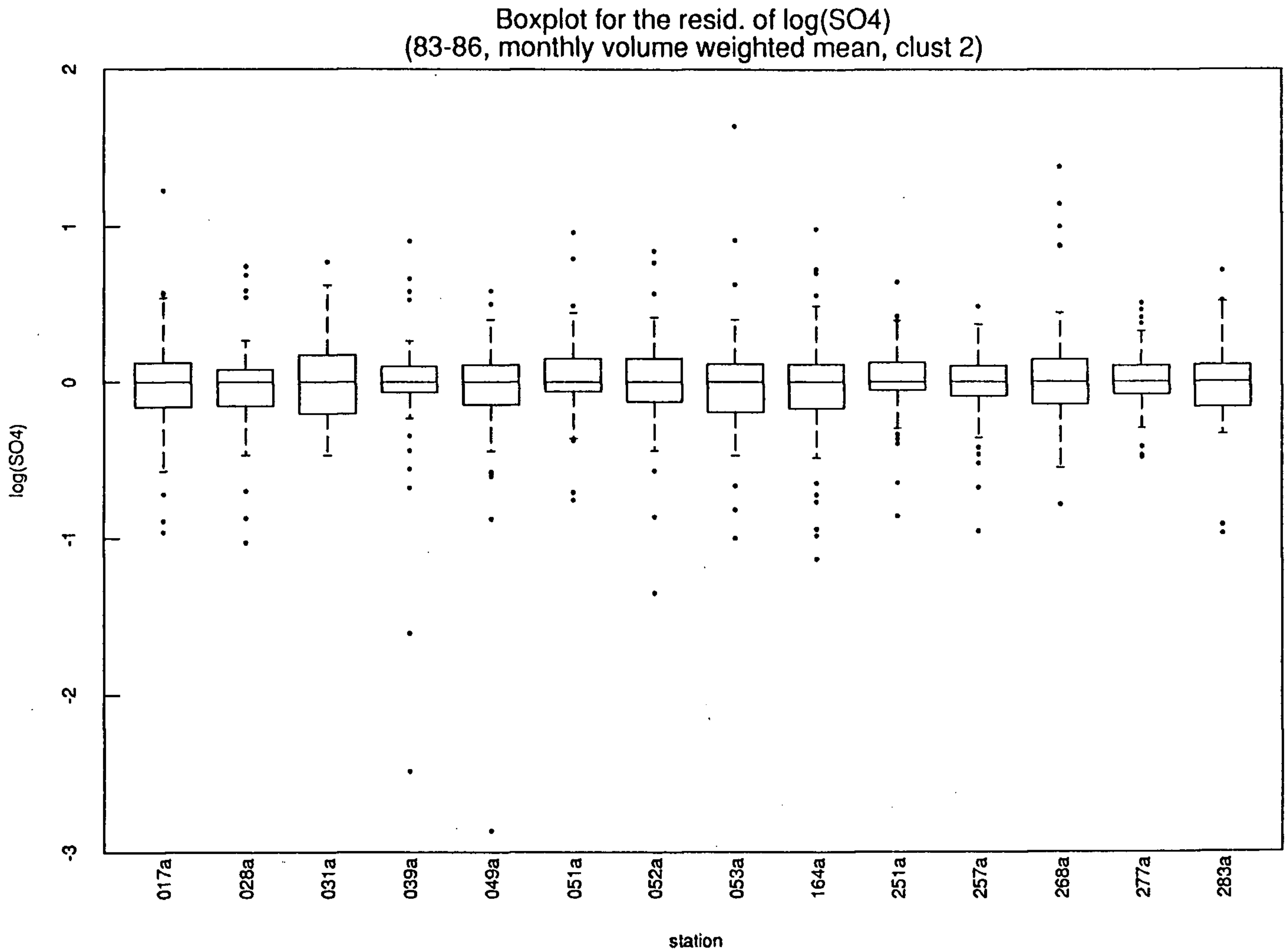


Figure 4.2.14(a)

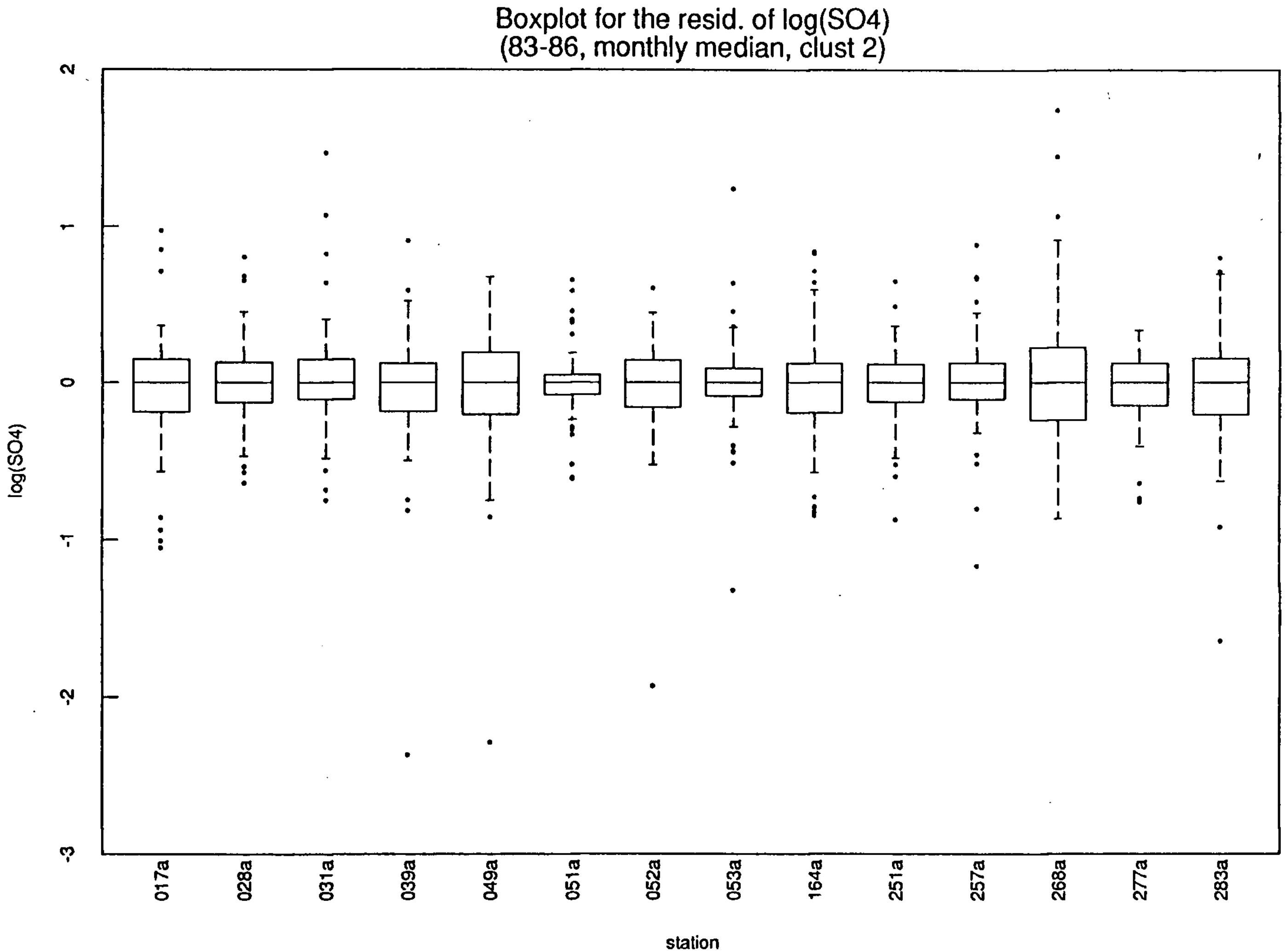


Figure 4.2.14(b)

Boxplot for the resid. of log(SO4)
(83-86, monthly volume weighted mean, clust 3)

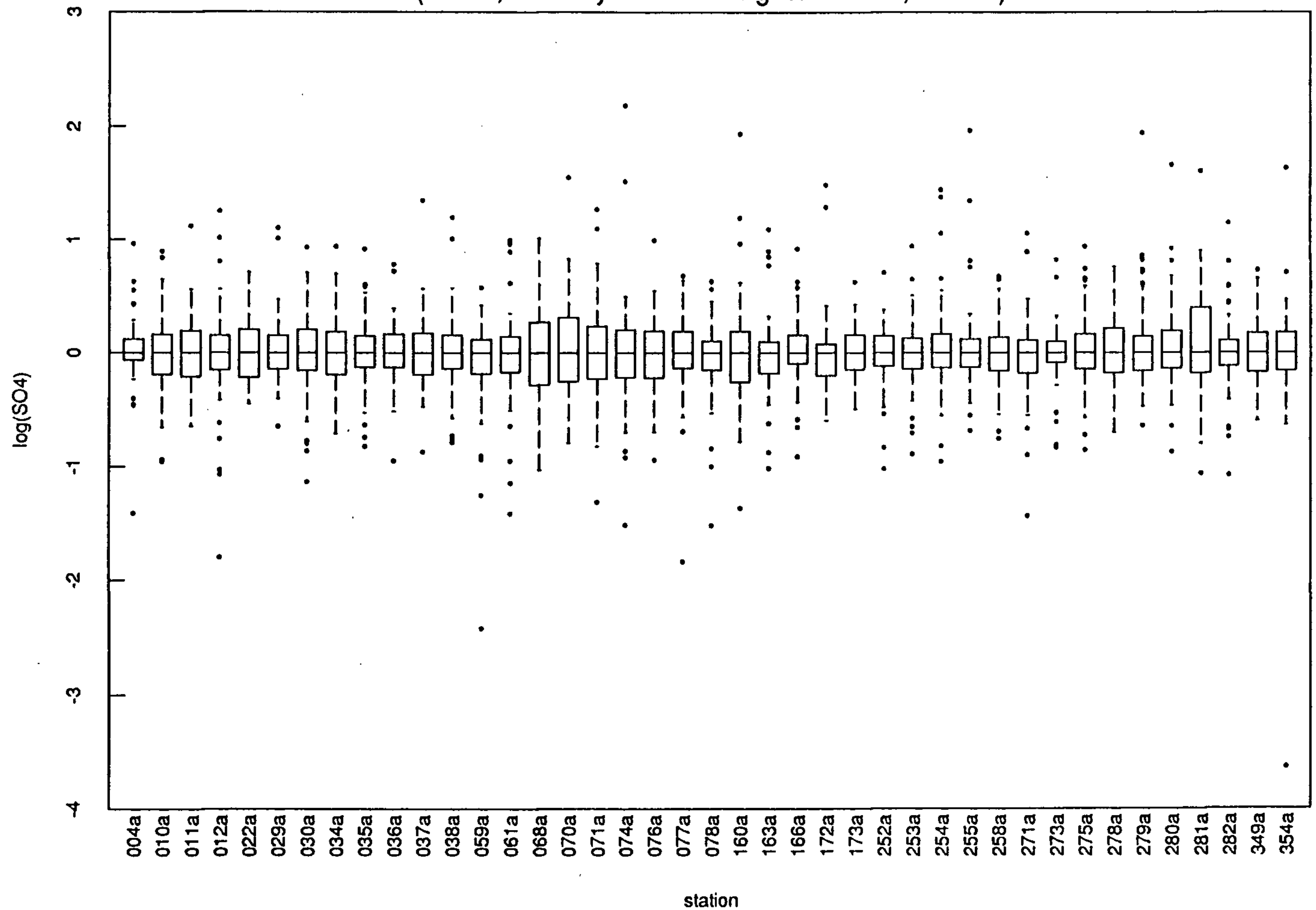


Figure 4.2.15(a)

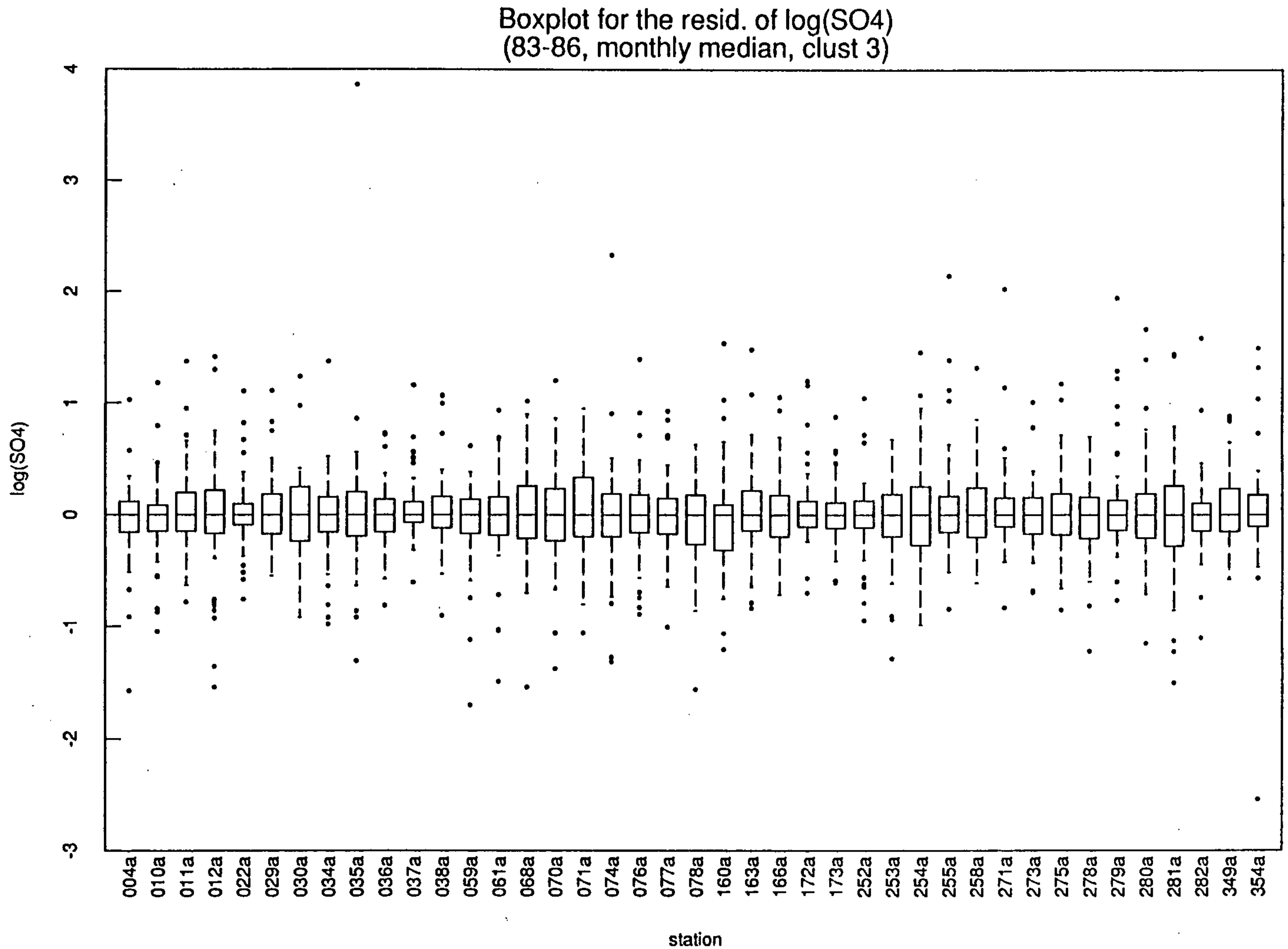


Figure 4.2.15(b)

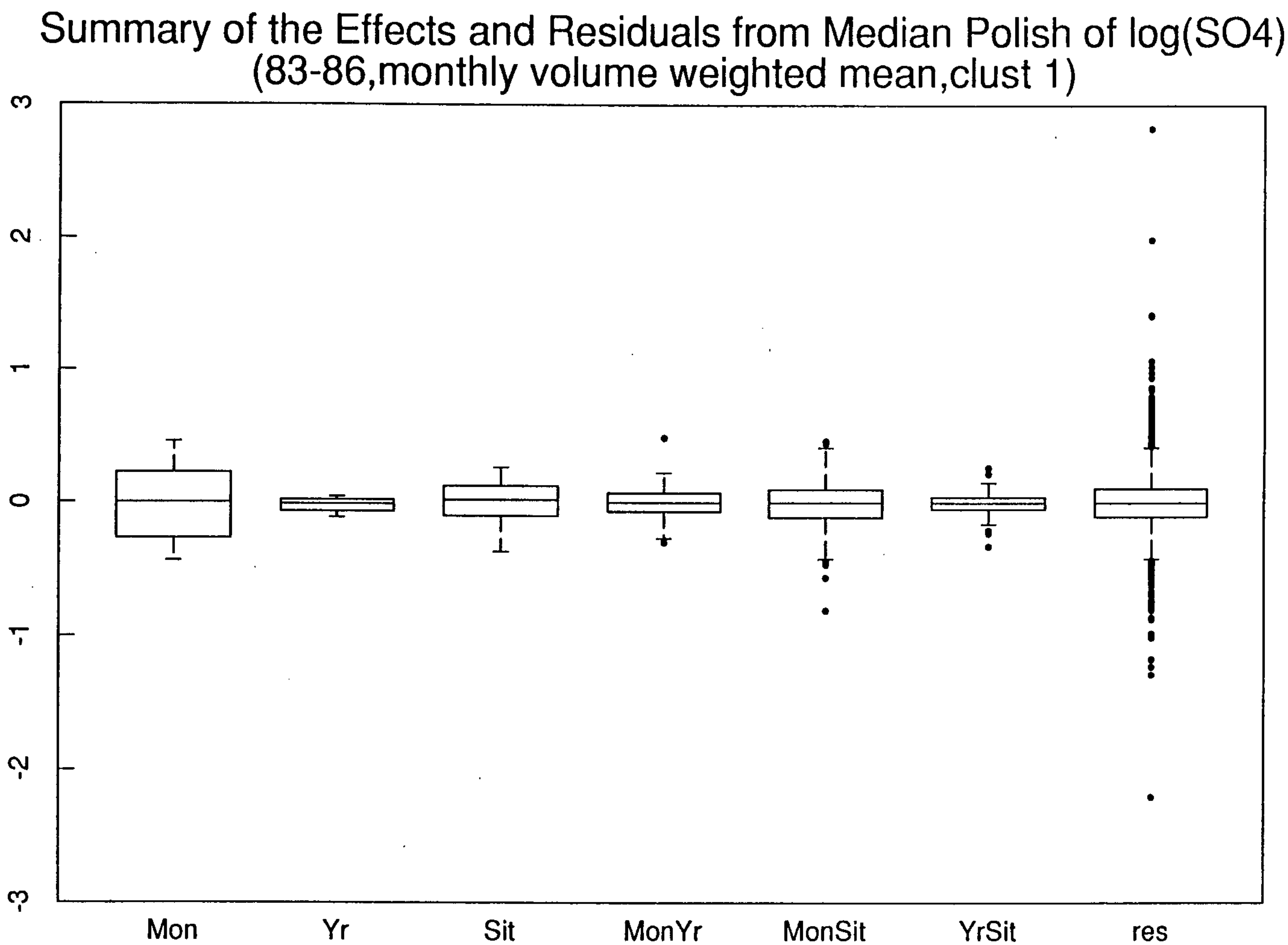


Figure 4.2.16(a)

Summary of the Effects and Residuals from Median Polish of log(SO4)
(83-86,monthly median,clust 1)

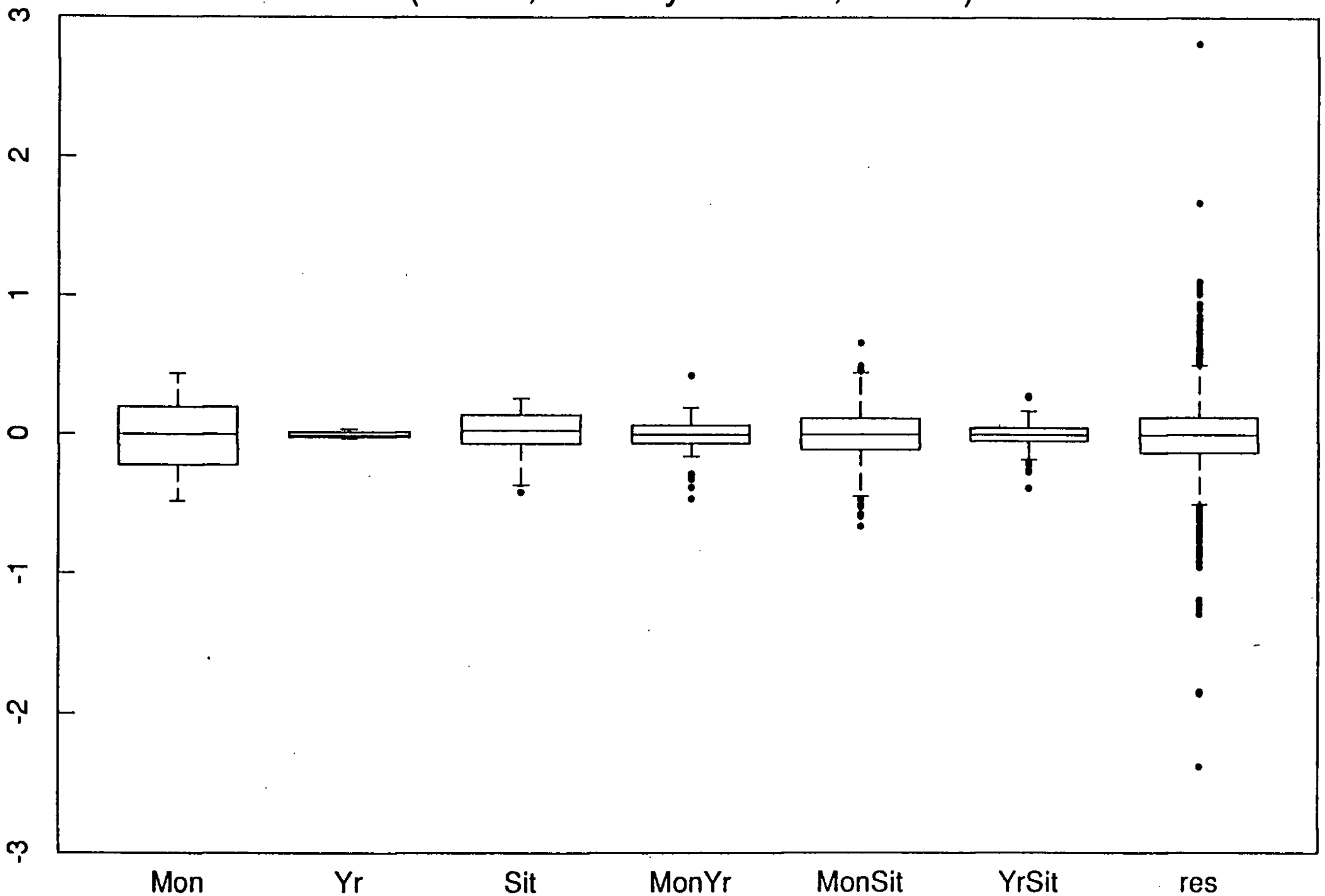


Figure 4.2.16(b)

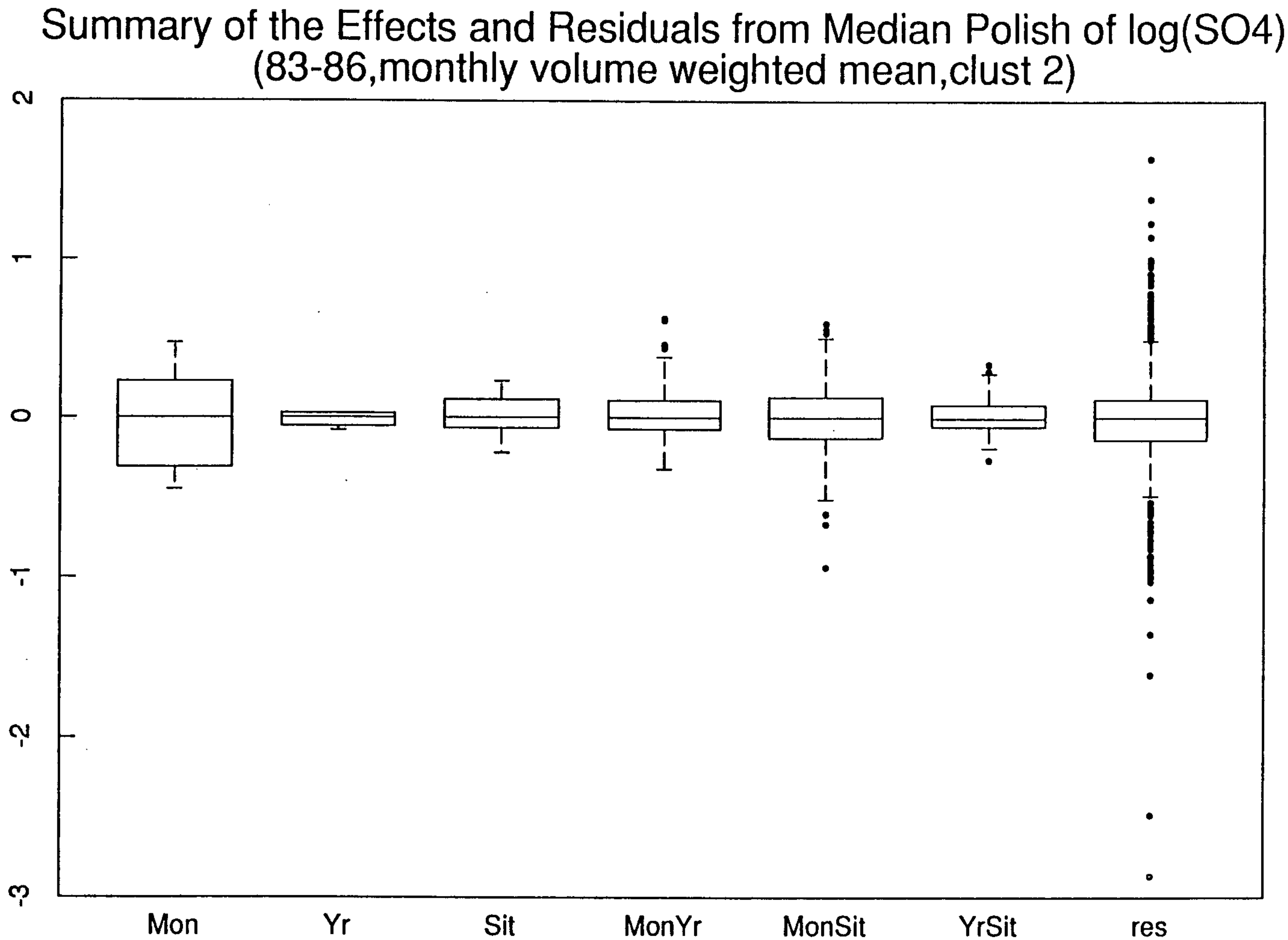


Figure 4.2.17(a)

Summary of the Effects and Residuals from Median Polish of log(SO4)
(83-86,monthly median,clust 2)

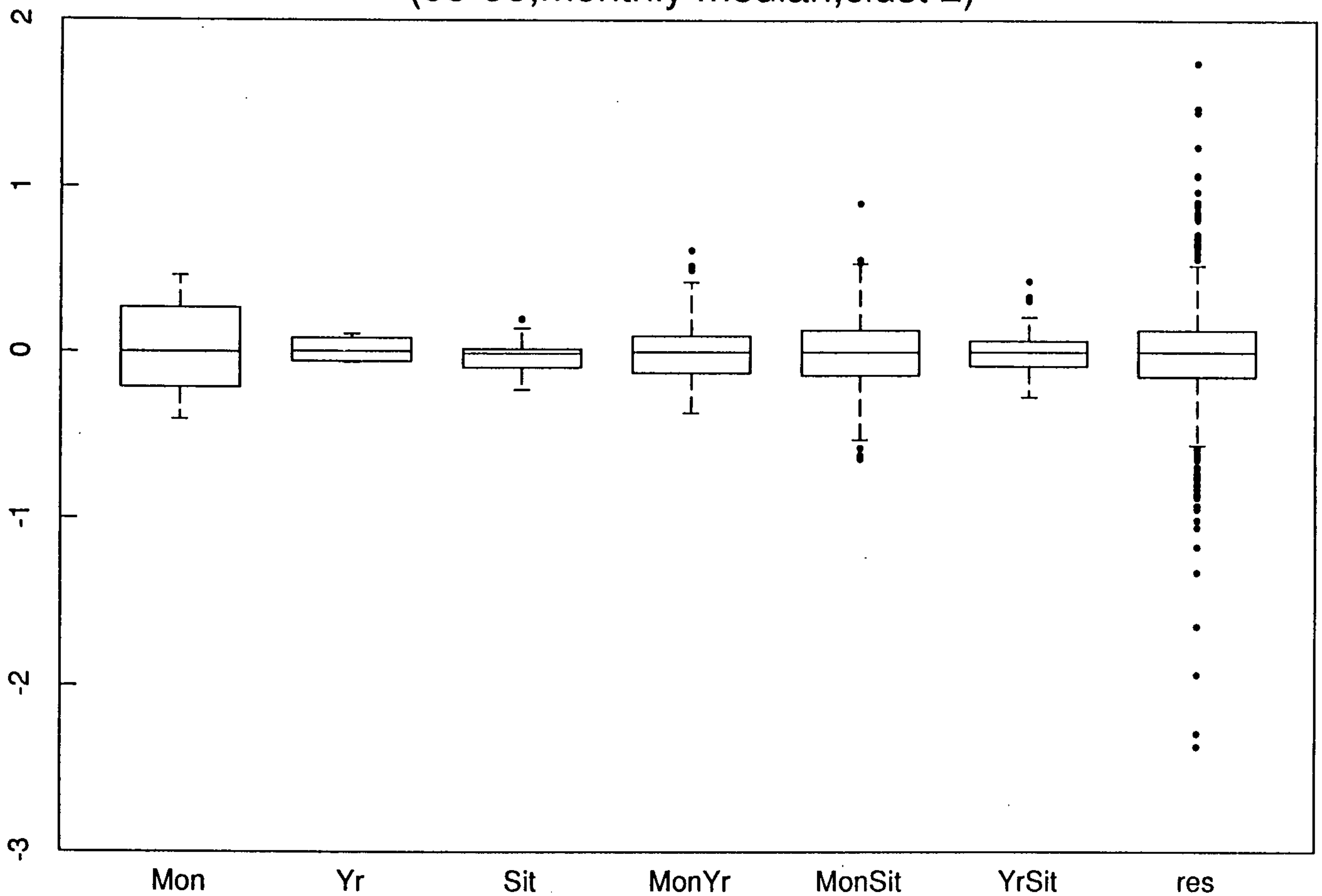


Figure 4.2.17(b)

Summary of the Effects and Residuals from Median Polish of log(SO4)
(83-86,monthly volume weighted mean,clust 3)

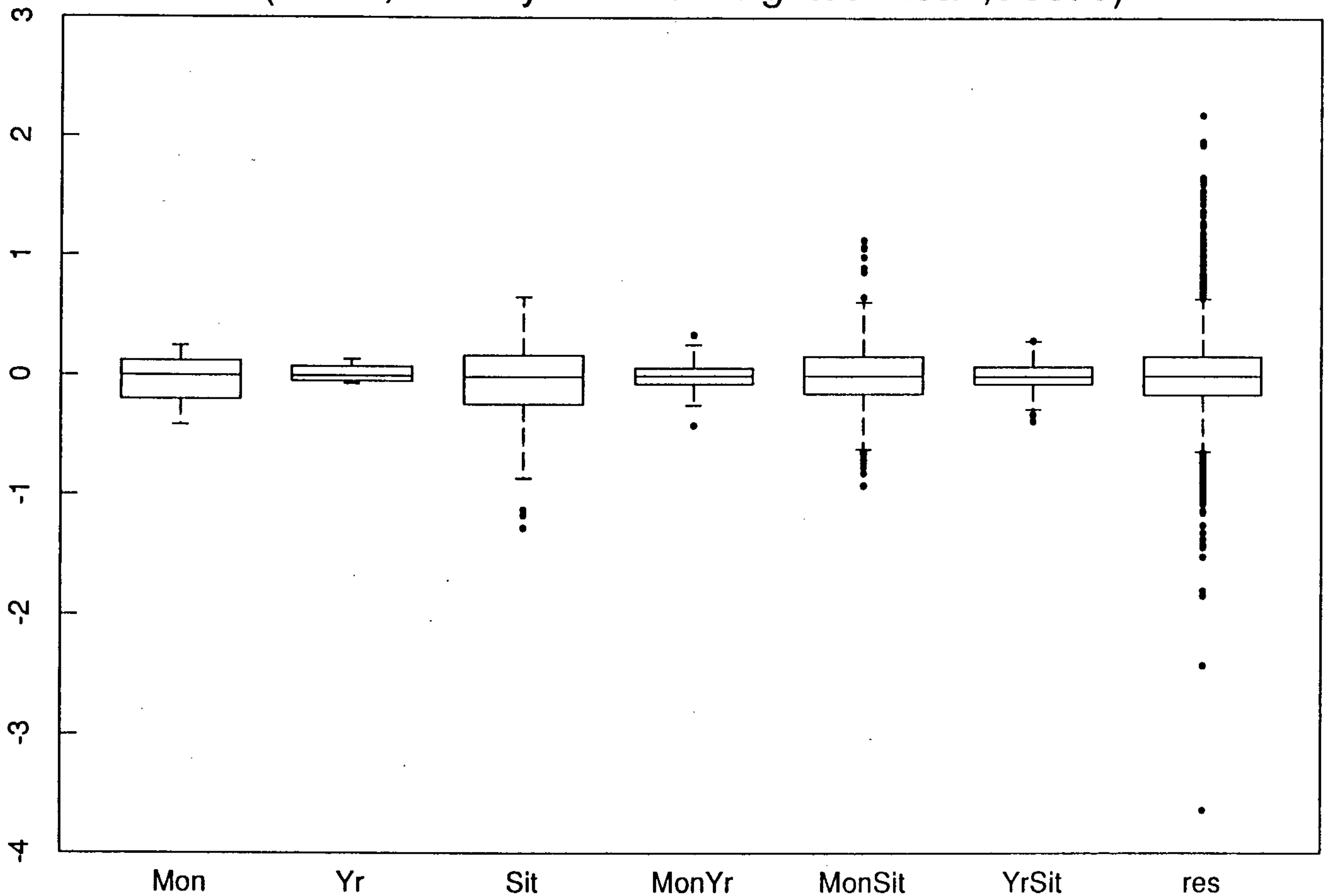


Figure 4.2.18(a)

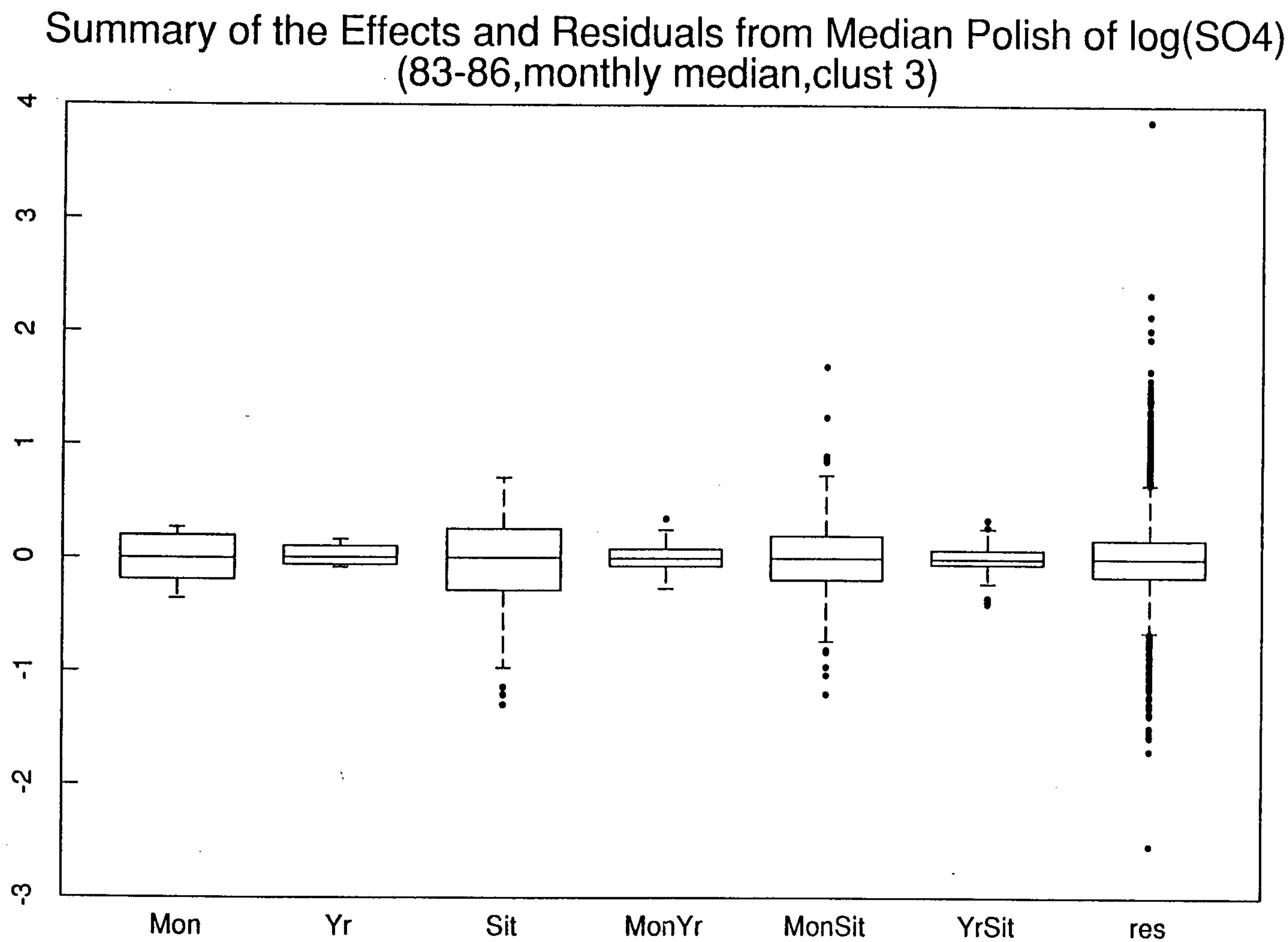


Figure 4.2.18(b)

Trend of log(SO4) from 1983 to 1986 at the 81 Stations
(calculated by monthly volume weighted mean)

Figure 4.2.19(a)

Trend of log(SO4) from 1983 to 1986 at the 81 Stations
(calculated by monthly median)

Figure 4.2.19(b)

Trend of log(SO4) from 1983 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

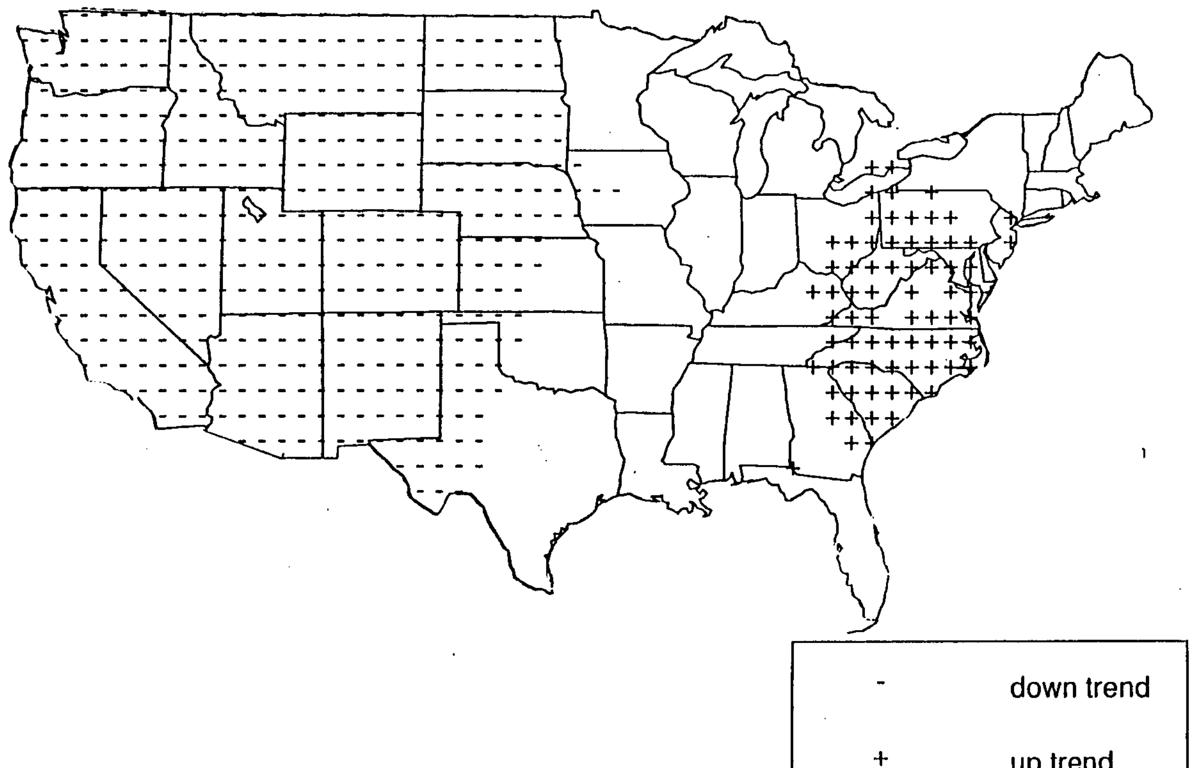


Figure 4.2.20(a)

Trend of log(SO4) from 1983 to 1986 in the USA
(calculated by Kriging from monthly median)

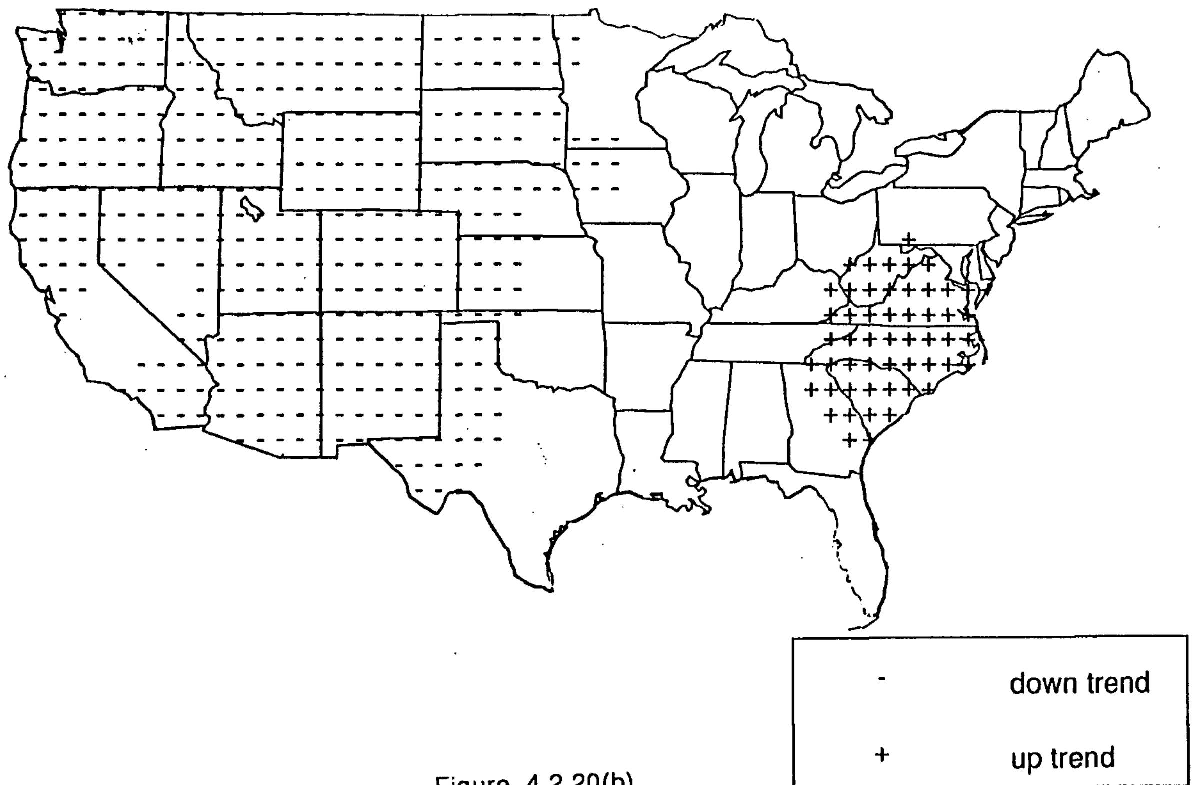


Figure 4.2.20(b)

TABLE 4.1.1

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES FOR THE CONCENTRATIONS OF SULPHATE (Monthly Volume Weighted Mean, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-3.047	0.002	-0.005	-0.007	-0.003
011a	-3.303	0.001	-0.007	-0.010	-0.004
017a	-0.750	0.453	-0.001	-0.004	0.001
020a	-2.516	0.012	-0.003	-0.004	-0.001
021a	-0.958	0.338	-0.001	-0.003	0.0
022a	-2.222	0.026	-0.003	-0.005	-0.001
023a	-2.391	0.017	-0.003	-0.005	-0.001
030a	-2.381	0.017	-0.004	-0.007	-0.002
031a	-2.728	0.006	-0.004	-0.006	-0.002
032a	-2.828	0.005	-0.003	-0.005	-0.002
034a	-2.104	0.035	-0.003	-0.005	-0.001
036a	-0.761	0.447	-0.001	-0.003	0.001
038a	-2.797	0.005	-0.005	-0.008	-0.003
039a	-2.906	0.004	-0.004	-0.006	-0.002
040a	-1.583	0.113	-0.002	-0.004	0.0
041a	-1.398	0.162	-0.002	-0.004	0.0
049a	-1.468	0.142	-0.002	-0.005	0.0
051a	-3.851	0.0	-0.007	-0.009	-0.004
052a	-2.048	0.041	-0.004	-0.006	-0.001
053a	-1.445	0.148	-0.003	-0.005	0.0
055a	-1.919	0.055	-0.002	-0.003	0.0
056a	-0.999	0.318	-0.001	-0.002	0.0
058a	0.213	0.831	0.0	-0.001	0.001
059a	-2.234	0.025	-0.003	-0.006	-0.001
064a	-1.141	0.254	-0.001	-0.003	0.0
074a	-1.758	0.079	-0.004	-0.007	-0.001
075a	-0.936	0.349	-0.001	-0.003	0.0
076a	-3.831	0.0	-0.006	-0.008	-0.004
168a	-2.610	0.009	-0.004	-0.006	-0.002
171a	-1.736	0.083	-0.002	-0.003	0.0
173a	-4.274	0.0	-0.009	-0.012	-0.007

TABLE 4.1.2

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF SULPHATE
(Monthly Median, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-2.445	0.014	-0.004	-0.006	-0.001
011a	-2.728	0.006	-0.006	-0.009	-0.003
017a	-1.148	0.251	-0.002	-0.004	0.0
020a	-2.107	0.035	-0.003	-0.005	-0.001
021a	-0.541	0.588	-0.001	-0.002	0.001
022a	-3.099	0.002	-0.005	-0.007	-0.003
023a	-2.000	0.045	-0.004	-0.007	-0.001
030a	-2.478	0.013	-0.004	-0.006	-0.002
031a	-3.286	0.001	-0.006	-0.008	-0.003
032a	-1.957	0.050	-0.002	-0.004	-0.001
034a	-2.420	0.016	-0.004	-0.006	-0.002
036a	-0.715	0.475	-0.001	-0.004	0.001
038a	-2.982	0.003	-0.005	-0.009	-0.003
039a	-0.646	0.518	0.0	-0.003	0.001
040a	-1.857	0.063	-0.003	-0.005	0.0
041a	-0.639	0.523	-0.001	-0.002	0.001
049a	-3.494	0.0	-0.006	-0.009	-0.004
051a	-4.306	0.0	-0.007	-0.009	-0.005
052a	-1.447	0.148	-0.003	-0.005	0.0
053a	-1.070	0.285	-0.002	-0.005	0.0
055a	-1.304	0.192	-0.001	-0.003	0.0
056a	-0.840	0.401	-0.001	-0.003	0.0
058a	-0.786	0.432	-0.001	-0.002	0.0
059a	-1.498	0.134	-0.003	-0.005	0.0
064a	-2.240	0.025	-0.002	-0.004	-0.001
074a	-2.479	0.013	-0.005	-0.009	-0.002
075a	-0.387	0.699	0.0	-0.002	0.001
076a	-3.608	0.0	-0.006	-0.008	-0.004
168a	-2.601	0.009	-0.004	-0.006	-0.002
171a	-1.639	0.101	-0.002	-0.004	0.0
173a	-3.443	0.001	-0.007	-0.010	-0.005

TABLE 4.2.1

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF SULPHATE
(Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	U-bd
004a	-3.045	0.002	-0.009	-0.013	-0.006
010a	-1.247	0.212	-0.006	-0.012	0.0
011a	-1.573	0.116	-0.009	-0.014	-0.001
012a	-0.460	0.646	-0.002	-0.011	0.003
017a	2.137	0.033	0.007	0.002	0.012
020a	0.605	0.545	0.001	-0.002	0.005
021a	1.200	0.230	0.003	-0.001	0.007
022a	0.098	0.922	0.001	-0.004	0.004
023a	1.164	0.244	0.003	0.0	0.007
024a	-1.040	0.298	-0.003	-0.006	0.001
025a	0.524	0.600	0.001	-0.002	0.004
028a	1.654	0.098	0.004	0.001	0.010
029a	-2.053	0.040	-0.006	-0.011	-0.002
030a	0.624	0.533	0.002	-0.003	0.007
031a	-0.752	0.452	-0.003	-0.008	0.003
032a	0.028	0.978	0.0	-0.004	0.003
033a	-0.498	0.619	-0.002	-0.007	0.002
034a	-2.102	0.036	-0.007	-0.011	-0.002
035a	-1.550	0.121	-0.010	-0.014	-0.002
036a	1.831	0.067	0.004	0.002	0.008
037a	-1.723	0.085	-0.007	-0.012	-0.002
038a	-0.138	0.891	0.0	-0.006	0.004
039a	0.169	0.866	0.001	-0.004	0.006
040a	1.395	0.163	0.004	0.0	0.007
041a	1.040	0.298	0.003	-0.001	0.007
046a	0.667	0.505	0.002	-0.002	0.006
047a	1.173	0.241	0.004	0.0	0.007
049a	-0.160	0.873	-0.001	-0.006	0.005
051a	0.660	0.509	0.002	-0.002	0.007
052a	-0.786	0.432	-0.003	-0.008	0.002
053a	1.733	0.083	0.006	0.002	0.010
055a	1.164	0.244	0.002	0.0	0.005
056a	1.752	0.080	0.005	0.001	0.008
058a	1.786	0.074	0.005	0.002	0.008
059a	-1.247	0.212	-0.004	-0.008	0.0
061a	-1.810	0.070	-0.009	-0.014	-0.003
063a	1.458	0.145	0.003	0.0	0.005
064a	1.067	0.286	0.003	-0.001	0.006
065b	1.194	0.232	0.003	0.0	0.006
068a	-1.032	0.302	-0.007	-0.014	0.002
070a	-0.492	0.623	-0.002	-0.008	0.004
071a	-0.667	0.505	-0.003	-0.008	0.003
073a	2.055	0.040	0.007	0.002	0.012

TABLE 4.2.1 (continued)

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF SULPHATE
(Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	-1.311	0.190	-0.007	-0.013	0.0
075a	4.035	0.0	0.012	0.008	0.016
076a	-1.211	0.226	-0.003	-0.008	0.0
077a	0.0	1.000	0.0	-0.005	0.004
078a	-3.025	0.002	-0.011	-0.015	-0.007
160a	-3.278	0.001	-0.013	-0.019	-0.008
161a	0.116	0.908	0.001	-0.004	0.005
163a	0.038	0.970	0.0	-0.004	0.005
164a	-0.956	0.339	-0.004	-0.009	0.001
166a	-1.229	0.219	-0.004	-0.008	0.0
168a	0.294	0.769	0.001	-0.003	0.004
171a	1.591	0.112	0.004	0.001	0.008
172a	-2.994	0.003	-0.011	-0.016	-0.006
173a	-2.427	0.015	-0.008	-0.012	-0.004
249a	1.644	0.100	0.007	0.001	0.015
251a	1.360	0.174	0.005	0.0	0.011
252a	-2.471	0.013	-0.008	-0.010	-0.004
253a	0.293	0.769	0.001	-0.003	0.004
254a	-1.468	0.142	-0.005	-0.010	-0.001
255a	-2.578	0.010	-0.007	-0.012	-0.004
257a	0.415	0.678	0.001	-0.003	0.008
258a	0.0	1.000	0.0	-0.004	0.004
268a	1.244	0.213	0.003	0.0	0.009
271a	-1.477	0.140	-0.005	-0.009	0.0
272a	0.436	0.663	0.001	-0.001	0.003
273a	-0.924	0.355	-0.002	-0.006	0.001
275a	0.836	0.403	0.003	-0.002	0.007
277a	2.384	0.017	0.008	0.004	0.011
278a	-2.641	0.008	-0.008	-0.014	-0.004
279a	-1.324	0.185	-0.007	-0.011	0.0
280a	-1.751	0.080	-0.006	-0.010	-0.002
281a	1.816	0.069	0.010	0.002	0.018
282a	1.635	0.102	0.005	0.001	0.009
283a	-1.577	0.115	-0.004	-0.008	-0.001
285a	1.401	0.161	0.006	0.0	0.013
349a	1.420	0.156	0.003	0.0	0.007
350a	2.113	0.035	0.010	0.004	0.016
354a	-3.539	0.0	-0.018	-0.025	-0.012

TABLE 4.2.2

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF SULPHATE
(Monthly Medain, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	U-bd
004a	-2.531	0.011	-0.008	-0.012	-0.004
010a	-1.385	0.166	-0.004	-0.011	0.0
011a	-1.209	0.227	-0.007	-0.014	0.0
012a	-0.029	0.977	0.0	-0.008	0.005
017a	1.865	0.062	0.007	0.001	0.012
020a	0.037	0.971	0.0	-0.003	0.003
021a	0.684	0.494	0.002	-0.002	0.006
022a	-0.836	0.403	-0.003	-0.008	0.001
023a	0.933	0.351	0.004	-0.002	0.009
024a	-1.173	0.241	-0.004	-0.008	0.0
025a	-0.133	0.894	0.0	-0.005	0.003
028a	0.333	0.739	0.001	-0.004	0.007
029a	-1.680	0.093	-0.008	-0.013	-0.002
030a	0.349	0.727	0.002	-0.003	0.007
031a	-1.082	0.279	-0.005	-0.011	0.001
032a	-1.082	0.279	-0.004	-0.011	0.001
033a	-0.560	0.576	-0.002	-0.006	0.002
034a	-2.509	0.012	-0.009	-0.013	-0.004
035a	-1.246	0.213	-0.006	-0.013	0.0
036a	1.262	0.207	0.004	0.0	0.007
037a	-2.160	0.031	-0.005	-0.010	-0.002
038a	0.147	0.883	0.0	-0.003	0.004
039a	1.076	0.282	0.005	0.0	0.010
040a	0.356	0.722	0.001	-0.003	0.005
041a	1.111	0.267	0.003	0.0	0.005
046a	-0.062	0.950	0.0	-0.005	0.004
047a	0.829	0.407	0.003	-0.001	0.007
049a	0.320	0.749	0.002	-0.005	0.007
051a	0.862	0.389	0.003	-0.002	0.007
052a	0.985	0.325	0.004	-0.001	0.009
053a	1.487	0.137	0.006	0.001	0.012
055a	1.644	0.100	0.004	0.001	0.007
056a	1.235	0.217	0.004	0.0	0.008
058a	2.507	0.012	0.005	0.002	0.008
059a	-0.541	0.588	-0.002	-0.006	0.003
061a	-1.576	0.115	-0.006	-0.012	-0.001
063a	0.468	0.640	0.001	-0.003	0.004
064a	0.720	0.472	0.002	-0.002	0.006
065b	0.0	1.000	0.0	-0.004	0.003
068a	0.047	0.962	0.0	-0.005	0.006
070a	-1.036	0.300	-0.005	-0.012	0.0
071a	-1.182	0.237	-0.006	-0.012	0.001
073a	2.143	0.032	0.009	0.004	0.015

TABLE 4.2.2 (continued)

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF SULPHATE
(Monthly Medain, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	-1.954	0.051	-0.007	-0.014	-0.003
075a	3.760	0.0	0.012	0.008	0.016
076a	-2.100	0.036	-0.008	-0.013	-0.003
077a	0.165	0.869	0.001	-0.004	0.006
078a	-1.409	0.159	-0.007	-0.015	0.0
160a	-3.992	0.0	-0.019	-0.028	-0.012
161a	0.213	0.831	0.001	-0.004	0.005
163a	-0.682	0.495	-0.003	-0.008	0.002
164a	-0.758	0.449	-0.004	-0.010	0.003
166a	-0.935	0.350	-0.004	-0.010	0.001
168a	-0.161	0.872	0.0	-0.005	0.003
171a	1.235	0.217	0.003	0.0	0.007
172a	-2.211	0.027	-0.006	-0.011	-0.003
173a	-2.089	0.037	-0.008	-0.014	-0.004
249a	2.505	0.012	0.008	0.003	0.016
251a	1.702	0.089	0.007	0.001	0.014
252a	-1.902	0.057	-0.004	-0.008	-0.001
253a	0.147	0.883	0.0	-0.005	0.005
254a	0.312	0.755	0.001	-0.006	0.007
255a	-1.804	0.071	-0.009	-0.015	-0.003
257a	0.526	0.599	0.002	-0.003	0.008
258a	0.104	0.917	0.0	-0.005	0.005
268a	0.560	0.576	0.002	-0.004	0.009
271a	-2.720	0.007	-0.007	-0.011	-0.004
272a	-0.862	0.389	-0.002	-0.005	0.001
273a	-0.987	0.324	-0.003	-0.007	0.001
275a	1.040	0.298	0.004	-0.001	0.009
277a	2.697	0.007	0.007	0.003	0.011
278a	-2.580	0.010	-0.010	-0.015	-0.005
279a	-1.440	0.150	-0.006	-0.012	-0.001
280a	-1.218	0.223	-0.005	-0.011	0.0
281a	1.816	0.069	0.010	0.003	0.019
282a	0.702	0.483	0.002	-0.002	0.007
283a	-1.825	0.068	-0.005	-0.010	-0.002
285a	1.808	0.071	0.008	0.002	0.013
349a	0.511	0.609	0.001	-0.002	0.005
350a	1.879	0.060	0.007	0.002	0.012
354a	-3.277	0.001	-0.013	-0.019	-0.009

Chapter 5 ANALYSES AND CONCLUSIONS FOR NITRATE

Monthly precipitation volume weighted mean and monthly median are are referred to as "weighted mean" and "median", respectively, below.

5.1 Results for the Historical Data

"Historical data" refers to the data collected from January, 1980 to December, 1986. For NO_3, there are 31 sites with 5 or fewer missing observations during the period. The locations of these 31 sites are plotted in Figure 5.1.1. From this figure we can see that more sites are located in the East than the West. The results of this section are based on the data obtained from these 31 sites.

5.1.1 Clustering and Transformation

Before clustering the data, one outlier was deleted from the data set, namely the observation at the site with ID=040a (see Olsen and Slavich, 1986), observed in September, 1986. Its value is 37.76, much larger than 6.07, the second largest observation at that site. Figure 5.1.2 (a) shows the hierarchical cluster structure of the weighted means. The sites are partitioned quite naturally into two clusters. The locations of the sites labelled by cluster are shown in Figure 5.1.3. The pattern is obvious: cluster 1 is located in the subregion of highest industrial concentration; and cluster 3 is spread over the rest of the United States. This indicates that the concentration of NO_3 in wet deposition is higher for the industrial areas than that for other areas. A cluster analysis was done without deleting outliers. The clusters are the same however, except for site 040a which is ruled out of any cluster (Figure 5.1.2 (b)).

The histograms of the original data and the transformed data under the three kinds of transformations described in Section 3.1, are shown in Figure 5.1.5. It appears that the histograms of the log transformed data and the 1/4 power transformed data fit the normal curve best. But they still have long left tails. The other two are right skewed. Since the log transformation is good for many other cases as well, it is adopted in this case and the remainder of our analysis of the nitrate data is based on the transformed data. The histograms of the log transformed data for each cluster are depicted in Figure 5.1.5. It turns out that both histograms have log left tails and the one for cluster 2 having the heavier tail. In general, neither histogram fits the normal curve well.

5.1.2 Trend, Seasonality and Spatial Patterns

The analysis described below is done for both log transformed monthly volume weighted mean and log transformed monthly median NO_3 data. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster, in the manner described in Section 3.2. Figures 5.1.6 and 5.1.7 show the plots of the yearly effects of the two clusters for weighted mean and median, respectively. They show similar patterns although the one for the medians seems more stable. Both clusters show down-trends before 1983. After 1983, the data of cluster 1 show an increasing trend while that for cluster 2 is basically flat. This suggests that the concentration of NO_3 in wet deposition increased after 1983 for the most industrialized subregion of the United States and had remained constant for the rest of the United States to 1986.

Figure 5.1.8 and Figure 5.1.9 show the monthly effects for the weighted means and medians, respectively. They have almost the same pattern . The patterns of the two clusters are the same as well. The effects are high from January to August and low from

September to December. Sudden jumps occurred around September and December. In addition, there is a gap around March. The fact that the seasonal patterns are very similar for all three clusters suggests a reassuring spatial stability in this seasonal pattern. Further study of this interesting pattern is needed.

Figures 5.1.10 to 5.1.11 display the site effects of log(NO_3) by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. The site effects of Station 041a, 055a and 056a for weighted means in Figure 5.1.10 are quite different from those of the medians. This may be caused by the fact that the distributions of the data from these stations are quite different from those of the other stations. In this figure, the stations located in the center of the industrial area tend to have large positive effects. In Figure 5.1.11, Stations 059a and 074a, the only two stations located along the West Coast, have remarkably large, negative effects. This fact are reflected in the structure of clustering analysis as well (see Figure 5.1.2).

Figures 5.1.12 to 5.1.13 are the boxplots of the residuals of the three-way median polish of the log(NO_3) data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order and comparable. The differences between the boxplots for the means and for the medians are even smaller. In Figures 5.1.12 ((a) and (b)), the boxes for Stations 020a and 038a are larger than the others. In fact, these two stations are the ones farthest from the center of the industrial area within cluster 1. In addition, we observe a few extreme outliers from Station 040a, and from Station 075a and 168a as well. For these sites, further detailed study is needed. In Figures 5.1.13 ((a), (b)), extreme outliers are observed in the boxplots for Station 034a, 039a, 049a and 059a and they need a more detailed study as well. But no analyses for specific site are included in this paper. The boxes for the two stations on West Coast, 059a and 074a, are relatively large indicating not only that they have large negative effects but also that they behaved differently from the others so that the model can not fit them well.

Figures 5.1.14 to 5.1.15 summarize the median polish for the two clusters, respectively where (a) is for the weighted means and (b) is for the medians. In each figure, boxes represent the main effects, interactions and residuals respectively. We regard interactions as non-negligible if the sizes of the boxes for the interactions are comparable with those for main effects. In Figures 5.1.14 (a) and (b), the boxes for yearly effects are relatively small, especially in (b). This suggests that the variation caused by trend is smaller than the variations caused by the other effects. The variation of the interactions between yearly effects and site effects is small as well. The two boxes for the interactions of month-by-year and month-by-site are relatively large suggesting that their contributions to the model should be taken into account. The pattern of the summary for cluster 2 is quite similar to that for cluster 1 except the effects of the two sites on the West Coast are ruled out of the box for site effects as two outliers (see Figures 5.1.15 (a) and (b)).

5.1.3 The Results of Trend Testing, Slope Estimation and Kriging

Mann-Kendall Test and Sen's nonparametric slope estimation procedure are applied to the deseasonalized data for each site in the manner described in Section 3.3. A Mann-Kendall Test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval are obtained for each site and for both weighted means and medians. These are shown in Tables 5.1.1 and 5.1.2 respectively. In Table 5.1.1, 14 out of the 31 sites have significant trends at the p=0.2 level. In Table 5.1.2, 15 out of the 31 sites have significant trends at the p=0.2 level.

The results of the nonparametric slope estimation are plotted in Figures 5.1.16 (a) and (b) with symbols defined in Section 3.3, for the weighted means and medians, respectively. If the lower 80% confidence bound of the slope estimate is greater than 0 at any site a "+" is plotted; if the upper 80% confidence bound of the slope estimate is less than 0 a "-" is plotted. Otherwise a "0" is plotted . "+", "-" and "0" represent respectively, an up-trend, a down-trend and no-trend. The patterns of these two figures are very similar. More than half of the sites show no-trends and they are mainly located in the Eastern United States. The data from rest of the sites located mainly in the other areas of the United States show a down-trend. This conclusion agrees with the previous finding that the wet deposition concentration of NO_3 has no trend or down-trend for most areas of the United States (Schertz and Hirsh,1985). However, as we mentioned before, this could be a misleading conclusion in the presences of a V-shaped trend. At the very least, this result says that the concentrations of nitrate for the latter years of the sampling period are about the same or smaller than that of early years of the same period.

The results of Kriging the slope estimates are displayed in Figures 5.1.17(a) and (b). As before, Figure 5.1.17(a) is for the weighted means and (b) is for medians. A "+" means an up-trend is estimated at the corresponding point and a "-" means a down-trend, estimated at the point. Again, these two plots are similar. Both Figures 5.1.17 (a) and (b) show that there is a no-trend for the Eastern United States and a down-trend for the rest of the United States. Considering that these results are obtained from the same set of data which produce Figures 5.1.16 (a) and (b), this is quite natural. The difference between these two figures are probably caused by the differences in magnitudes of the slope estimates of weighted means and of medians. Note that these two figures can only be used as a guideline for the spatial distribution of the trends and should be read in conjunction with Figure 5.1.16. At any particular point where the observations were not made, interpolation may not be reliable. This is because (a) estimation is based on the estimated slopes, not on the observations directly and (b) the spatial resolution of the measurements is low and therefore, the correlations between sites are small.

5.2 Results for the Recent Data

"Recent data" refers to the data collected from January, 1983 to December, 1986. For NO_3, there are 81 sites with 5 or fewer missing observations during the period. The locations of these 81 sites are plotted in Figure 5.2.1. From this figure we can see that more sites are located in the Northeast than the rest of the areas. The results of this section are based on the data obtained from these 81 sites.

5.2.1 Clustering and Transformation

Before clustering the weighted mean data, a outlier was deleted from the data set, namely the observation for the Station 040a (see Olsen and Slavich, 1986). It was observed in September, 1986, the same one mentioned in Section 5.1.1. Figure 5.2.2 (a) shows the hierarchical cluster structure of the weighted means. The sites first partition naturally into two clusters: a large one on the left and a smaller one on the right. If we ignore Stations 035a, 268a and 283a, then the large cluster can be partitioned into two subgroups: cluster 2 and the group of the rest of the stations, as shown in Figure 5.2.2. For technical convenience, Stations 035a, 268a and 283a were assigned to the group closest to these three stations. We label this group as cluster 1, as shown in Figure 5.2.2. The locations of the sites labelled by clusters are shown in Figure 5.2.3. Keeping in mind that clusters 1 and 2 are subgroups of one big cluster, the pattern of Figure 5.2.3 is similar to that of Figure 5.1.3. In fact, cluster 3 corresponds to cluster 1 in the latter, located in the subregion of highest industrial concentration; and clusters 1 and 2 in the former correspond to cluster 2 in the latter, located throughout the rest of the United States. This gives us the same indications we mentioned in Section 5.1.1, namely, that the concentration of NO_3 in wet deposition is higher for the industrial areas than that for other areas. A cluster analysis was done without deleting outliers. The clusters are the same except for Station 040a which is excluded (Figure 5.2.2 (b)).

The histograms of the original data and the transformed data under the three kinds of transformations described in Section 3.1, are shown in Figure 5.2.5. It appears that the histograms of the log transformed data and the 1/4 power transformed data fit the normal curve best. But they still have long tails. The other two are right skewed. We adopt the log transformation in this case since it is the best transformation in many other cases and, particularly, since it was adopted in the case of "historical data". The remaining analyses of this section are based on the log transformed data. The histograms of the log transformed data for each cluster are shown in Figure 5.2.5. It turns out that all the histograms have long tails while the tail in the data for cluster 1 is heavy.

5.2.2 Trend, Seasonality and Spatial Patterns

The analysis described below is applied to both log transformed monthly volume weighted means and log transformed monthly medians. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster. Figures 5.2.6 and 5.2.7 show the plots of the yearly effects of the three clusters for weighted mean and median respectively. Combining these two figures, it seems that the sites in cluster 1 reveal a degree of down trend while the trends for clusters 2 and 3 are not obvious. Observe that here cluster 3 and cluster 1 in Figure 5.1.3 cover essentially the same geographic subregion while clusters 1 and 2 correspond to cluster 2 in Figure 5.1.3; the pattern of the trends observed here are consistent with the pattern of the trends in Figures 5.1.6 and 5.1.7.

Figures 5.2.8 and 5.2.9 show the monthly effects for the weighted means and medians, respectively. Although more than half of the data used in the present analyses were not in the historical data, the patterns seen here are almost the same as those displayed in Figures 5.1.8 and 5.1.9 for historical case: the effects are high from January to August and low from September to December; two sudden jumps around September and December and a gap around March. This convinces us again of spatial stability in the seasonal pattern. Cluster 1 seems have a larger variation than the other clusters.

Figures 5.2.10 to 5.2.12 show the site effects of log(NO_3) by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. For cluster 1, the stations located in the middle of the United States have large positive effects while the stations located along the West Coast have large negative effects (see Figure 5.2.10). In cluster 3, the stations close to the center of the industrial region tend to have a positive effects and the stations located far from the center of the industrial region tend to have a negative effects (see Figure 5.2.12). In general, the closer a station is located to the center of the industrial region, the more likely it has a large positive effect.

Figures 5.2.13 to 5.2.15 are the boxplots of the residuals of the three-way median polish of the log(NO_3) data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order and are comparable to each other and to those in Figures 5.1.12 to 5.1.13 as well. The differences between the boxplots for the means and for the medians are even smaller. Extreme outliers are found in Stations 035a, 074a, 078a and 281a in Figures 5.2.13 (a) or (b), 034a, 039a, 049a and 349a in Figures 5.2.14(a) or (b) and 040a, 075a,168a and 350a in Figures 5.2.15 (a) or (b). Among these stations, 034a, 039a, 040a, 049a, 075a, 168a, have the extreme outliers observed in Figures 5.2.12 to 5.1.13 and those outliers may be the same. Some other stations do have outliers as well. And all these stations need further study.

Figures 5.2.16 to 5.2.18 summarize the median polish for the three clusters respectively where (a) is for weighted means and (b) is for medians. In each figure, boxes represent the main effects, interaction and residuals. In Figures 5.2.16 (a) and (b), we note that the variation of the station effects are quite large. The reason for this is that the Station 035a, 268a and 283a have large positive effects, which can be seen from Figure 5.2.10. Actually these 3 stations were assigned to this cluster only for technical convenience (see Figure 5.2.2(a)). This causes the large variation of the station effects in this cluster. Among the other effects, the variations in year effects and the year-by-station interactions are relatively small as in the situation portrayed in Figure 5.1.15 (a) and (b). For cluster 2, the variation in the yearly effects and the interactions of yearly effects with other effects are all small. This suggests that for the stations in this cluster there was little year to year variation in their data patterns (see Figures 5.2.17 (a) and (b)). Figures 5.2.18 (a) and (b) show the summaries for cluster 3. The pattern in this figure is similar to that in Figure 5.1.14.

5.2.3 The Results of Trend Testing, Slope Estimation and Kriging

The Mann-Kendall Test and Sen's slope estimation procedure which were applied to "historical data" are used for "recent data" as well. A Mann-Kendall test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval were obtained for each site for both the weighted means and medians. These are shown in Tables 5.2.1 and 5.2.2 respectively. In Table 5.2.1, 28 out of the 81 sites have significant trends at the p=0.2 level. In Table 5.2.2, 27 out of the 81 sites have significant trends at the p=0.2 level.

Figures 5.2.19 (a) and (b) are obtained by the same method used to produce Figures 5.1.16 for weighted means and medians respectively. Again, "+", "-" and "0" represent an up-trend, a down-trend and no-trend, respectively. In Figure 5.2.19 (a), 12 of the sites show up-trends and are mainly located in the Eastern United States. 15 of the sites show down-trend and are located mainly in the West and middle of the United States. The rest of the sites show no-trend and are located throughout the United States. The pattern of Figure 5.2.19 (b) is similar to that in Figure 5.2.19 (a) with fewer "up-trend" stations. This probably is due to the fact that the estimates based on medians are typically more conservative than those based on means. Comparing Figures 5.2.19 ((a) and (b)) with Figures 5.1.16 ((a) and (b)), we can see that some of the "-'s" and "0's" in former become "0's" and "+'s" in latter and such changes happened mainly in the Eastern United States. This suggests that although the trends for the stations located in the Western United States continued to decline, most of the stations located in the Eastern United States no longer have down-trends and some of them had an up-trend during 1983 to 1986. This implies that some of the stations do have a V-shaped trend.

The results of Kriging the slope estimates are displayed in Figures 5.2.20 (a) and (b) where (a) is for the weighted means and (b) is for the medians. As before, a "+" means an up-trend and a "-" means a down-trend estimated at the corresponding point, respectively. Again, these two plots are similar except that both the "+" area and "-" areas in (b) are smaller than those in (a). Both Figures 5.2.20 (a) and (b) show that there is a down-trend for some areas in the Western United States, an up-trend in the East and no-trend in the rest of the subregions. Comparing these two figures with Figures 5.1.17 ((a) and (b)), we can see that the areas with "-" shrink significantly towards the West and the areas with "+" appear in the East. The difference in the patterns of the trends is quite remarkable. For the reason given in Section 5.1.3, these two figures can only be used as indicators of the spatial distribution of the trends as the interpolation error for a specific point could be large.

The Locations Of The 31 Monitoring Stations From 1980 To 1986
(For Nitrate)

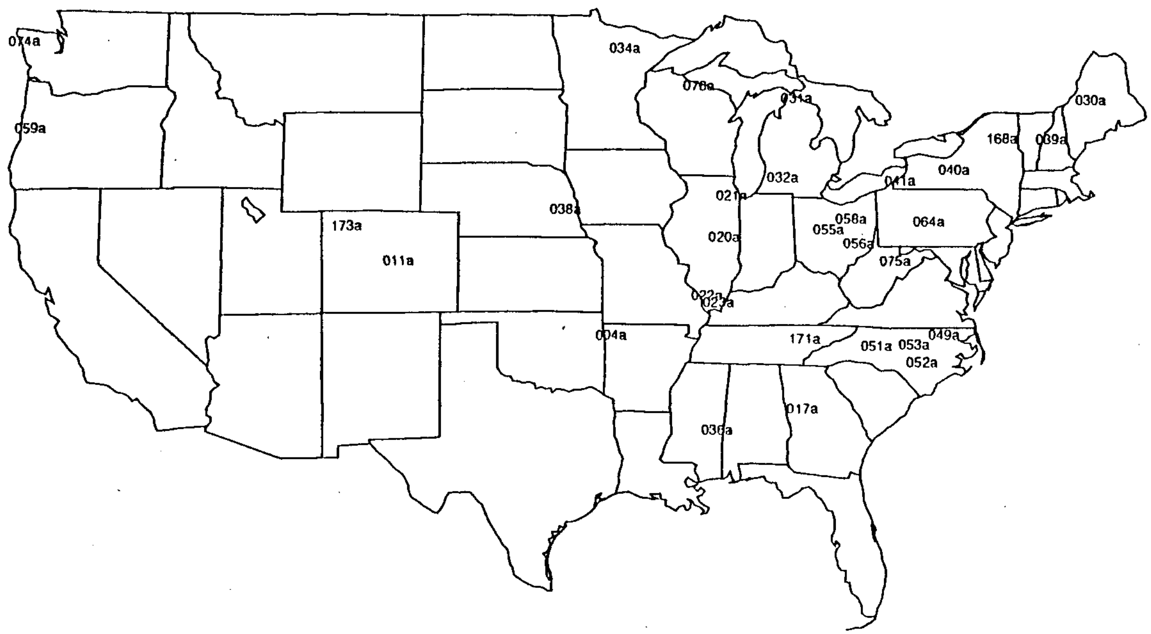


Figure 5.1.1

Clustering of NO3 monthly volume weighted mean based on sqrt(MSE)

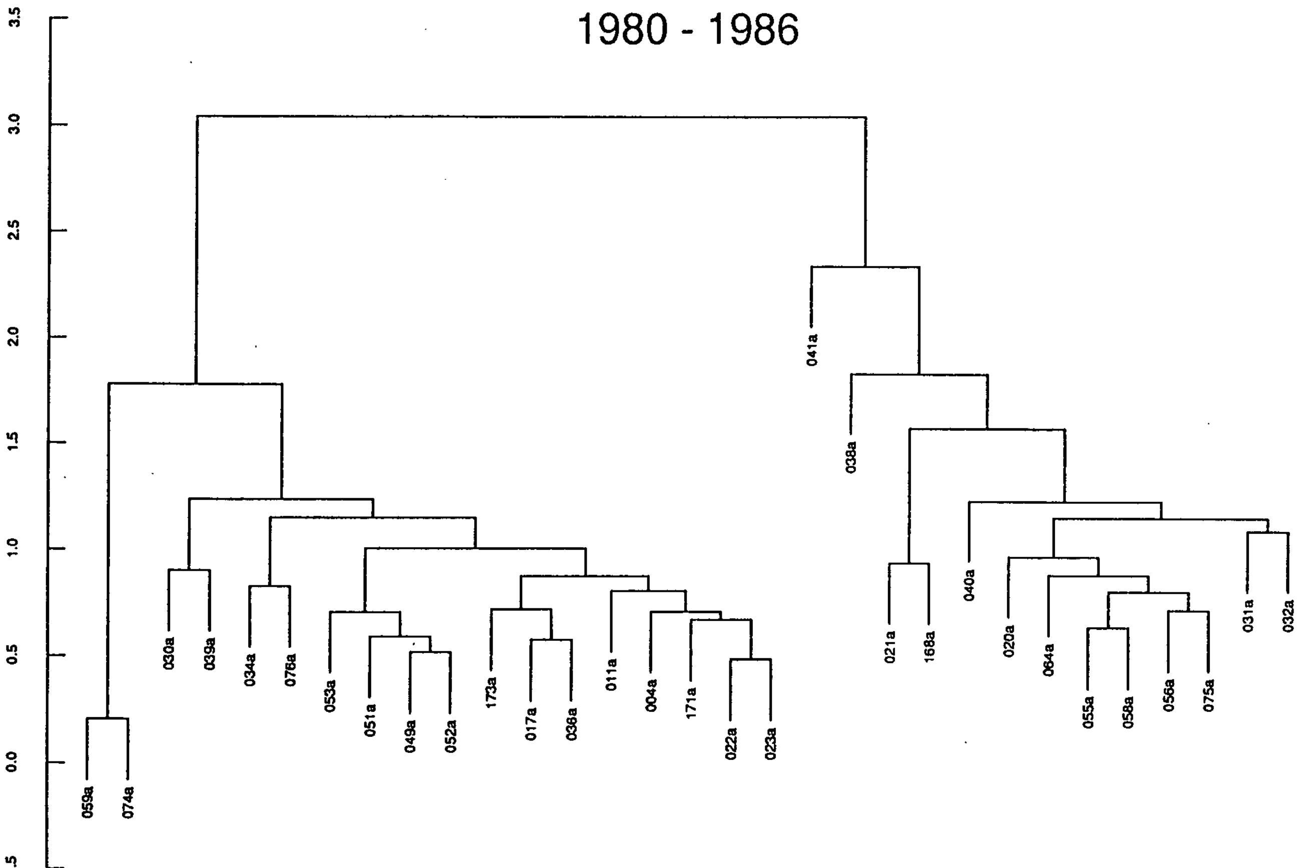


Figure 5.1.2(a)

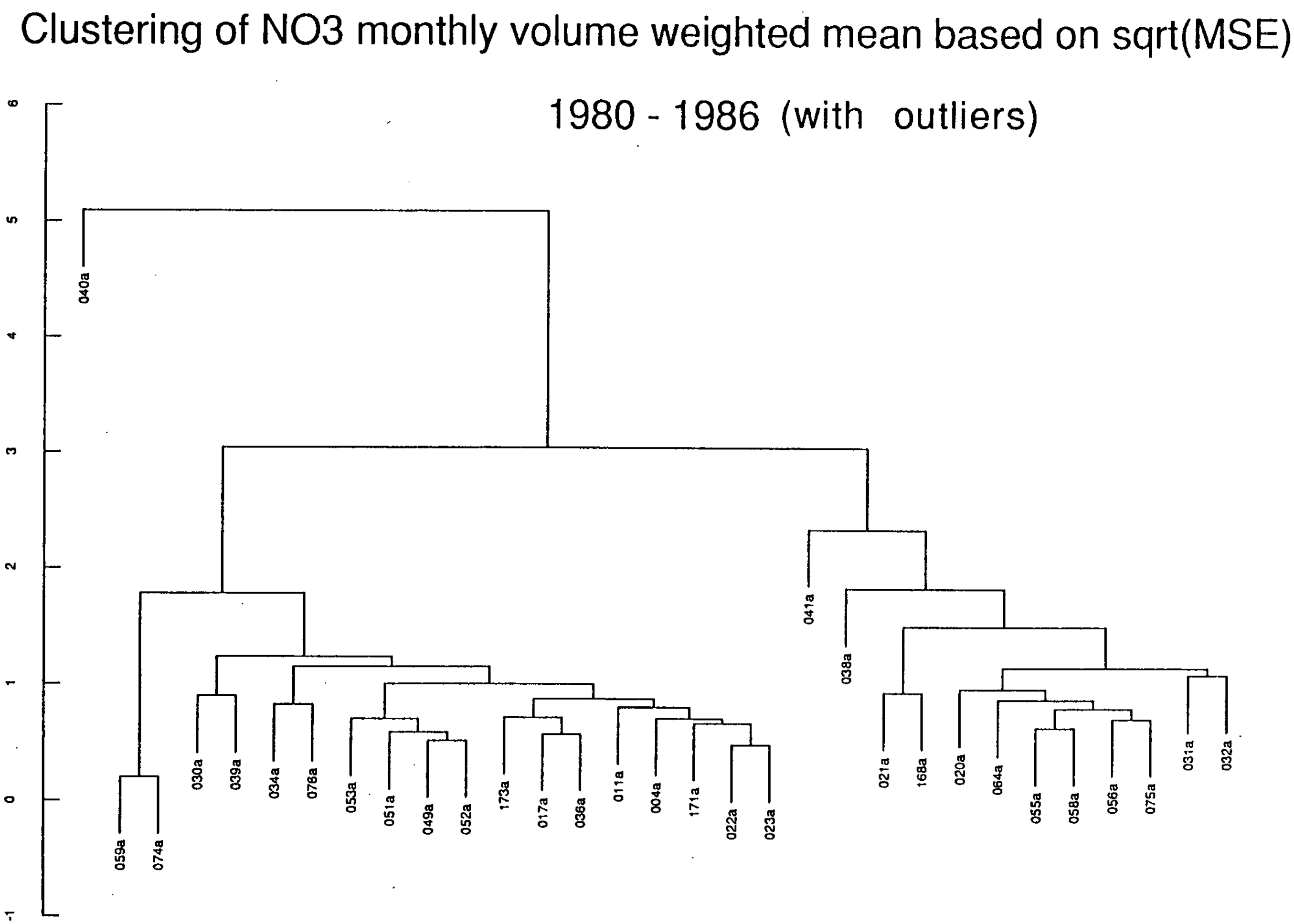

Clustering of NO3 monthly volume weighted mean based on sqrt(MSE)
1980 - 1986 (with outliers)
6
5
4
3
2
1
0
-1
040a
059a
074a
030a
039a
034a
076a
053a
051a
049a
052a
173a
017a
036a
011a
004a
171a
022a
023a
041a
038a
021a
168a
020a
064a
055a
058a
056a
075a
031a
032a

Figure 5.1.2(b)

Clusters of NO3 monthly volume weighted mean based on sqrt(MSE)

1980 - 1986 (k=2)

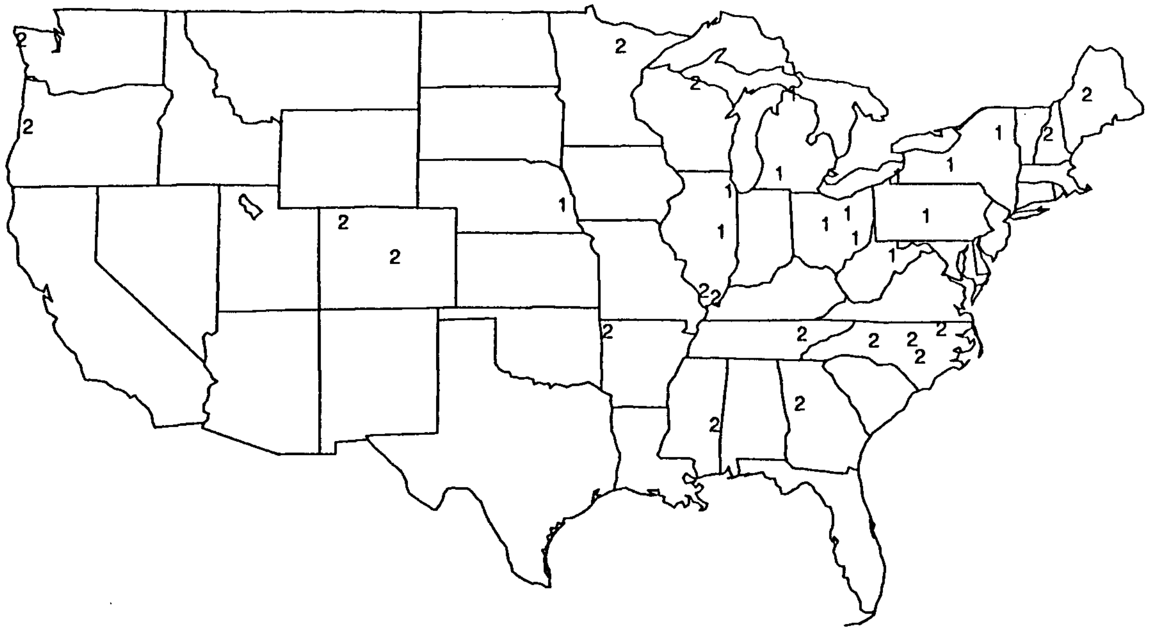


Figure 5.1.3

Histograms of Transformed NO3 (Volume Weighted Mean, 80-86)

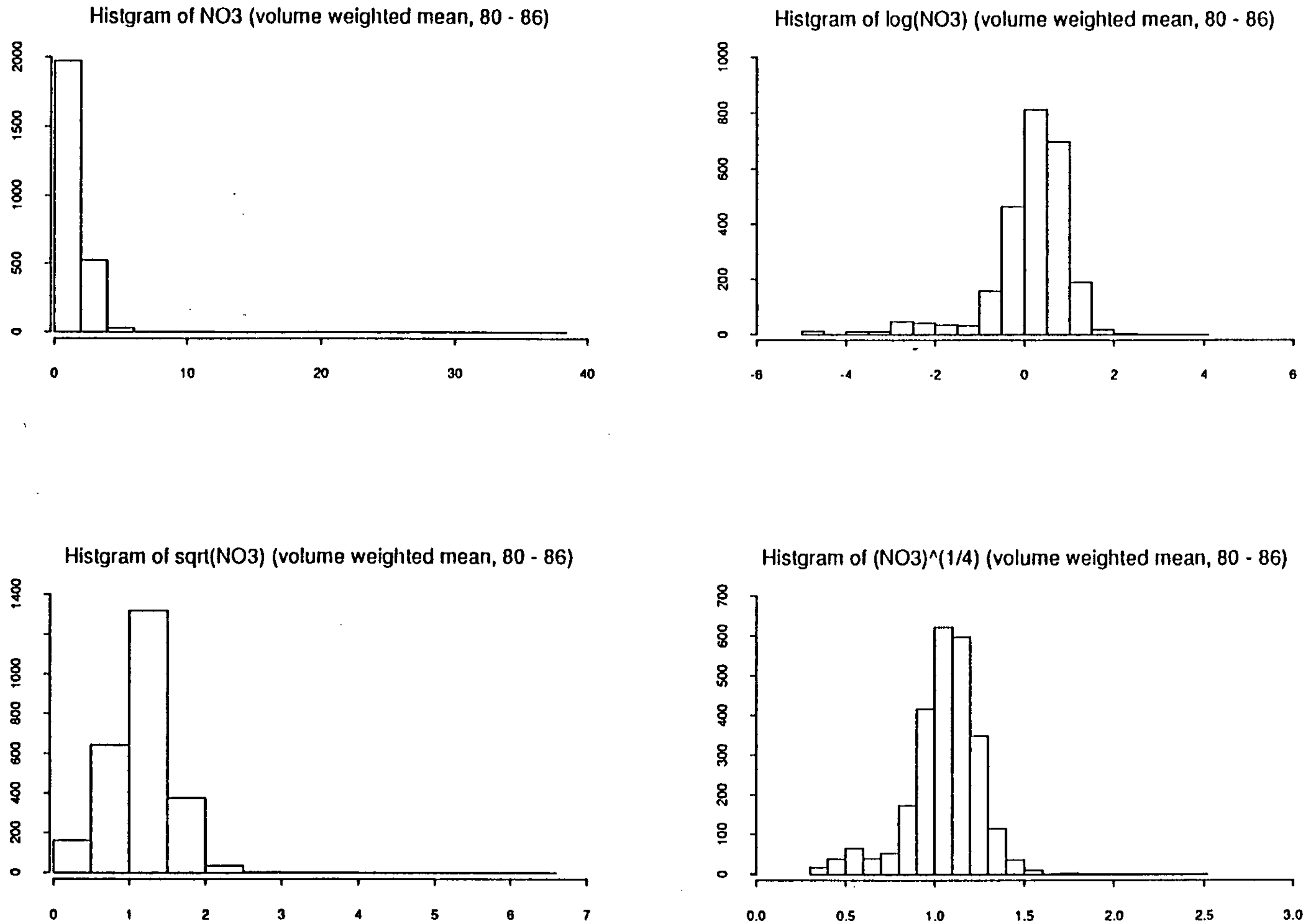


Figure 5.1.4

Histograms of log(NO3) by Clusters (Volume Weighted Mean, 80-86)

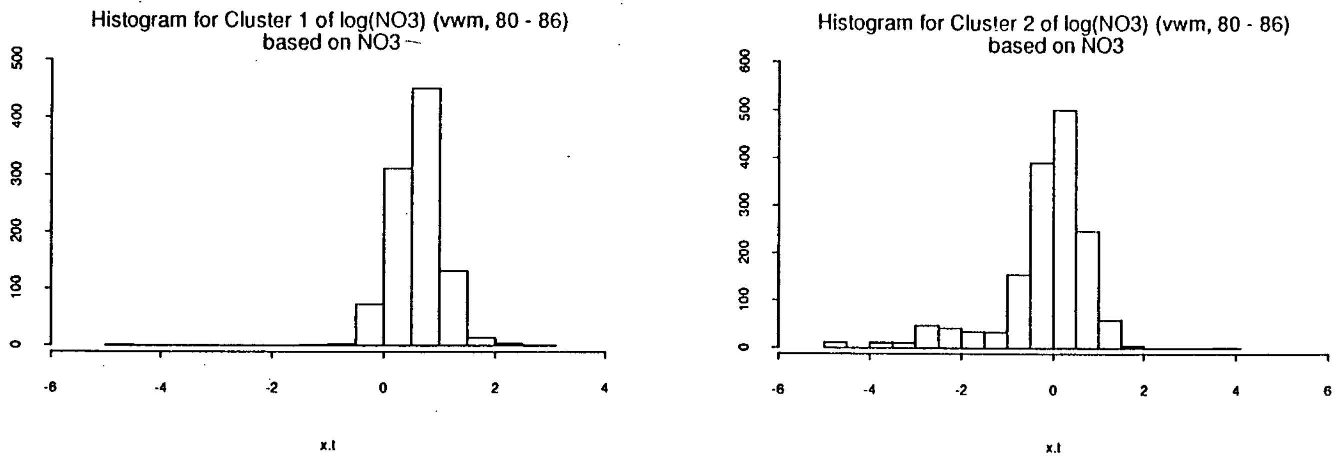


Figure 5.1.5

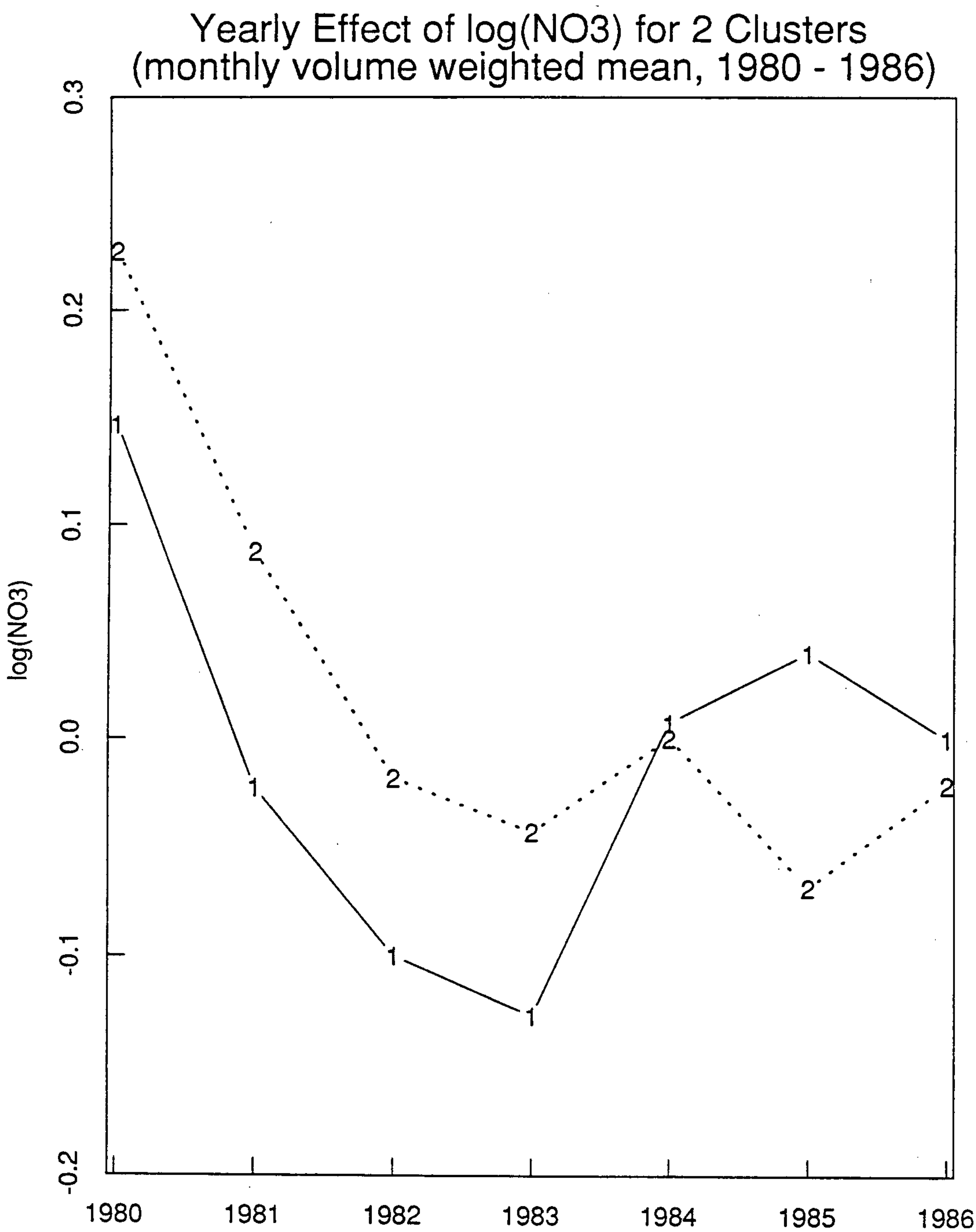

Yearly Effect of log(NO3) for 2 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(NO3)
0.3
0.2
0.1
0.0
-0.1
-0.2
1980
1981
1982
1983
1984
1985
1986

Figure 5.1.6

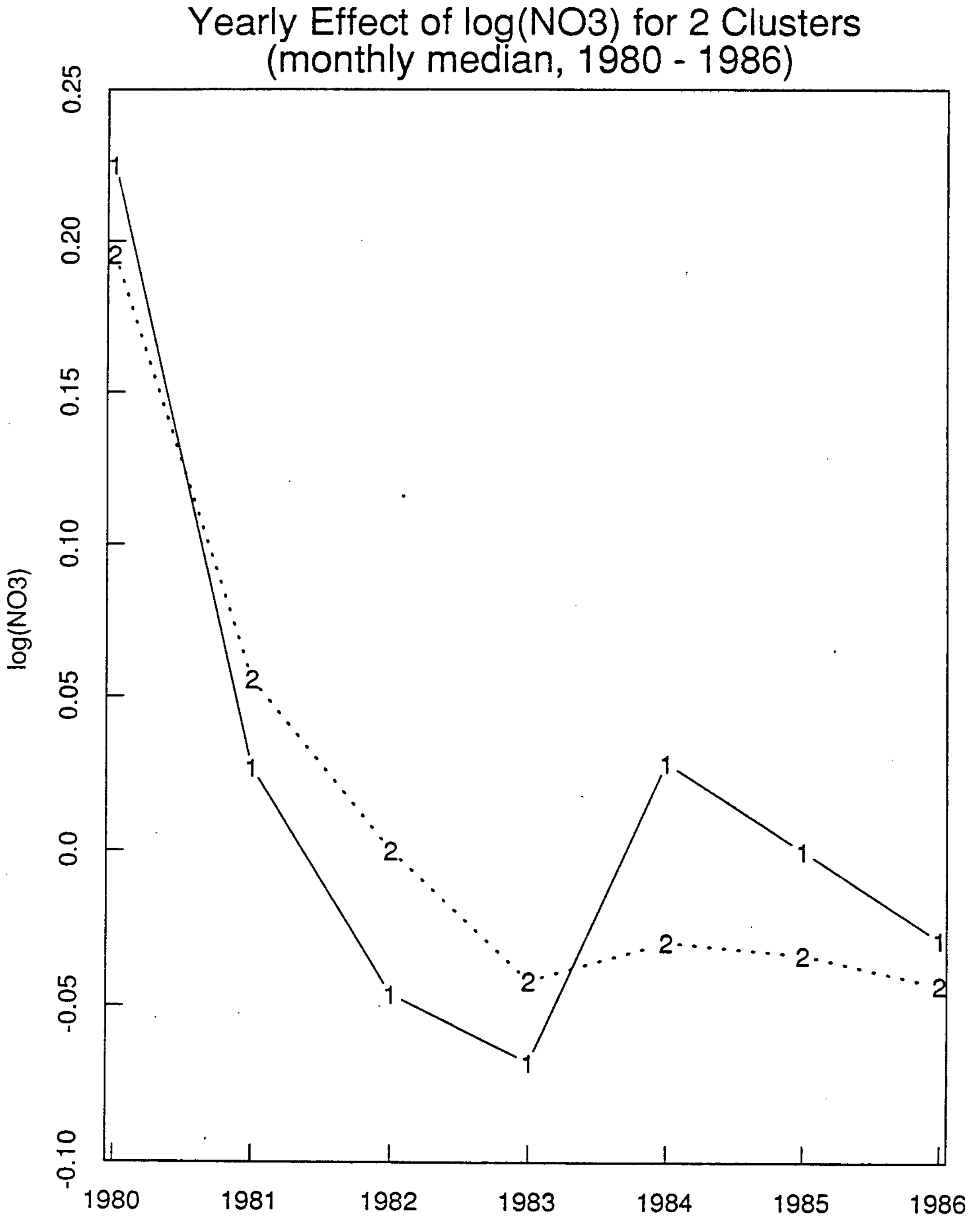
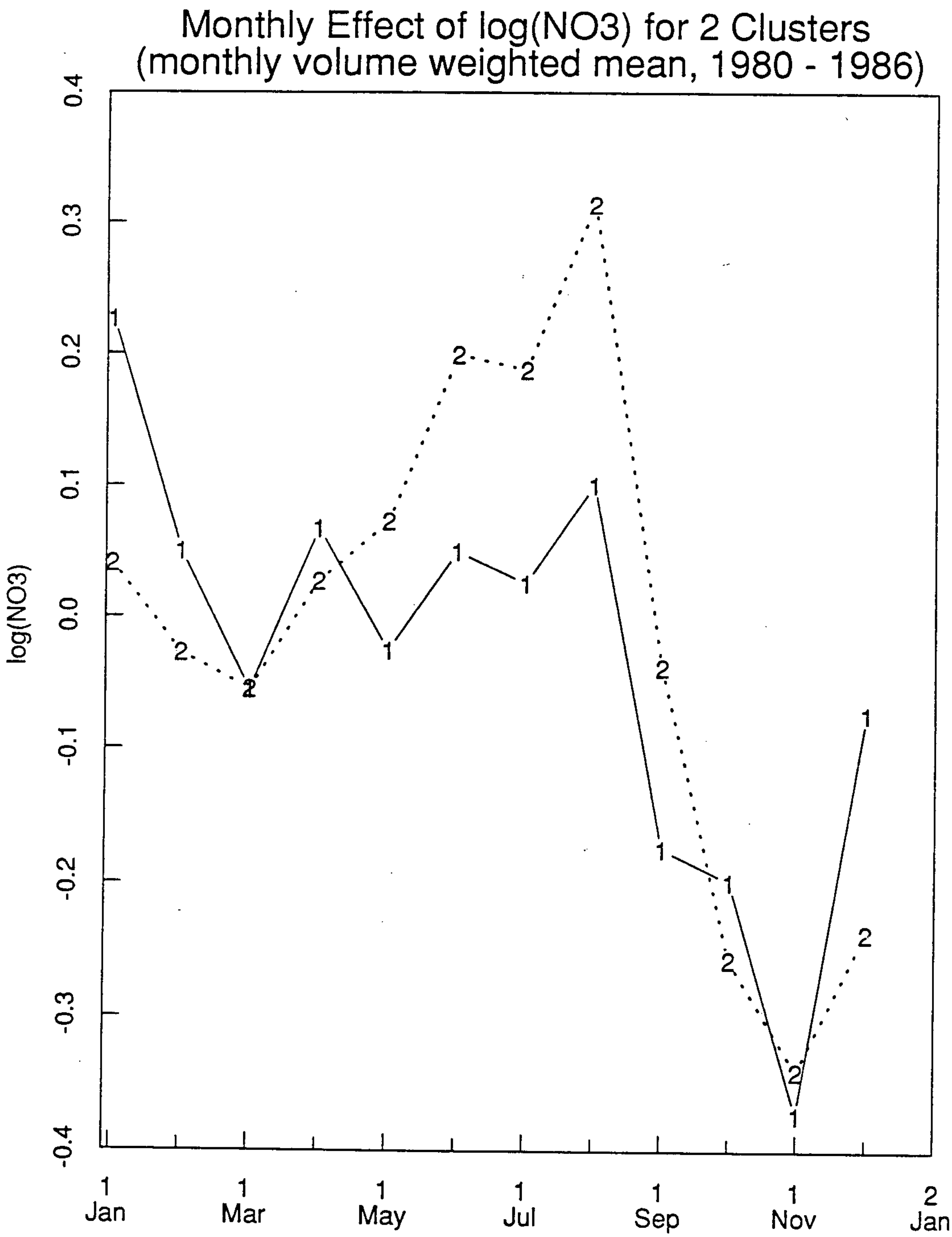

Yearly Effect of log(NO3) for 2 Clusters
(monthly median, 1980 - 1986)
log(NO3)
0.25
0.20
0.15
0.10
0.05
0.0
-0.05
-0.10
1980
1981
1982
1983
1984
1985
1986

Figure 5.1.7



Monthly Effect of log(NO3) for 2 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(NO3)
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 5.1.8

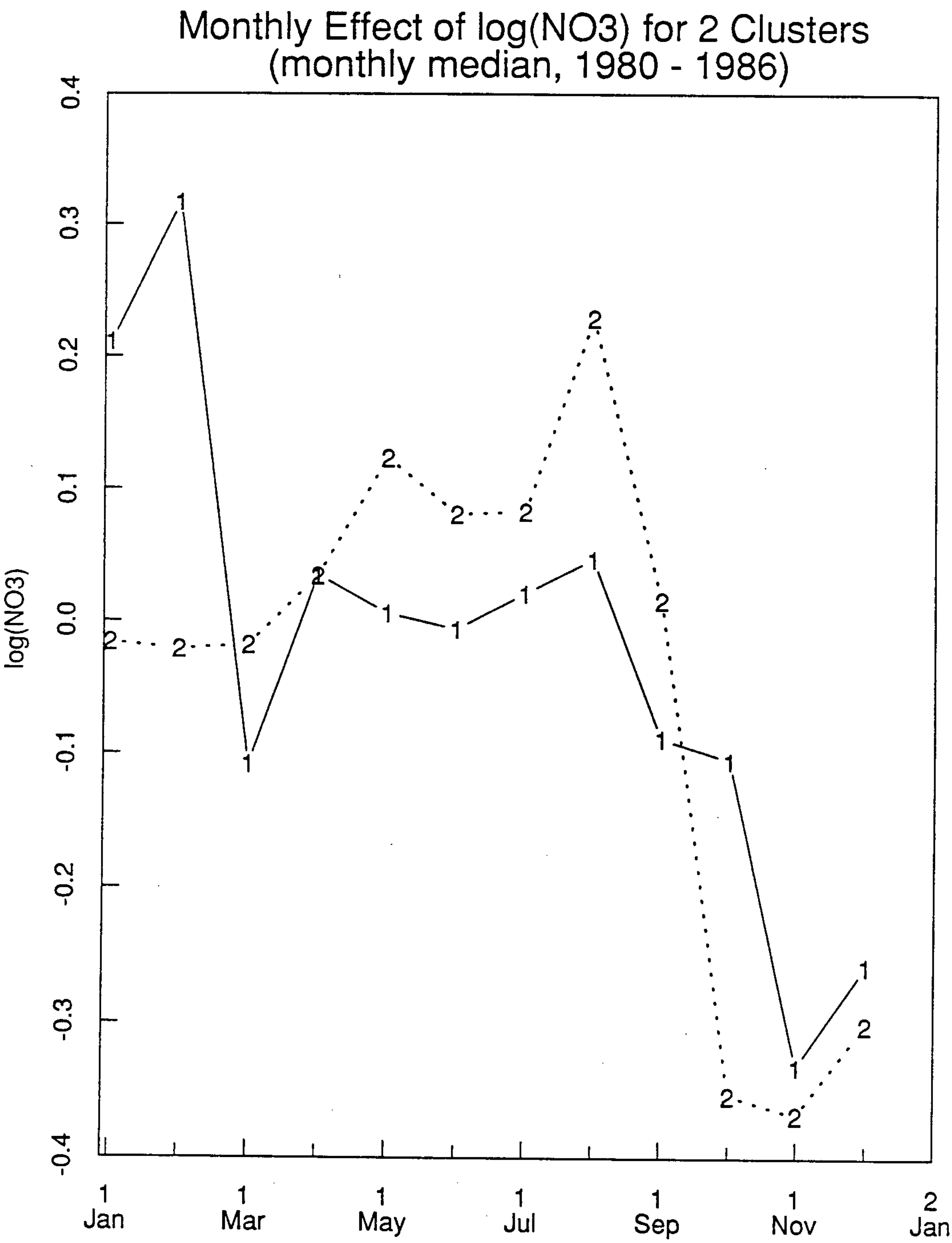

Monthly Effect of log(NO3) for 2 Clusters
(monthly median, 1980 - 1986)
log(NO3)
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 5.1.9

Station Effect of log(NO3) for Cluster 1
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)

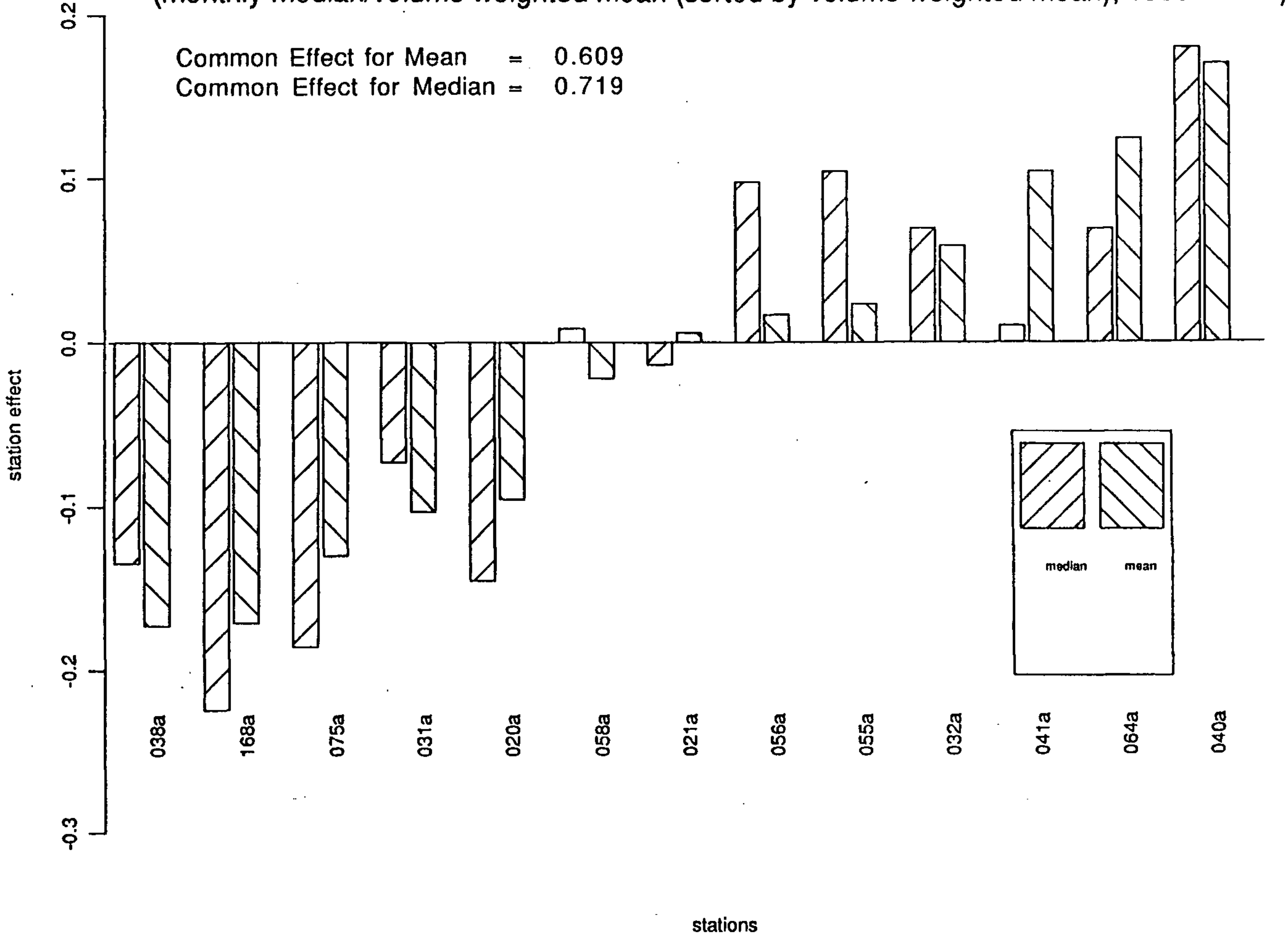


Figure 5.1.10

Station Effect of log(NO3) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)

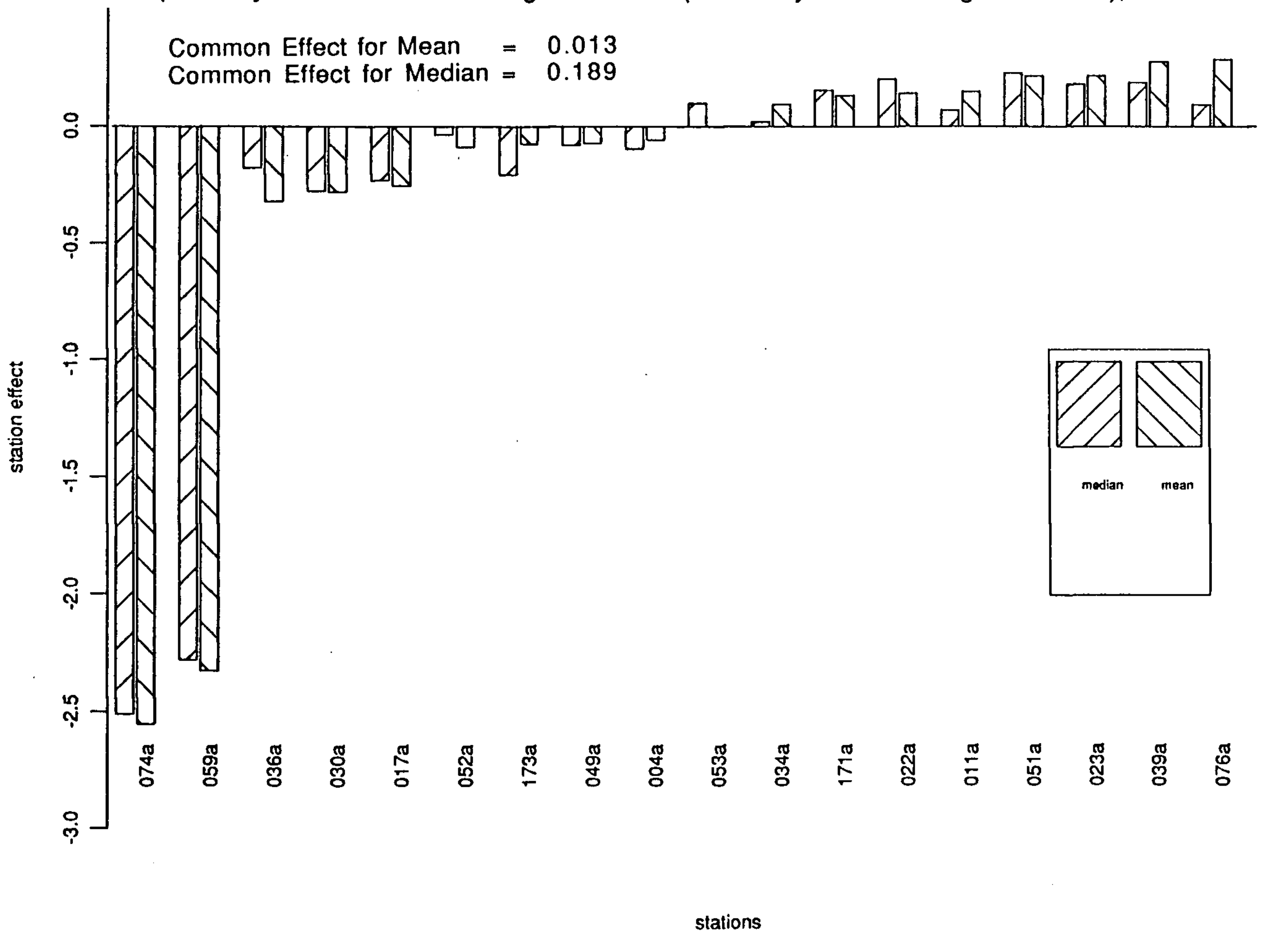


Figure 5.1.11

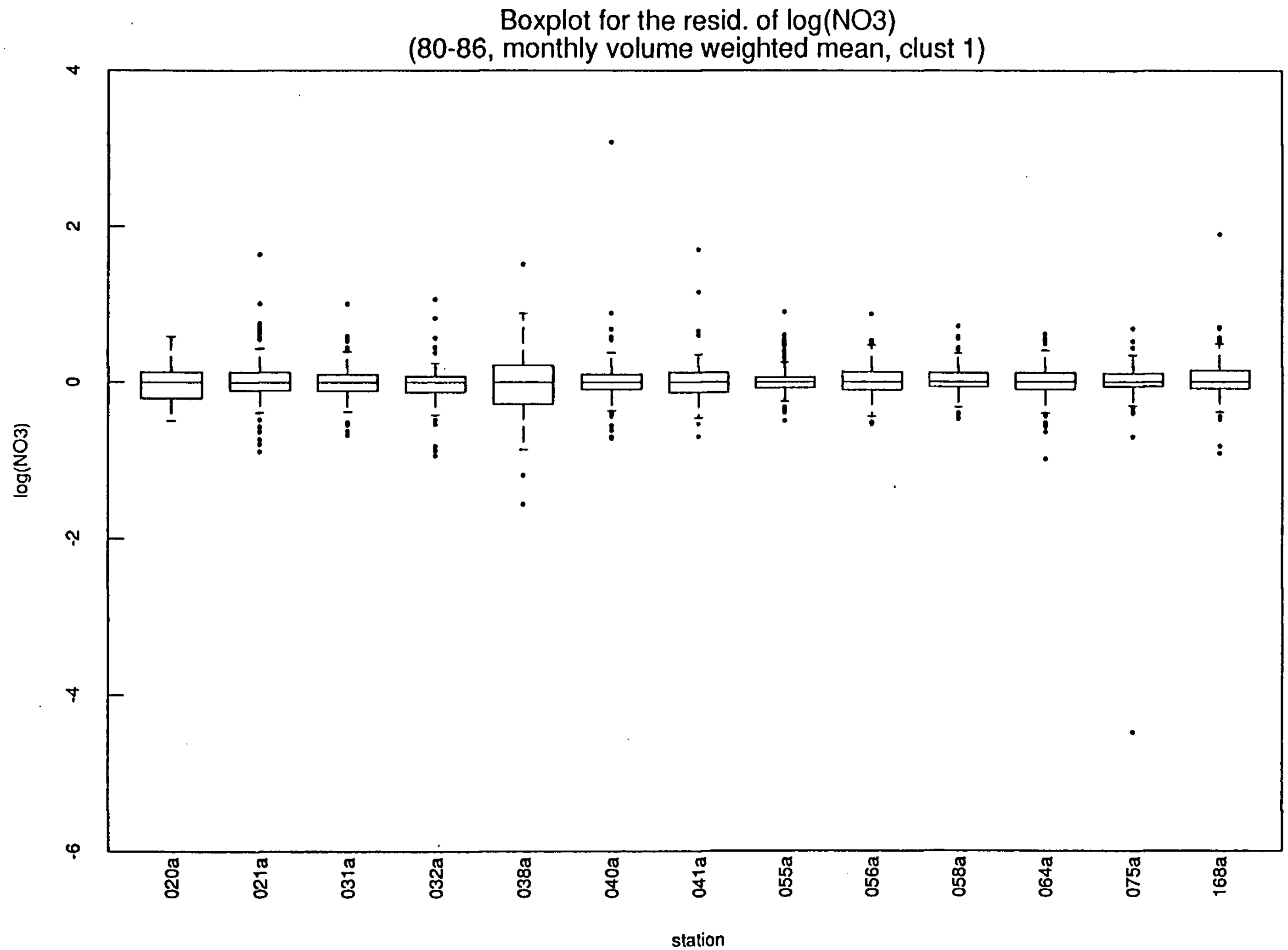


Figure 5.1.12(a)

Boxplot for the resid. of log(NO3)
(80-86, monthly median, clust 1)

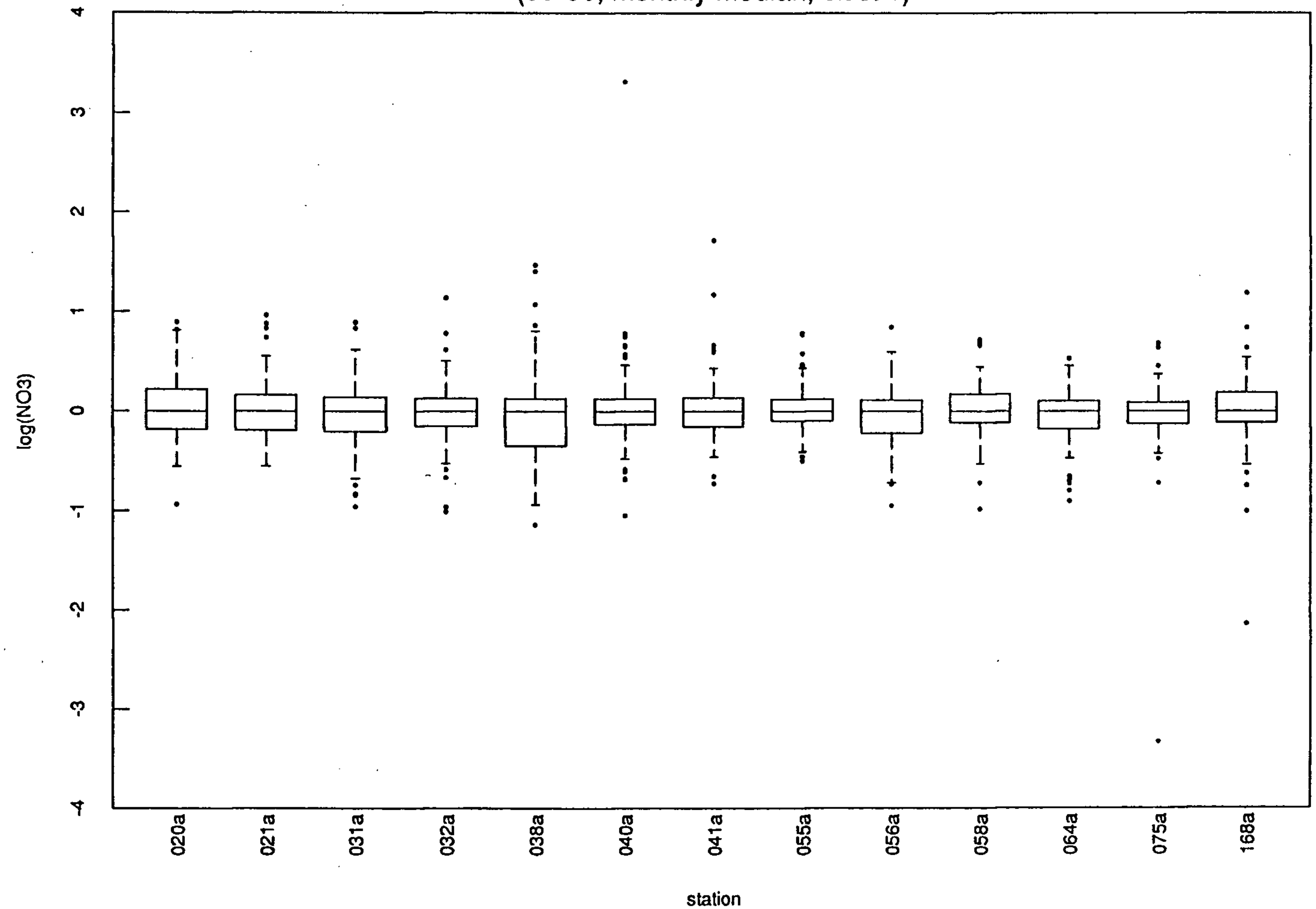


Figure 5.1.12(b)

Boxplot for the resid. of log(NO3)
(80-86, monthly volume weighted mean, clust 2)

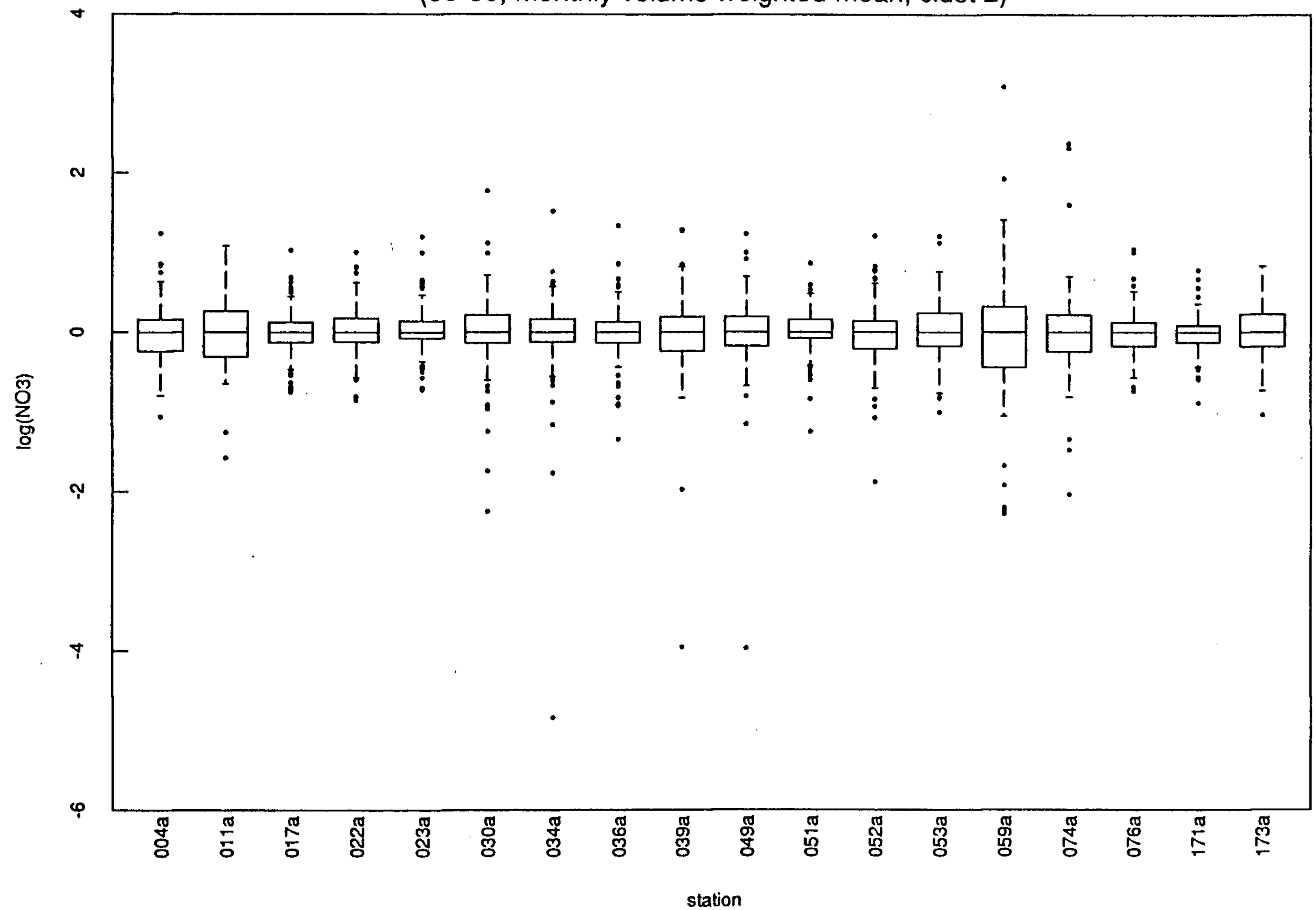


Figure 5.1.13(a)

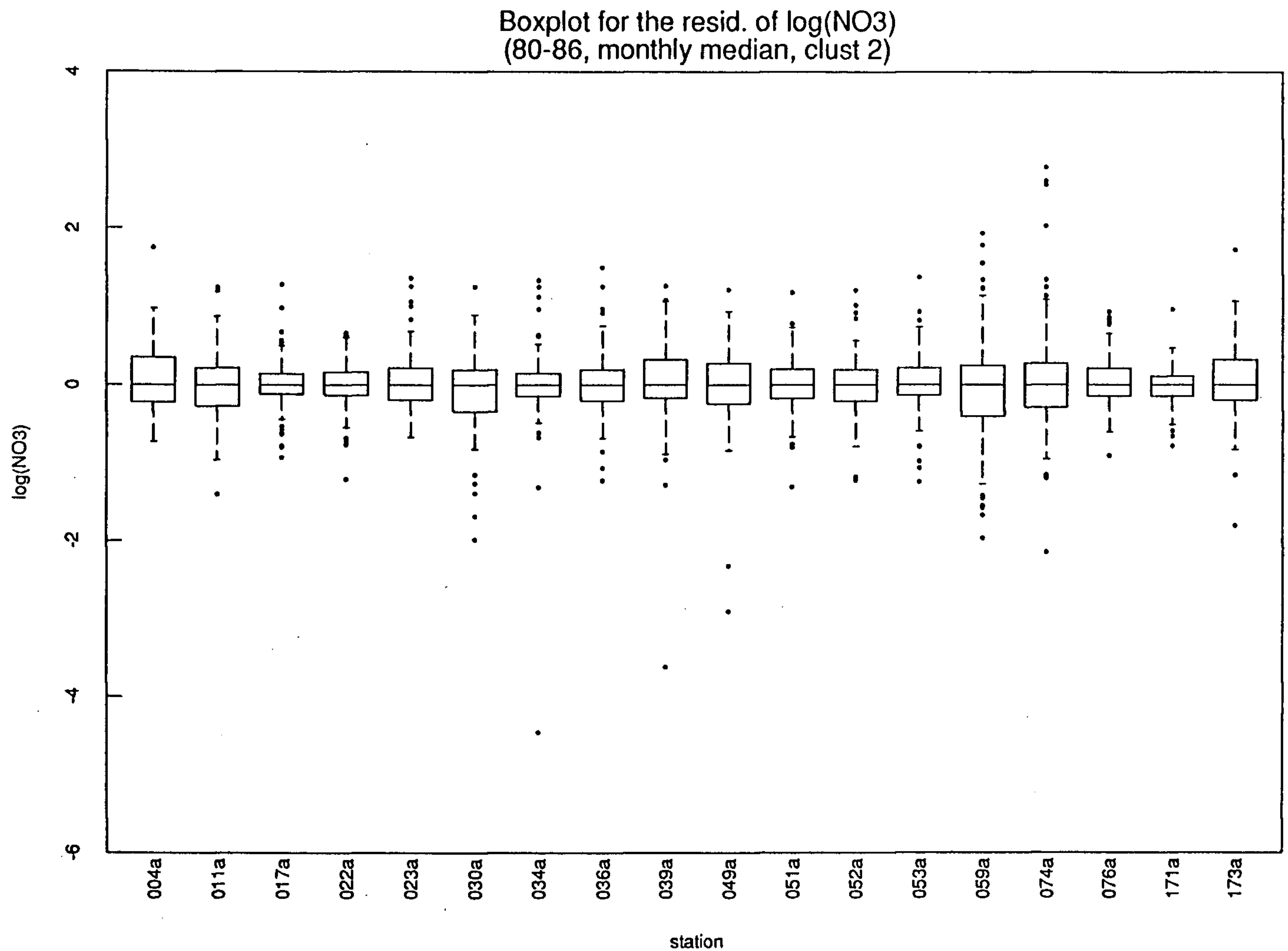


Figure 5.1.13(b)

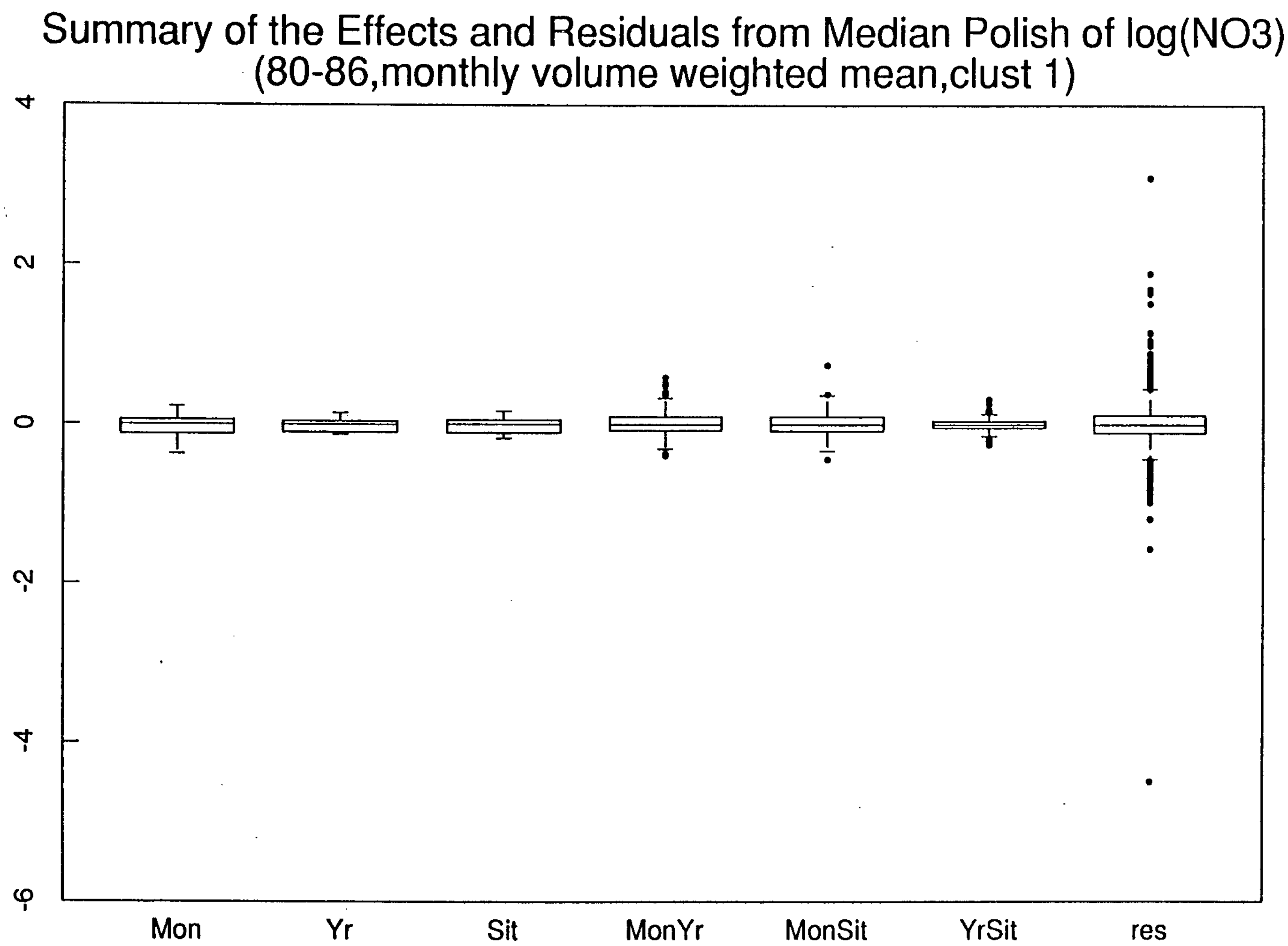


Figure 5.1.14(a)

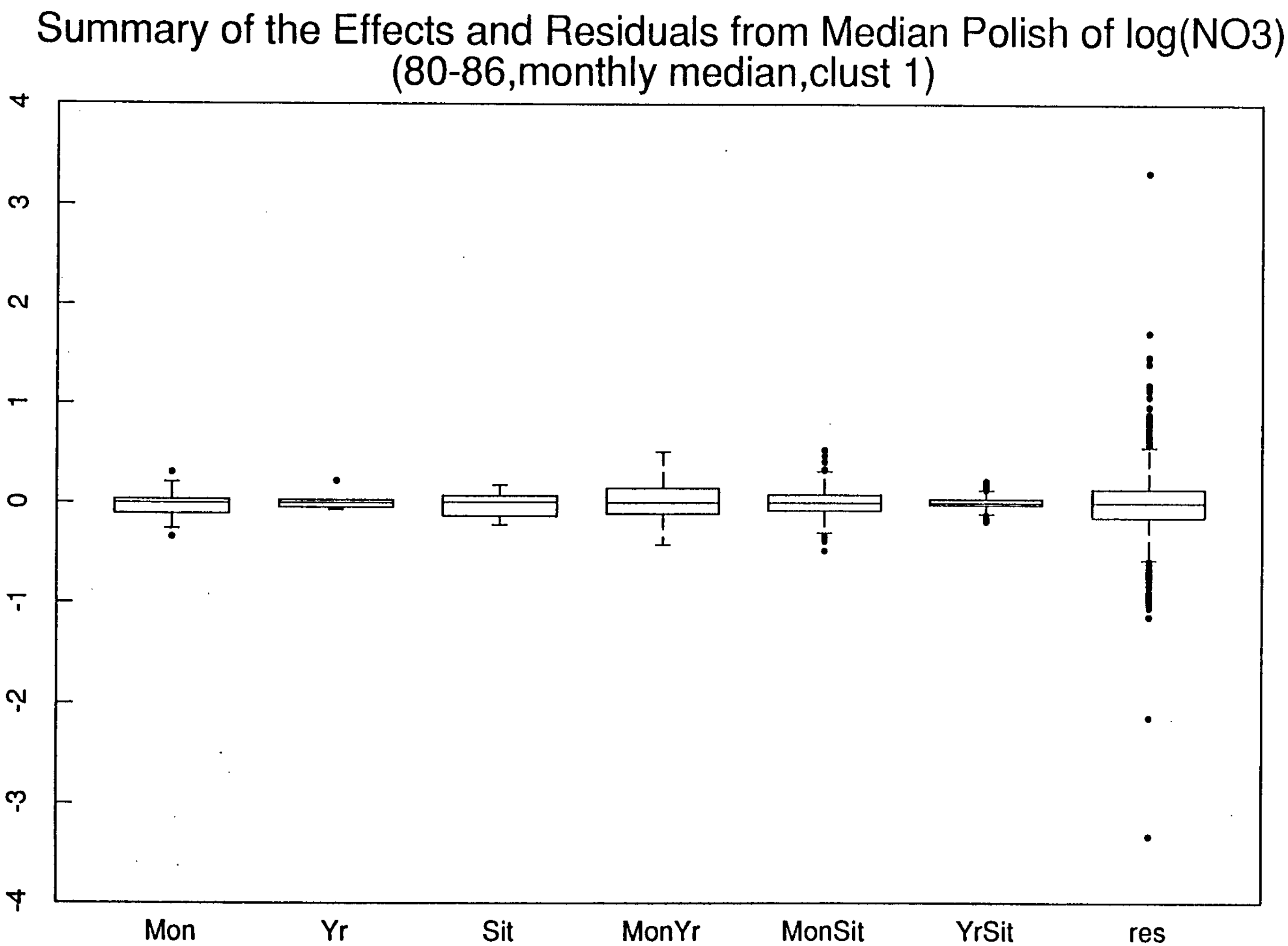


Figure 5.1.14(b)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(80-86,monthly volume weighted mean,clust 2)

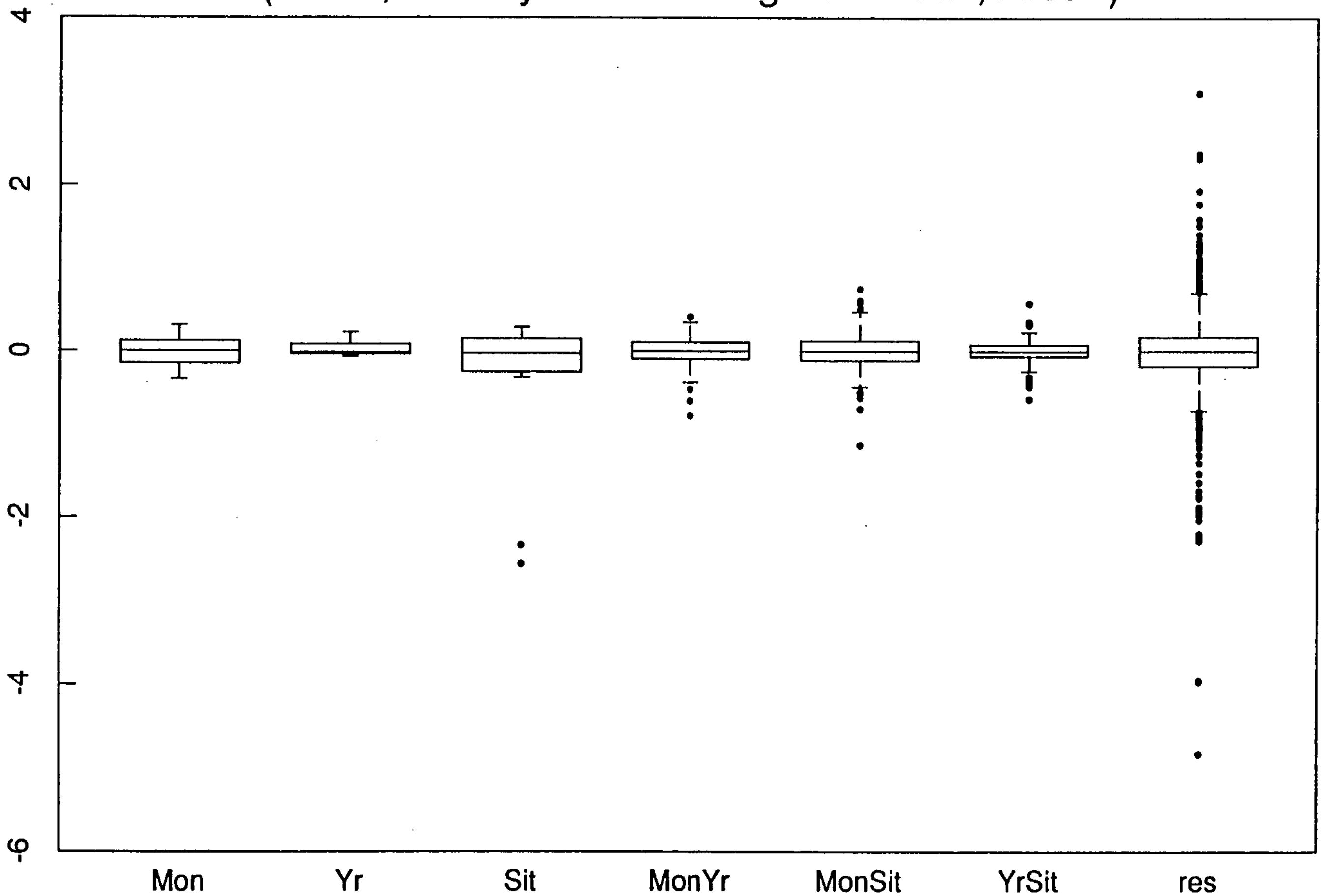


Figure 5.1.15(a)

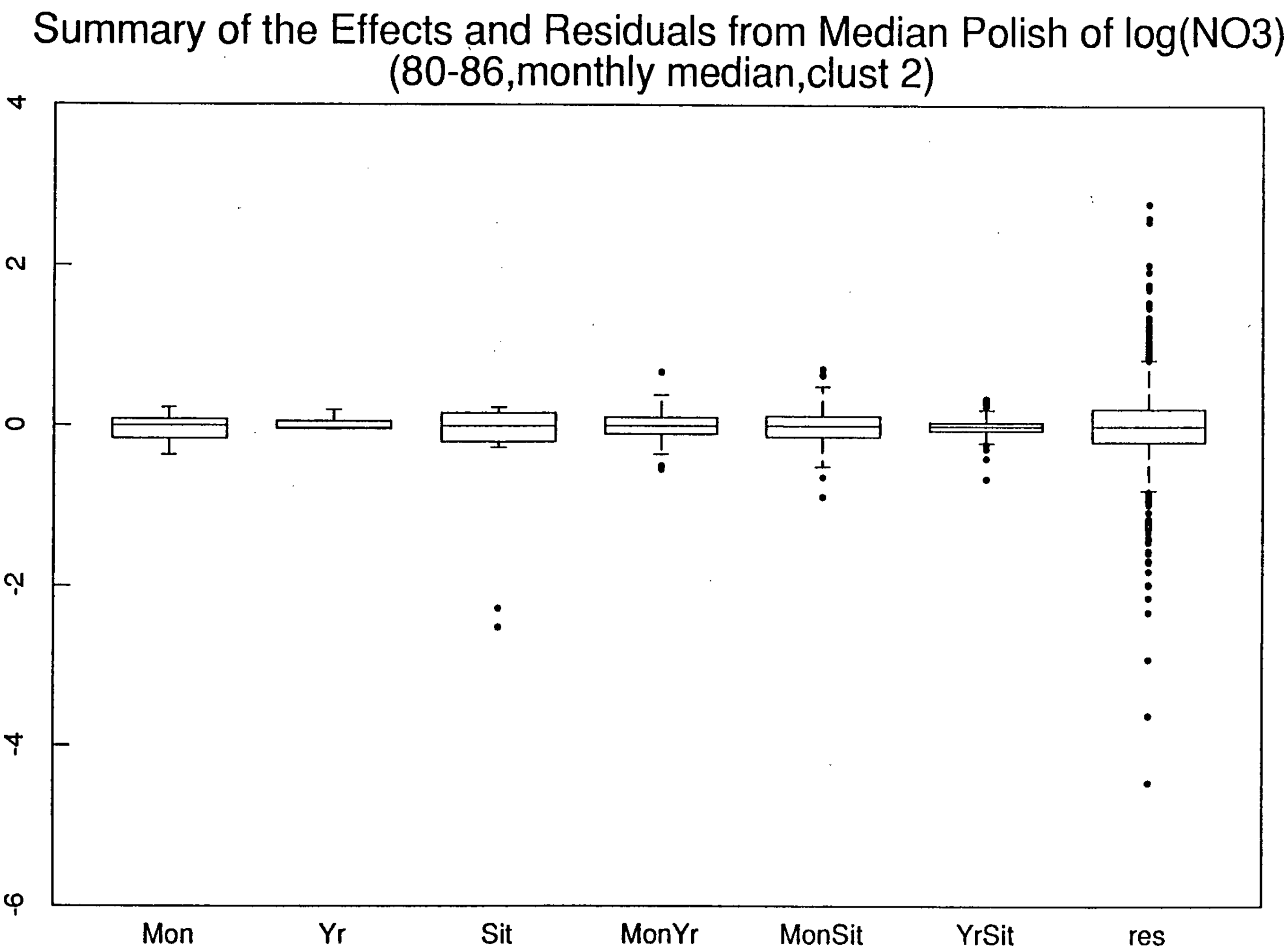


Figure 5.1.15(b)

Trend of log(NO3) from 1980 to 1986 at the 31 Stations
(calculated by monthly volume weighted mean)

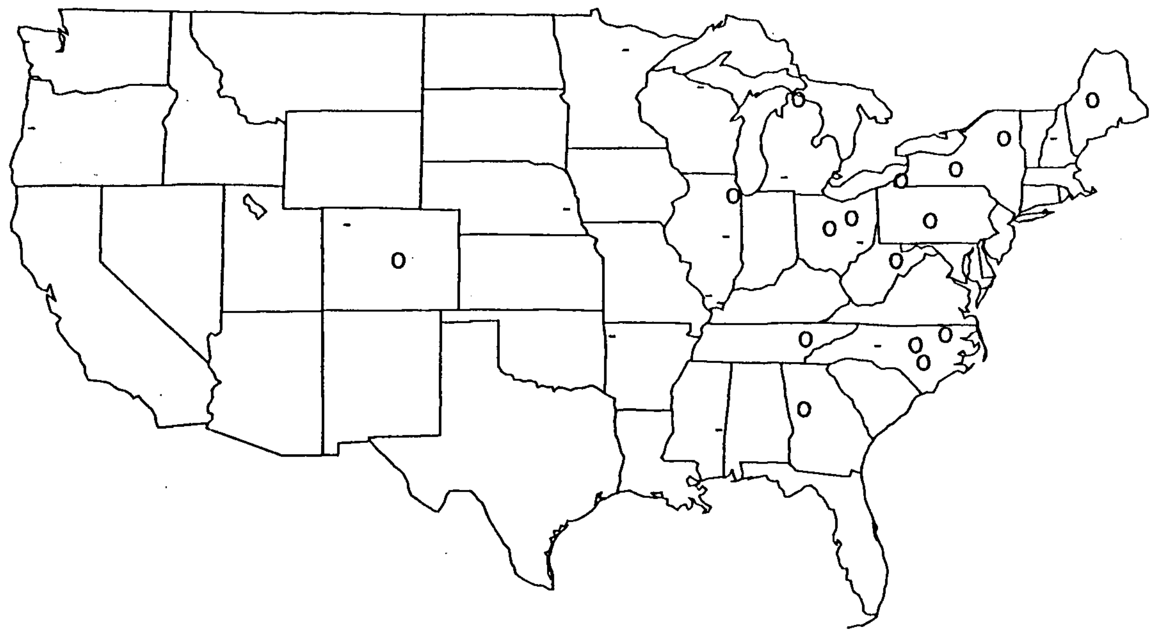


Figure 5.1.16(a)

Trend of log(NO3) from 1980 to 1986 at the 31 Stations
(calculated by monthly median)

Figure 5.1.16(b)

Trend of log(NO3) from 1980 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

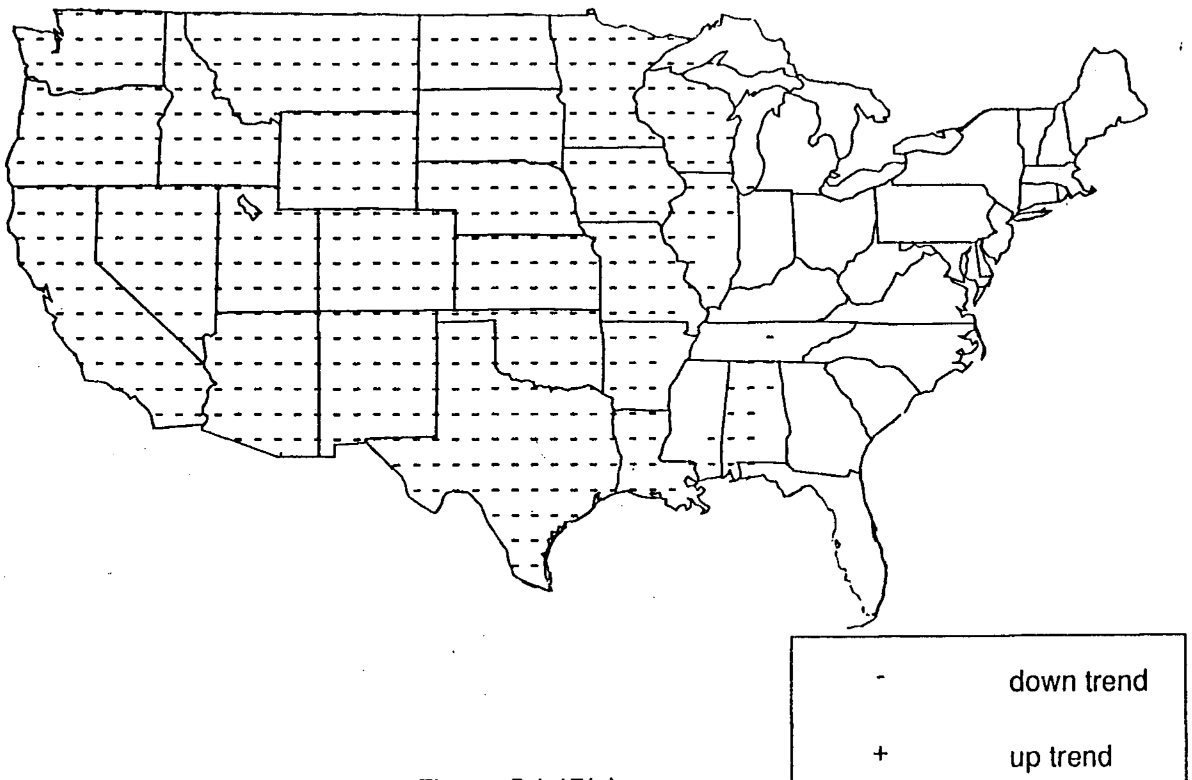


Figure 5.1.17(a)

Trend of log(NO3) from 1980 to 1986 in the USA
(calculated by Kriging from monthly median)

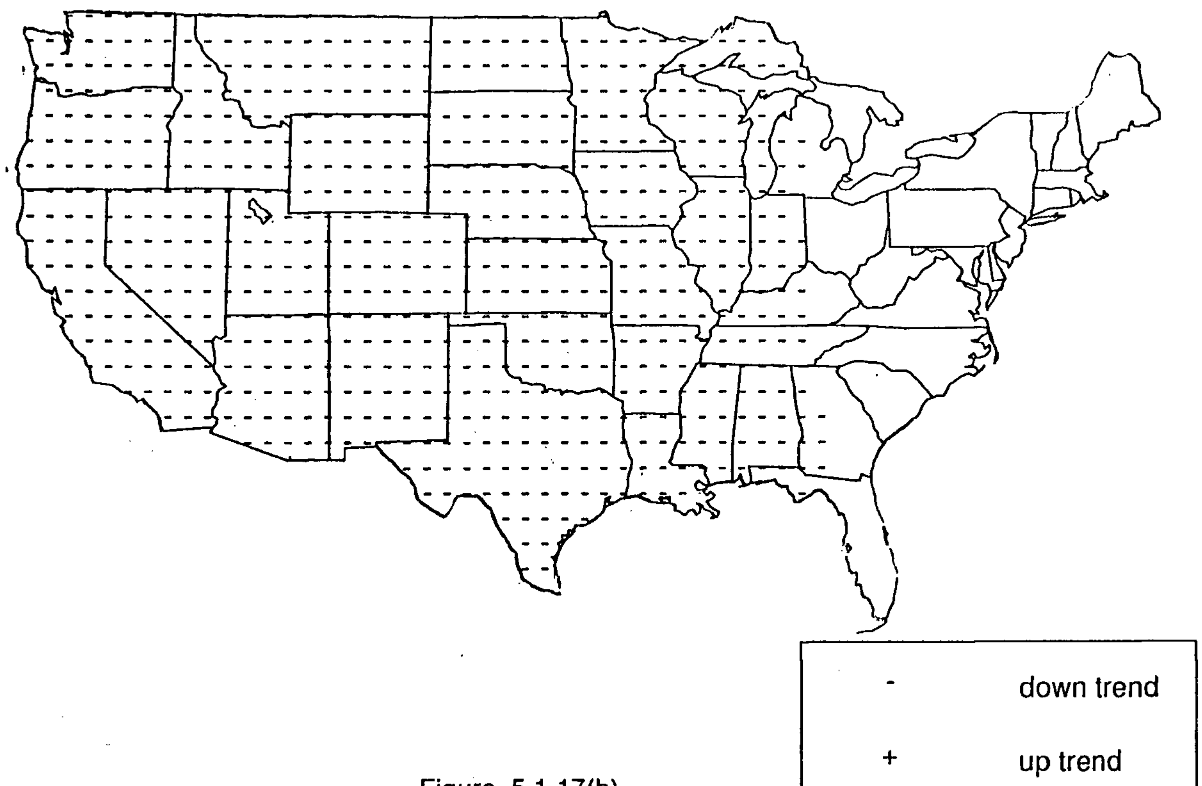


Figure 5.1.17(b)

The Locations Of The 81 Monitoring Stations From 1983 To 1986
(For Nitrate)

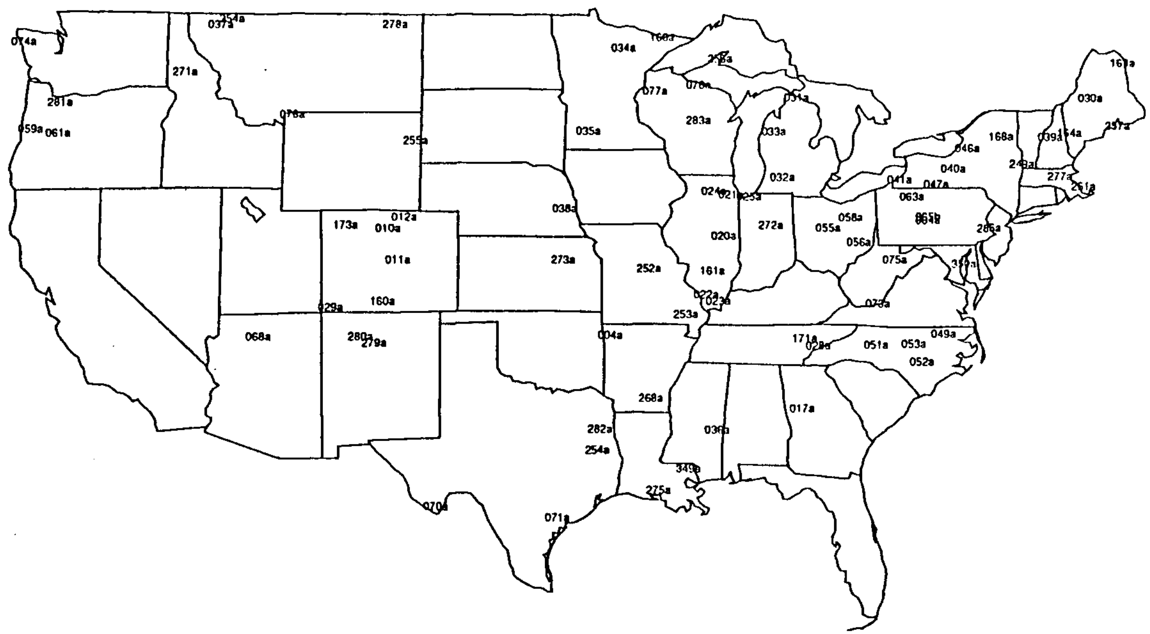


Figure 5.2.1

Clustering of NO3 monthly volume weighted mean based on sqrt(MSE)

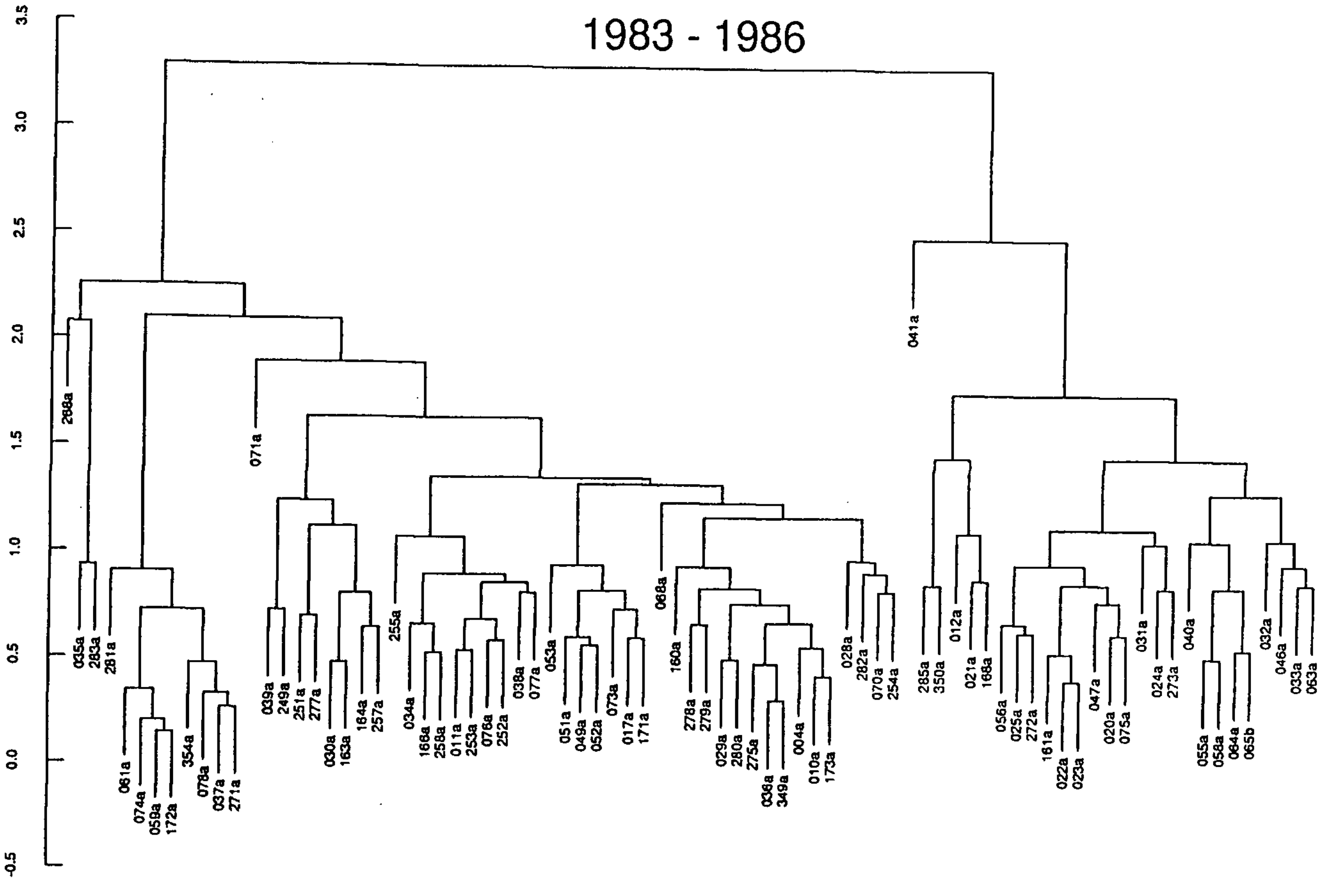


Figure 5.2.2(a)

Clustering of NO3 monthly volume weighted mean based on sqrt(MSE)

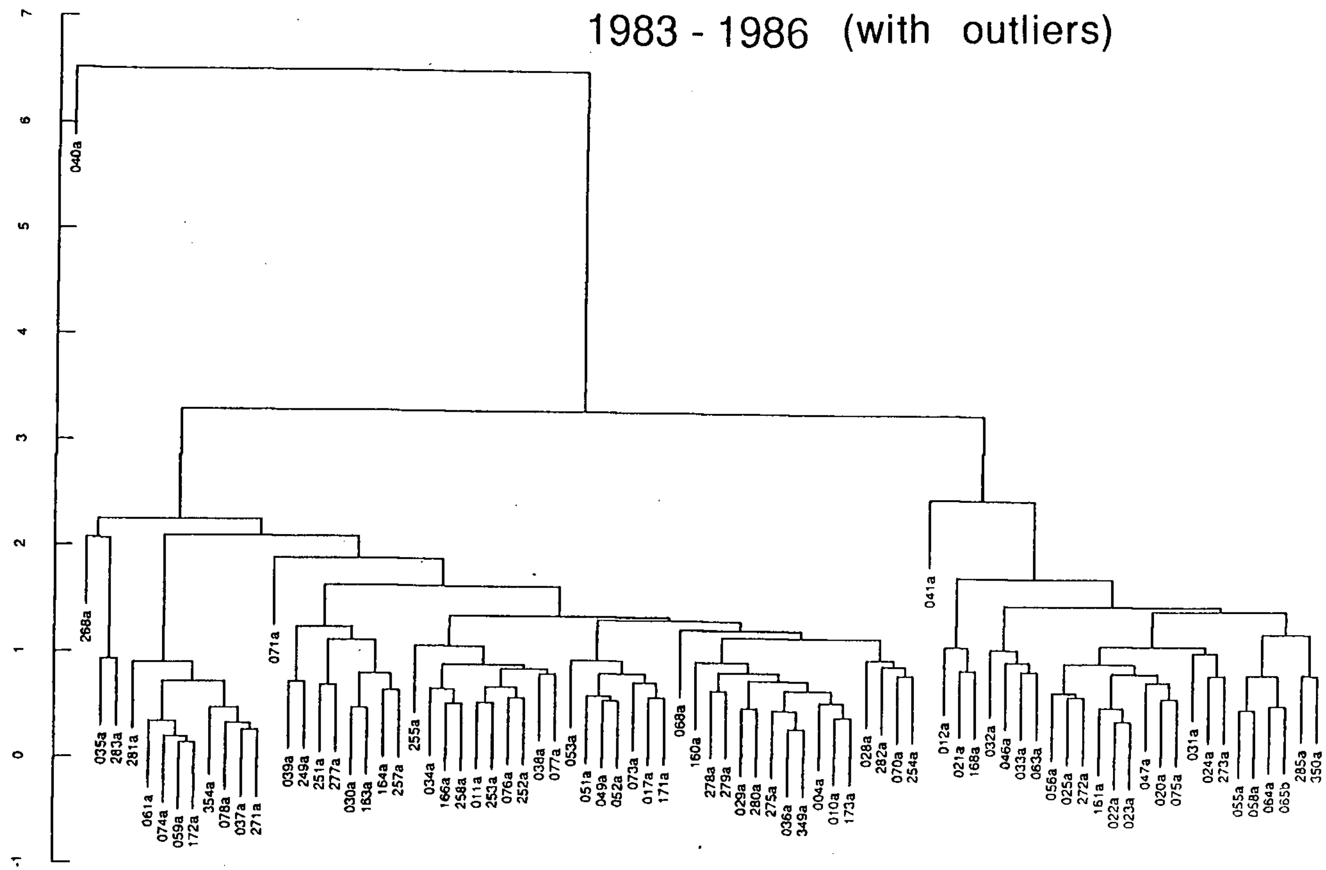


Figure 5.2.2(b)

Clusters of NO3 monthly volume weighted mean based on sqrt(MSE)

1983 - 1986 (k=3)

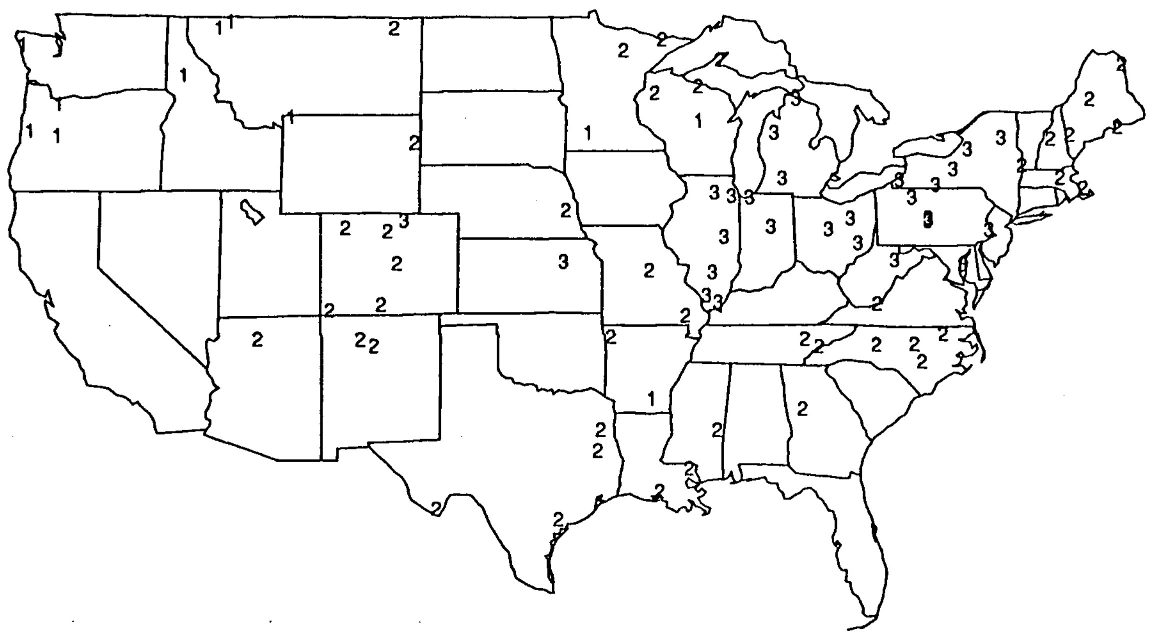


Figure 5.2.3

Histograms of Transformed NO3 (Volume Weighted Mean, 83-86)

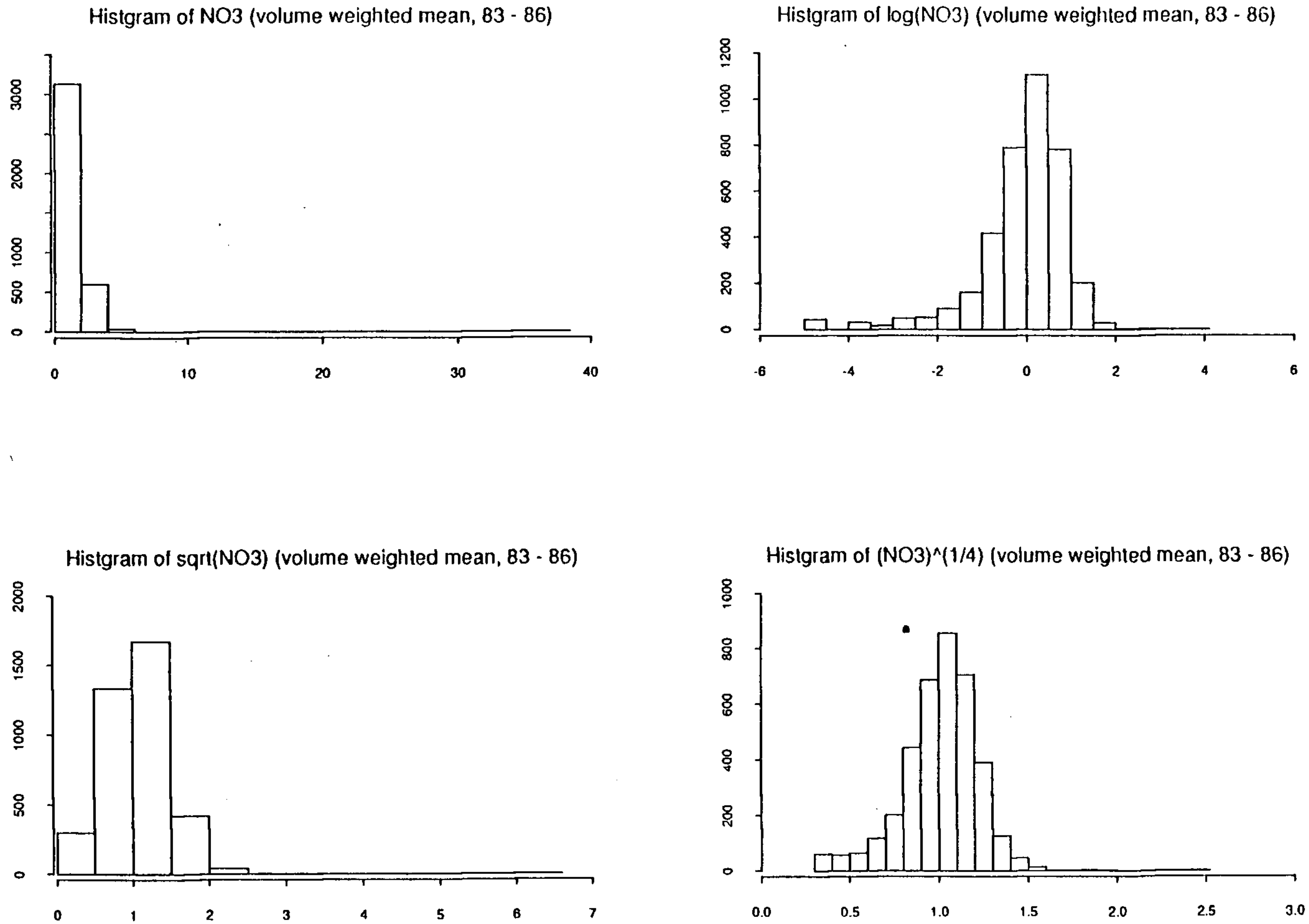


Figure 5.2.4

Histograms of log(NO3) by Clusters (Volume Weighted Mean, 83-86)

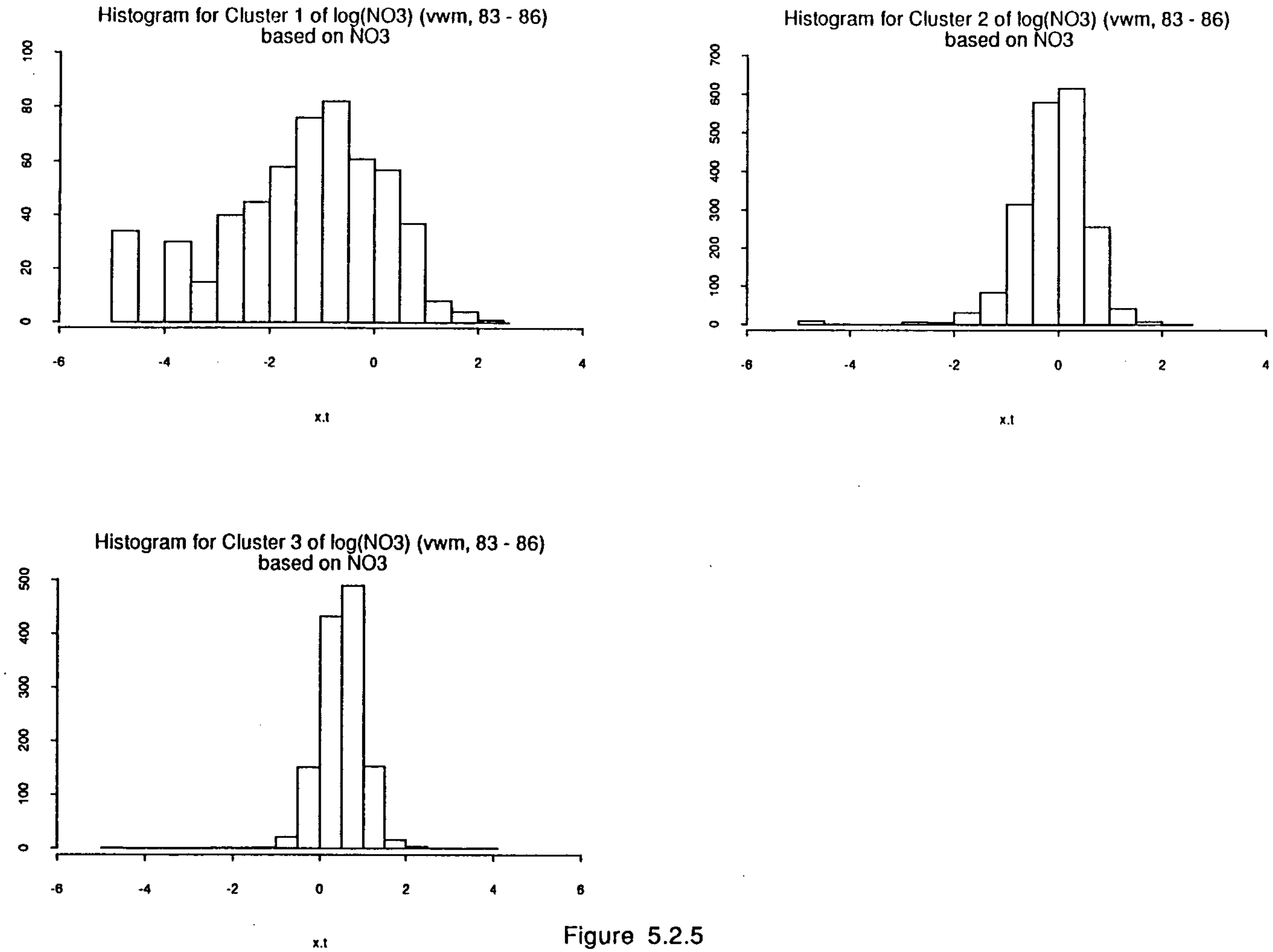


Figure 5.2.5

Yearly Effect of log(NO3) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)

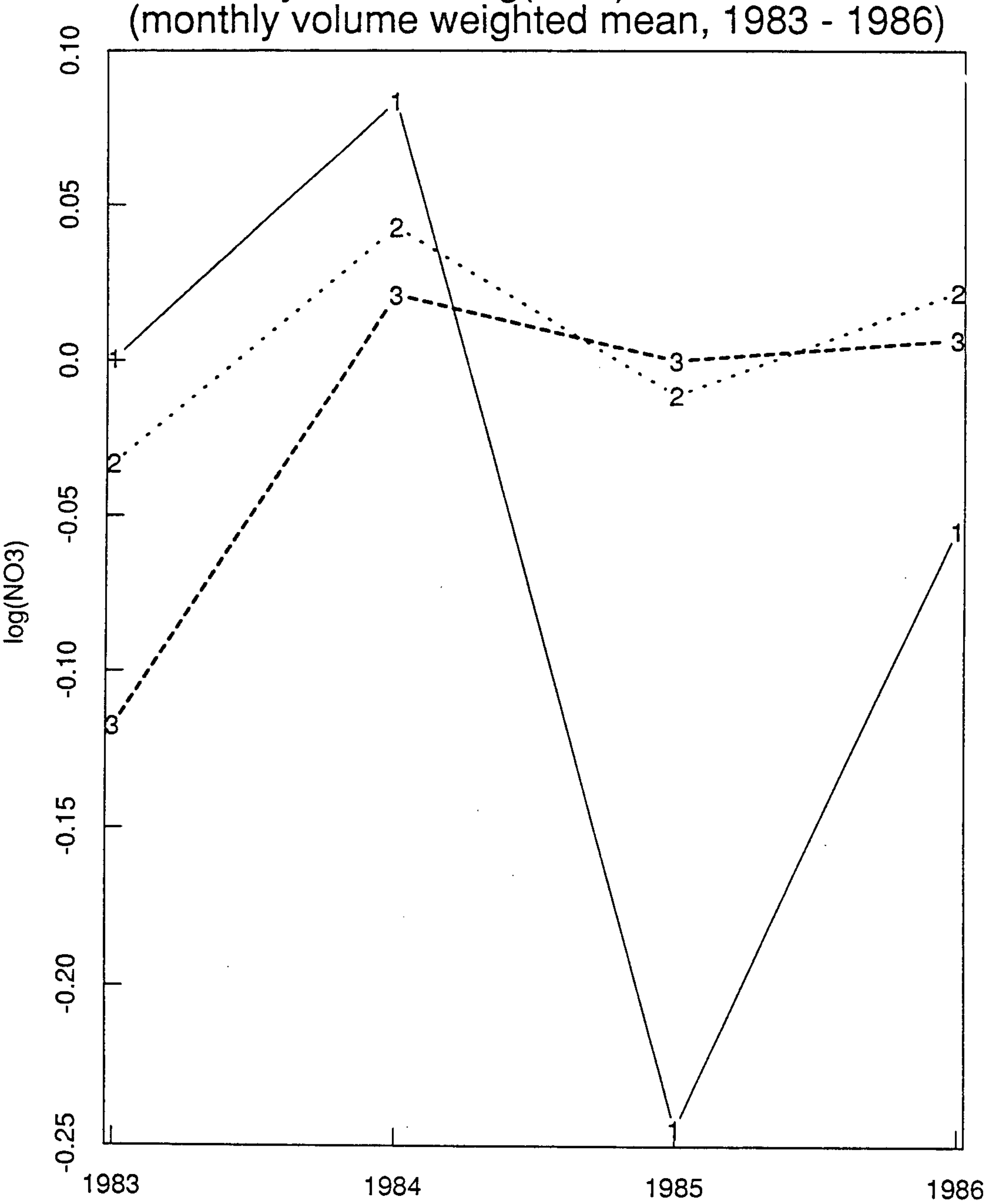


Figure 5.2.6

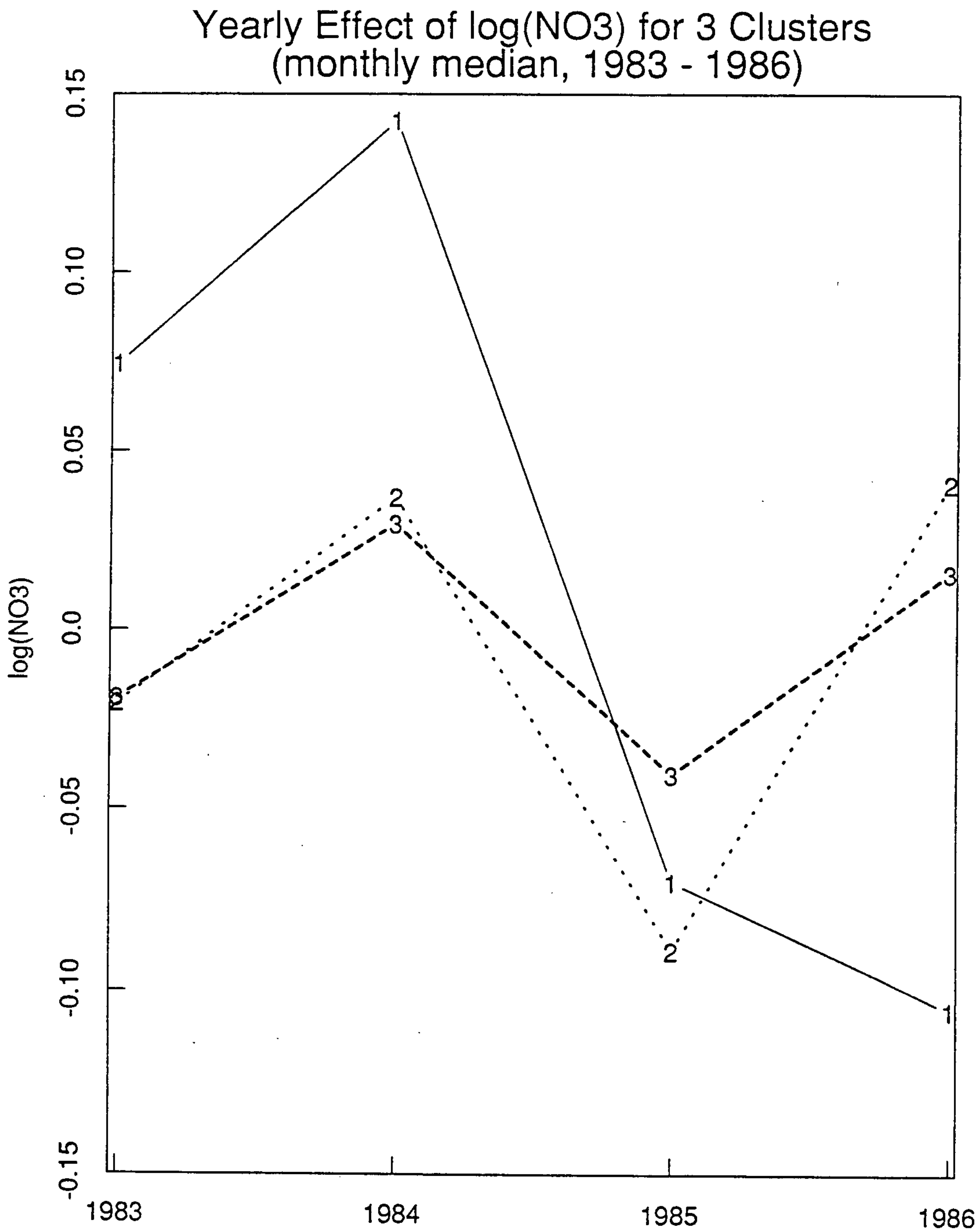
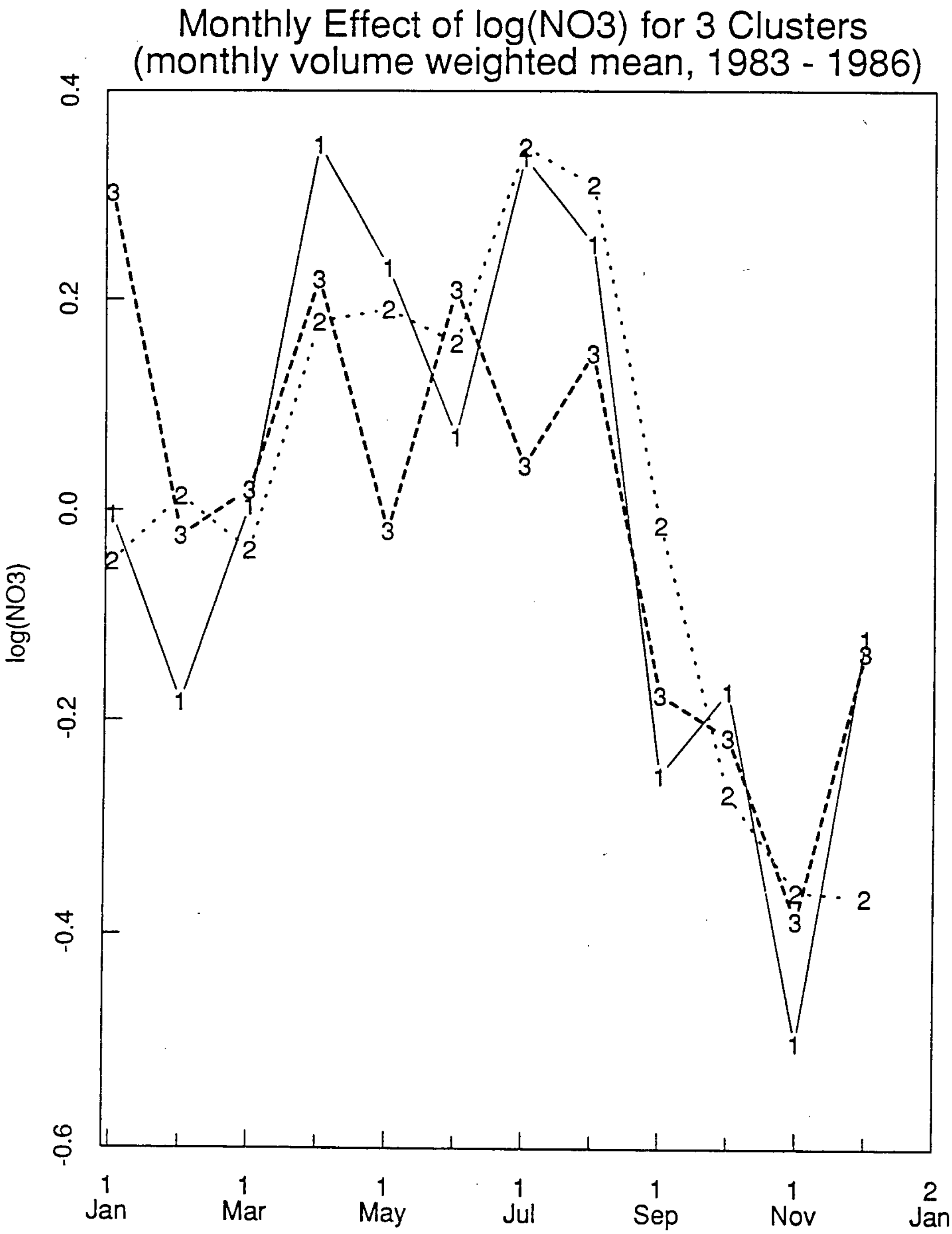

Yearly Effect of log(NO3) for 3 Clusters
(monthly median, 1983 - 1986)
log(NO3)
0.15
0.10
0.05
0.0
-0.05
-0.10
-0.15
1983
1984
1985
1986

Figure 5.2.7



Monthly Effect of log(NO3) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)
log(NO3)
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 5.2.8

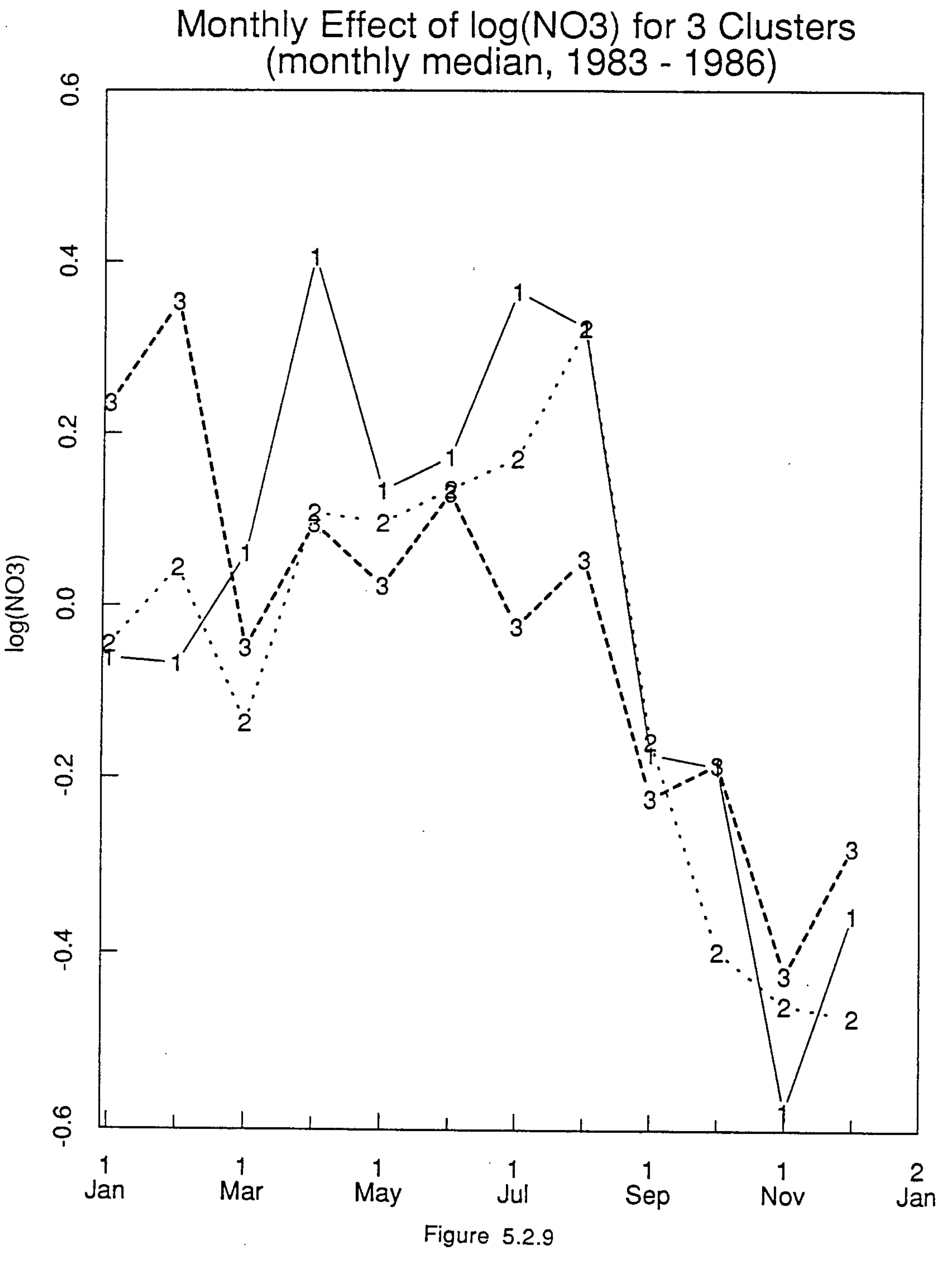

Monthly Effect of log(NO3) for 3 Clusters
(monthly median, 1983 - 1986)
log(NO3)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 5.2.9

Station Effect of log(NO3) for Cluster 1
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)

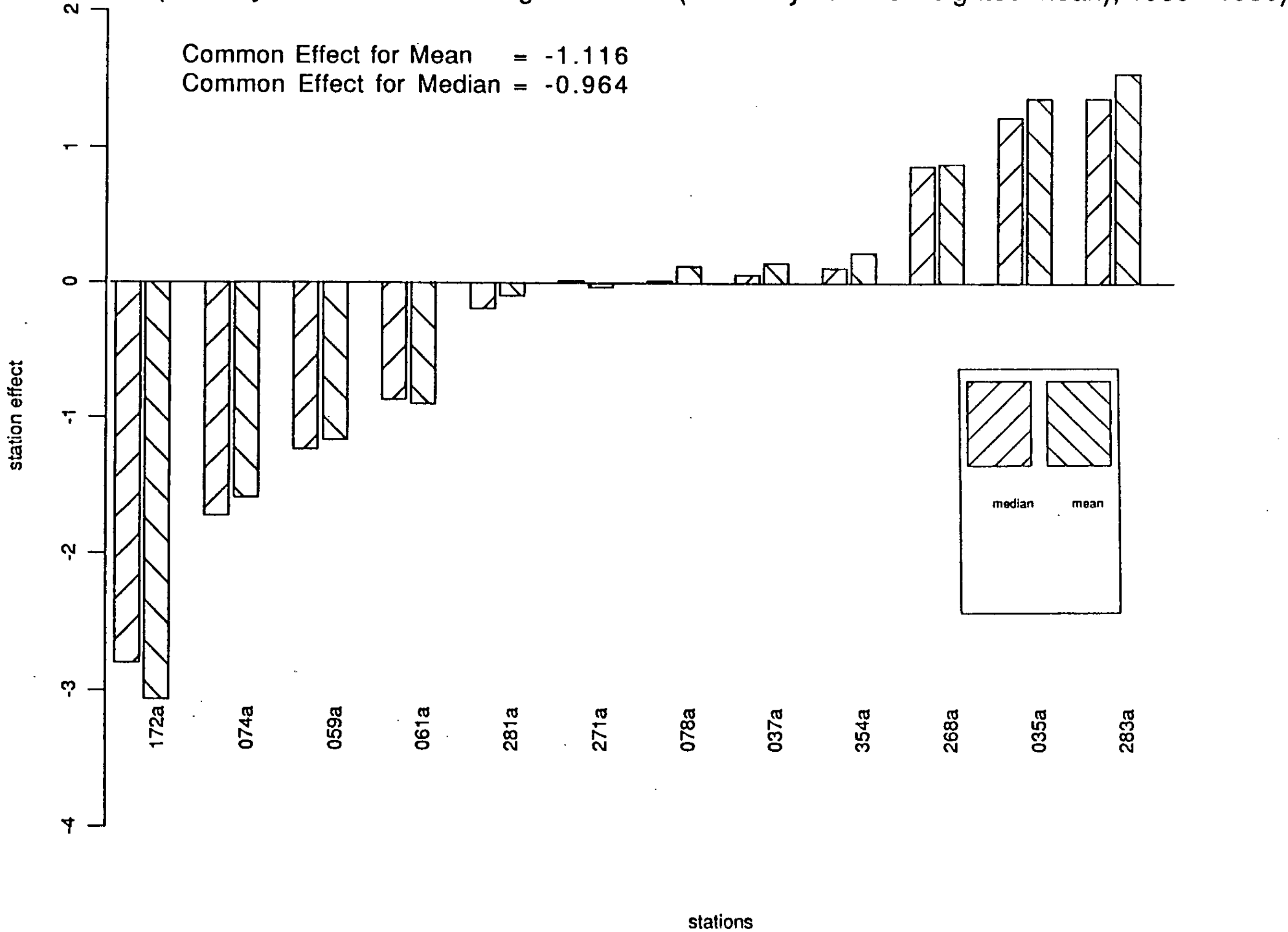
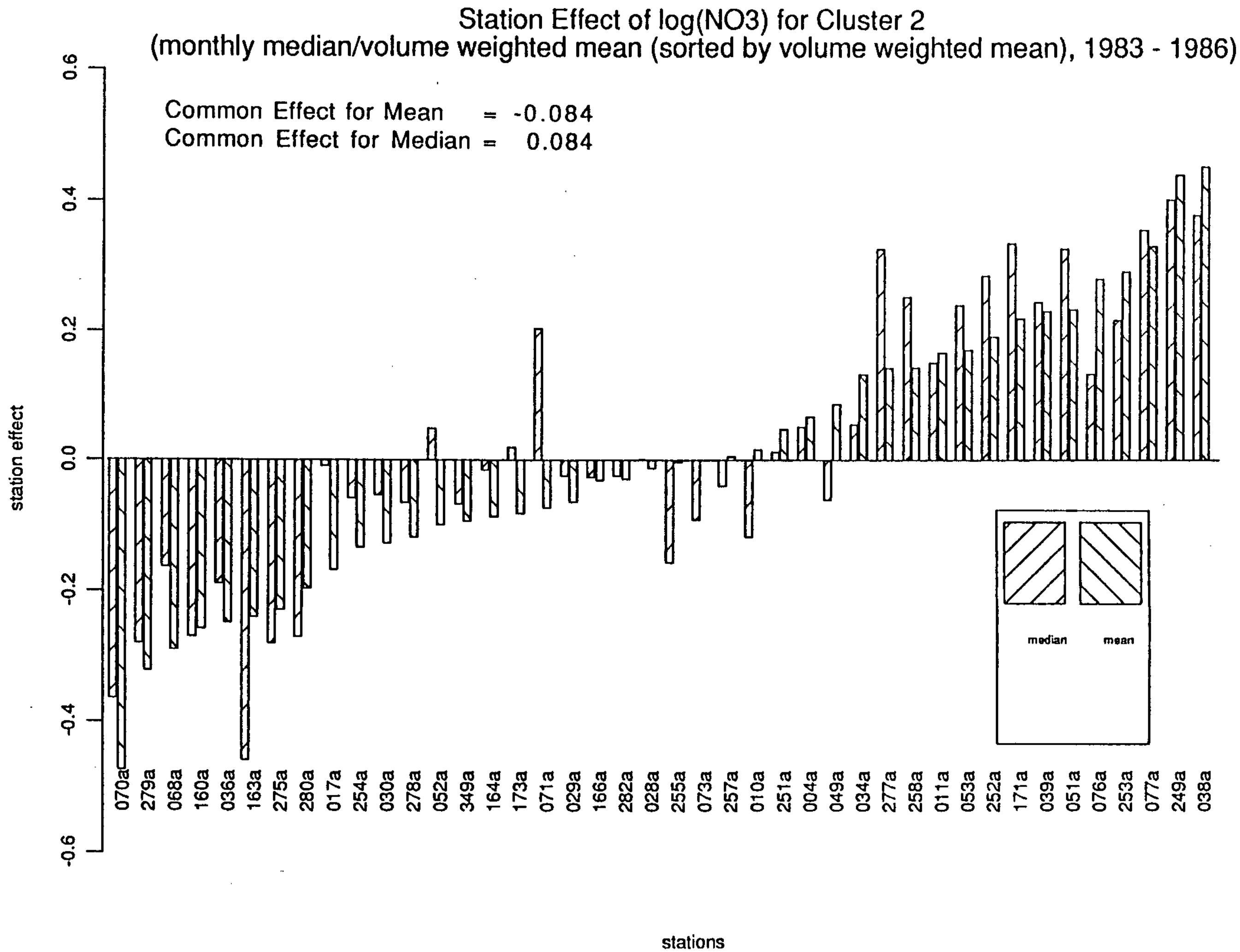


Figure 5.2.10


Station Effect of log(NO3) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = -0.084
Common Effect for Median = 0.084
station effect
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
070a 279a 068a 160a 036a 163a 275a 280a 017a 254a 030a 278a 052a 349a 164a 173a 071a 029a 166a 282a 028a 255a 073a 257a 010a 251a 004a 049a 034a 277a 258a 011a 053a 252a 171a 039a 051a 076a 253a 077a 249a 038a
median
mean
stations
Figure 5.2.11

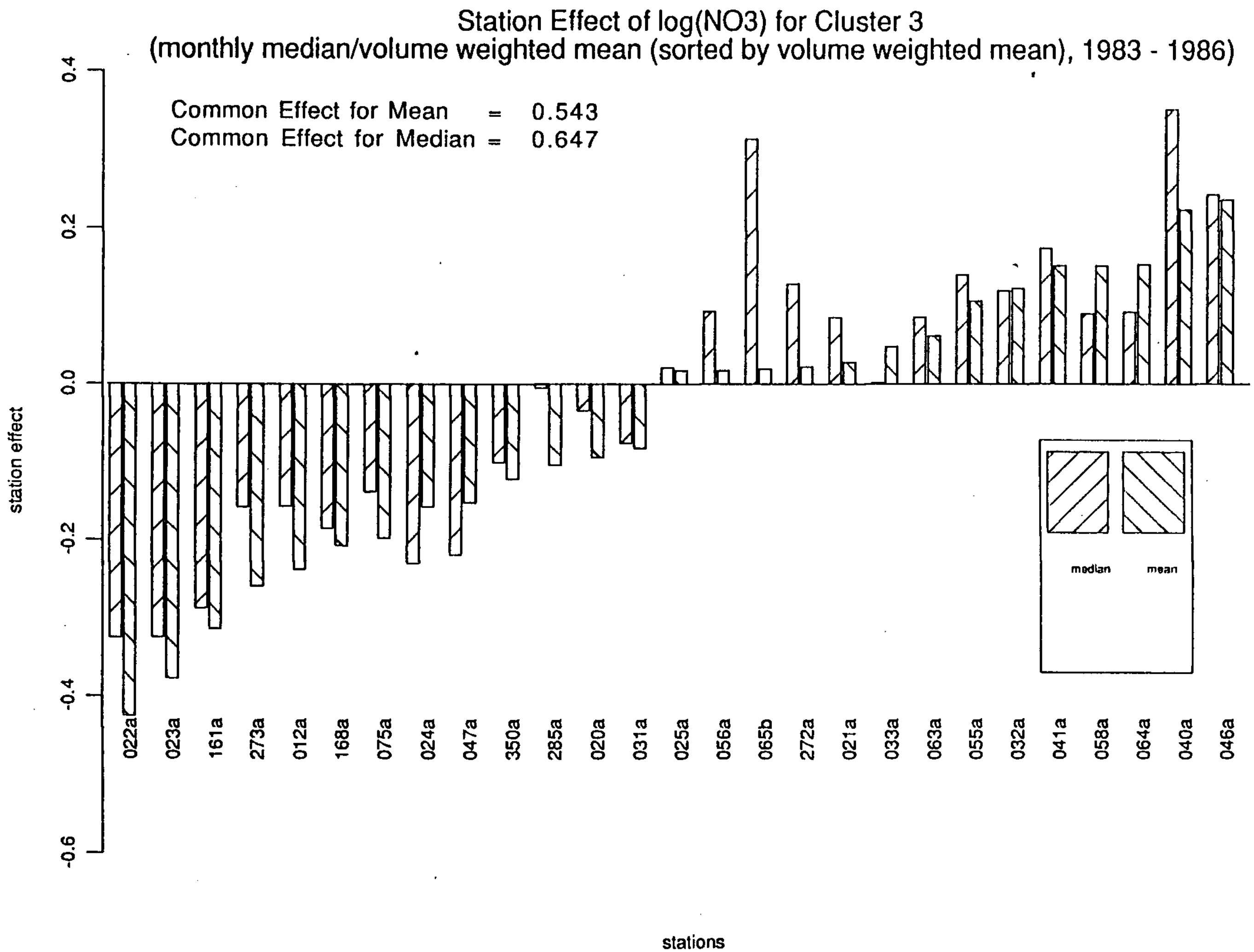

Station Effect of log(NO3) for Cluster 3
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 0.543
Common Effect for Median = 0.647
station effect
0.4
0.2
0.0
-0.2
-0.4
-0.6
022a
023a
161a
273a
012a
168a
075a
024a
047a
350a
285a
020a
031a
025a
056a
065b
272a
021a
033a
063a
055a
032a
041a
058a
064a
040a
046a
median
mean
stations

Figure 5.2.12

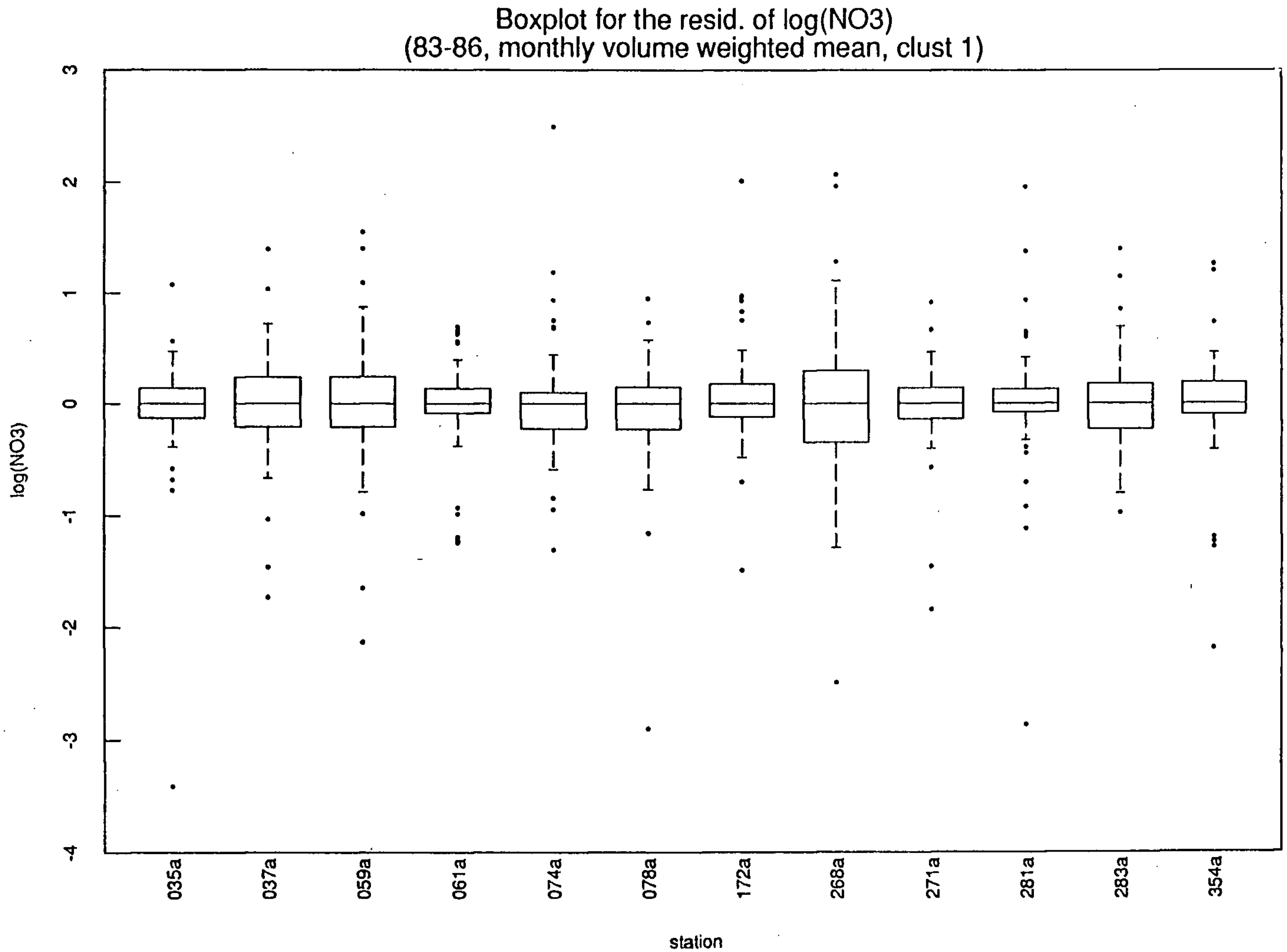


Figure 5.2.13(a)

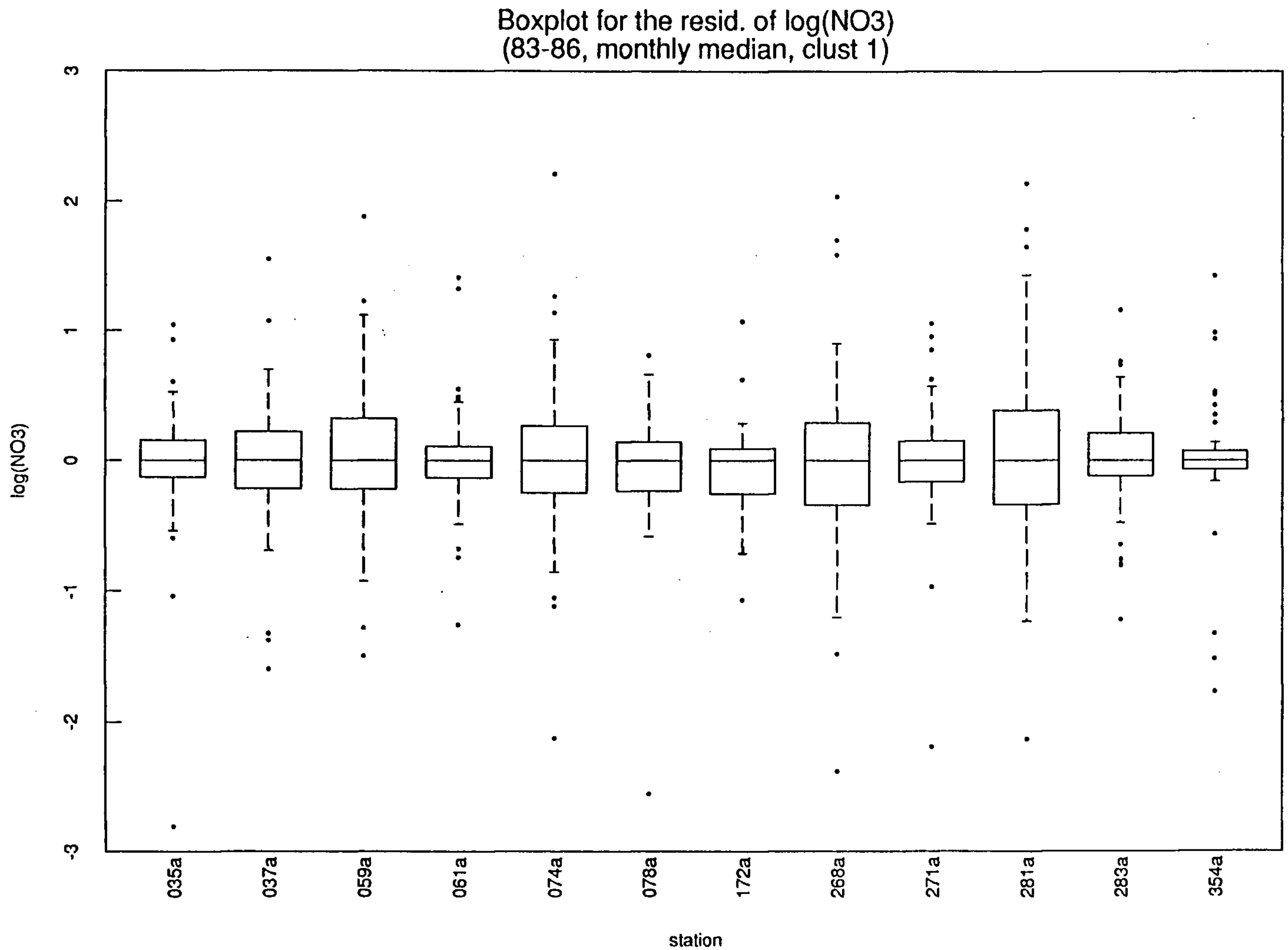


Figure 5.2.13(b)

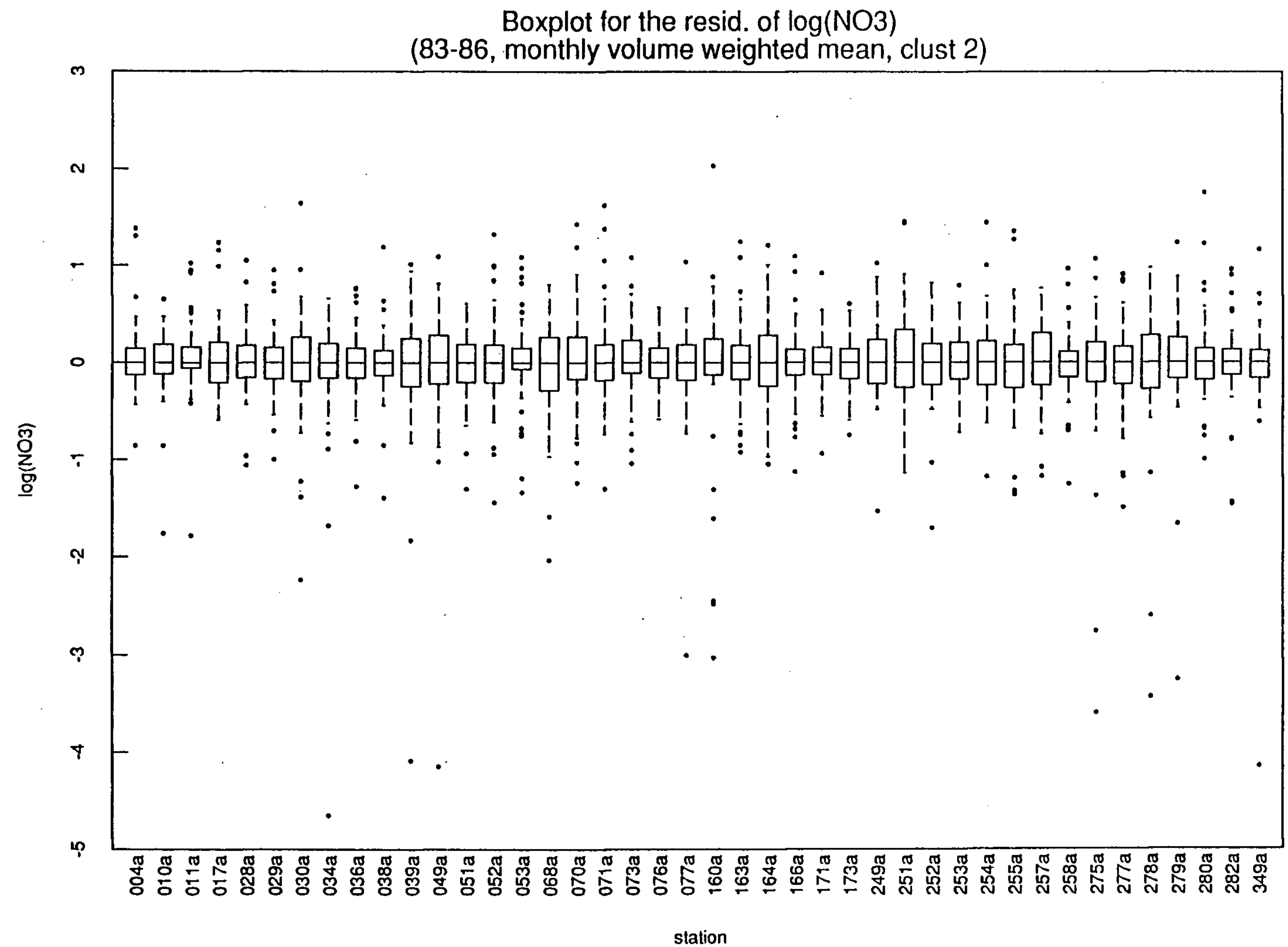


Figure 5.2.14(a)

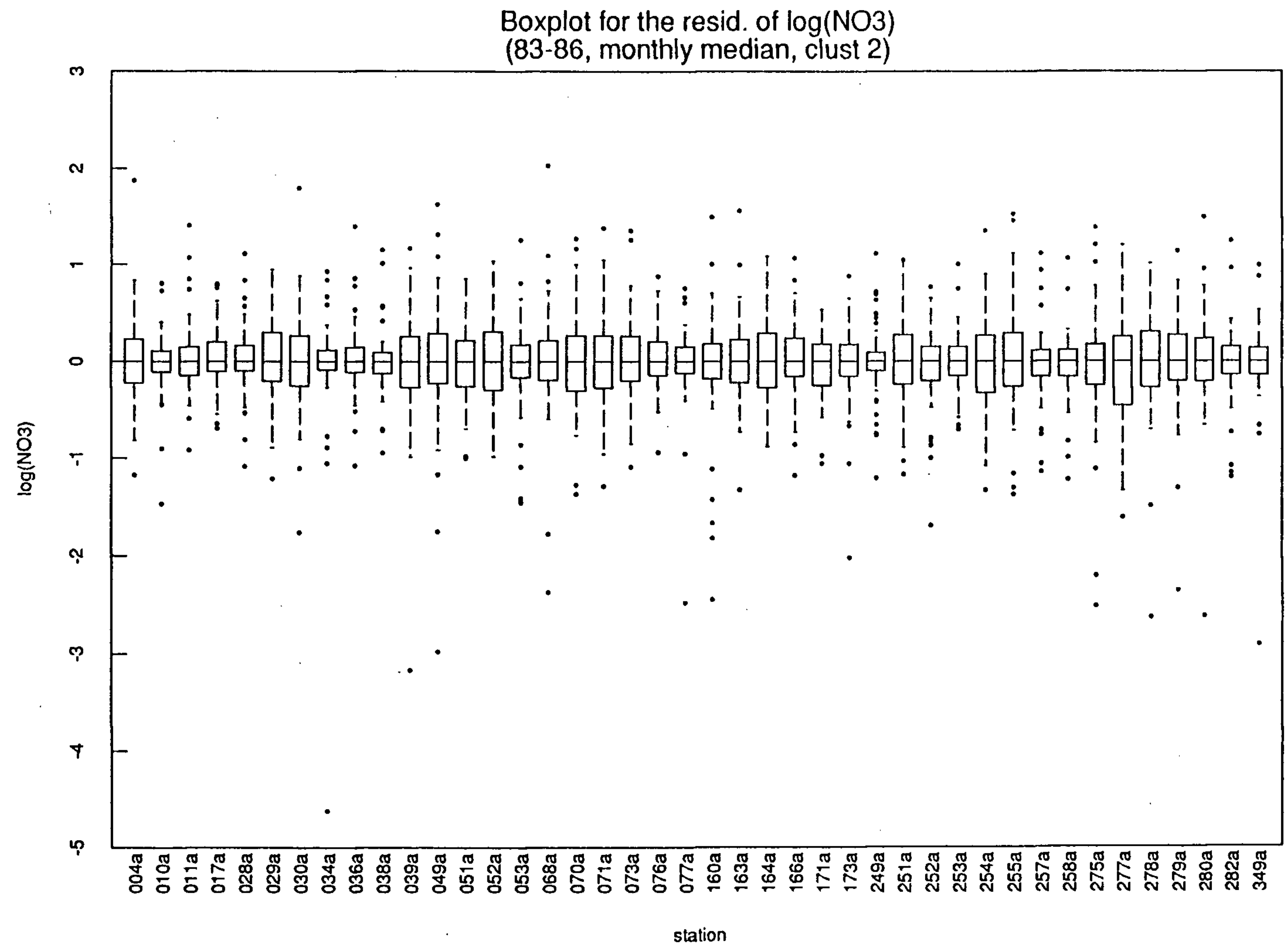


Figure 5.2.14(b)

Boxplot for the resid. of log(NO3)
(83-86, monthly volume weighted mean, clust 3)

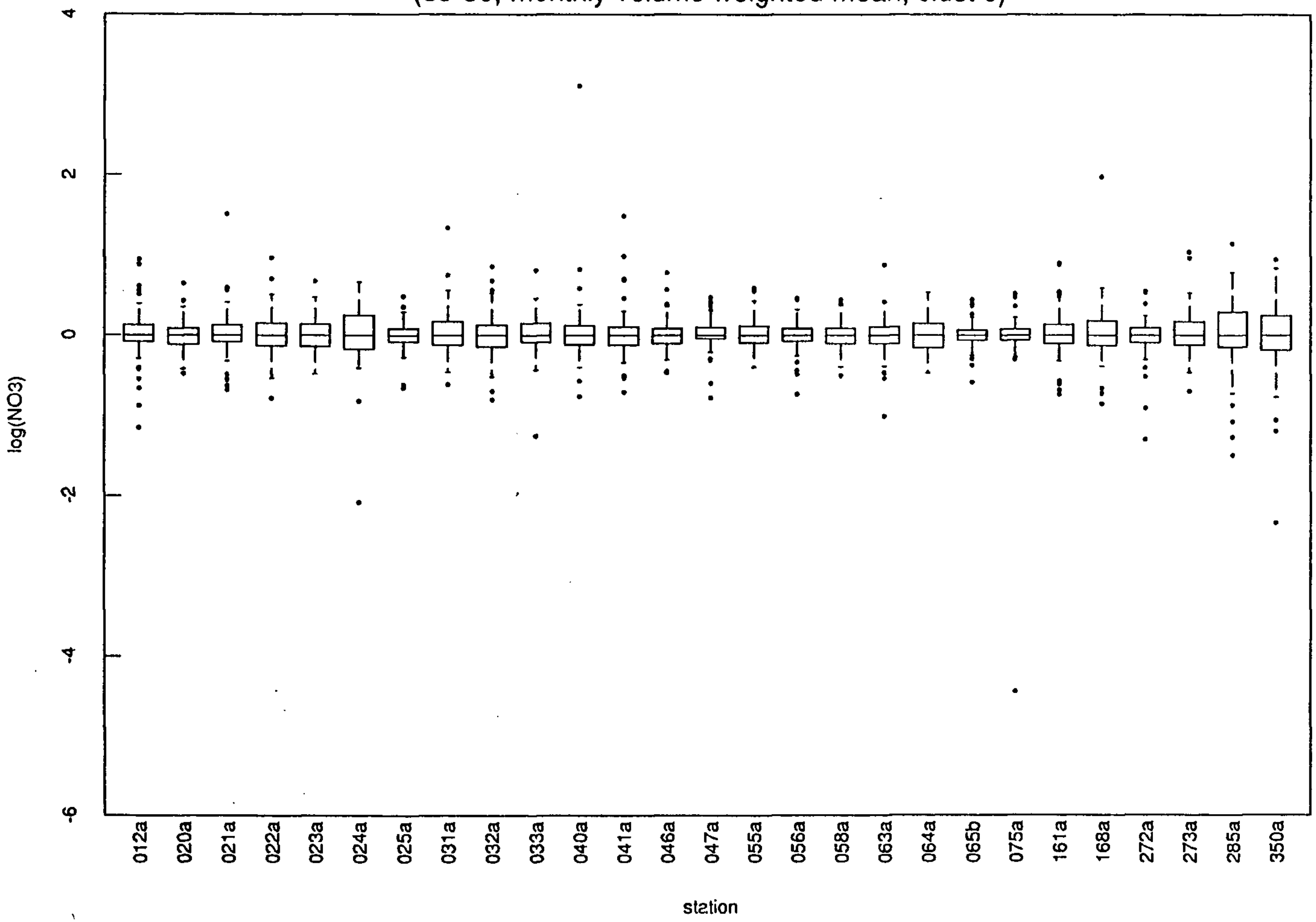


Figure 5.2.15(a)

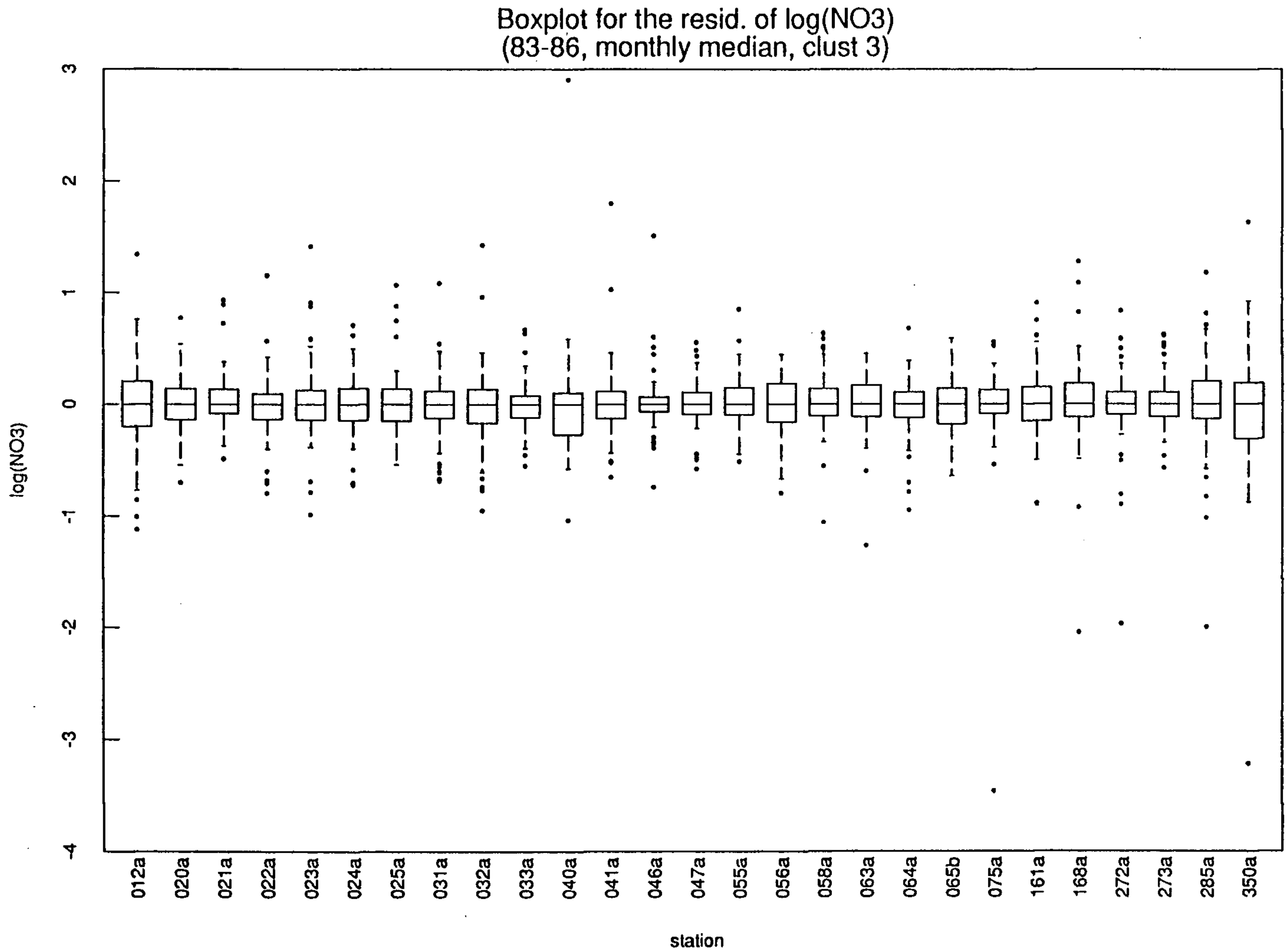


Figure 5.2.15(b)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(83-86,monthly volume weighted mean,clust 1)

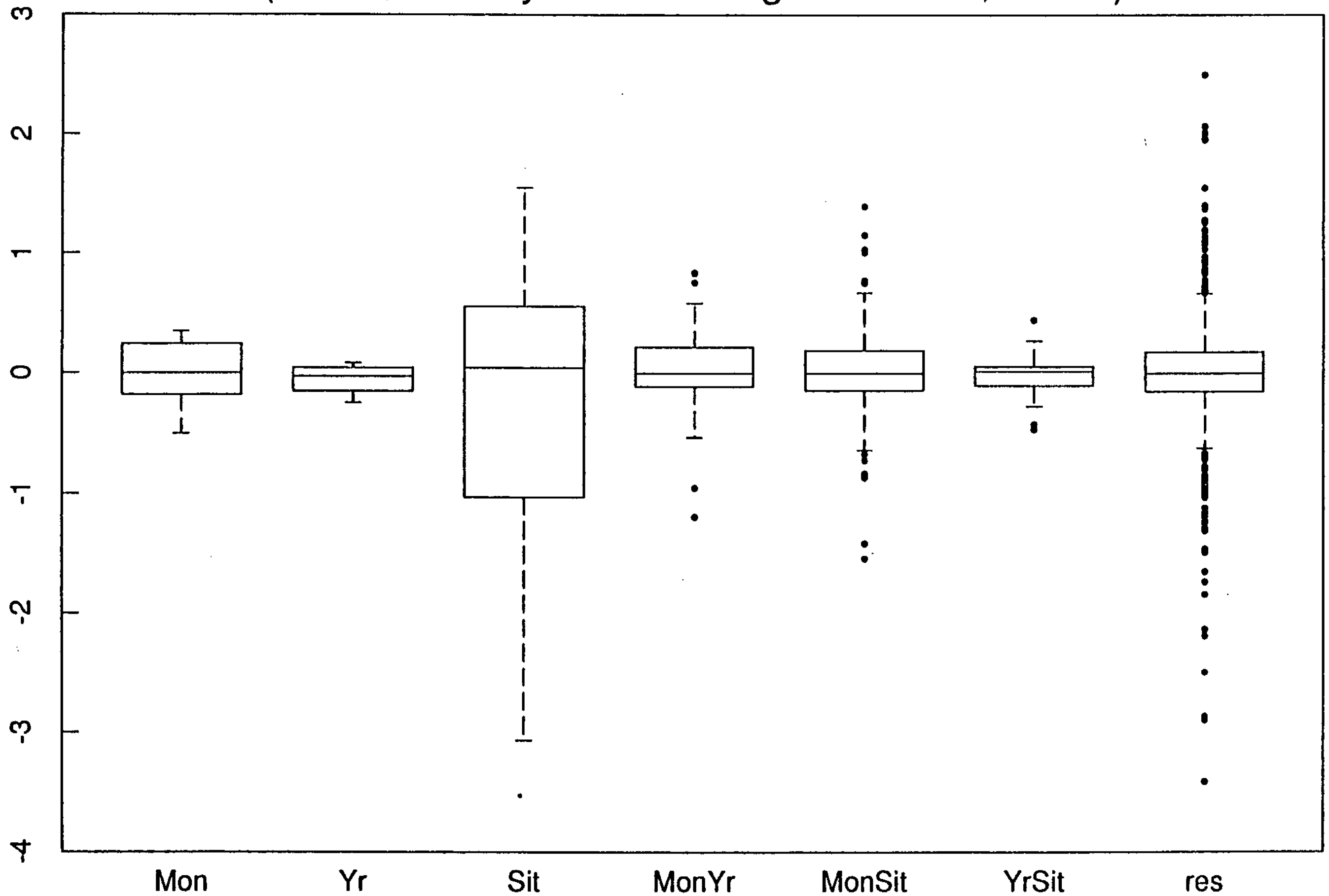


Figure 5.2.16(a)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(83-86,monthly median,clust 1)

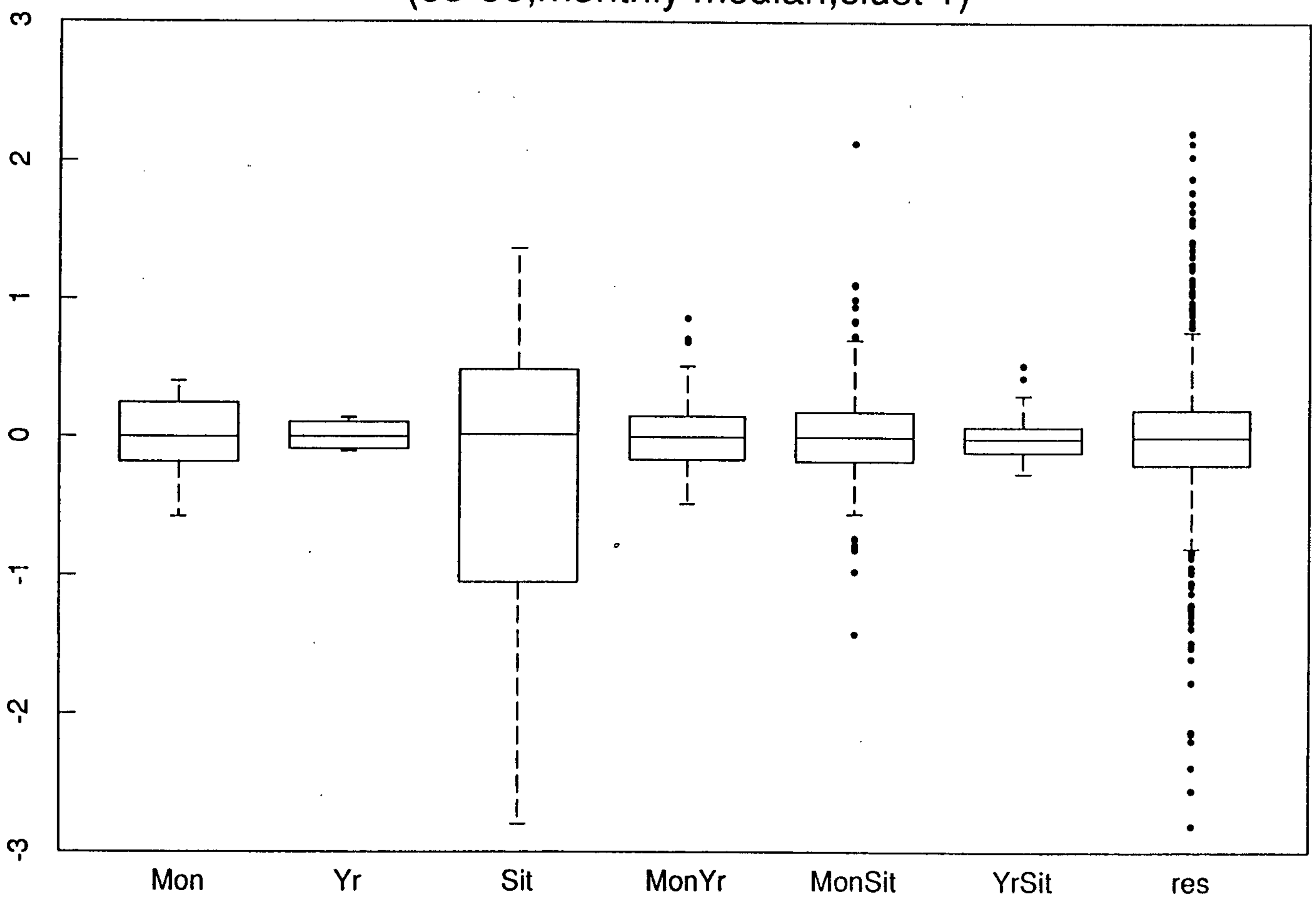


Figure 5.2.16(b)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(83-86,monthly volume weighted mean,clust 2)

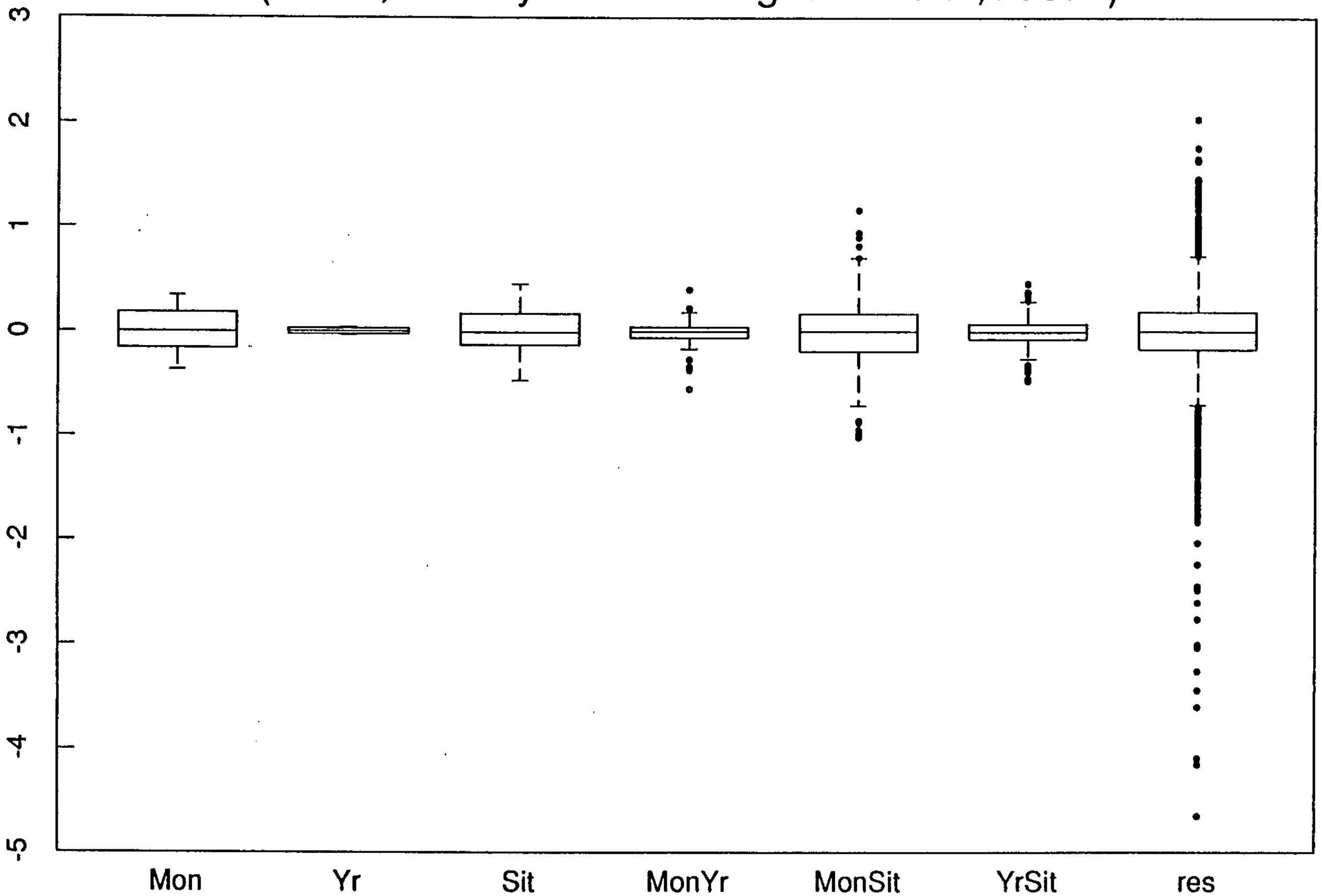


Figure 5.2.17(a)

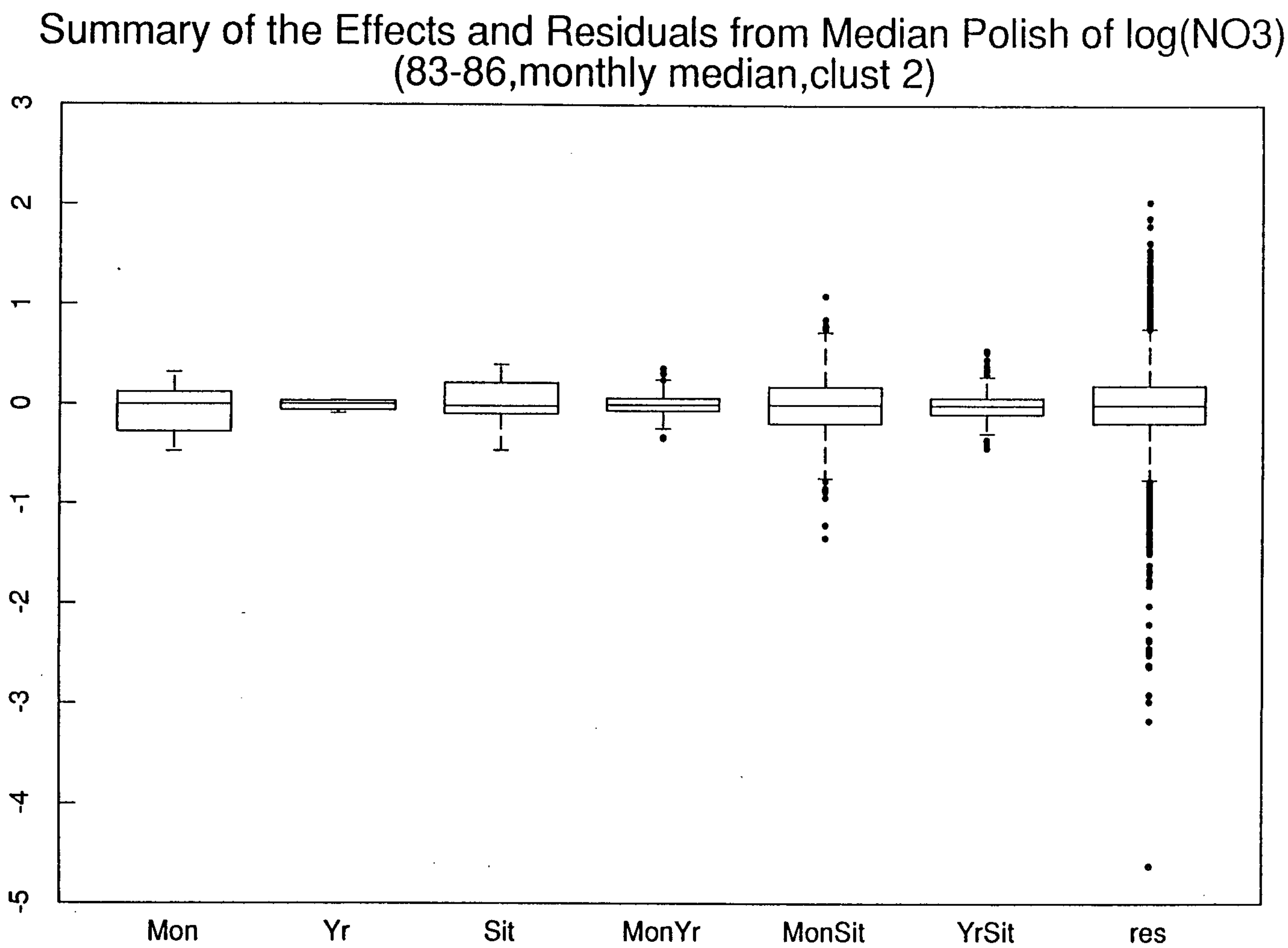


Figure 5.2.17(b)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(83-86,monthly volume weighted mean,clust 3)

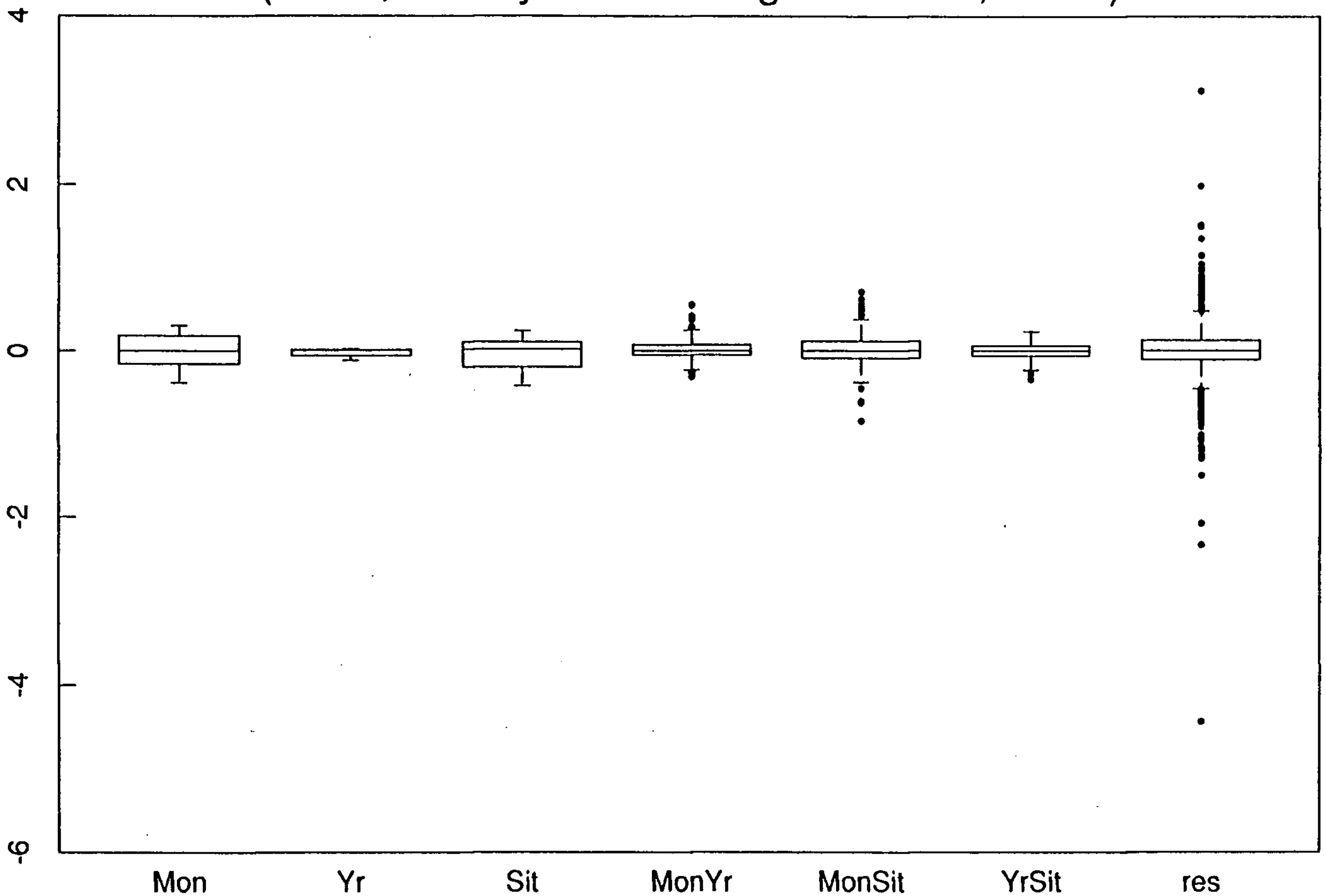


Figure 5.2.18(a)

Summary of the Effects and Residuals from Median Polish of log(NO3)
(83-86,monthly median,clust 3)

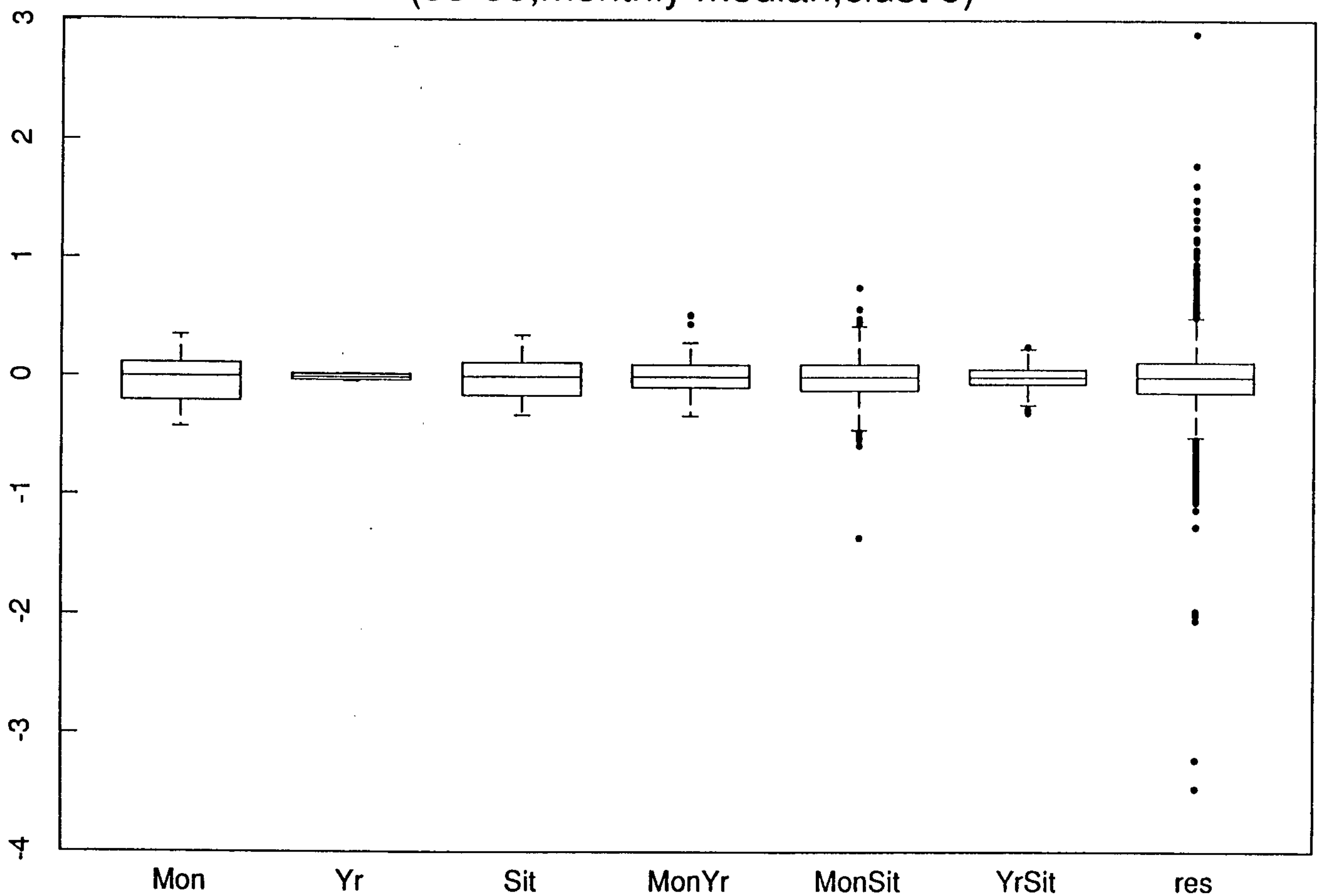


Figure 5.2.18(b)

Trend of log(NO3) from 1983 to 1986 at the 81 Stations
(calculated by monthly volume weighted mean)

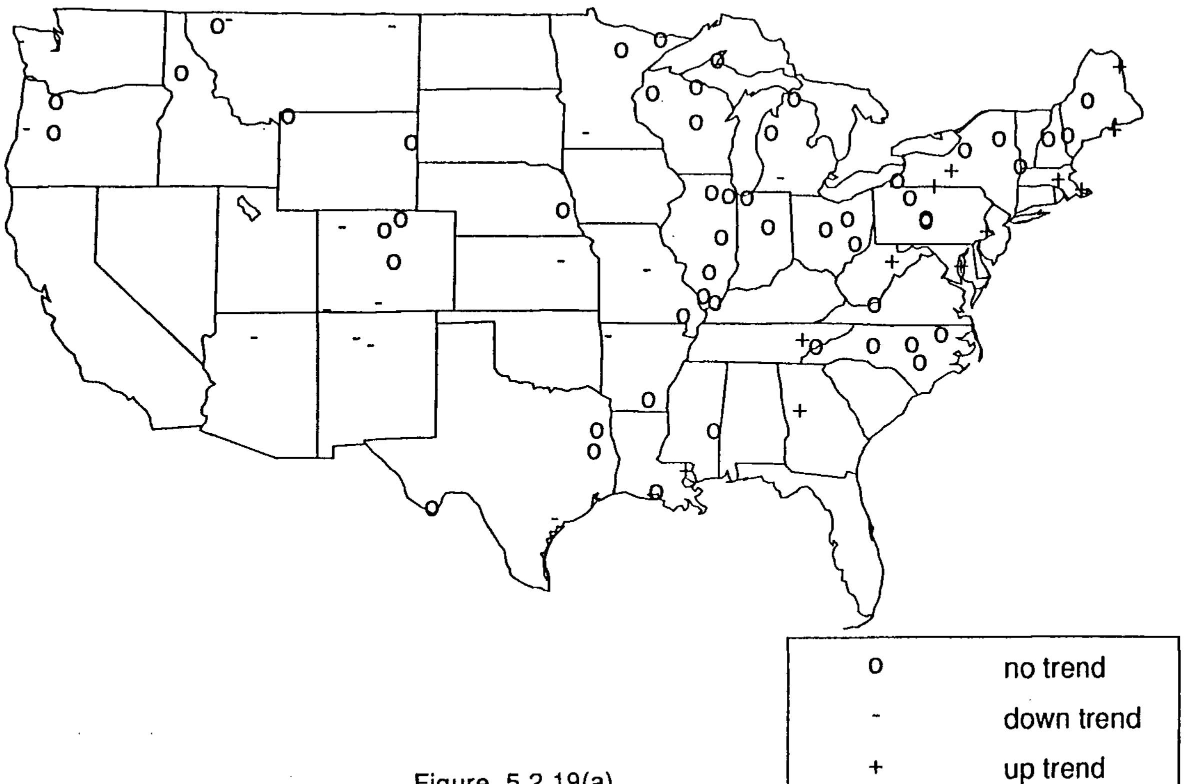


Figure 5.2.19(a)

Trend of log(NO3) from 1983 to 1986 at the 81 Stations
(calculated by monthly median)

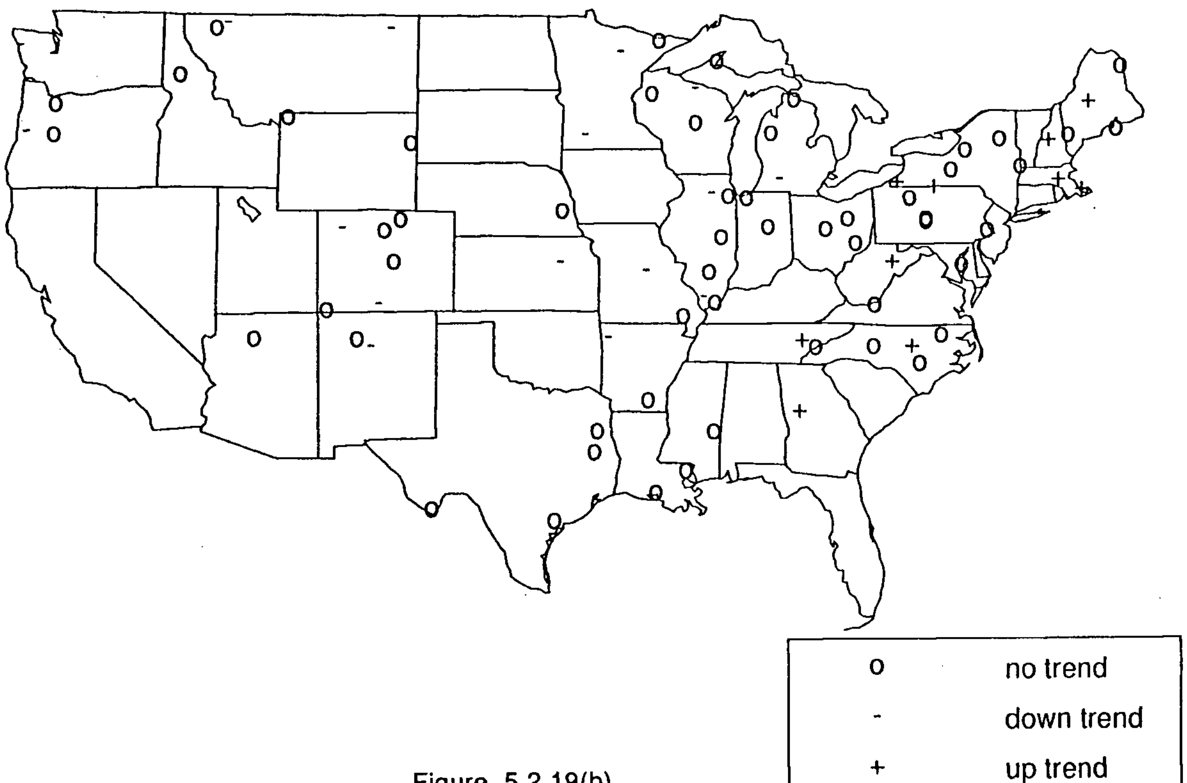


Figure 5.2.19(b)

Trend of log(NO3) from 1983 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

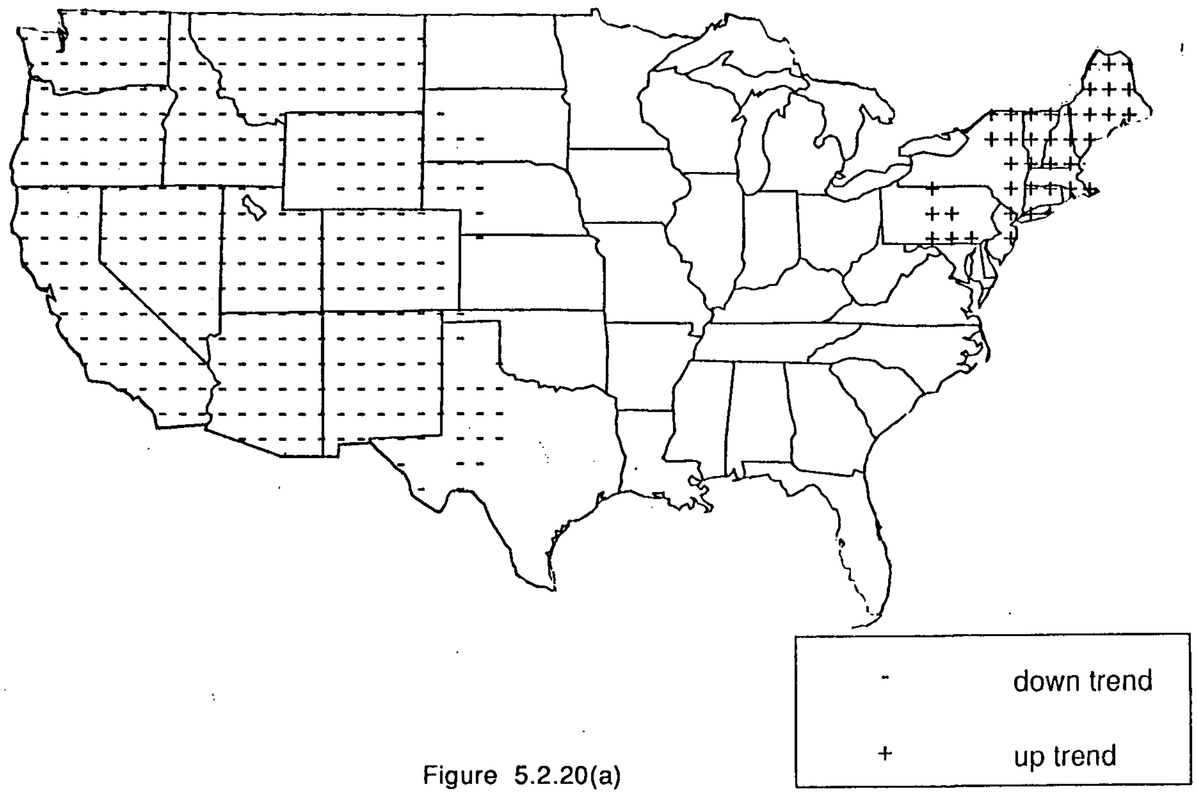


Figure 5.2.20(a)

Trend of log(NO3) from 1983 to 1986 in the USA
(calculated by Kriging from monthly median)

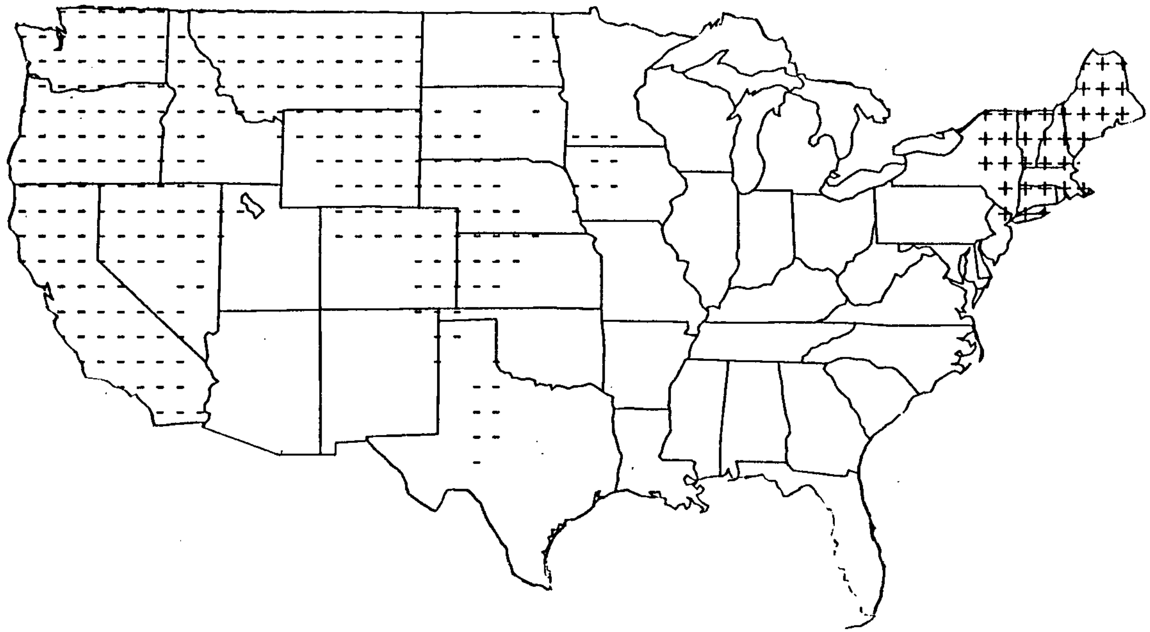


Figure 5.2.20(b)

TABLE 5.1.1

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRARTIONS OF NITRATE
(Monthly Volume Weighted Mean, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.697	0.090	-0.003	-0.005	0.0
011a	-1.021	0.307	-0.002	-0.004	0.001
017a	0.152	0.879	0.0	-0.002	0.002
020a	-2.801	0.005	-0.003	-0.006	-0.002
021a	-0.333	0.739	0.0	-0.003	0.001
022a	-2.059	0.039	-0.003	-0.006	-0.001
023a	-2.252	0.024	-0.003	-0.005	-0.001
030a	-1.135	0.256	-0.003	-0.006	0.0
031a	-1.007	0.314	-0.001	-0.004	0.0
032a	-1.831	0.067	-0.002	-0.004	0.0
034a	-1.970	0.049	-0.002	-0.004	0.0
036a	-1.684	0.092	-0.002	-0.004	0.0
038a	-2.607	0.009	-0.005	-0.007	-0.002
039a	-1.613	0.107	-0.003	-0.006	0.0
040a	-0.604	0.546	0.0	-0.003	0.001
041a	-0.711	0.477	-0.001	-0.003	0.001
049a	-0.890	0.373	-0.001	-0.005	0.0
051a	-2.269	0.023	-0.004	-0.006	-0.001
052a	-1.279	0.201	-0.002	-0.005	0.0
053a	0.449	0.653	0.001	-0.002	0.004
055a	-0.732	0.464	-0.001	-0.002	0.0
056a	-1.025	0.305	-0.001	-0.004	0.0
058a	0.406	0.685	0.0	-0.001	0.002
059a	-1.913	0.056	-0.005	-0.010	-0.001
064a	-0.290	0.772	0.0	-0.003	0.001
074a	-4.169	0.0	-0.011	-0.015	-0.008
075a	-0.178	0.859	0.0	-0.002	0.001
076a	-3.099	0.002	-0.004	-0.007	-0.003
168a	-1.114	0.265	-0.002	-0.005	0.0
171a	-0.237	0.813	0.0	-0.003	0.002
173a	-3.122	0.002	-0.006	-0.009	-0.003

TABLE 5.1.2

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRARTIONS OF NITRATE
(Monthly Median, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.954	0.051	-0.003	-0.006	-0.001
011a	-0.746	0.456	-0.002	-0.004	0.001
017a	0.538	0.591	0.001	-0.001	0.003
020a	-2.575	0.010	-0.004	-0.006	-0.002
021a	-1.383	0.167	-0.002	-0.004	0.0
022a	-3.236	0.001	-0.006	-0.008	-0.003
023a	-1.660	0.097	-0.003	-0.006	0.0
030a	-0.428	0.669	-0.001	-0.004	0.002
031a	-1.053	0.293	-0.002	-0.004	0.0
032a	-1.489	0.136	-0.002	-0.004	0.0
034a	-1.994	0.046	-0.003	-0.005	-0.001
036a	-1.207	0.228	-0.002	-0.004	0.0
038a	-2.363	0.018	-0.004	-0.007	-0.001
039a	-0.925	0.355	-0.002	-0.004	0.0
040a	-0.062	0.951	0.0	-0.002	0.002
041a	0.381	0.703	0.001	-0.002	0.002
049a	-1.240	0.215	-0.002	-0.006	0.0
051a	-2.248	0.025	-0.003	-0.006	-0.001
052a	-1.146	0.252	-0.003	-0.005	0.0
053a	-0.253	0.800	0.0	-0.003	0.002
055a	-1.223	0.221	-0.002	-0.003	0.0
056a	-1.265	0.206	-0.002	-0.005	0.0
058a	-0.937	0.349	-0.001	-0.002	0.0
059a	-1.533	0.125	-0.004	-0.007	0.0
064a	-1.367	0.172	-0.002	-0.004	0.0
074a	-3.654	0.0	-0.011	-0.016	-0.007
075a	-0.592	0.554	-0.001	-0.003	0.001
076a	-3.064	0.002	-0.004	-0.006	-0.002
168a	-2.394	0.017	-0.004	-0.007	-0.002
171a	-0.850	0.395	-0.001	-0.004	0.0
173a	-2.080	0.037	-0.004	-0.007	-0.001

TABLE 5.2.1

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRARTIONS OF NITRATE
(Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-2.111	0.035	-0.006	-0.010	-0.002
010a	0.514	0.608	0.001	-0.003	0.006
011a	-0.382	0.702	-0.001	-0.006	0.004
012a	0.029	0.977	0.0	-0.005	0.005
017a	2.757	0.006	0.009	0.004	0.014
020a	-1.256	0.209	-0.004	-0.008	0.0
021a	-0.453	0.650	-0.001	-0.005	0.002
022a	0.258	0.797	0.001	-0.004	0.004
023a	-0.222	0.824	-0.001	-0.005	0.004
024a	-0.880	0.379	-0.003	-0.009	0.002
025a	-1.093	0.274	-0.002	-0.005	0.0
028a	0.597	0.550	0.002	-0.002	0.007
029a	-1.502	0.133	-0.007	-0.011	-0.001
030a	1.202	0.230	0.007	0.0	0.014
031a	1.119	0.263	0.005	-0.001	0.009
032a	-2.036	0.042	-0.006	-0.011	-0.002
033a	-0.018	0.986	0.0	-0.004	0.003
034a	-1.193	0.233	-0.003	-0.007	0.0
035a	-2.062	0.039	-0.008	-0.015	-0.004
036a	0.284	0.776	0.001	-0.003	0.005
037a	-0.767	0.443	-0.005	-0.014	0.004
038a	-0.422	0.673	-0.001	-0.004	0.002
039a	0.507	0.612	0.003	-0.005	0.010
040a	1.840	0.066	0.008	0.003	0.012
041a	1.218	0.223	0.003	0.0	0.007
046a	0.960	0.337	0.003	-0.001	0.007
047a	1.730	0.084	0.005	0.001	0.008
049a	0.018	0.986	0.0	-0.008	0.008
051a	0.183	0.854	0.001	-0.005	0.006
052a	0.445	0.656	0.002	-0.004	0.007
053a	0.841	0.400	0.005	-0.003	0.013
055a	-0.364	0.716	-0.001	-0.004	0.002
056a	1.022	0.307	0.003	-0.001	0.008
058a	-0.436	0.663	-0.001	-0.004	0.003
059a	-2.194	0.028	-0.016	-0.026	-0.006
061a	-0.716	0.474	-0.003	-0.011	0.002
063a	1.229	0.219	0.004	0.0	0.008
064a	0.595	0.552	0.002	-0.002	0.006
065b	-1.133	0.257	-0.003	-0.007	0.0
068a	-1.392	0.164	-0.006	-0.014	-0.001
070a	-0.544	0.586	-0.002	-0.008	0.004
071a	-1.680	0.093	-0.006	-0.013	-0.001
073a	0.734	0.463	0.003	-0.002	0.008

TABLE 5.2.1 (continued)

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRARTIONS OF NITRATE
(Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	-3.799	0.0	-0.021	-0.030	-0.014
075a	3.511	0.0	0.009	0.006	0.012
076a	-0.770	0.441	-0.003	-0.006	0.002
077a	-0.257	0.797	-0.001	-0.007	0.004
078a	-1.204	0.229	-0.005	-0.012	0.0
160a	-2.974	0.003	-0.014	-0.019	-0.008
161a	0.116	0.908	0.001	-0.005	0.005
163a	1.941	0.052	0.005	0.001	0.010
164a	1.013	0.311	0.006	-0.002	0.014
166a	0.0	1.000	0.0	-0.004	0.005
168a	0.975	0.329	0.003	-0.001	0.008
171a	1.902	0.057	0.007	0.002	0.011
172a	0.914	0.361	0.0	0.0	0.005
173a	-3.013	0.003	-0.009	-0.014	-0.006
249a	1.096	0.273	0.005	-0.001	0.012
251a	1.702	0.089	0.011	0.003	0.020
252a	-1.698	0.090	-0.007	-0.011	-0.002
253a	-0.761	0.447	-0.003	-0.009	0.002
254a	-1.013	0.311	-0.005	-0.012	0.001
255a	-0.667	0.505	-0.003	-0.011	0.003
257a	1.598	0.110	0.008	0.001	0.017
258a	-0.502	0.616	-0.002	-0.006	0.004
268a	0.293	0.769	0.002	-0.003	0.006
271a	-0.479	0.632	-0.003	-0.013	0.003
272a	-1.253	0.210	-0.004	-0.009	0.0
273a	-1.902	0.057	-0.005	-0.010	-0.002
275a	-0.107	0.915	0.0	-0.006	0.005
277a	1.816	0.069	0.012	0.004	0.018
278a	-3.329	0.001	-0.014	-0.022	-0.009
279a	-2.329	0.020	-0.011	-0.018	-0.004
280a	-1.351	0.177	-0.005	-0.011	-0.001
281a	-0.431	0.666	-0.003	-0.013	0.006
282a	1.129	0.259	0.004	-0.001	0.009
283a	-1.247	0.212	-0.003	-0.005	0.0
285a	1.411	0.158	0.008	0.0	0.016
349a	1.506	0.132	0.004	0.001	0.006
350a	1.879	0.060	0.011	0.003	0.019
354a	-4.106	0.0	-0.028	-0.040	-0.018

TABLE 5.2.2

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES FOR THE CONCENTRARTIONS OF NITRATE (Monthly Median, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	U-bd
004a	-1.577	0.115	-0.006	-0.012	-0.001
010a	0.138	0.891	0.0	-0.003	0.003
011a	0.124	0.901	0.001	-0.004	0.005
012a	0.127	0.899	0.001	-0.006	0.006
017a	1.781	0.075	0.006	0.001	0.012
020a	-0.935	0.350	-0.003	-0.007	0.001
021a	-0.711	0.477	-0.003	-0.007	0.002
022a	-1.467	0.143	-0.005	-0.010	-0.001
023a	0.667	0.505	0.003	-0.002	0.007
024a	-2.391	0.017	-0.009	-0.016	-0.004
025a	-0.667	0.505	-0.002	-0.005	0.002
028a	0.020	0.984	0.0	-0.005	0.006
029a	-0.649	0.516	-0.003	-0.010	0.003
030a	1.431	0.153	0.007	0.001	0.014
031a	0.248	0.804	0.001	-0.004	0.005
032a	-2.054	0.040	-0.008	-0.014	-0.003
033a	0.071	0.943	0.0	-0.004	0.003
034a	-1.827	0.068	-0.008	-0.014	-0.003
035a	-1.874	0.061	-0.010	-0.016	-0.004
036a	0.258	0.797	0.001	-0.003	0.004
037a	-0.388	0.698	-0.002	-0.010	0.006
038a	-0.165	0.869	0.0	-0.004	0.003
039a	1.786	0.074	0.009	0.003	0.017
040a	1.084	0.278	0.004	-0.001	0.009
041a	2.062	0.039	0.006	0.002	0.010
046a	0.560	0.576	0.002	-0.002	0.006
047a	1.760	0.078	0.006	0.001	0.009
049a	0.471	0.638	0.003	-0.005	0.011
051a	0.954	0.340	0.003	-0.001	0.007
052a	0.767	0.443	0.004	-0.003	0.011
053a	1.350	0.177	0.007	0.0	0.014
055a	-0.009	0.993	0.0	-0.003	0.004
056a	1.022	0.307	0.004	-0.001	0.010
058a	0.827	0.408	0.003	-0.002	0.007
059a	-1.992	0.046	-0.011	-0.022	-0.003
061a	-1.243	0.214	-0.009	-0.016	0.0
063a	1.100	0.271	0.003	0.0	0.007
064a	0.782	0.434	0.002	-0.001	0.008
065b	-0.734	0.463	-0.003	-0.008	0.002
068a	-0.511	0.609	-0.002	-0.010	0.004
070a	-0.680	0.496	-0.002	-0.008	0.004
071a	-1.147	0.252	-0.006	-0.014	0.001
073a	0.724	0.469	0.003	-0.003	0.008

TABLE 5.2.2 (continued)

RESULTS OF THE MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRARTIONS OF NITRATE
(Monthly Median, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	-2.039	0.041	-0.013	-0.022	-0.004
075a	2.711	0.007	0.010	0.005	0.016
076a	-1.495	0.135	-0.004	-0.008	-0.001
077a	-0.330	0.741	-0.001	-0.004	0.002
078a	-0.489	0.625	-0.003	-0.011	0.004
160a	-3.933	0.0	-0.018	-0.025	-0.012
161a	0.338	0.736	0.001	-0.003	0.004
163a	0.407	0.684	0.002	-0.005	0.006
164a	1.165	0.244	0.008	0.0	0.015
166a	-0.917	0.359	-0.003	-0.008	0.001
168a	0.104	0.917	0.0	-0.006	0.006
171a	1.627	0.104	0.006	0.002	0.011
172a	3.227	0.001	0.010	0.006	0.013
173a	-1.964	0.049	-0.009	-0.015	-0.003
249a	1.252	0.210	0.003	0.0	0.011
251a	1.772	0.076	0.013	0.005	0.021
252a	-1.733	0.083	-0.006	-0.011	-0.002
253a	-0.715	0.474	-0.002	-0.007	0.002
254a	-0.218	0.828	-0.002	-0.010	0.004
255a	-0.844	0.398	-0.005	-0.013	0.003
257a	0.850	0.396	0.005	-0.002	0.011
258a	-0.739	0.460	-0.003	-0.008	0.002
268a	-0.373	0.709	-0.002	-0.007	0.003
271a	-0.108	0.914	0.0	-0.009	0.006
272a	-0.080	0.936	0.0	-0.005	0.003
273a	-2.195	0.028	-0.008	-0.013	-0.003
275a	0.142	0.887	0.001	-0.005	0.006
277a	2.623	0.009	0.019	0.011	0.027
278a	-3.299	0.001	-0.015	-0.023	-0.009
279a	-1.964	0.050	-0.010	-0.017	-0.004
280a	-1.129	0.259	-0.005	-0.012	0.001
281a	-0.138	0.891	-0.001	-0.008	0.011
282a	0.702	0.483	0.003	-0.003	0.009
283a	-0.972	0.331	-0.002	-0.006	0.001
285a	1.155	0.248	0.007	-0.001	0.011
349a	1.278	0.201	0.004	0.0	0.008
350a	1.106	0.269	0.007	0.0	0.015
354a	-3.486	0.0	-0.018	-0.026	-0.012

Chapter 6 ANALYSES AND CONCLUSIONS FOR HYDROGEN ION

Monthly precipitation volume weighted mean and monthly median are referred to as "weighted mean" and "median", respectively, below.

6.1 Results for the Historical Data

"Historical data" refers to the data collected from January, 1980 to December, 1986. For H^+ (calculated from pH), there are 32 sites with 5 or fewer missing observations during the period. The locations of these 32 sites are plotted in Figure 6.1.1. From this figure we can see that more sites are located in the East than the West. The results of this section are based on the data obtained from these 32 sites.

6.1.1 Clustering and Transformation

Figure 6.1.2 shows the hierarchical cluster structure of the weighted means. The sites are partitioned quite naturally into three clusters. The locations of the sites labelled by clusters are shown in Figure 6.1.3. The pattern is obvious: clusters 3 and 1 are located in the subregion of highest industrial concentration and the surrounding regions; cluster 2 is spread over the rest of the United States. This indicates that the pattern of wet acid deposition in the industrial areas is different from that of the other areas. Note that cluster 2 can be split into two subgroups according to Figure 6.1.2, one subgroup is located along the East Coast and the other spread over the West and the middle of the United State. This suggests again that the geographical location and the manner of wet acid deposition are closely related.

The histograms of the original data and the transformed data under the three kinds of transformations described in Section 3.1, are shown in Figure 6.1.4. None of the four histograms appears symmetric. The ones for 1/4 power transformation and for log transformation seem better than the other two. Since the log transformation of H^+ is equivalent to the pH value, which gives us a familiar measure of acidity, the log transformation is adopted in this case and the remainder of our analysis of the hydrogen ion data is based on the transformed data. The histograms of the log transformed data for each cluster are depicted in Figure 6.1.5. It seems that the histogram for cluster 3 fits the normal curve well and that for cluster 1 it is approximately symmetric. The one for cluster 2 is left skewed.

6.1.2 Trend, Seasonality and Spatial Patterns

The analysis described below is done for both log transformed monthly volume weighted mean and log transformed monthly median H^+ data. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster, in the manner described in Section 3.2. Figures 6.1.6 and 6.1.7 show the plots of the yearly effects of the three clusters for weighted mean and median, respectively. In general they both show a V-shaped pattern where the one for weighted means is more obvious. In Figure 6.1.6, clusters 1 and 3 show down-trends before 1983 and up-trends after 1983. Cluster 2 behaved slightly differently, namely, the suggested change point is 1984 instead of 1983. Figure 6.1.7 shows a pattern like the one in Figure 6.1.6 except that the estimated effect of 1982 for cluster 2 is different. The reason for this needs further study. Generally speaking, Figures 6.1.6 and 6.1.7 suggest that the concentration of H^+ in wet deposition decreased before 1983 and increased after 1983 for the most industrialized subregion of the United States. They also suggest that for the other regions of the United States, there is a similar V-shaped trend.

Figure 6.1.8 and Figure 6.1.9 show the monthly effects for the weighted means and medians, respectively. They have almost the same pattern. The patterns for all three clusters are similar as well. The effects are high from June to September and low from November to April. The transitions from month to month are quite smooth and there is a peak in August. The fact that the seasonal patterns are very similar for all three clusters suggests a reassuring spatial stability in this seasonal pattern.

Figures 6.1.10 to 6.1.12 display the site effects of $\log(H^+)$ by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. In Figure 6.1.10, the stations located close to the center of the industrial area tend to have large positive effects and the stations located far from the center of the industrial area tend to have large negative effects. In Figure 6.1.11, there is an interesting pattern: all the stations with positive effects are located in the Eastern United States and all the stations with negative effects are located in the West or the middle of the United States.

Figures 6.1.13 to 6.1.15 are the boxplots of the residuals of the three-way median polish of the $\log(H^+)$ data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order. The differences between the boxplots for the means and for the medians are even smaller. The variation of the residuals is relatively small for cluster 3 and relatively large for cluster 2. In Figures 6.1.13 ((a) and (b)), we observe a few extreme outliers from Station 021a, 032a and 051a. In Figures 6.1.14 ((a), (b)), extreme outliers are observed at Stations 011a, 017a and 053a. In Figures 6.1.15 ((a), (b)), the boxes for Stations 041a have extreme outliers. Note that some other stations have outliers as well. All the stations with outliers should be carefully checked and further studied. But no analyses for specific site are included in this paper.

Figures 6.1.16 to 6.1.18 summarize the median polish for the three clusters, respectively where (a) is for the weighted means and (b) is for the medians. In each figure, boxes represent the main effects, interactions and residuals respectively. We regard interactions as non-negligible if the sizes of the boxes for the interactions are comparable with those for main effects. In these figures, the boxes for yearly effects are smaller indicating that compared with seasonal variation and site variation, the yearly changes are smaller. The boxes for the interactions of month-by-year and month-by-site have the sizes comparable to those of the main effects, which suggests that they have to be taken into account. For cluster 2, the box for site effect is very large. This may be caused by the fact that cluster 2 is the largest cluster spreading over the United States, and hence the variation of site effects is large (see Figures 6.1.17 (a) and (b)). Figures 6.1.18 (a) and (b) show the summaries for cluster 3. The fact that the boxes for site effects and their interactions are very small suggests that the four sites behaved very similarly.

6.1.3 The Results of Trend Testing, Slope Estimation and Kriging

Mann-Kendall Test and Sen's nonparametric slope estimation procedure are applied to the deseasonalized data for each site in the manner described in Section 3.3. A Mann-Kendall Test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval are obtained for each site for both weighted means and medians. These are shown in Tables 6.1.1 and 6.1.2 respectively. In Table 6.1.1, 11 out of the 32 sites have significant trends at the p=0.2 level. In Table 6.1.2, 10 out of the 32 sites have significant trends at the p=0.2 level.

The results of the nonparametric slope estimation are plotted in Figures 6.1.19 (a) and (b) with symbols defined in Section 3.3, for the weighted means and medians respectively.

If the lower 80% confidence bound of the slope estimate is greater than 0 at any site a "+" is plotted; if the upper 80% confidence bound of the slope estimate is less than 0 a "-" is plotted. Otherwise, a "0" is plotted . "+", "-" and "0" represent respectively, an up-trend, a down-trend and no-trend. The patterns of these two figures are very similar. In Figure 6.1.19 (a), 7 of the sites show down-trends and they are mainly located in the Eastern United States. The two located along the West Coast show up-trend. Other sites show no-trend. Figure 6.1.19 (b) shows a slightly different pattern in which two sites located in the South show up-trend. The main result is that during the period of 1980 - 1986, most of the sites have no-trend. However, keeping Figures 6.1.6 and 6.1.7 in mind, the results of trend tests and slope estimation may be misleading in case of the presences of a V-shaped trend.

The results of Kriging the slope estimates are displayed in Figures 6.1.20(a) and (b). As before, Figure 6.1.20(a) is for the weighted means and (b) is for medians. A "+" means an up-trend is estimated at the corresponding point and a "-" means a down-trend, estimated at the point. Since the estimated variograms used for these two plots are both constant, Figures 6.1.20 (a) and (b) are actually produced by moving averages. Hence these two figures are relatively rough and should be read in conjunction with Figure 6.1.19. At any particular point where the observations were not made, interpolation may not be reliable. This is because (a) estimation is based on the estimated slopes, not on the observations directly and (b) the spatial resolution of the measurements is low and therefore, the correlations between sites are small.

6.2 Results for the Recent Data

"Recent data" refers to the data collected from January, 1983 to December, 1986. For H^+, there are 86 sites with 5 or fewer missing observations during the period. The locations of these 86 sites are plotted in Figure 6.2.1. From this figure we can see that more sites are located in the Northeast than the rest of the areas. The results of this section are based on the data observed at these 86 sites.

6.2.1 Clustering and Transformation

Figure 6.2.2 shows the hierarchical cluster structure of the weighted means. The sites are partitioned naturally into three clusters. The locations of the sites labelled by cluster are shown in Figure 6.2.3. It gives a pattern similar to the one in Figure 6.1.3: cluster 2 is located in the subregion of highest industrial concentration; cluster 1 is located around that region; and cluster 3 is located throughout the rest of the United States, corresponding to cluster 3, 1 and 2 in Figure 6.3.1 respectively. This gives us the same indications that we mentioned in Section 6.1.1, namely, that the wet acid deposition in the industrial areas has a different manner from that of the other area.

The histograms of the original data and the transformed data under the three kinds of transformation described in Section 3.1, are shown in Figure 6.2.4. It appears that the histogram of the 1/4 power transformed data fits the normal curve slightly better than that of the log transformed data. For the reason given in Section 6.1.1, also for the reason that the log transformation is used in the case of "historical data", the log transformation is adopted for our analysis so that comparisons can be easily made. The histograms of the log transformed data for each cluster are depicted in Figure 6.2.5. It appears that the histogram for cluster 2 fits the normal curve well. The histogram of cluster 1 fit the normal curve too but it has a light long left tail. The histogram of cluster 3 is left skewed.

6.2.2 Trend, Seasonality and Spatial Patterns

The analysis described below is applied to both log transformed monthly volume weighted means and log transformed monthly medians. The analysis is applied separately to each cluster obtained from clustering the volume weighted mean data.

Three-way median polishes are applied to both weighted means and medians in each cluster. Figures 6.2.6 and 6.2.7 show the plots of the yearly effects of the three clusters for weighted mean and median respectively. They show similar patterns. In general, all three clusters show up-trends. This conclusion, obtained from 86 stations instead of 32, is consistent with what we got from Figures 6.1.7 and 6.1.8. That is, during the period of 1983 to 1986, the concentration of H^+ in wet deposition increased for the most industrialized subregion of the United States, and for most of the other regions of the United States as well.

Figures 6.2.8 and 6.2.9 show the monthly effects for the weighted means and medians, respectively. Although more than half of the data used in the present analyses were not included in the historical data, the patterns are quite close to those for the latter displayed in Figures 6.1.8 and 6.1.9: high from June to September; low from November to April with a peak in August. This convinces us again of spatial stability in the seasonal pattern. Besides, it seems more clearly that there is a gap in April for cluster 3.

Figures 6.2.10 to 6.2.12 show the site effects of $\log(H^+)$ by cluster; the weighted means and medians are plotted together and sorted by weighted means within each cluster. In Figure 6.2.10, the difference between the effects of the weighted means and the medians at Station 021a is large. To see how this occurred we need to study the original data further. For cluster 2, the stations located in the center of the industrial region tend to have larger effects.

Figures 6.2.13 to 6.2.15 are the boxplots of the residuals of the three-way median polish of the $\log(H^+)$ data where (a) is for weighted means and (b) is for medians and each box represents one site. In general, the variations of the residuals in each cluster are of the same order and are comparable to those in Figures 6.1.13 to 6.1.15. The differences between the boxplots for the means and for the medians are even smaller. The boxes for Station 035a and 038a in both Figures 6.1.15 (a) and (b) are larger indicating that the data from these two station do not fit the model well. Extreme outliers are found in Stations 021a, 023a, 025a,039a, 051a, 053a and 161a in Figures 6.2.13 (a) or (b), 058a in Figures 6.2.14 (b) and 015a, 024a, 029a, 068a, 071a, 077a, 225a and 339a in Figures 6.2.15 (a) or (b). Among these stations, 021a, and 053a have extreme outliers observed in Figures 6.1.13 to 6.1.15 and those outliers may be the same ones. All these sites and the other stations with outliers need further study.

Figures 6.2.16 to 6.2.18 summarize the result of the median polish for the three clusters respectively where (a) is for weighted means and (b) is for medians. In each figure, boxes represent the main effects, interaction and residuals. These figures are similar to those in Figures 6.1.16, 6.1.18 and 6.1.17 respectively. The interpretation for them can be applied here correspondingly. Generally speaking, cluster 2 fit the model best.

6.2.3 The Results of Trend Testing, Slope Estimation and Kriging

The Mann-Kendall Test and Sen's slope estimation procedure which were applied to "historical data" are used for "recent data" as well. A Mann-Kendall test statistic with the corresponding p-value and Sen's (1968b) nonparametric slope estimate with its 80% confidence interval are obtained for each site for both weighted means and medians. These are shown in Tables 6.2.1 and 6.2.2 respectively. In Table 6.2.1, 39 out of the 86 sites have significant trends at the p=0.2 level. In Table 6.2.2, 40 out of the 86 sites have significant trends at the p=0.2 level.

Figures 6.2.19 (a) and (b) are obtained by the same method used to produce Figures 6.1.19, for weighted means and medians respectively. Again, "+", "-" and "0" represent an up-trend, a down-trend and no-trend, respectively. In Figure 6.2.19 (a), 10 sites show down-trends and are located in the middle east of the United States. Slightly more than half of the sites show no-trend and the rest of the sites show up-trend. Those sites are located throughout the United States. The pattern of Figure 6.2.19 (b) is quite similar to Figure 6.2.19 (a) except that there are a few more "up-trend" stations and fewer "down-trend" stations. Comparing Figures 6.2.19 ((a) and (b)) with Figures 6.1.19 ((a) and (b)), we can see that a lot more stations with "+" appear in the former. Some of the stations with "-" in Figure 6.1.19 have "+" in Figures 6.2.19. This implies that some stations do have a V-shaped trend during the period of 1980 - 1986.

The results of Kriging the slope estimates are displayed in Figures 6.2.20(a) and (b) where (a) is for the weighted means and (b) is for the medians. As before, a "+" means an up-trend and a "-" means a down-trend estimated at the corresponding point, respectively. Again, these two plots are similar except that both the "+" area and "-" areas in Figure 6.2.20 (b) are smaller than those in Figure 6.2.20 (a). Both figures show that there is an up-trend for the East Coast area, a down-trend in certain areas of the middle of the United States and no-trend in the rest of the areas. For the reason given in Section 6.1.3, these two figures can only be used as indicators of the spatial distribution of the trends as the interpolation error for a specific point could be large.

The Locations Of The 32 Monitoring Stations From 1980 To 1986
(For Hydrogen ion)

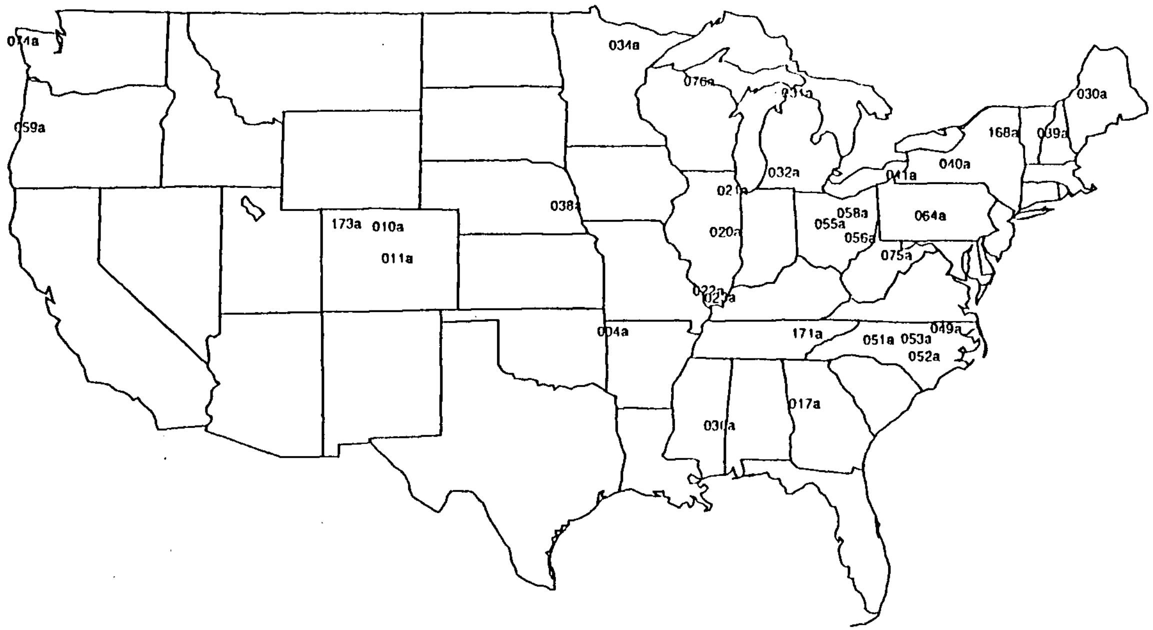


Figure 6.1.1

Clustering of H+ monthly volume weighted mean based on sqrt(MSE)

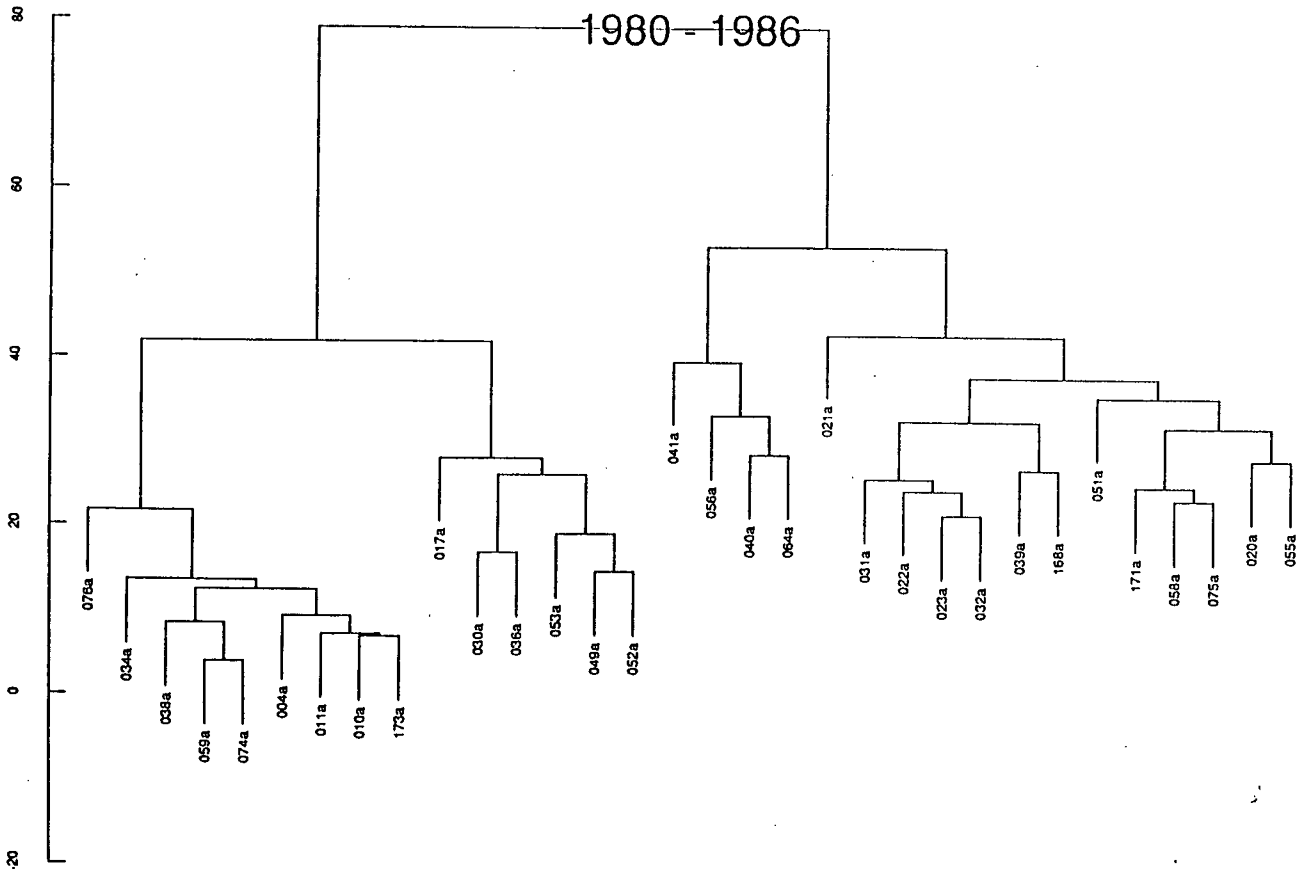


Figure 6.1.2

Clusters of H+ monthly volume weighted mean based on sqrt(MSE)

1980 - 1986 (k=3)

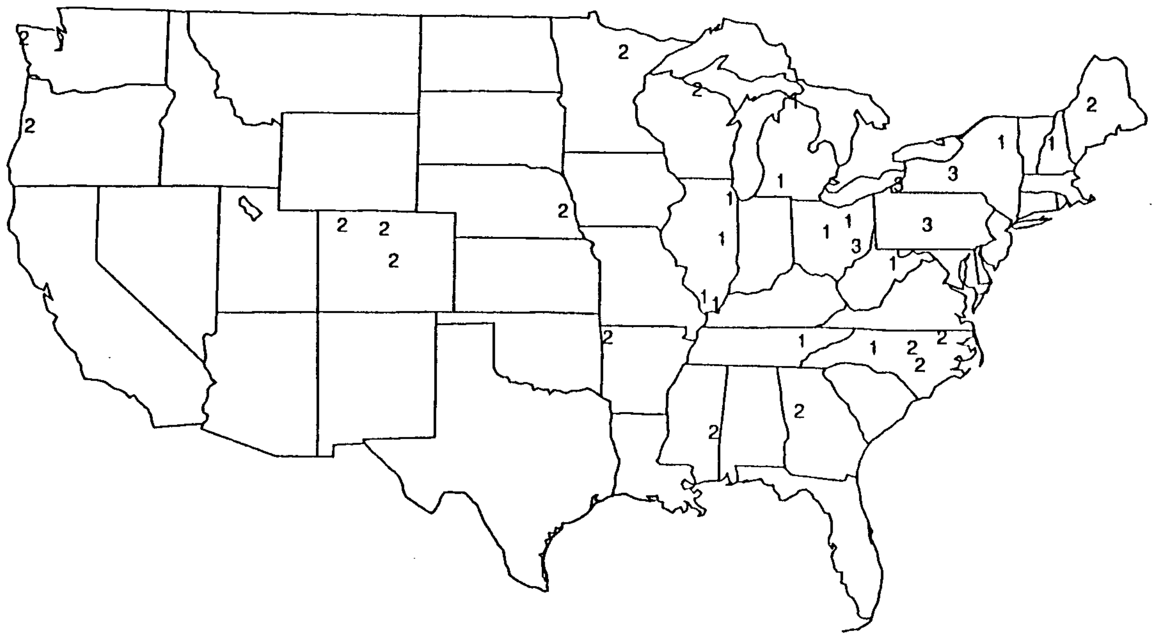


Figure 6.1.3

Histograms of Transformed H+ (Volume Weighted Mean, 80-86)

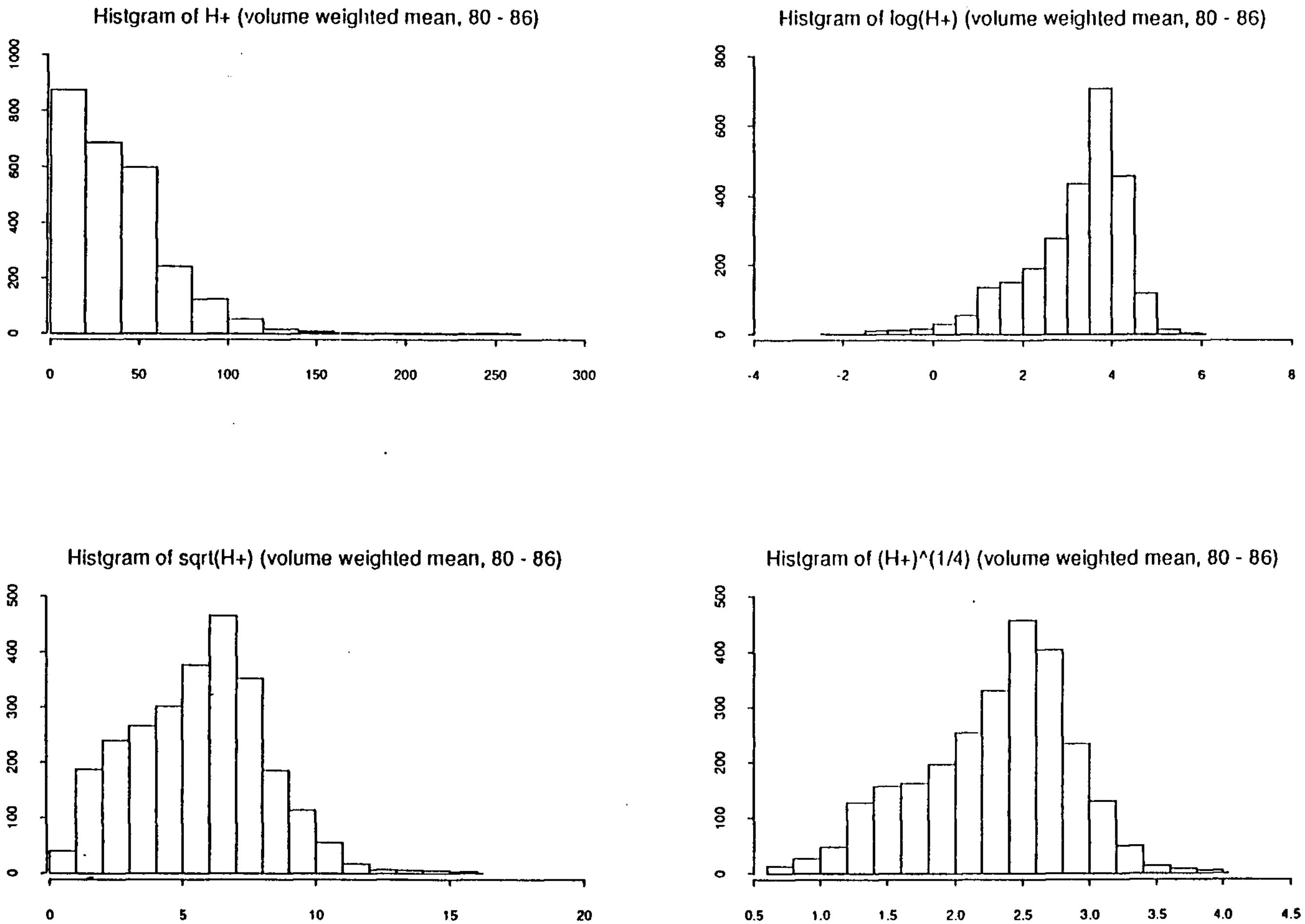


Figure 6.1.4

Histograms of log(H+) by Clusters (Volume Weighted Mean, 80-86)

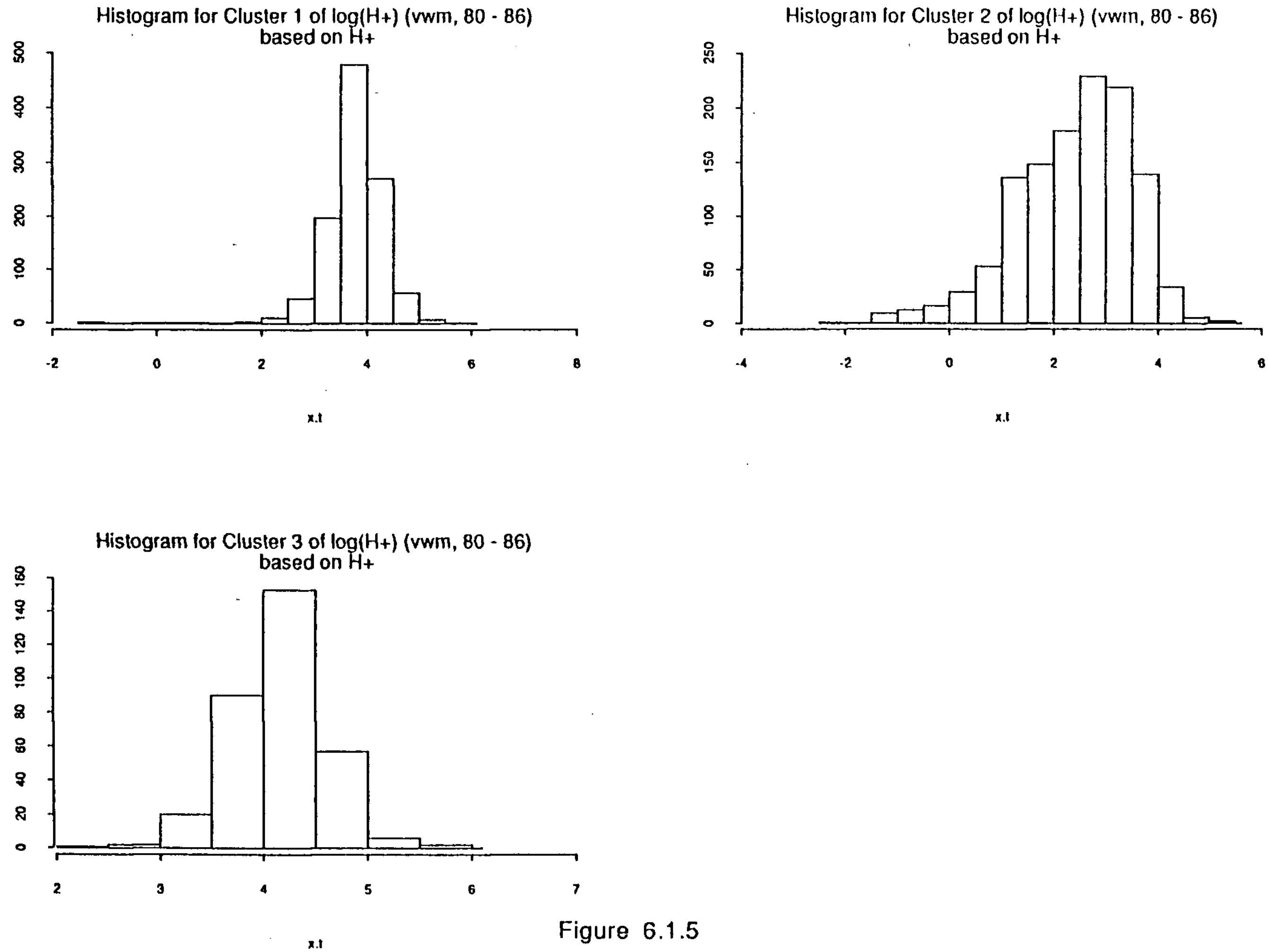


Figure 6.1.5

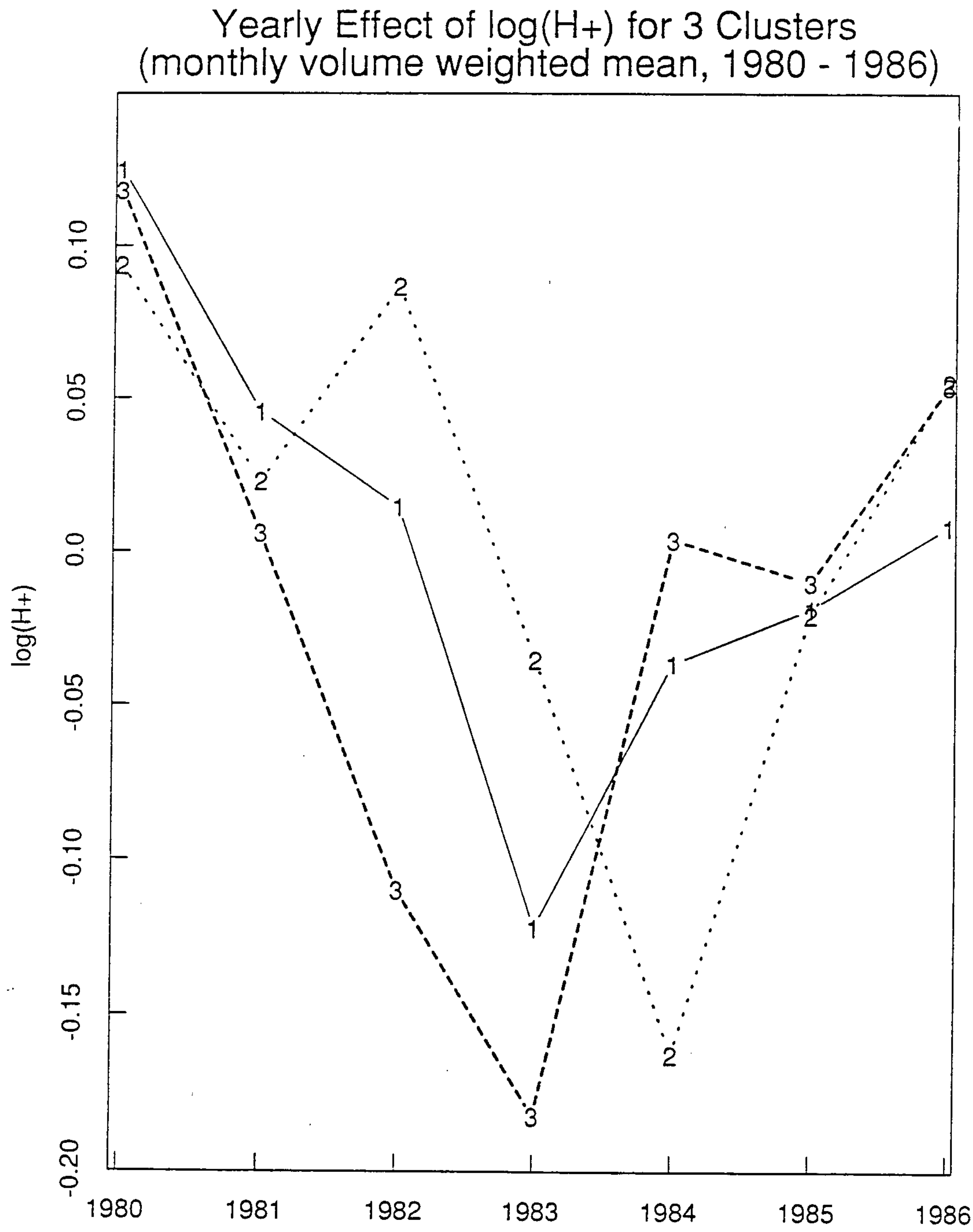

Yearly Effect of log(H+) for 3 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(H+)
0.10
0.05
0.0
-0.05
-0.10
-0.15
-0.20
1980
1981
1982
1983
1984
1985
1986

Figure 6.1.6

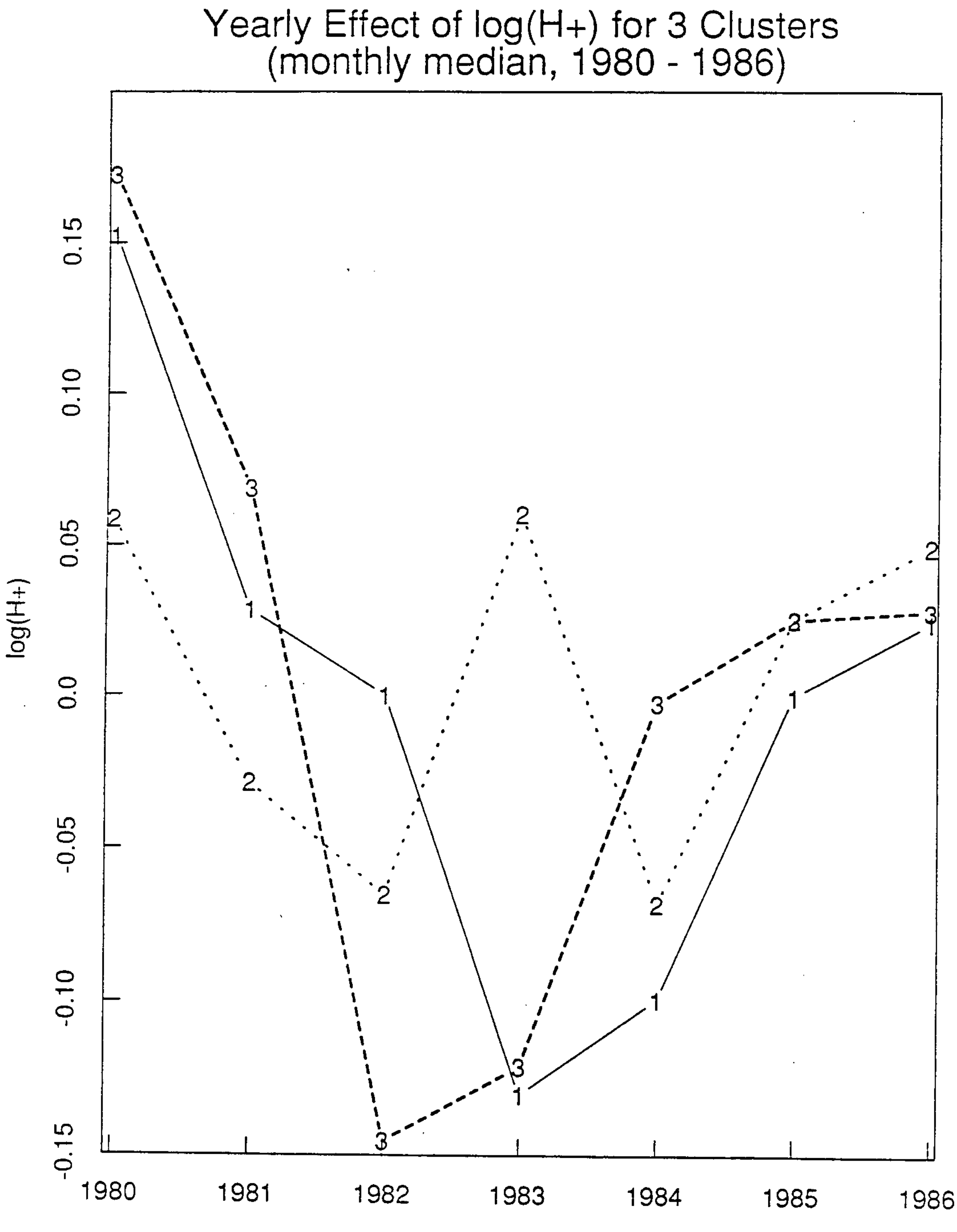
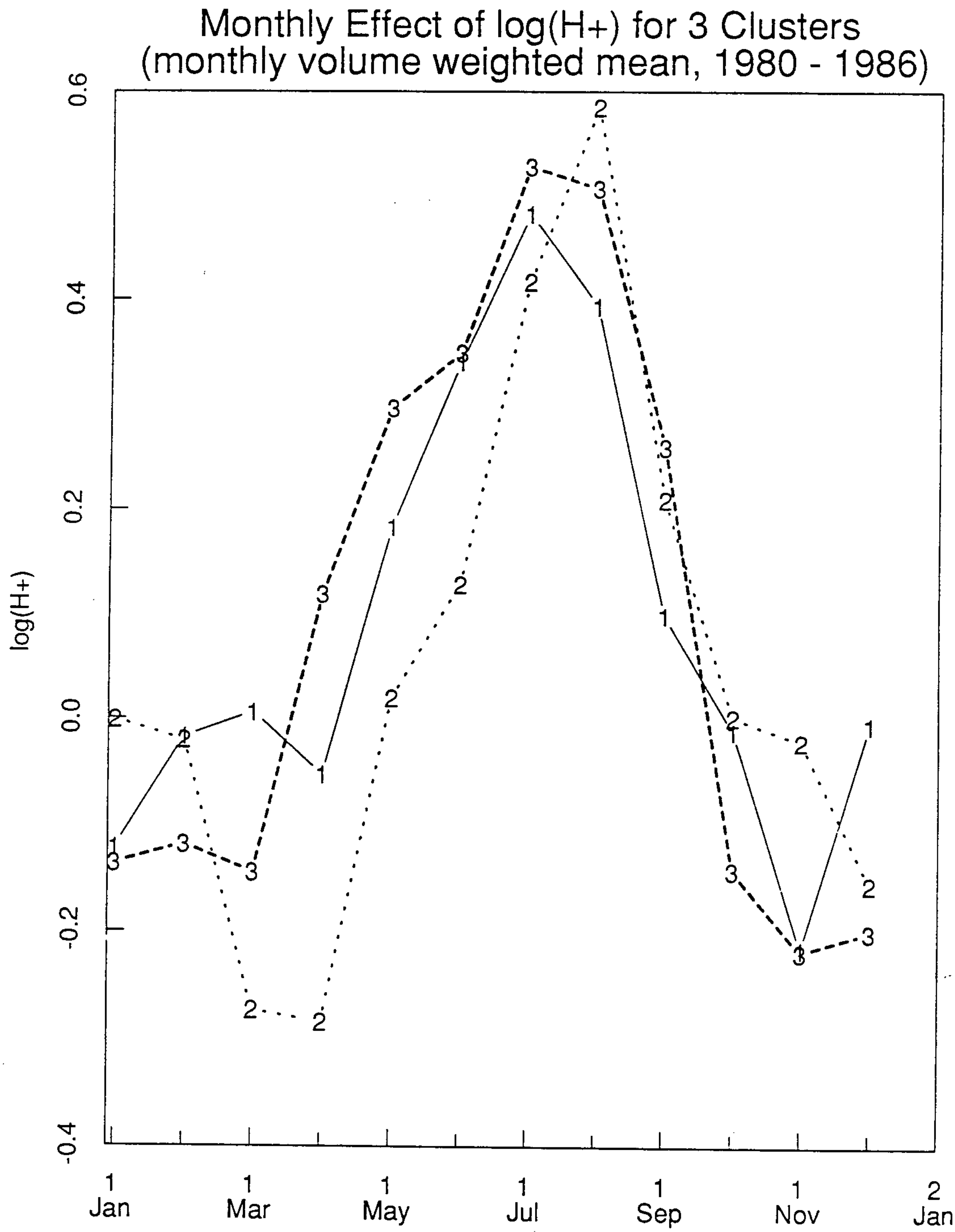

Yearly Effect of log(H+) for 3 Clusters
(monthly median, 1980 - 1986)
log(H+)
0.15
0.10
0.05
0.0
-0.05
-0.10
-0.15
1980
1981
1982
1983
1984
1985
1986

Figure 6.1.7



Monthly Effect of log(H+) for 3 Clusters
(monthly volume weighted mean, 1980 - 1986)
log(H+)
0.6
0.4
0.2
0.0
-0.2
-0.4
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 6.1.8

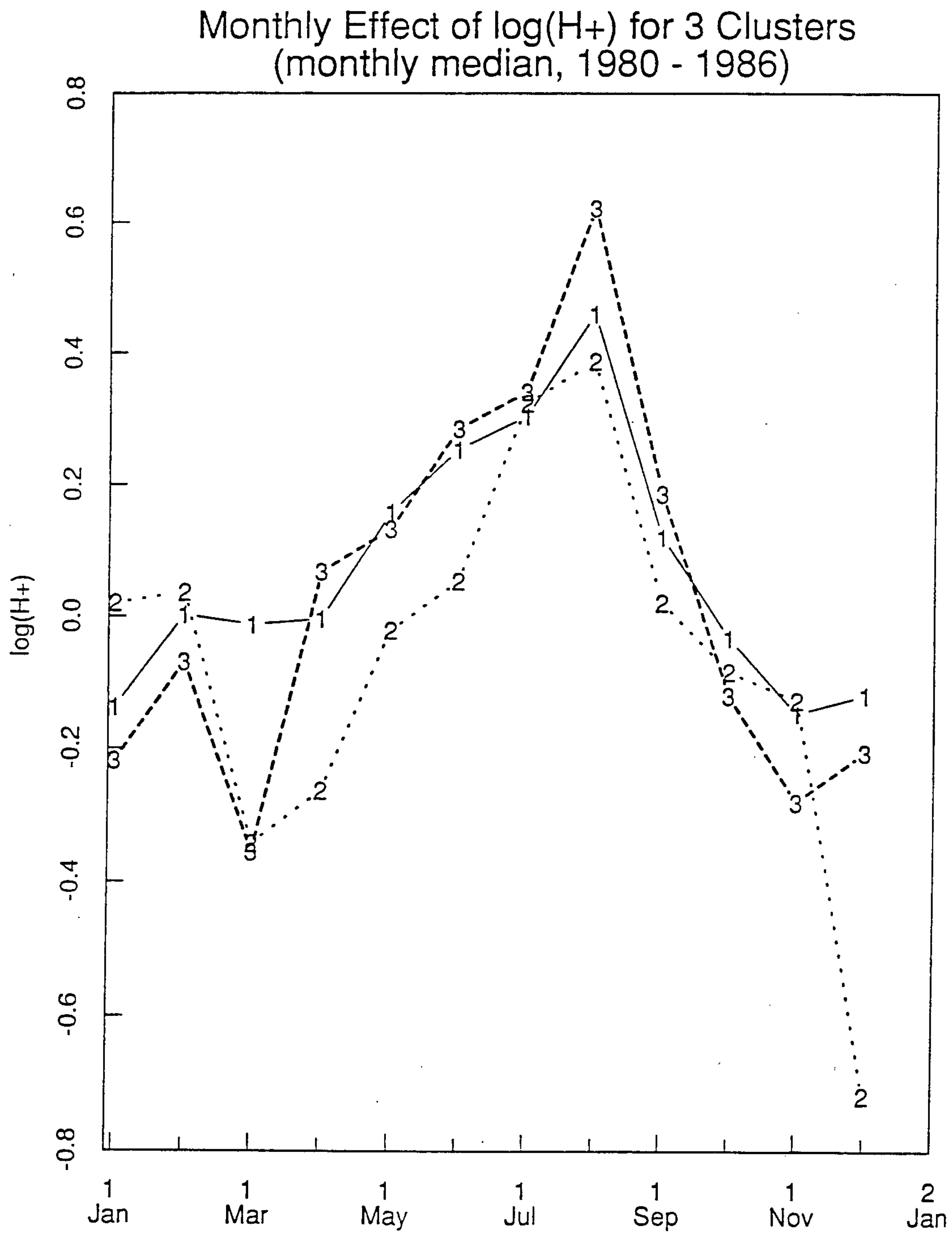

Monthly Effect of log(H+) for 3 Clusters
(monthly median, 1980 - 1986)
log(H+)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 6.1.9

Station Effect of log(H+) for Cluster 1
(monthly median/vclume weighted mean (sorted by volume weighted mean), 1980 - 1986)

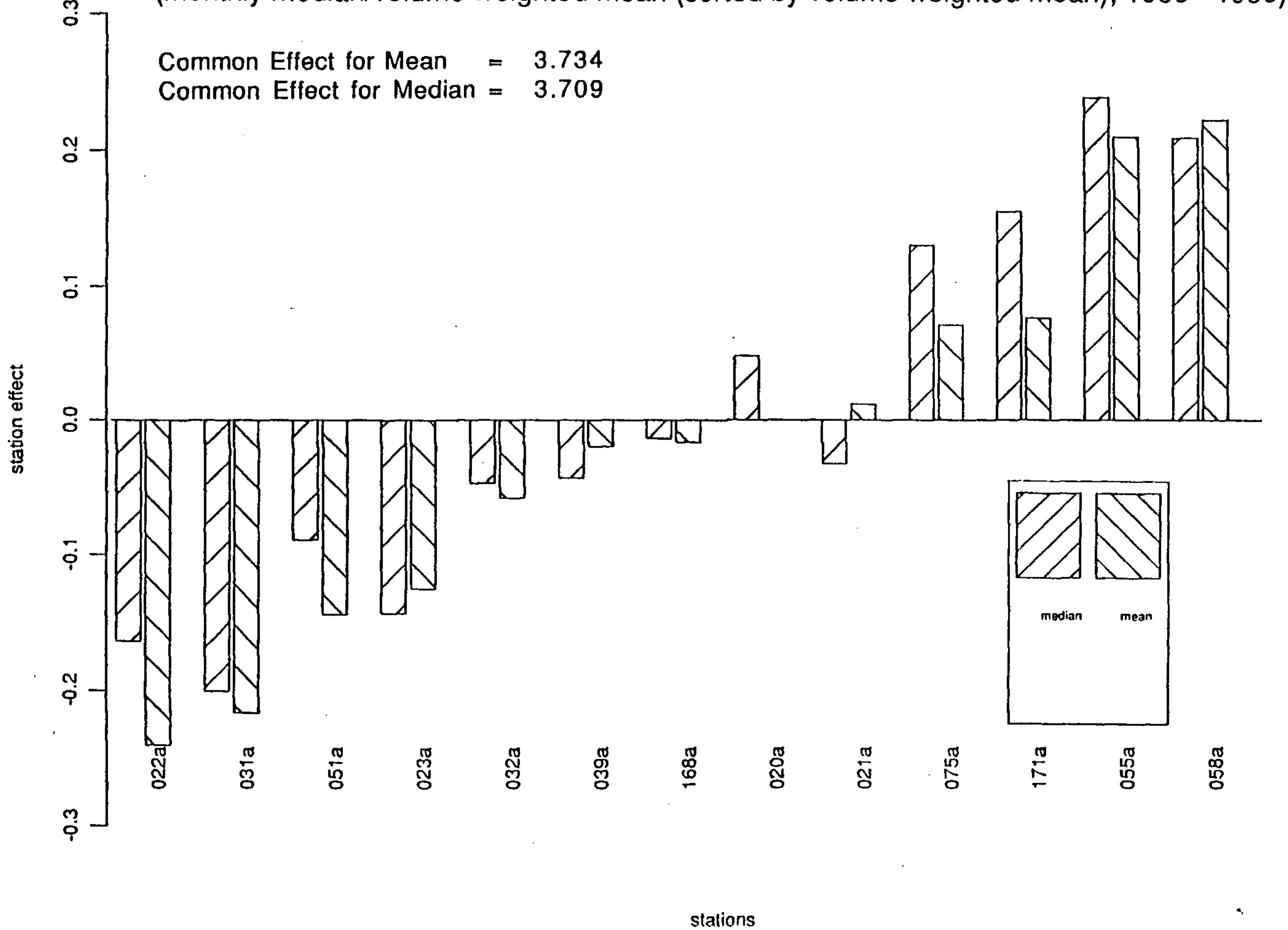


Figure 6.1.10

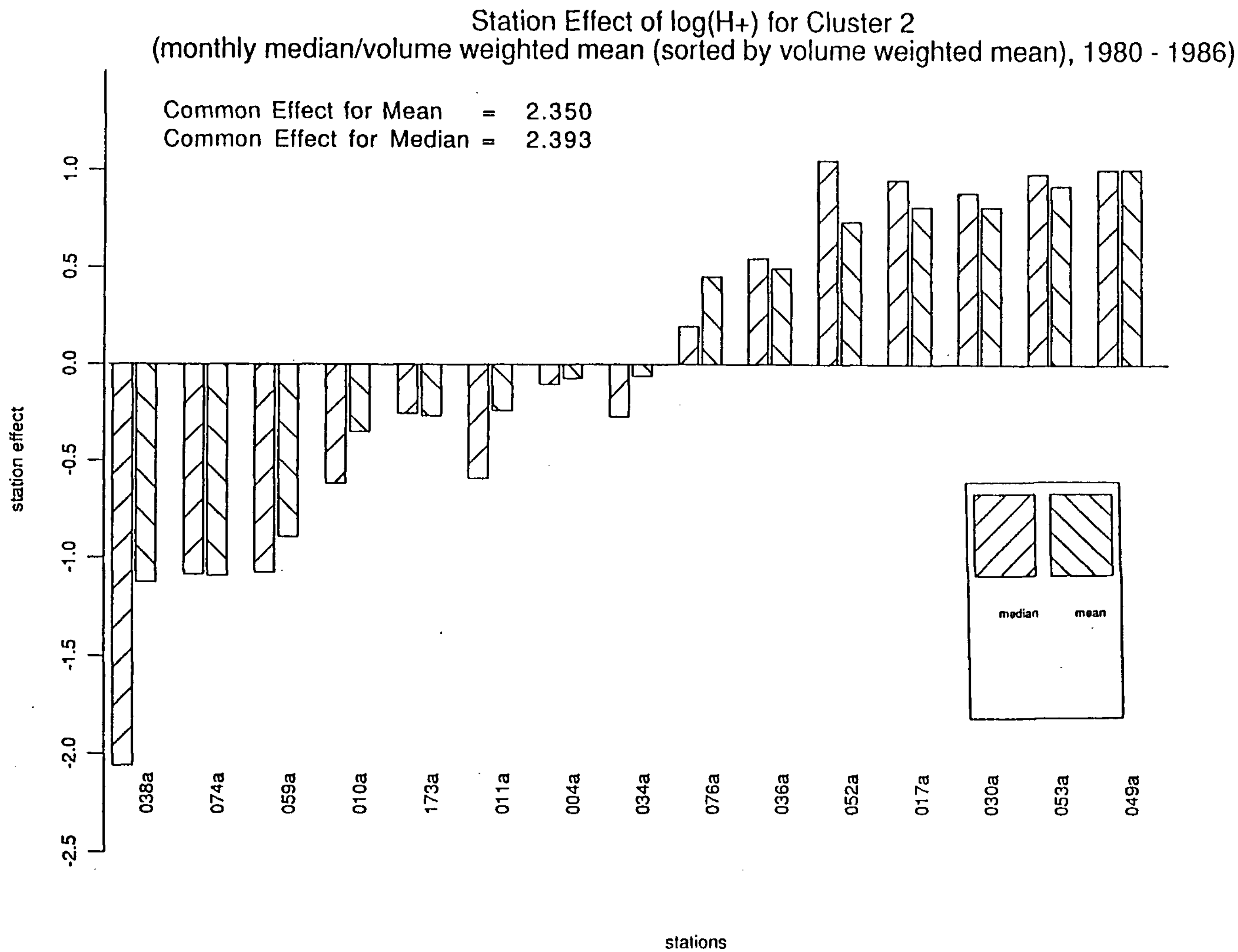
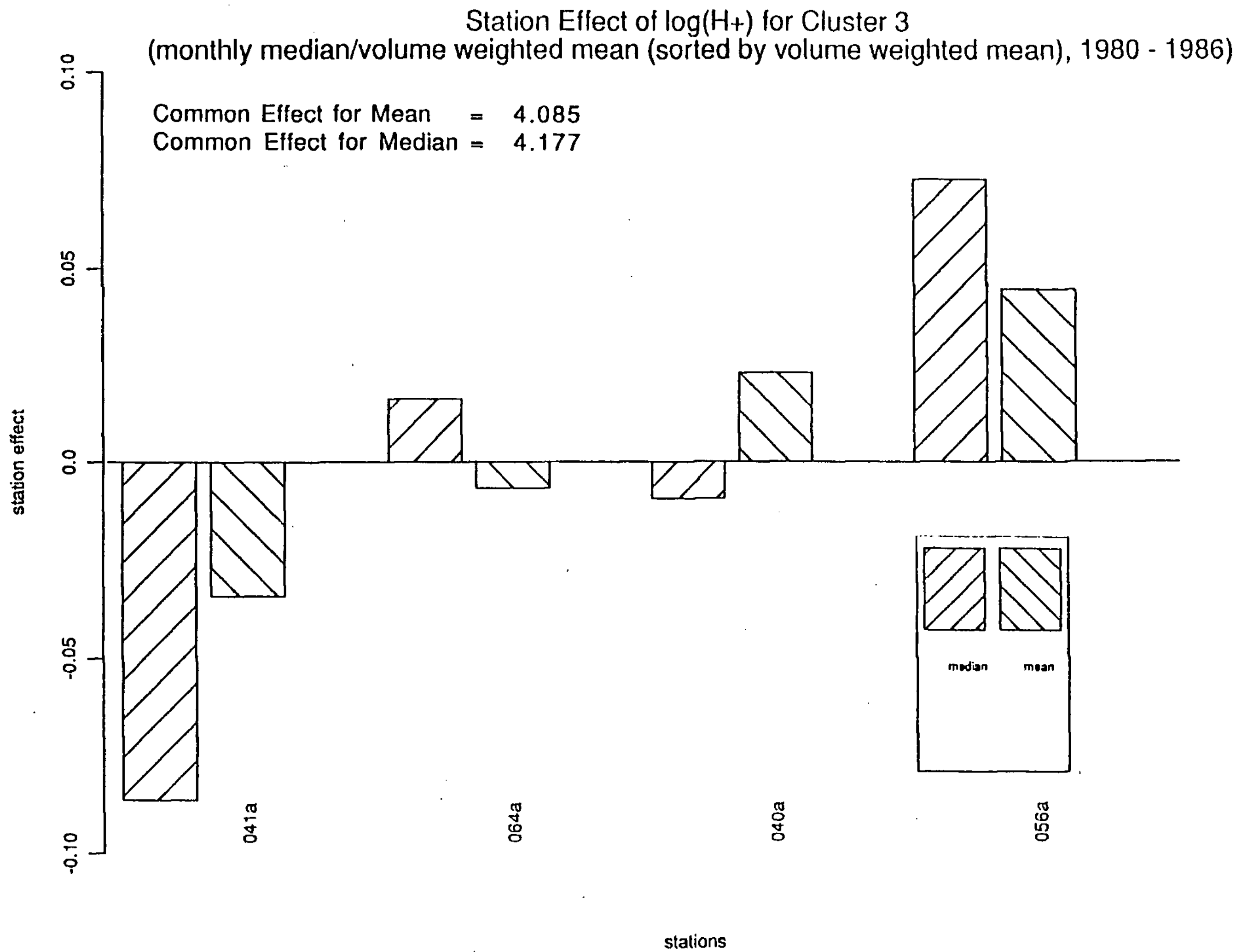

Station Effect of log(H+) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)
Common Effect for Mean = 2.350
Common Effect for Median = 2.393
station effect
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
038a
074a
059a
010a
173a
011a
004a
034a
076a
036a
052a
017a
030a
053a
049a
median
mean
stations

Figure 6.1.11



Station Effect of log(H+) for Cluster 3
(monthly median/volume weighted mean (sorted by volume weighted mean), 1980 - 1986)
Common Effect for Mean = 4.085
Common Effect for Median = 4.177
station effect
0.10
0.05
0.0
-0.05
-0.10
median
mean
041a
064a
040a
056a
stations

Figure 6.1.12

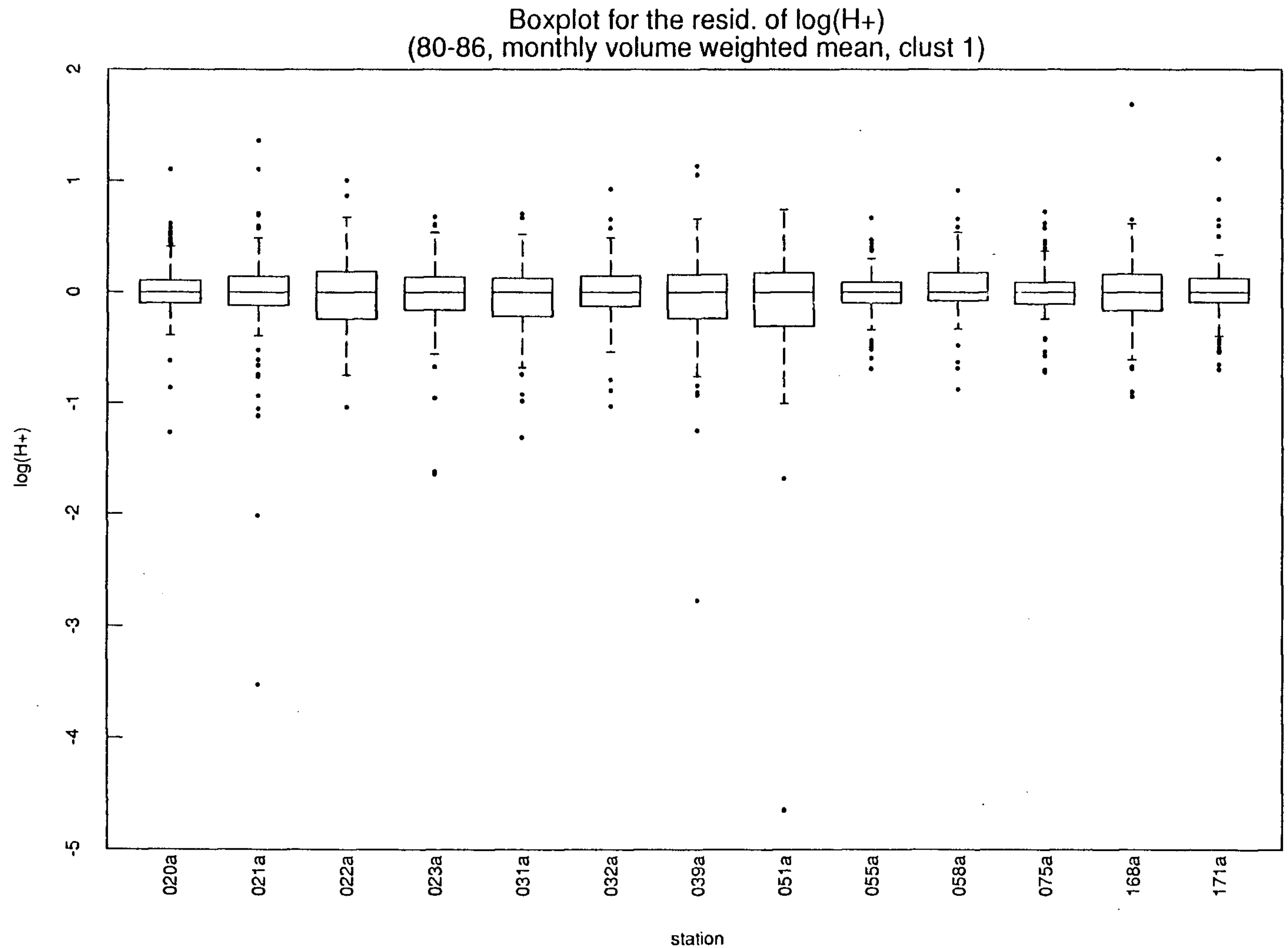


Figure 6.1.13(a)

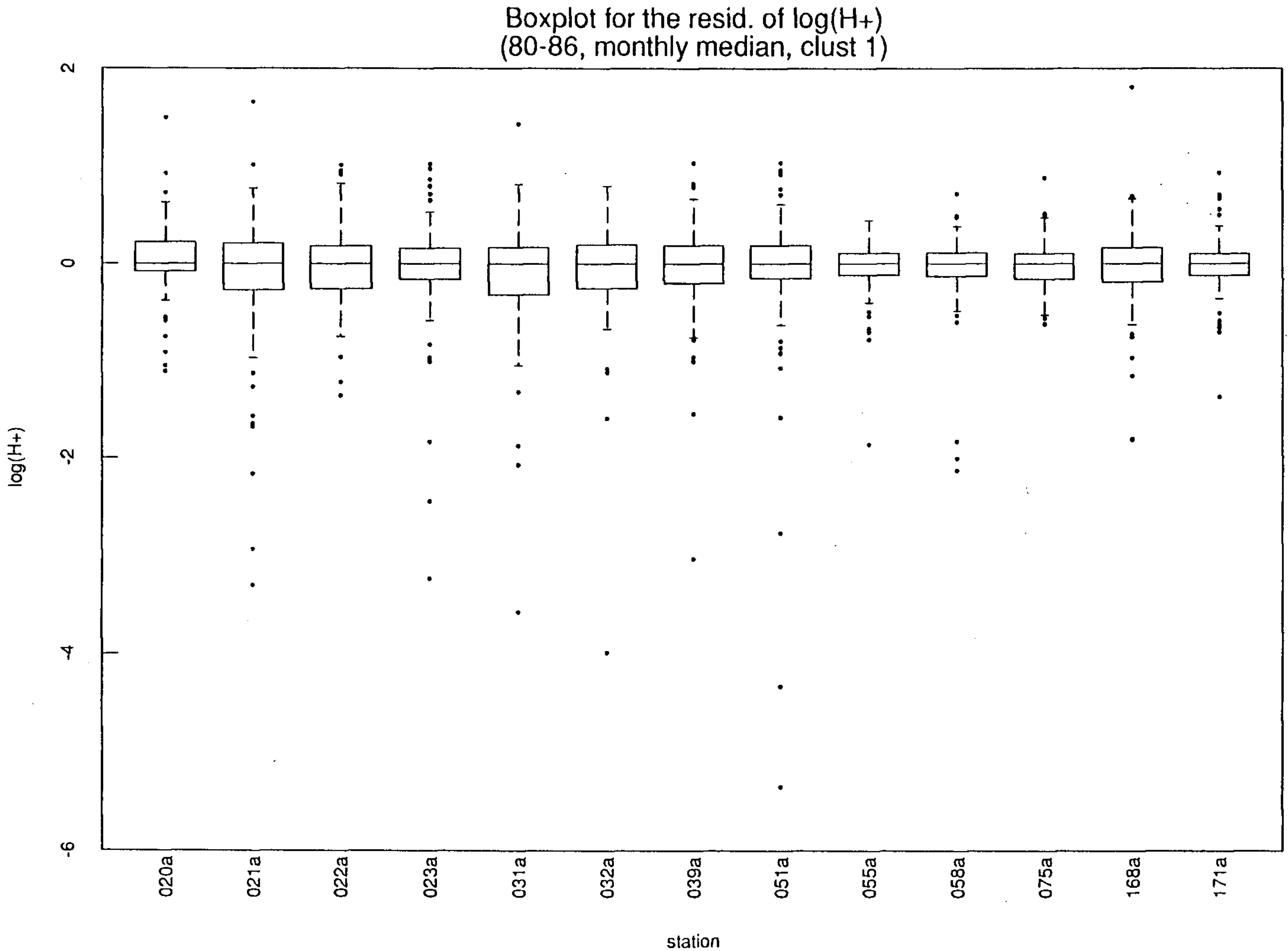


Figure 6.1.13(b)

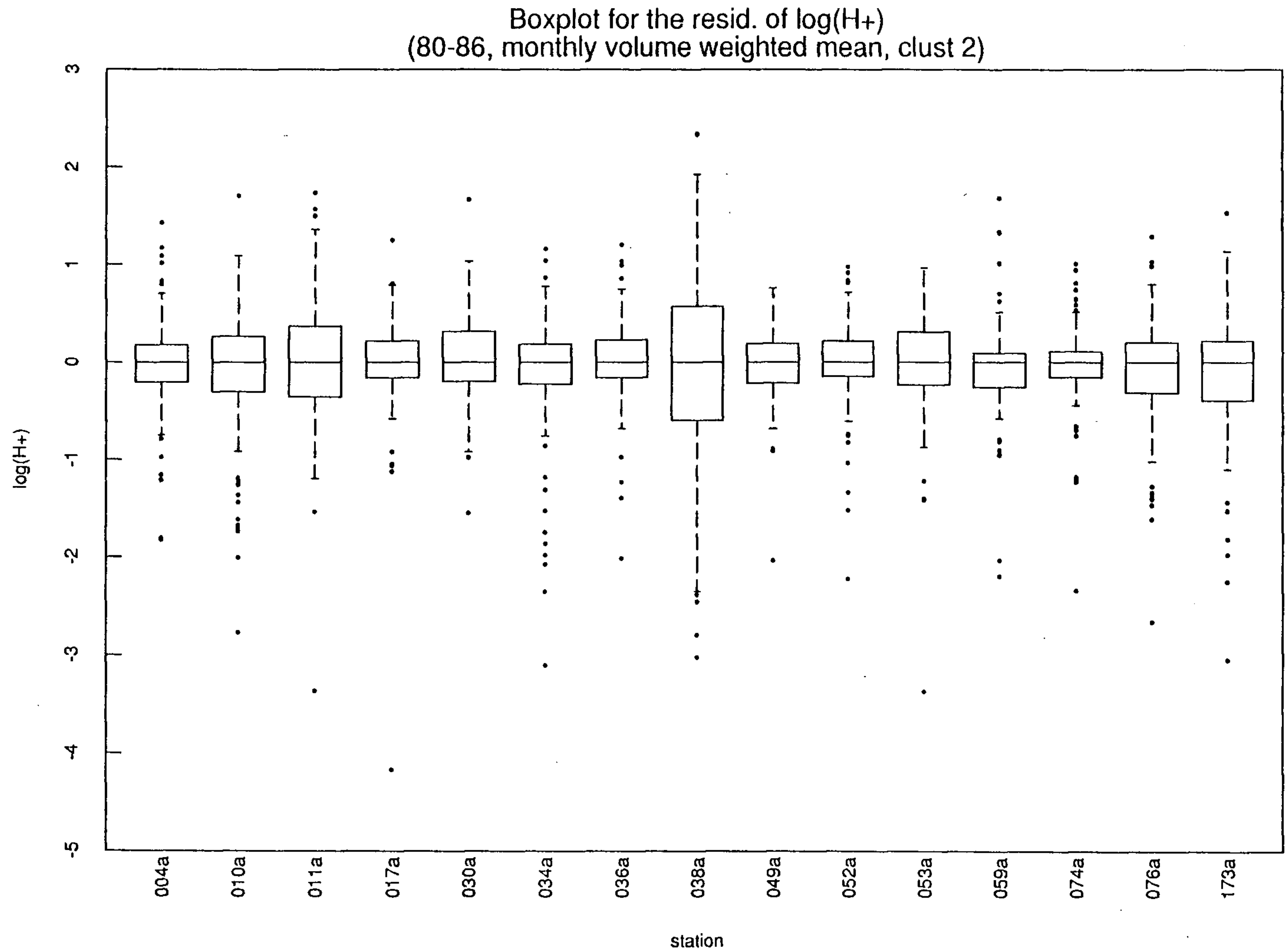


Figure 6.1.14(a)

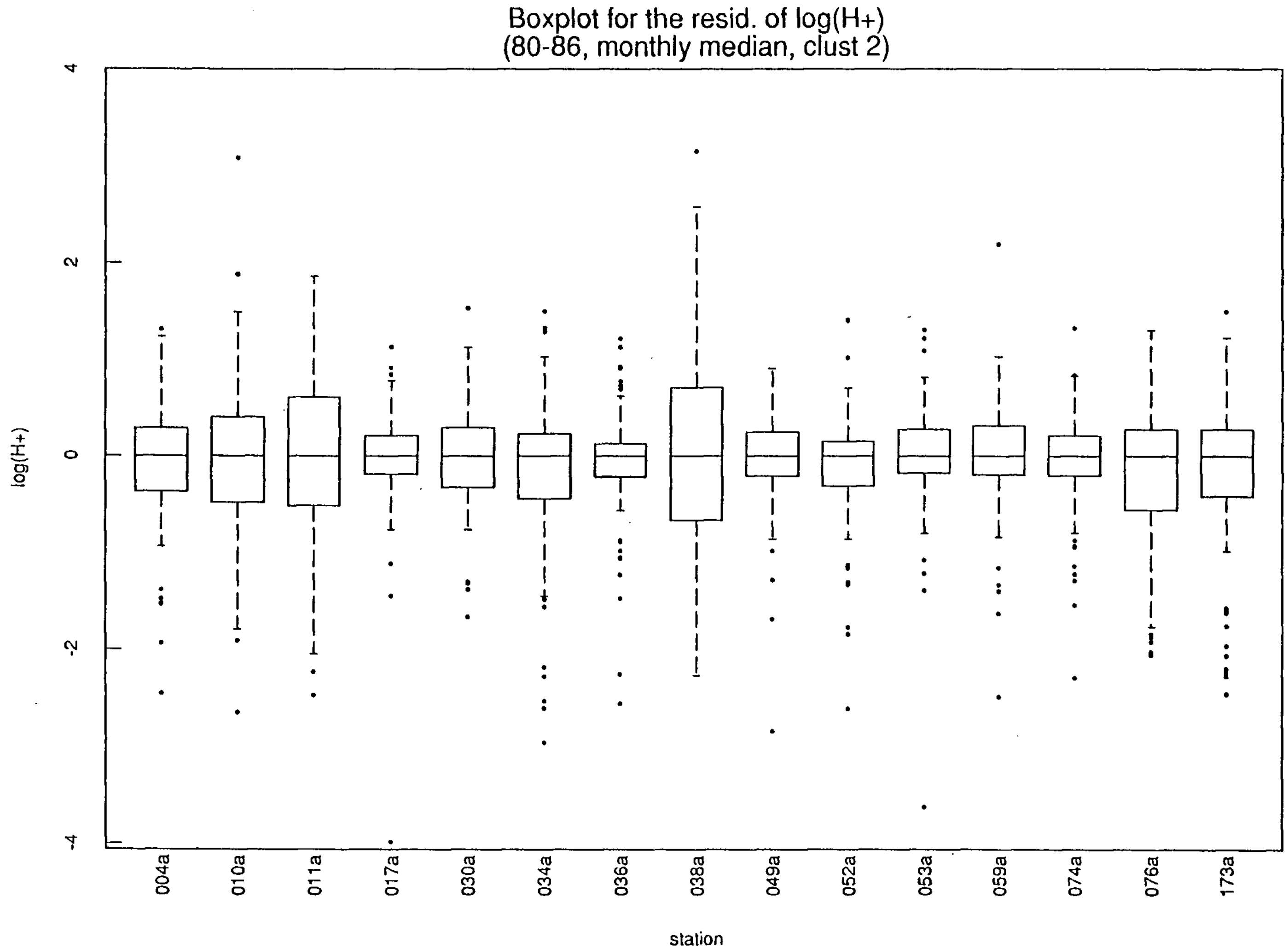


Figure 6.1.14(b)

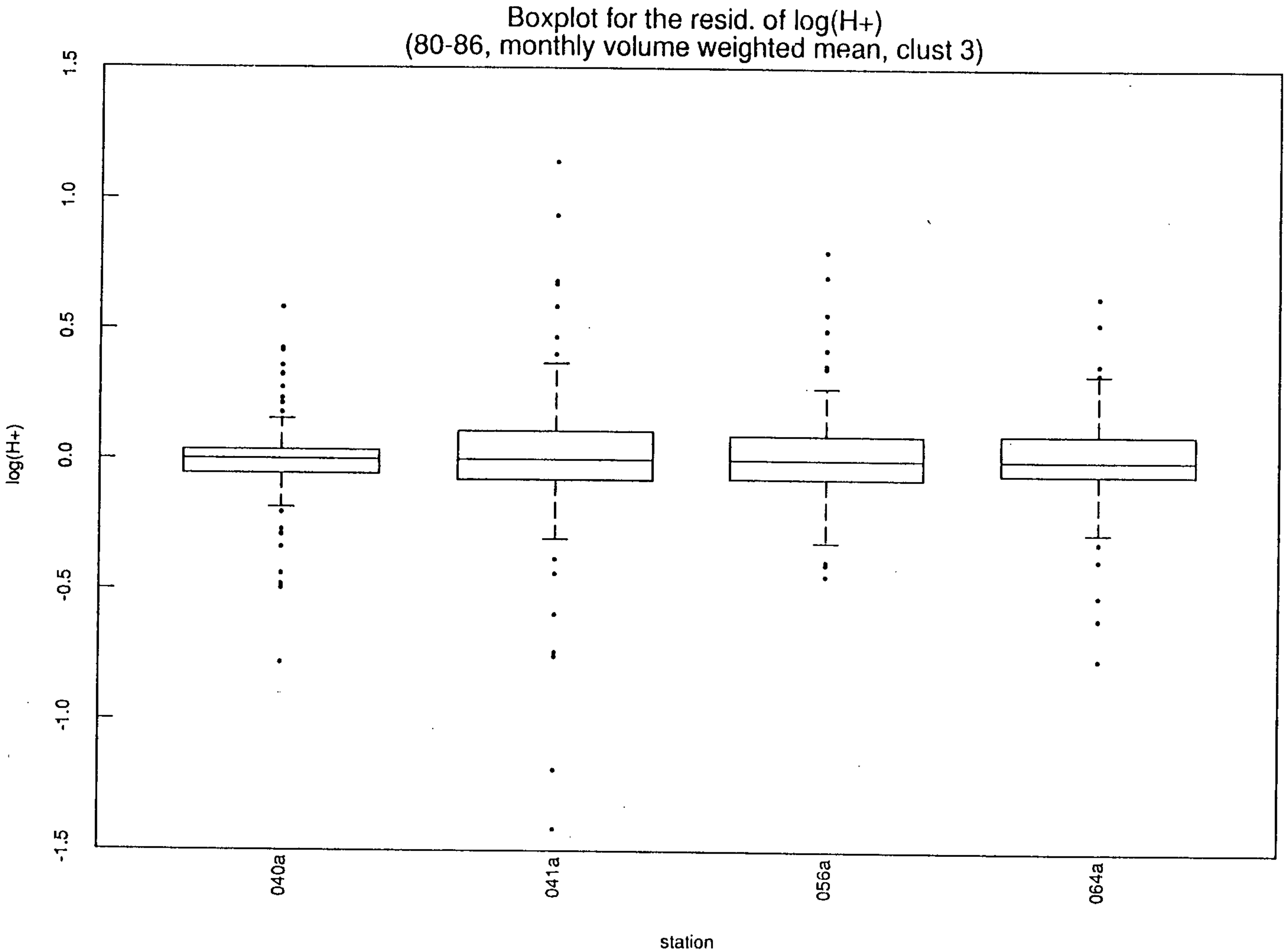


Figure 6.1.15(a)

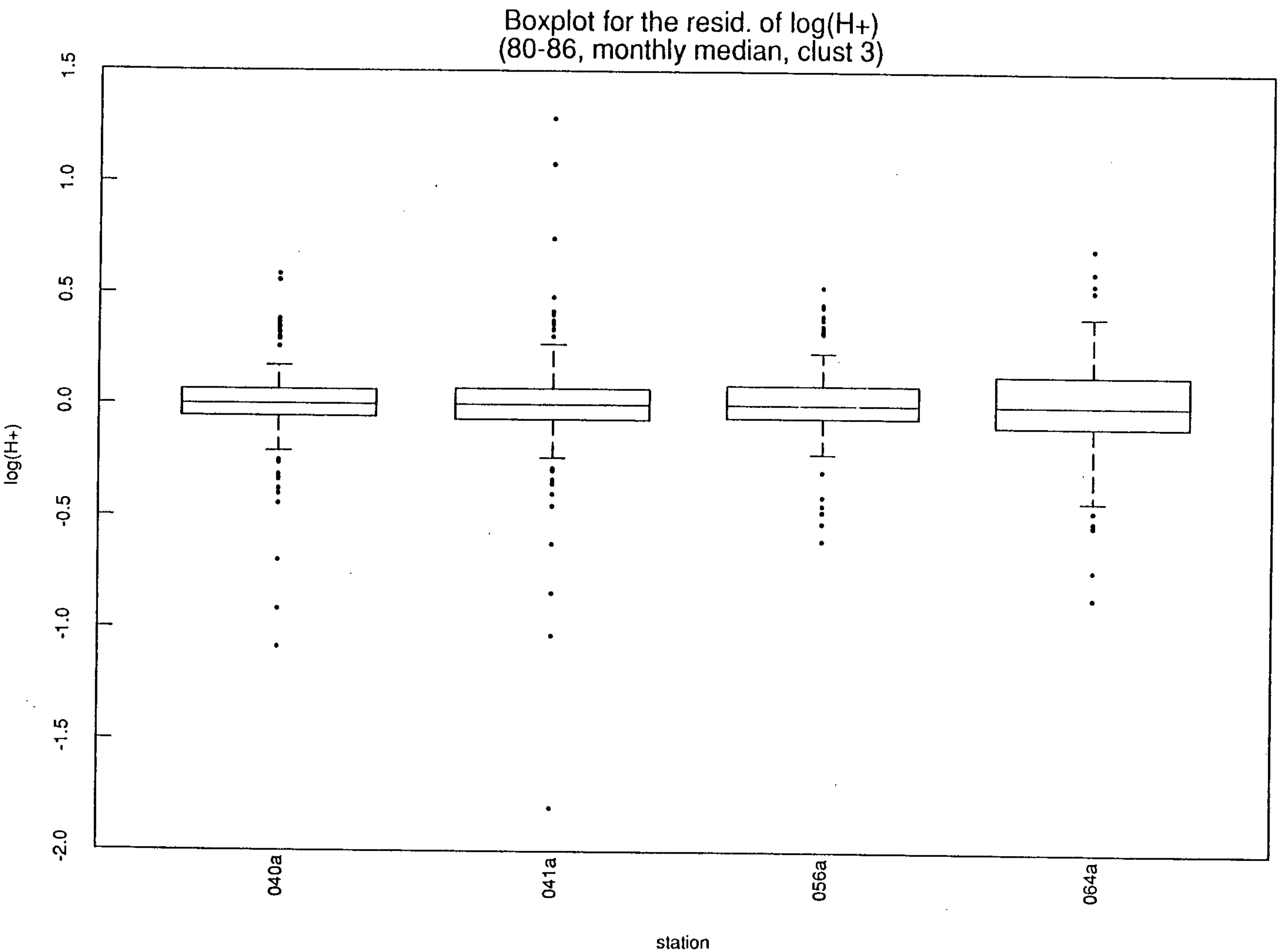


Figure 6.1.15(b)

Summary of the Effects and Residuals from Median Polish of log(H+)
(80-86,monthly volume weighted mean,clust 1)

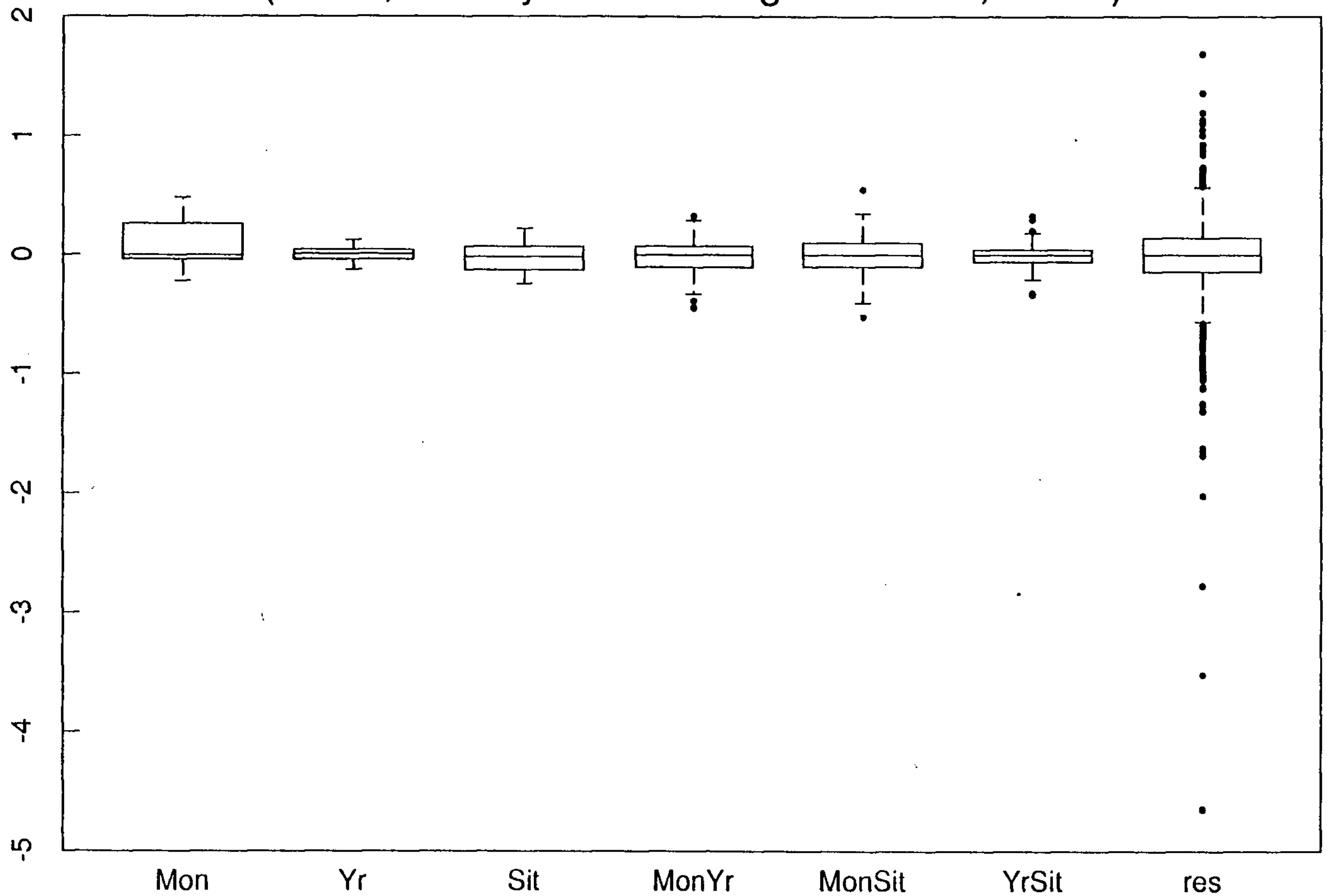


Figure 6.1.16(a)

Summary of the Effects and Residuals from Median Polish of log(H+)
(80-86,monthly median,clust 1)

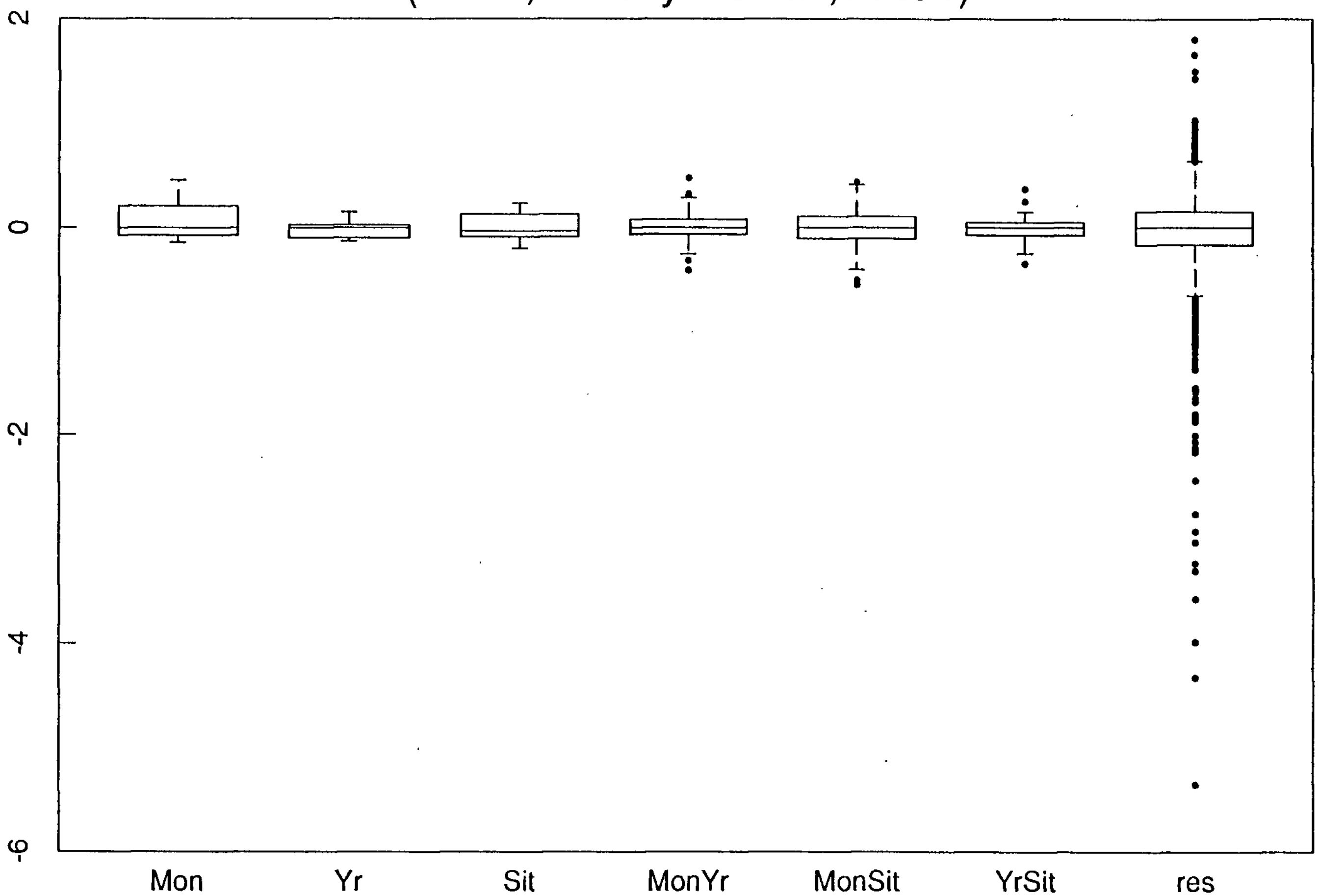


Figure 6.1.16(b)

Summary of the Effects and Residuals from Median Polish of log(H+)
(80-86,monthly volume weighted mean,clust 2)

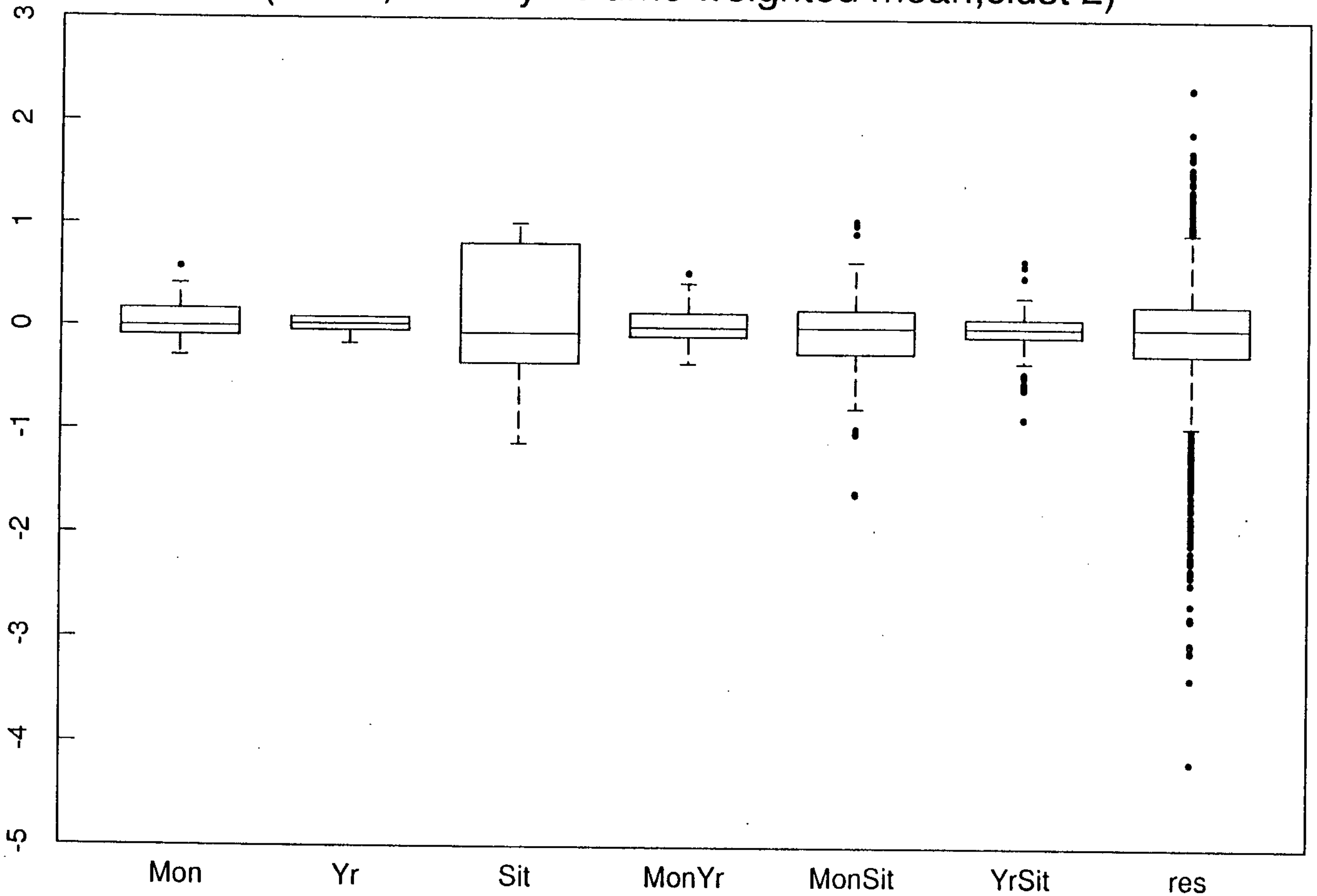


Figure 6.1.17(a)

Summary of the Effects and Residuals from Median Polish of log(H+)
(80-86,monthly median,clust 2)

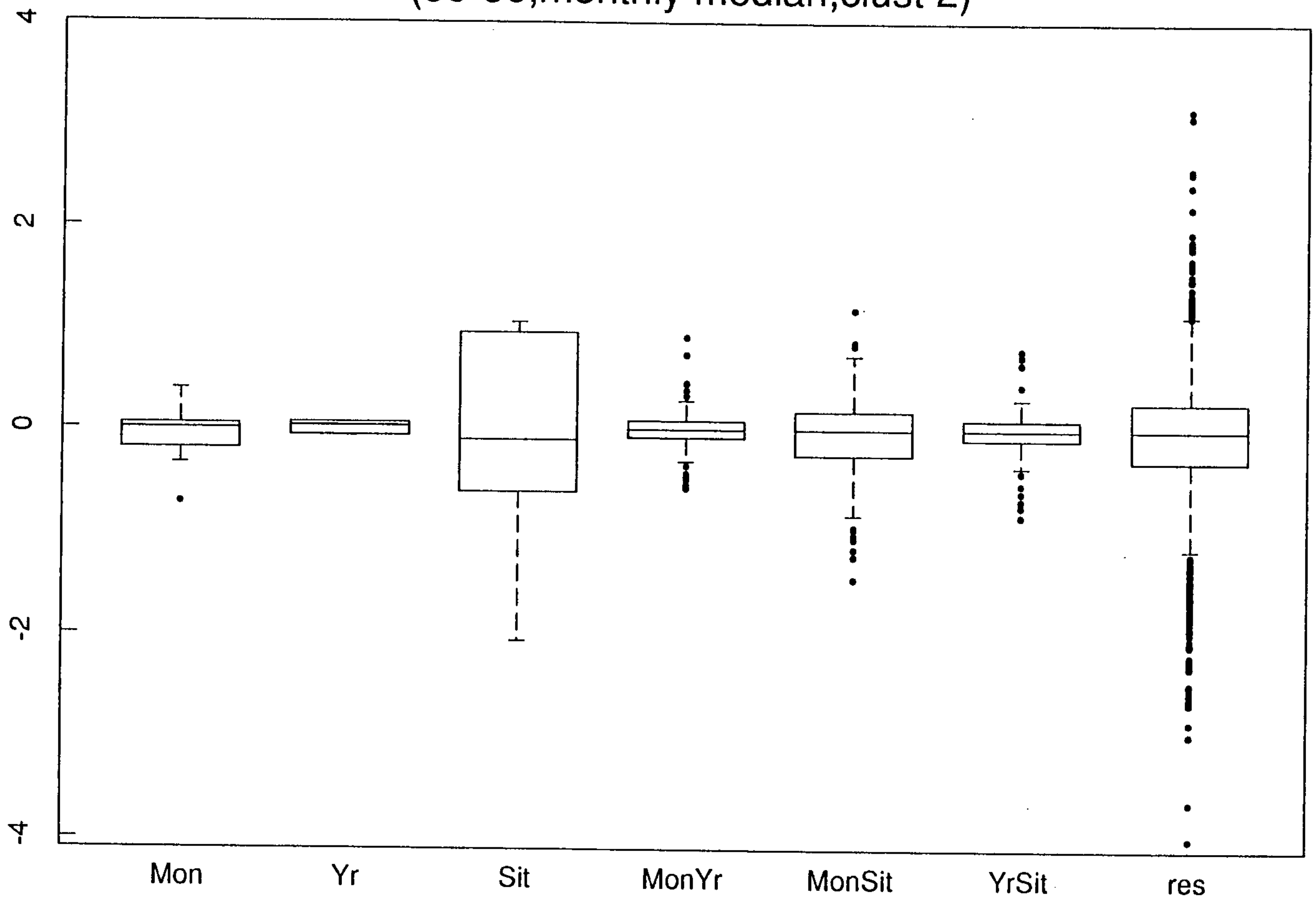


Figure 6.1.17(b)

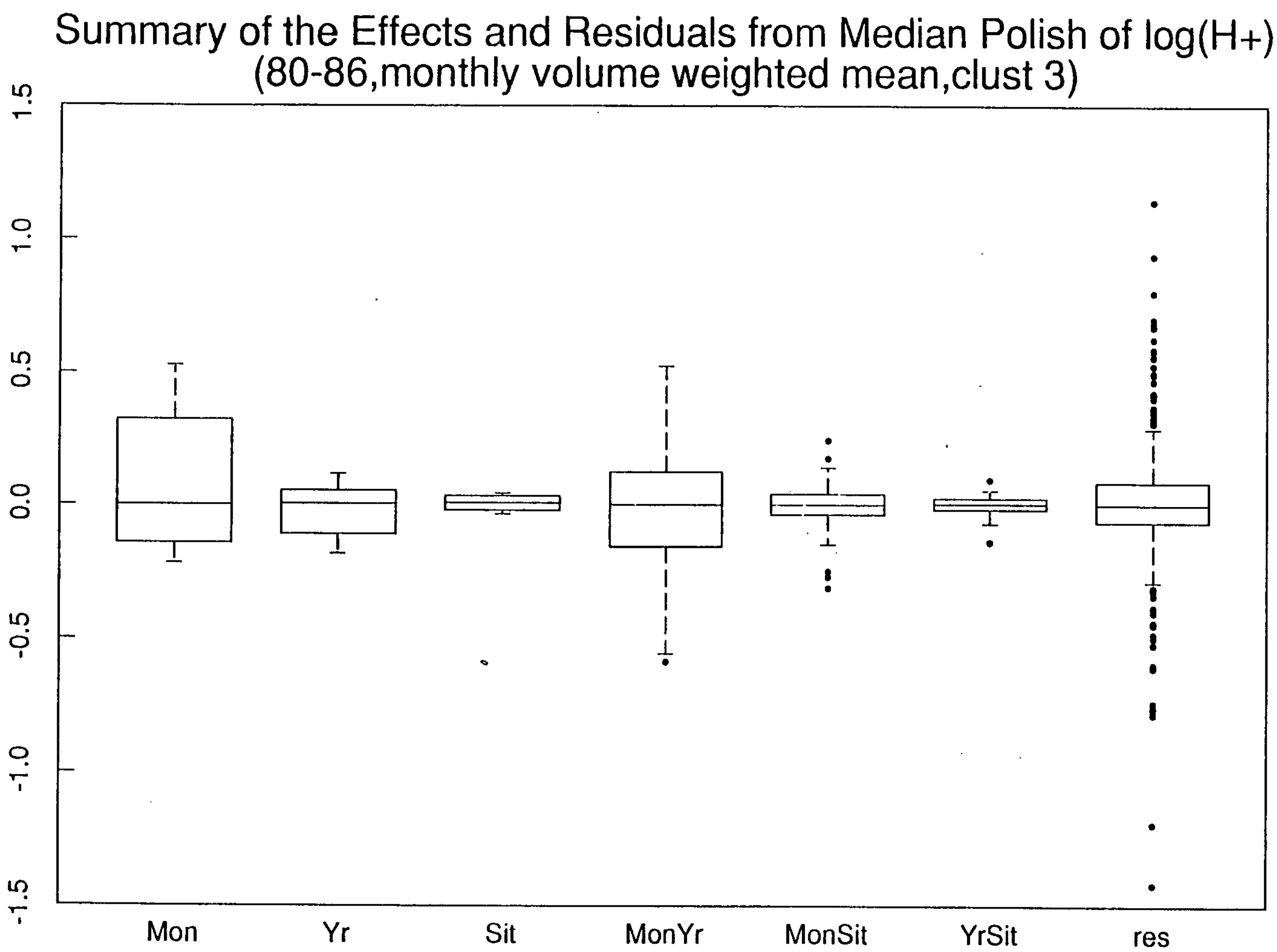


Figure 6.1.18(a)

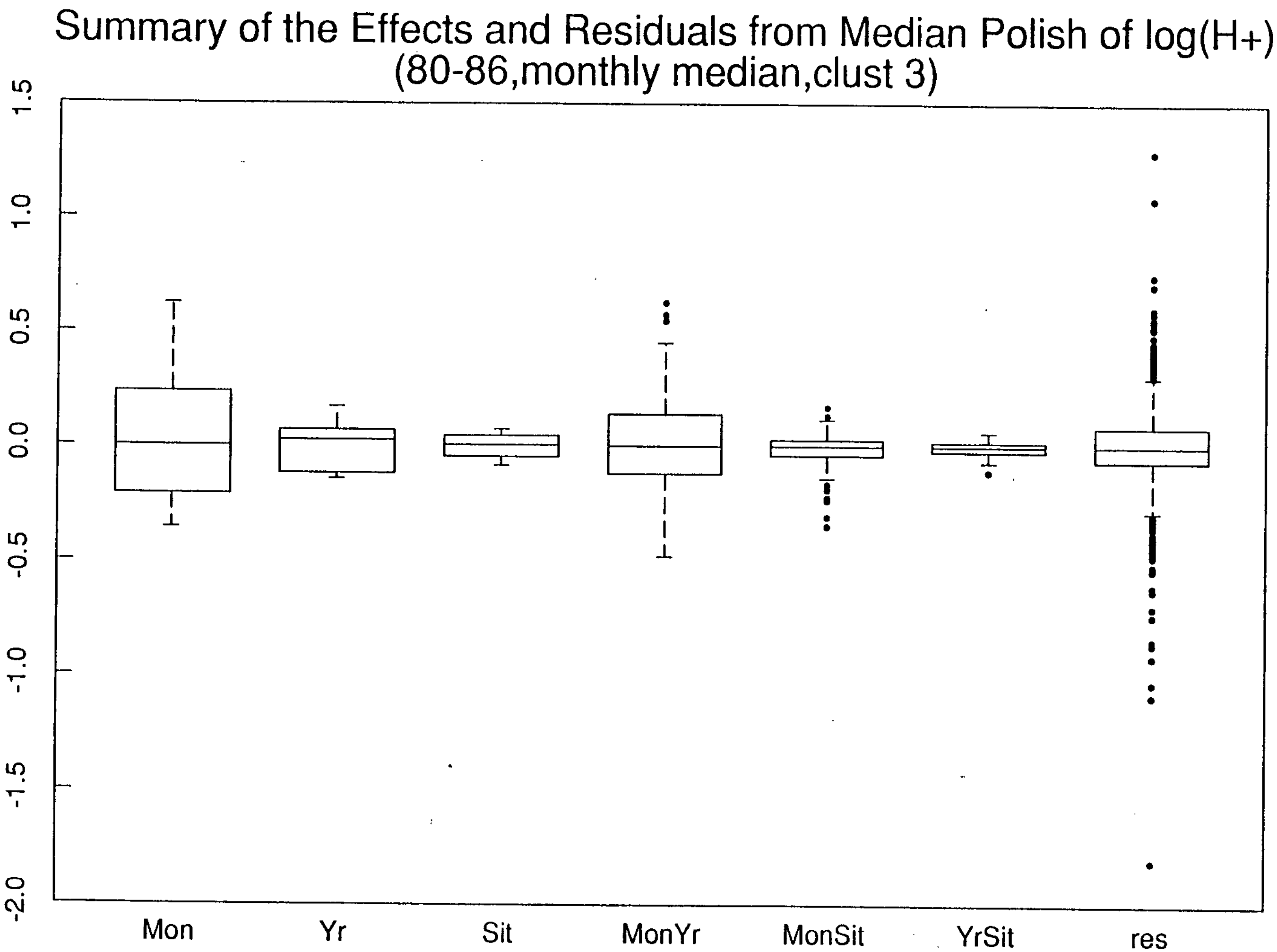
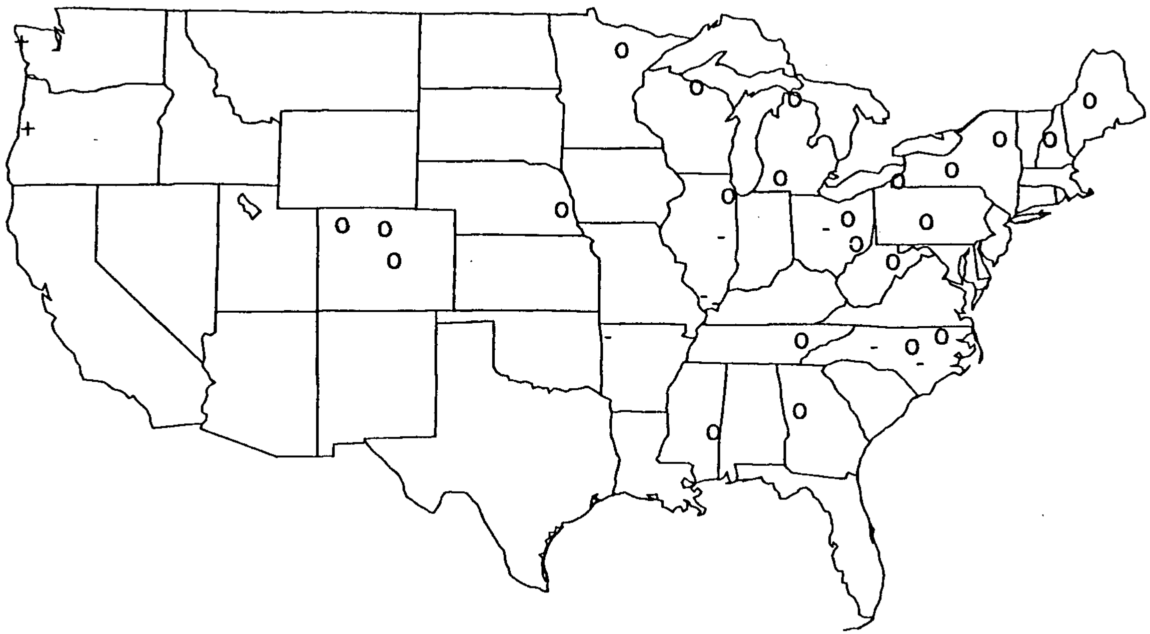


Figure 6.1.18(b)

Trend of log(H+) from 1980 to 1986 at the 32 Stations
(calculated by monthly volume weighted mean)



o	no trend
-	down trend
+	up trend

Figure 6.1.19(a)

Trend of log(H+) from 1980 to 1986 at the 32 Stations
(calculated by monthly median)

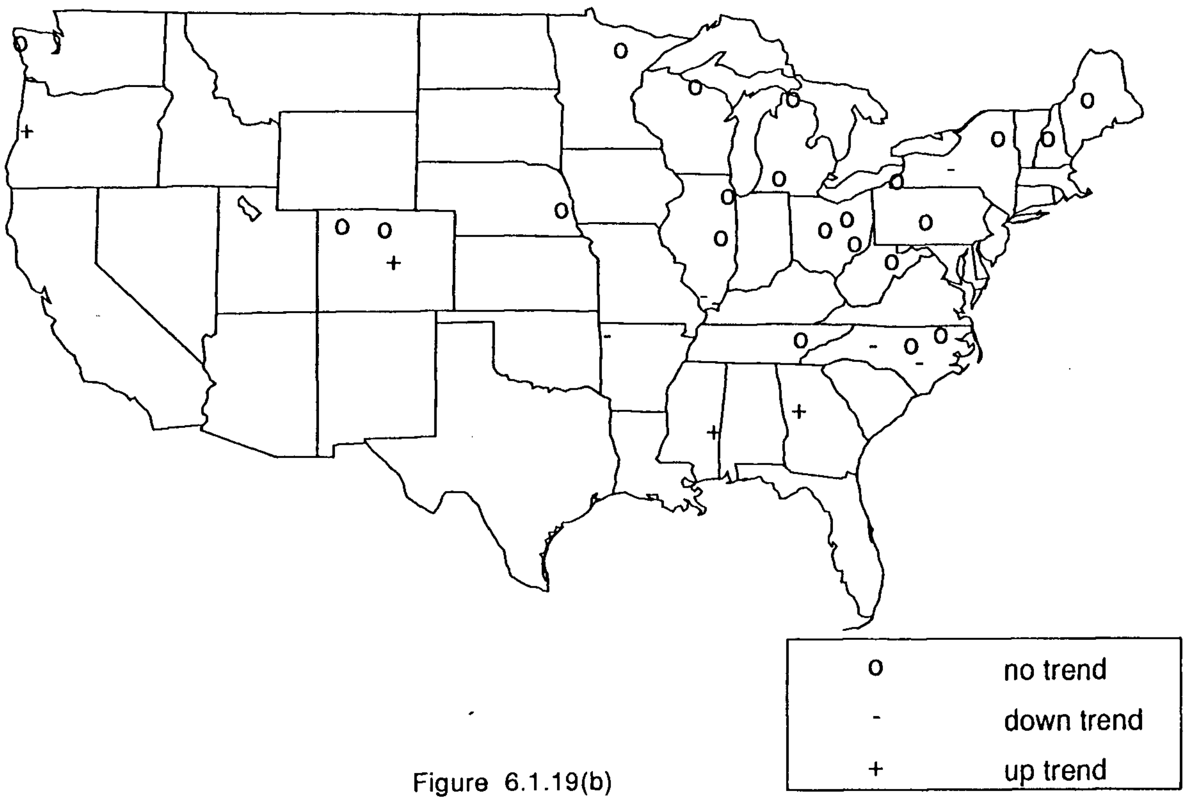


Figure 6.1.19(b)

Trend of log(H+) from 1980 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

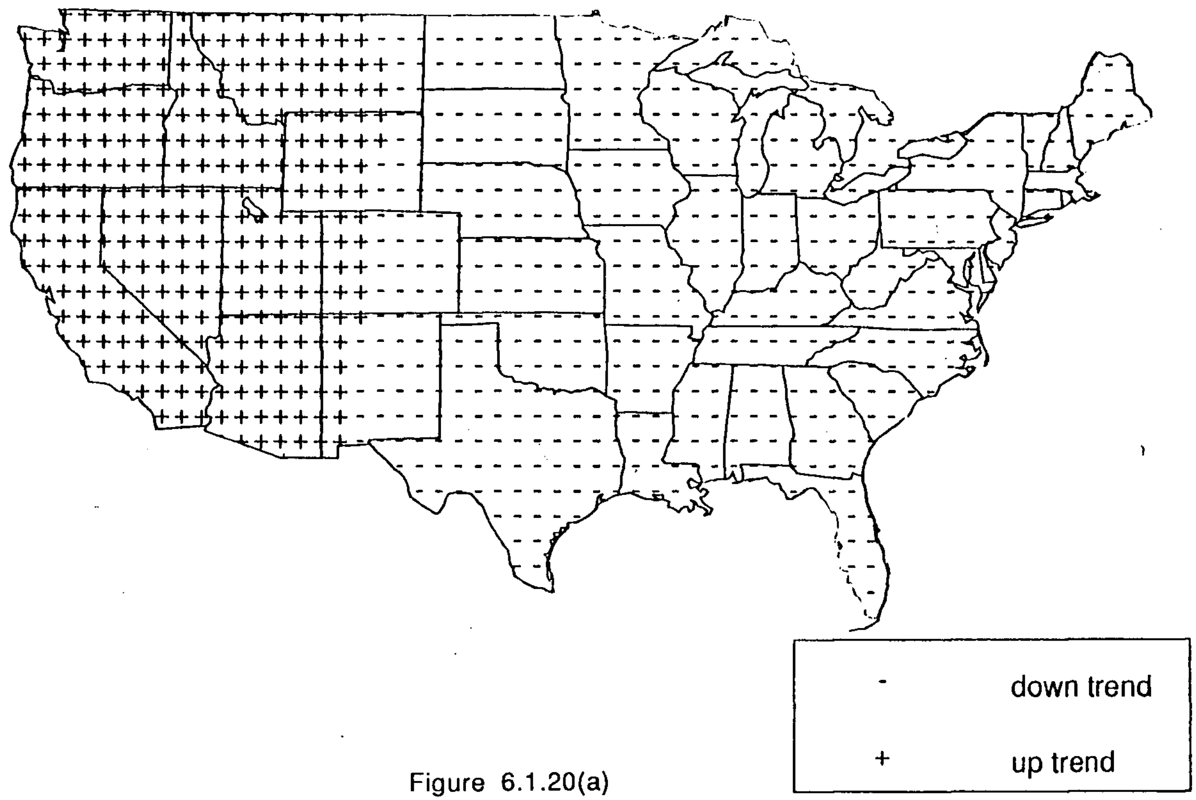


Figure 6.1.20(a)

Trend of log(H+) from 1980 to 1986 in the USA
(calculated by Kriging from monthly median)

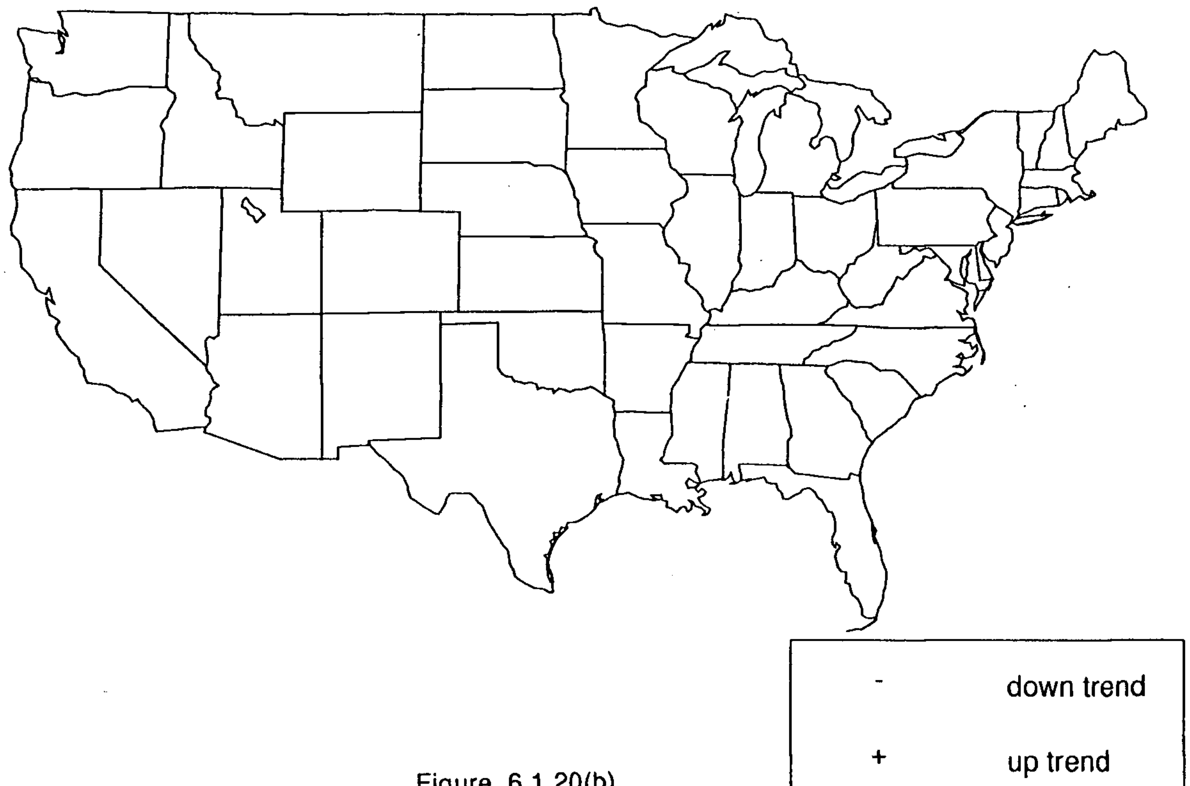


Figure 6.1.20(b)

The Locations Of The 86 Monitoring Stations From 1983 To 1986
(For Hydrogen ion)

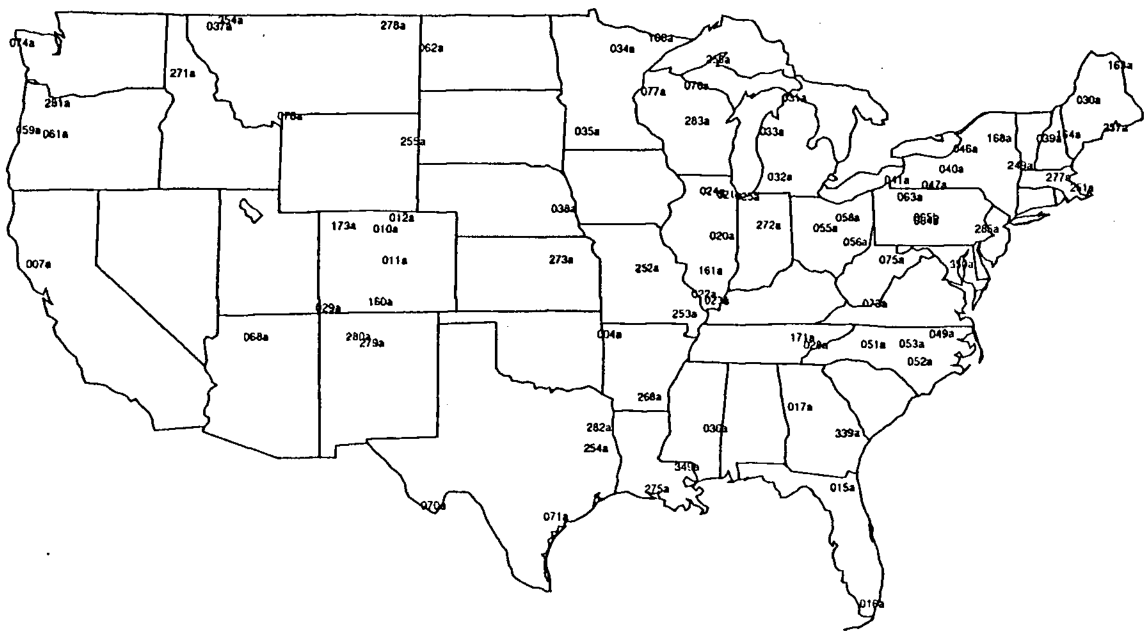


Figure 6.2.1

Clustering of H+ monthly volume weighted mean based on sqrt(MSE)

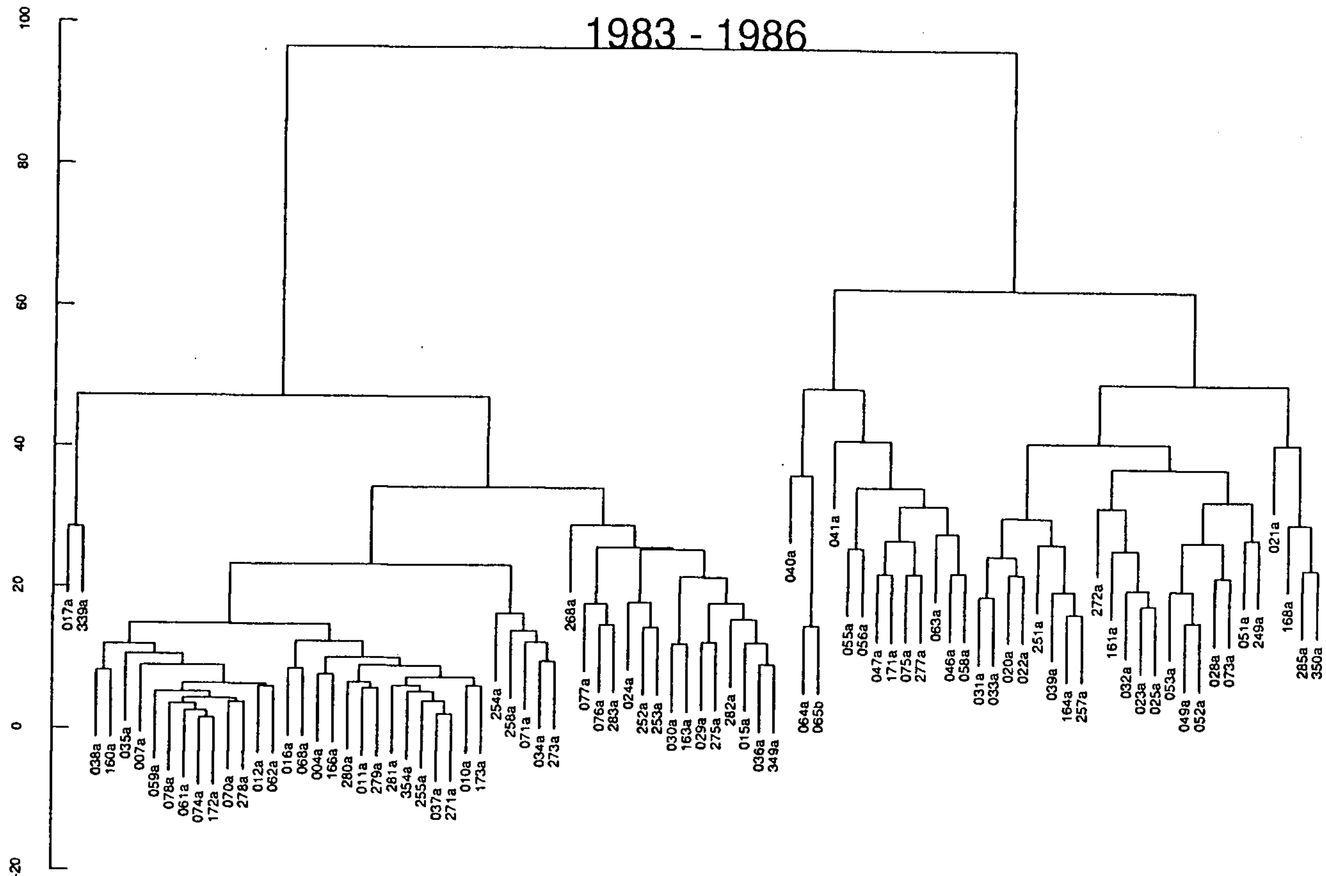


Figure 6.2.2

Clusters of H+ monthly volume weighted mean based on sqrt(MSE)

1983 - 1986 (k=3)

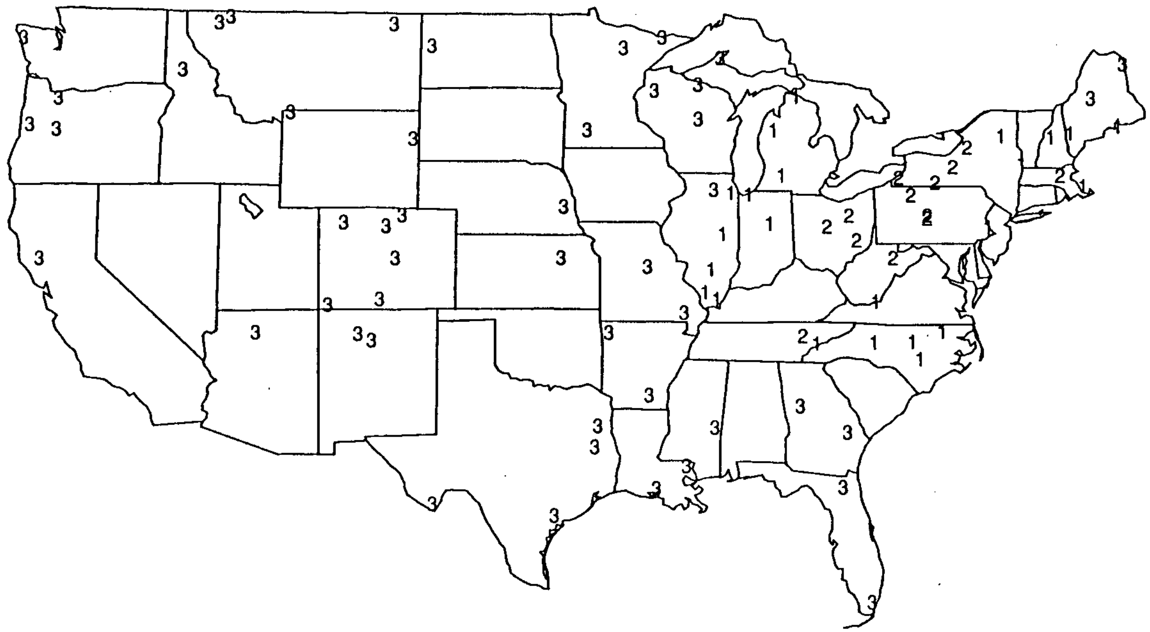


Figure 6.2.3

Histograms of Transformed H+ (Volume Weighted Mean, 83-86)

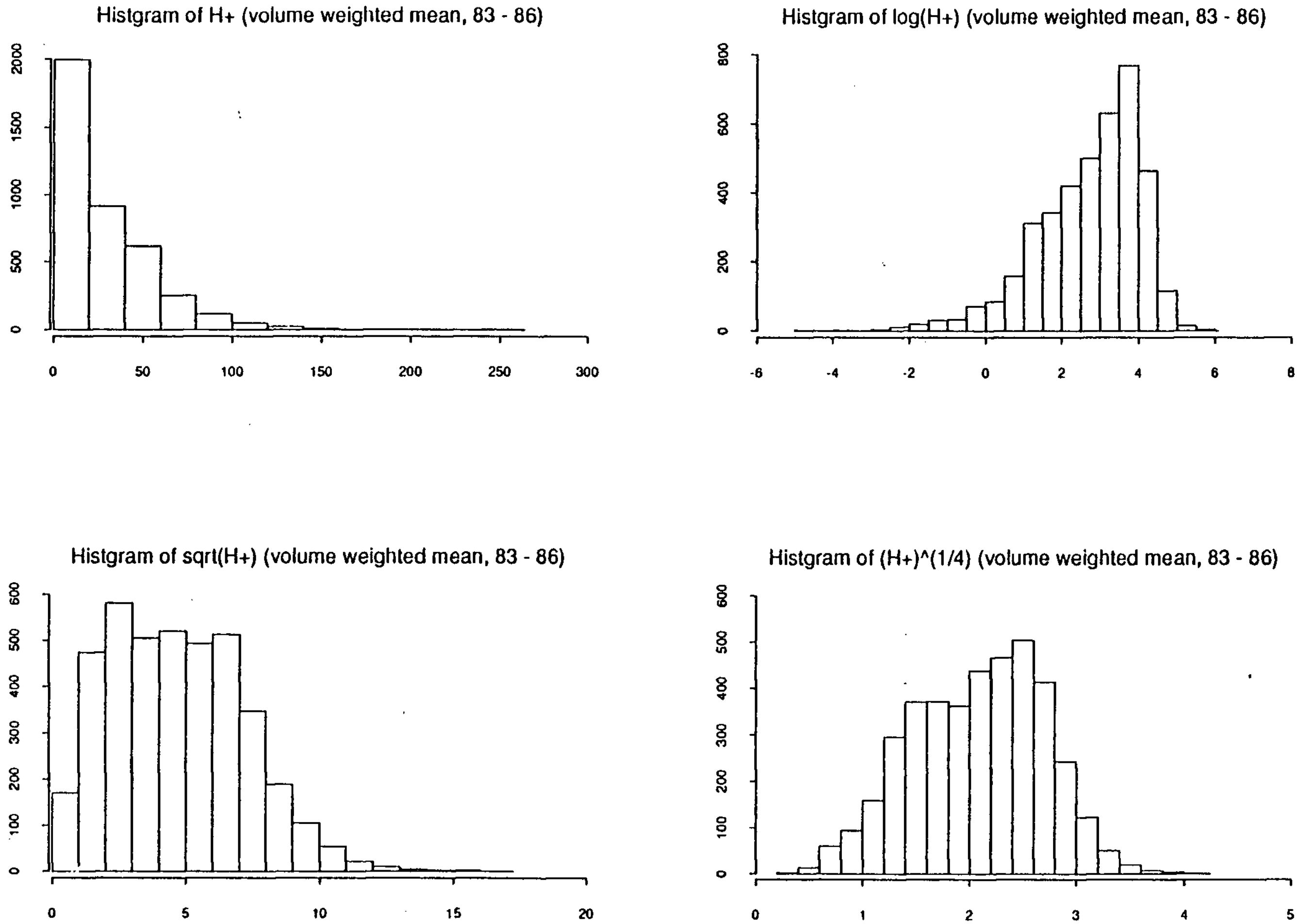
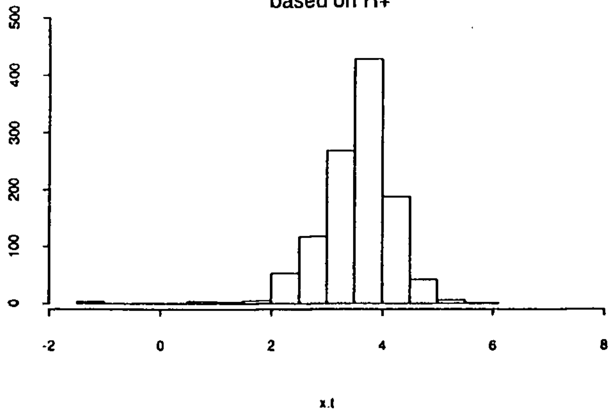


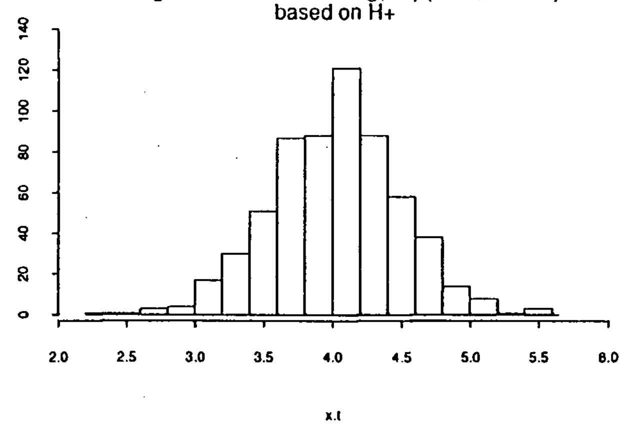
Figure 6.2.4

Histograms of log(H+) by Clusters (Volume Weighted Mean, 83-86)

Histogram for Cluster 1 of log(H+) (vwm, 83 - 86)
based on H+



Histogram for Cluster 2 of log(H+) (vwm, 83 - 86)
based on H+



Histogram for Cluster 3 of log(H+) (vwm, 83 - 86)
based on H+

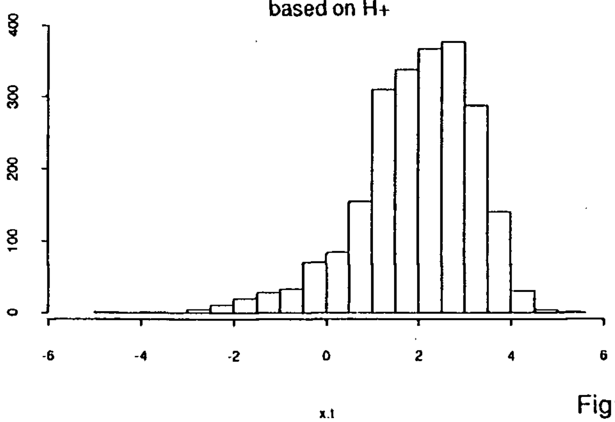
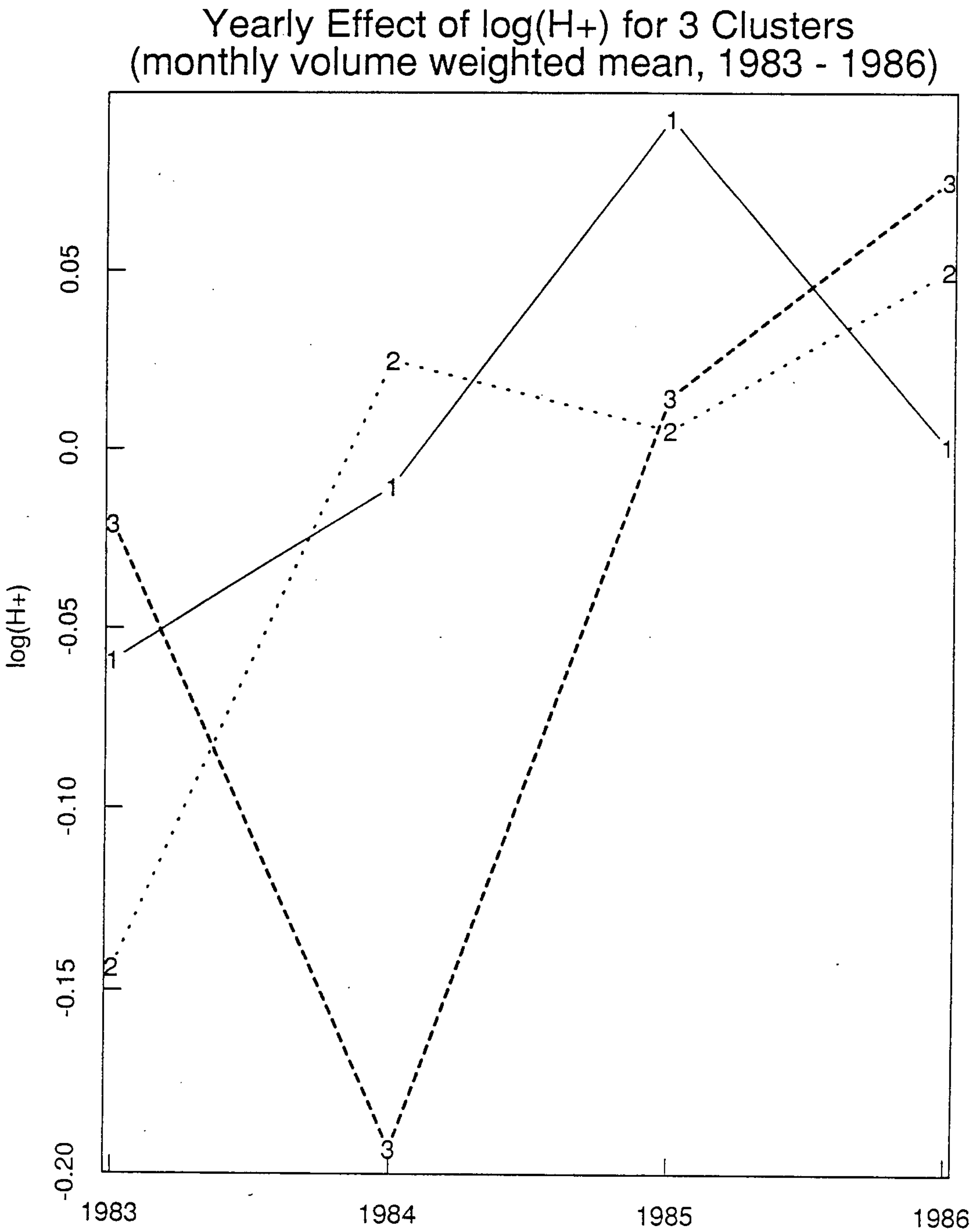
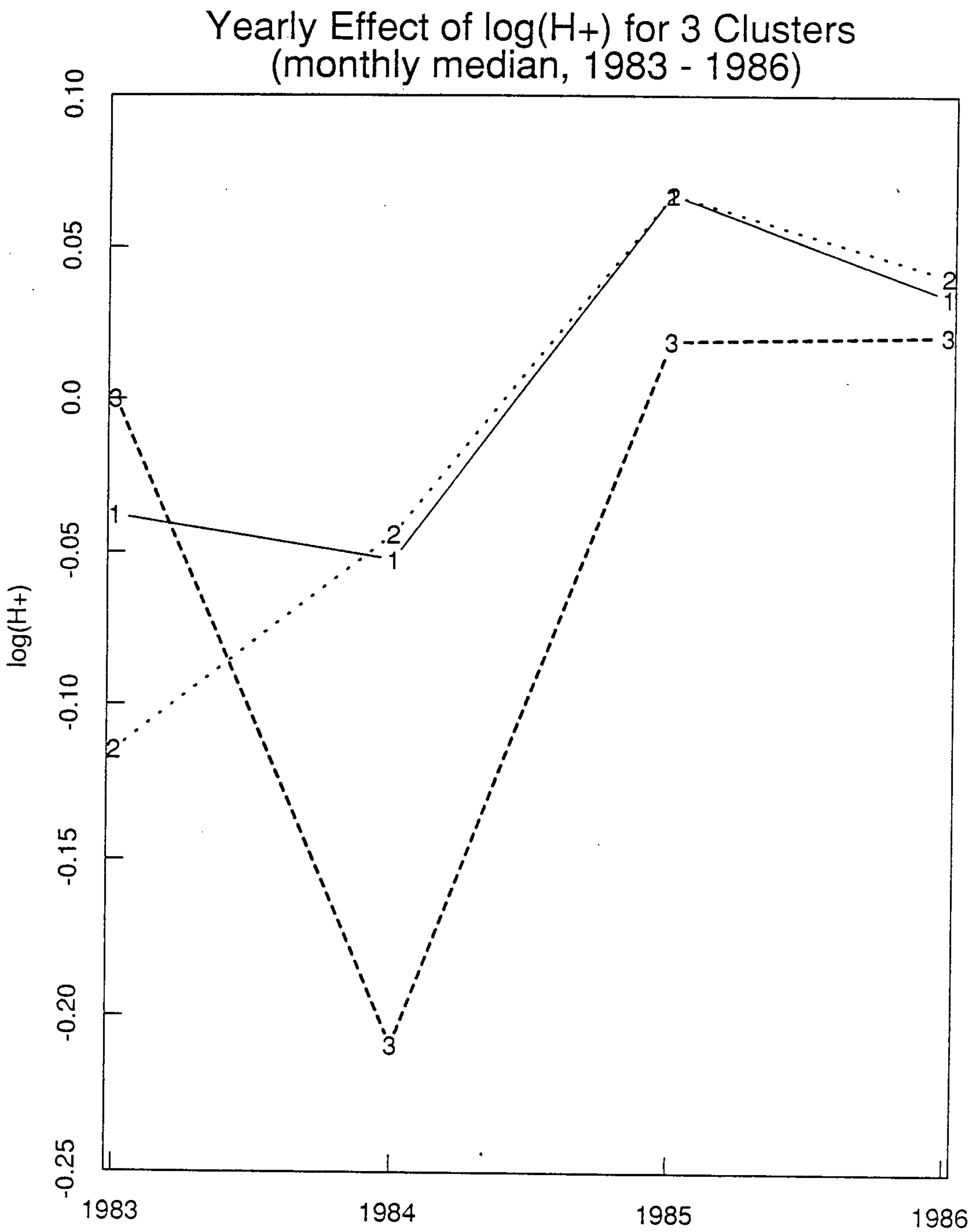


Figure 6.2.5



Yearly Effect of log(H+) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)
log(H+)
0.05
0.0
-0.05
-0.10
-0.15
-0.20
1983
1984
1985
1986

Figure 6.2.6



Yearly Effect of log(H+) for 3 Clusters
(monthly median, 1983 - 1986)
log(H+)
0.10
0.05
0.0
-0.05
-0.10
-0.15
-0.20
-0.25
1983
1984
1985
1986

Figure 6.2.7

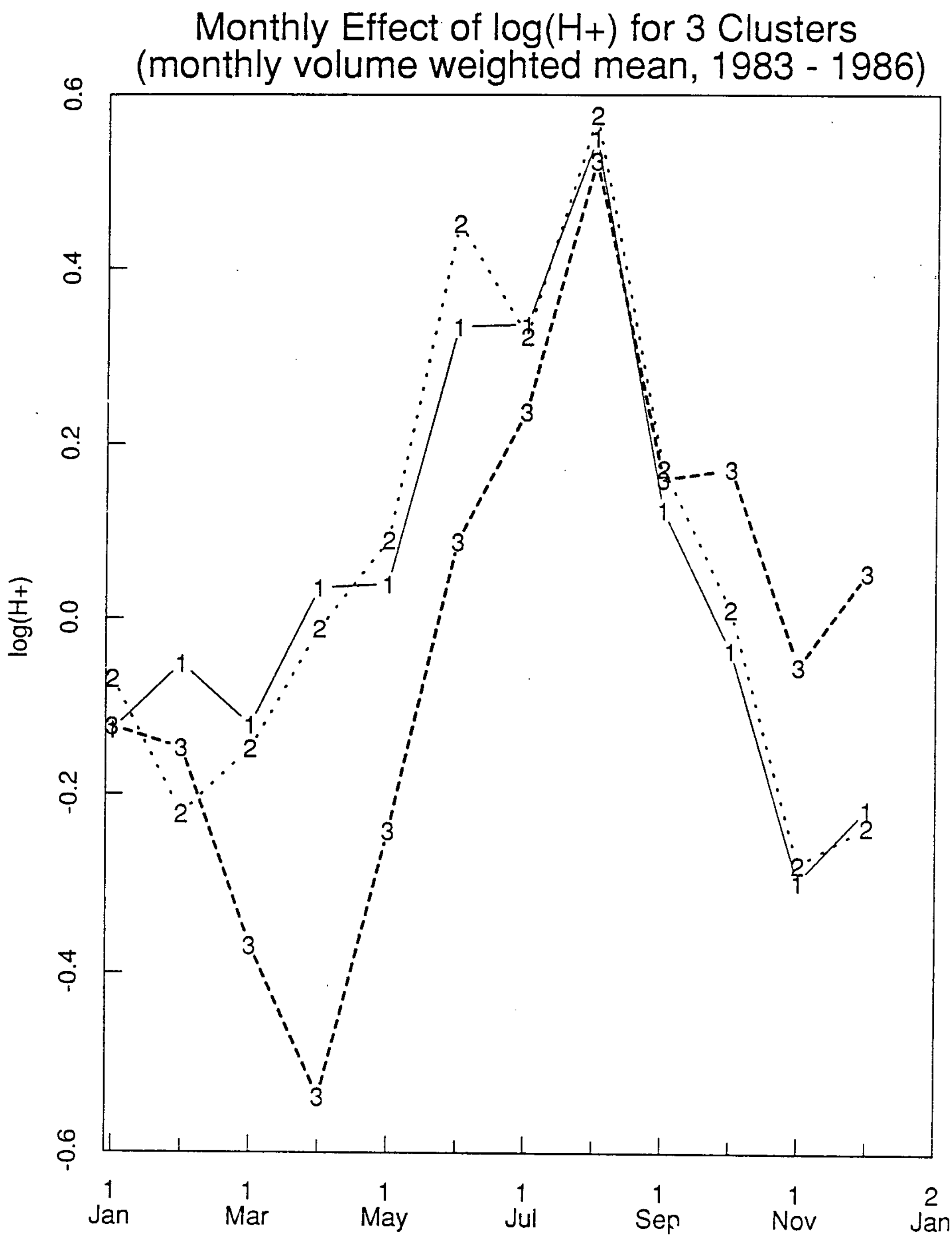

Monthly Effect of log(H+) for 3 Clusters
(monthly volume weighted mean, 1983 - 1986)
log(H+)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 6.2.8

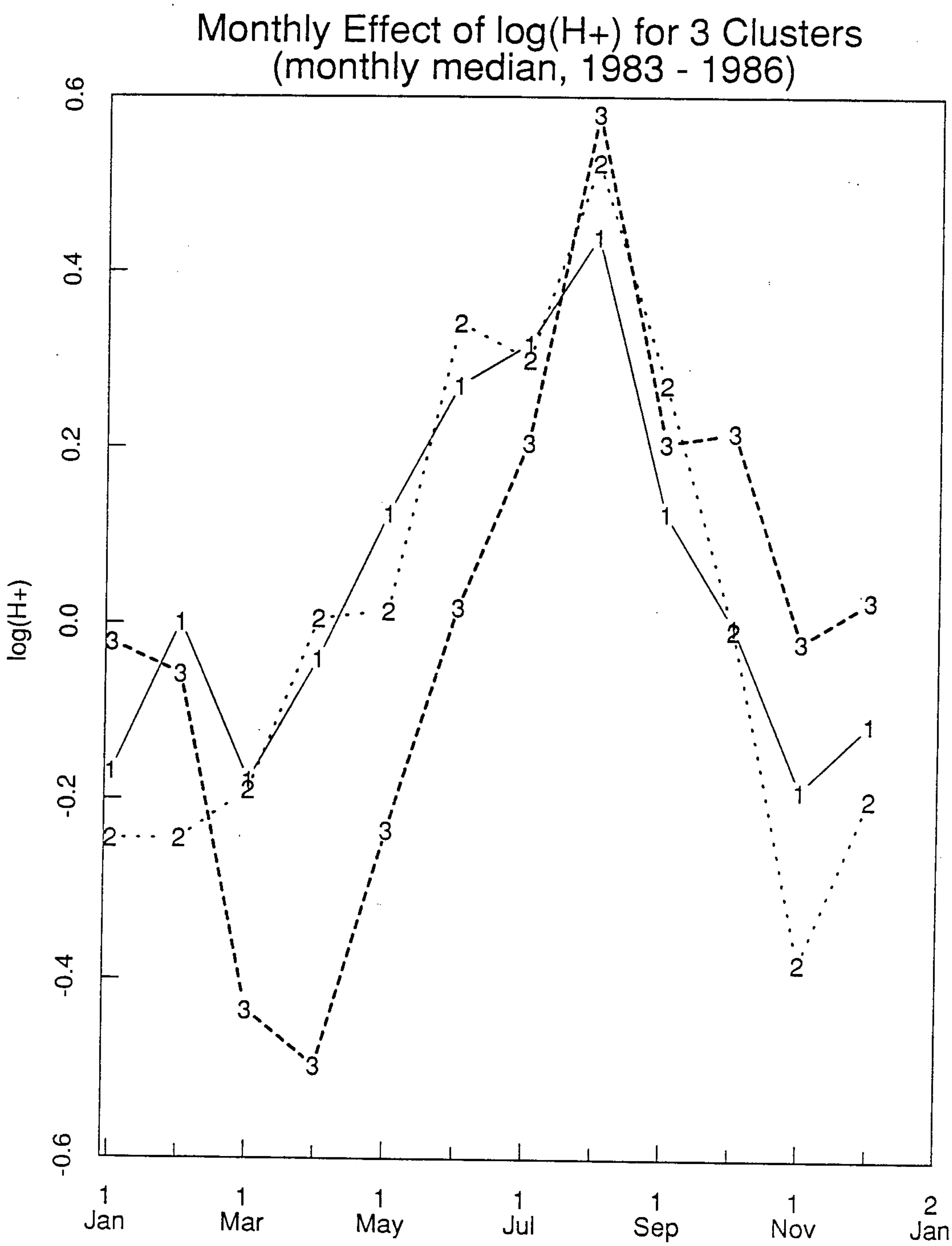

Monthly Effect of log(H+) for 3 Clusters
(monthly median, 1983 - 1986)
log(H+)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1 Jan
1 Mar
1 May
1 Jul
1 Sep
1 Nov
2 Jan

Figure 6.2.9

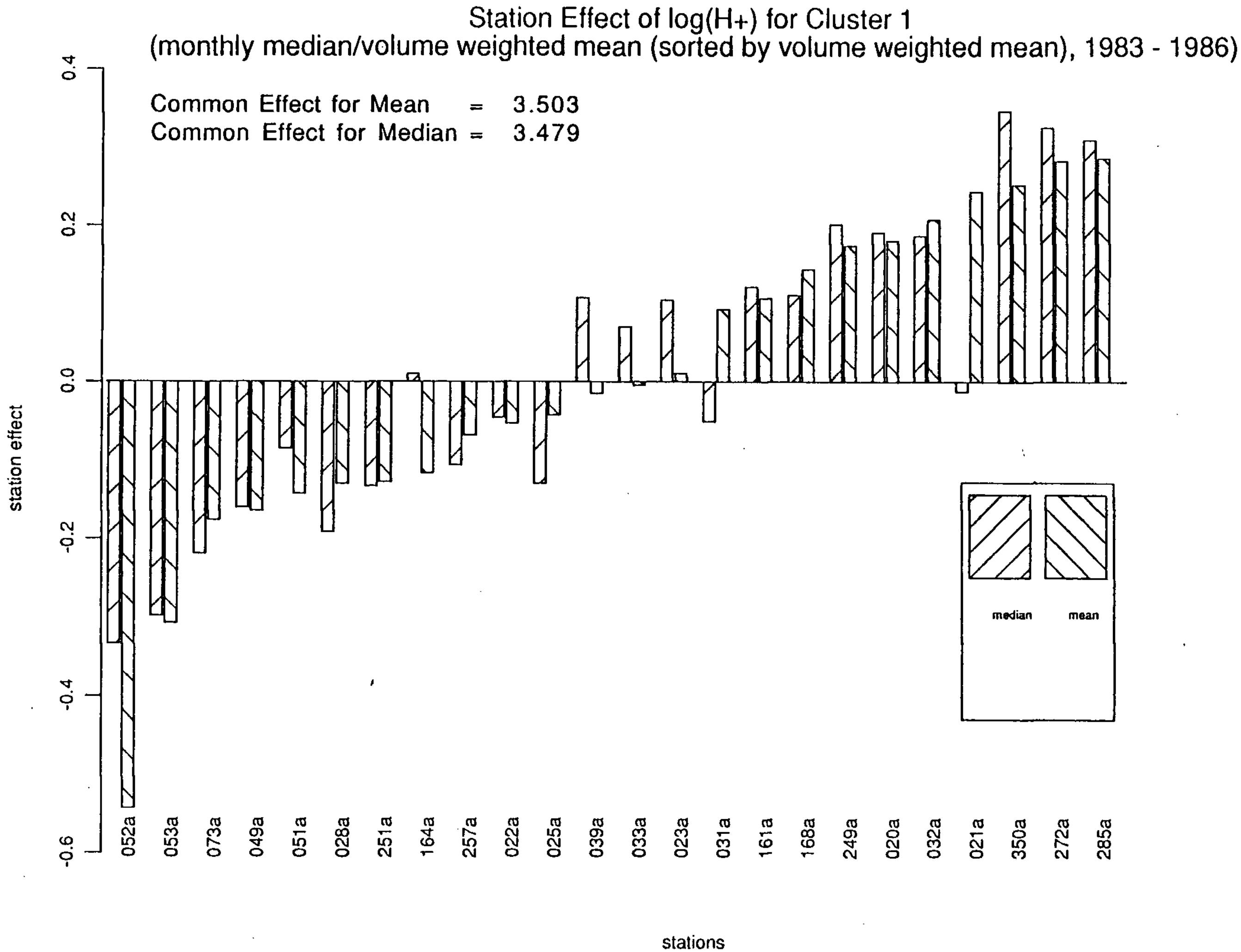

Station Effect of log(H+) for Cluster 1
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 3.503
Common Effect for Median = 3.479
station effect
0.4
0.2
0.0
-0.2
-0.4
-0.6
052a
053a
073a
049a
051a
028a
251a
164a
257a
022a
025a
039a
033a
023a
031a
161a
168a
249a
020a
032a
021a
350a
272a
285a
median
mean
stations

Figure 6.2.10

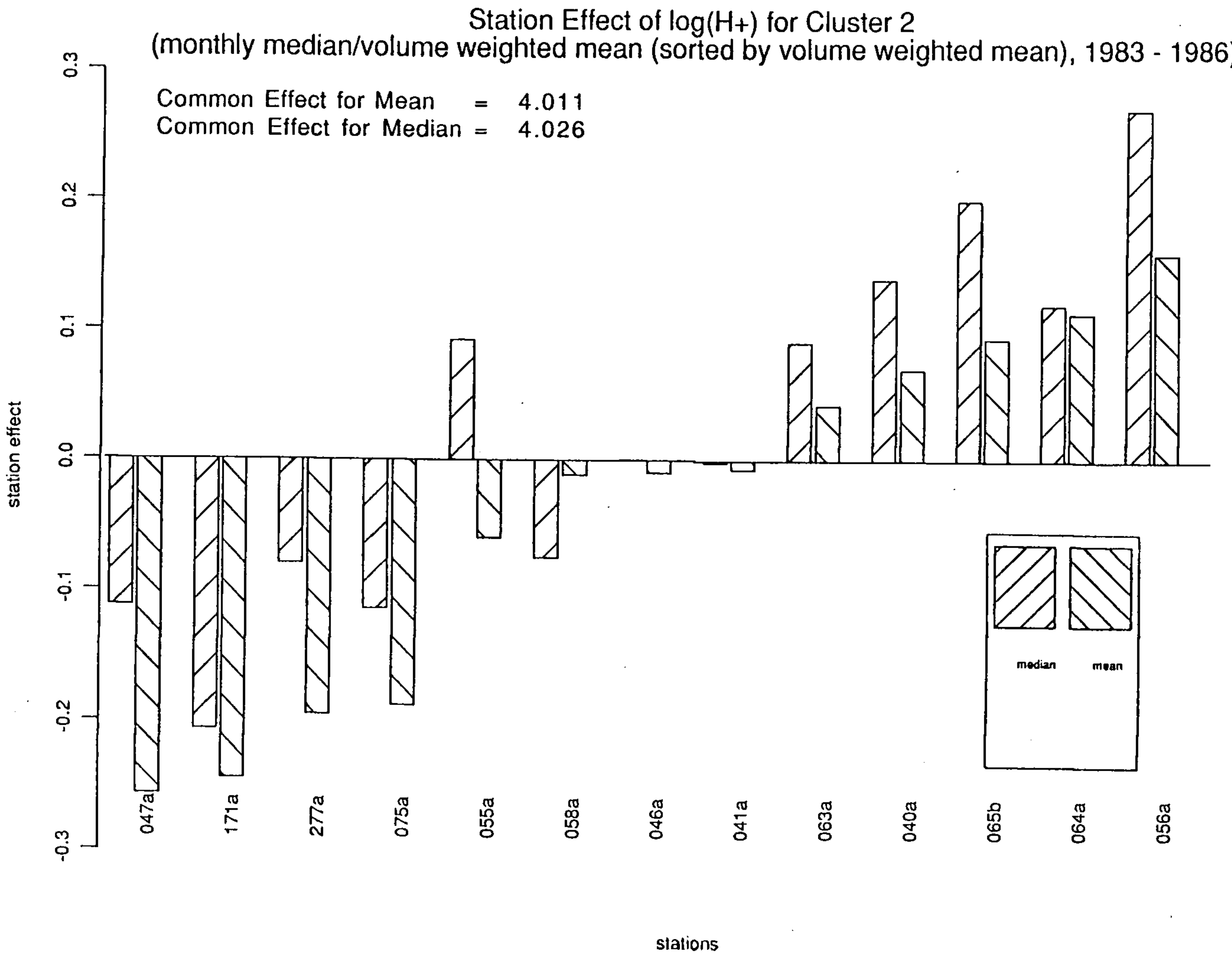

Station Effect of log(H+) for Cluster 2
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 4.011
Common Effect for Median = 4.026
station effect
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
047a
171a
277a
075a
055a
058a
046a
041a
063a
040a
065b
064a
056a
median
mean
stations

Figure 6.2.11

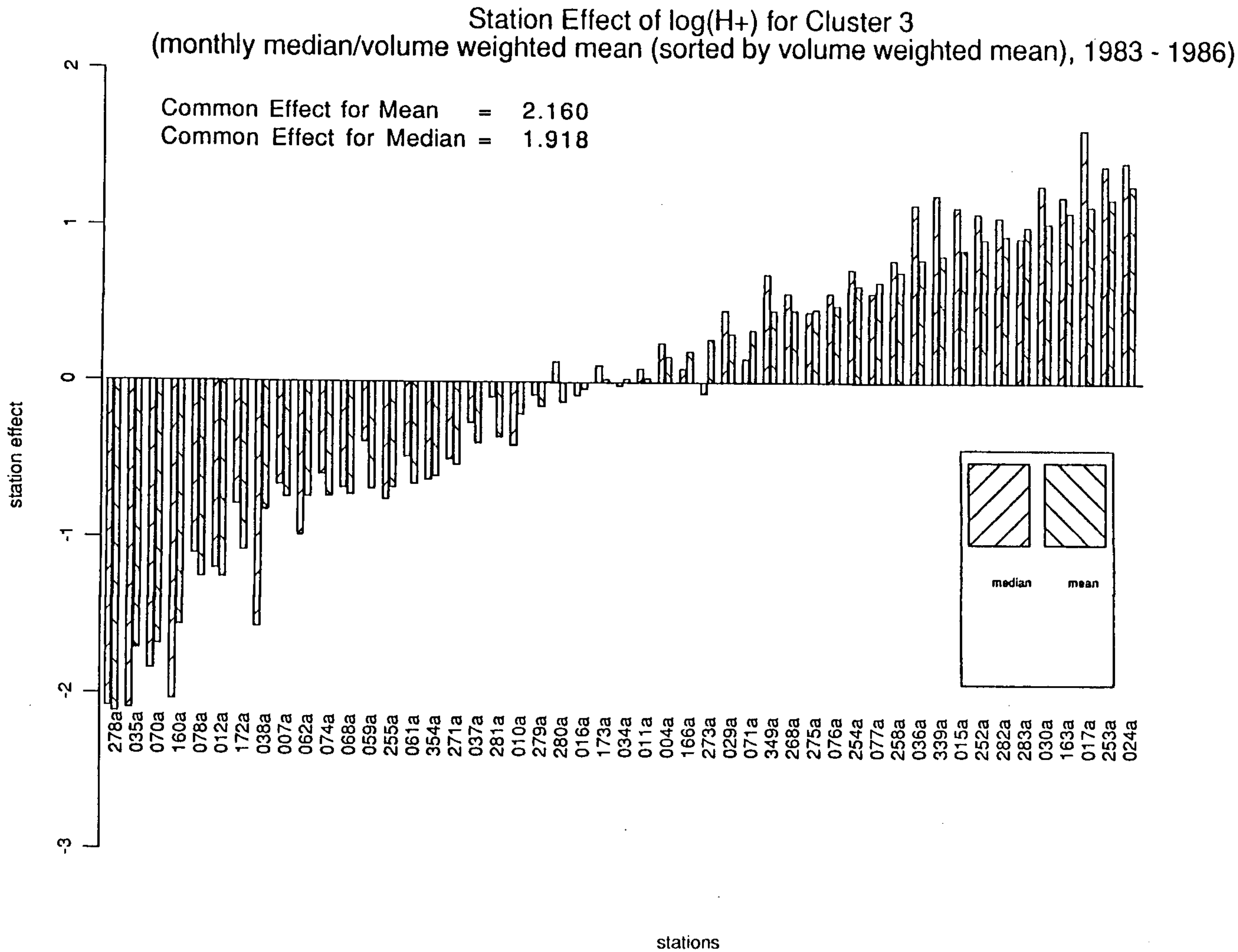

Station Effect of log(H+) for Cluster 3
(monthly median/volume weighted mean (sorted by volume weighted mean), 1983 - 1986)
Common Effect for Mean = 2.160
Common Effect for Median = 1.918
station effect
2
1
0
-1
-2
-3
median
mean
278a 035a 070a 160a 078a 012a 172a 038a 007a 062a 074a 068a 059a 255a 061a 354a 271a 037a 281a 010a 279a 280a 016a 173a 034a 011a 004a 166a 273a 029a 071a 349a 268a 275a 076a 254a 077a 258a 036a 339a 015a 252a 282a 283a 030a 163a 017a 253a 024a
stations

Figure 6.2.12

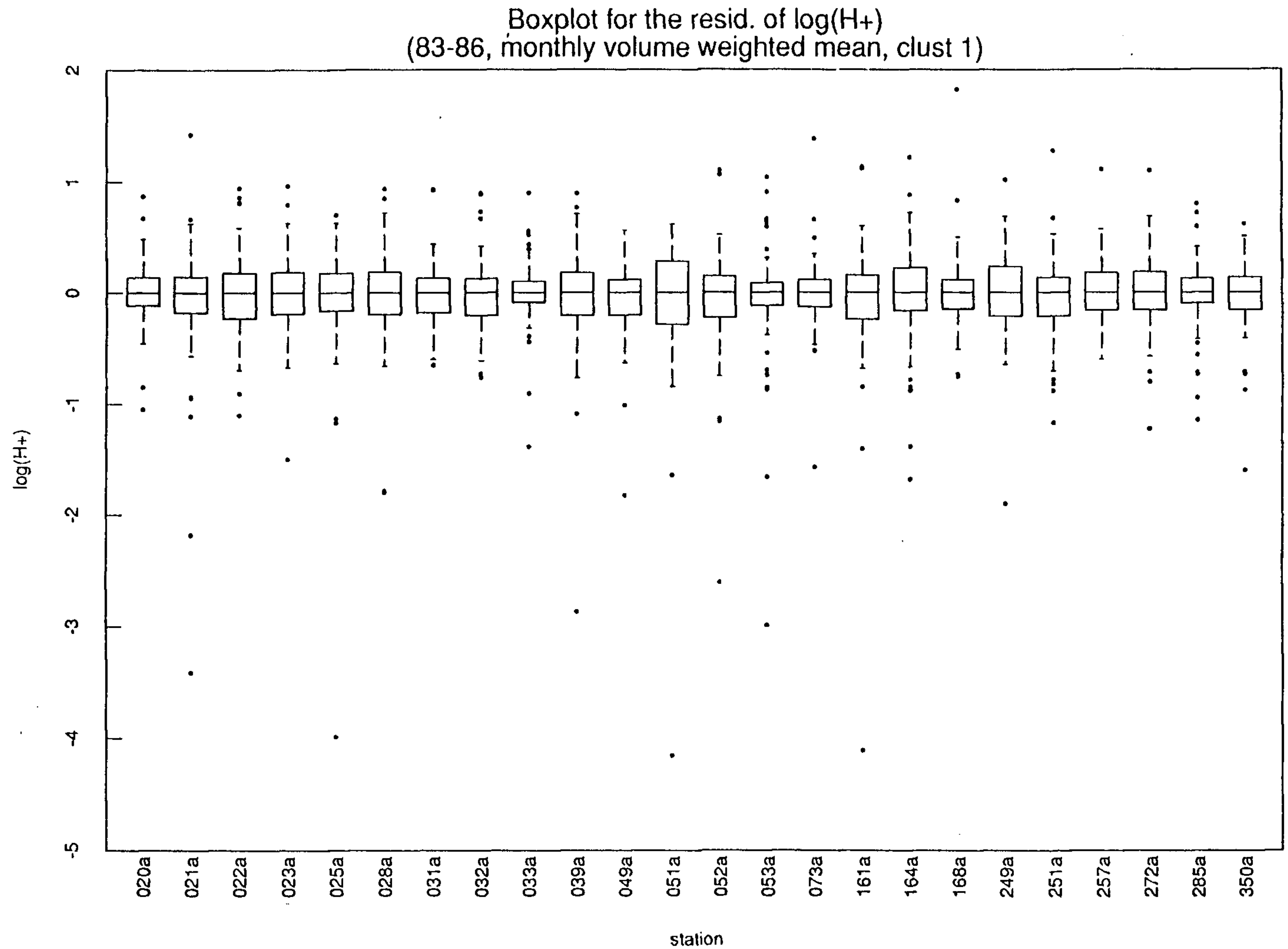


Figure 6.2.13(a)

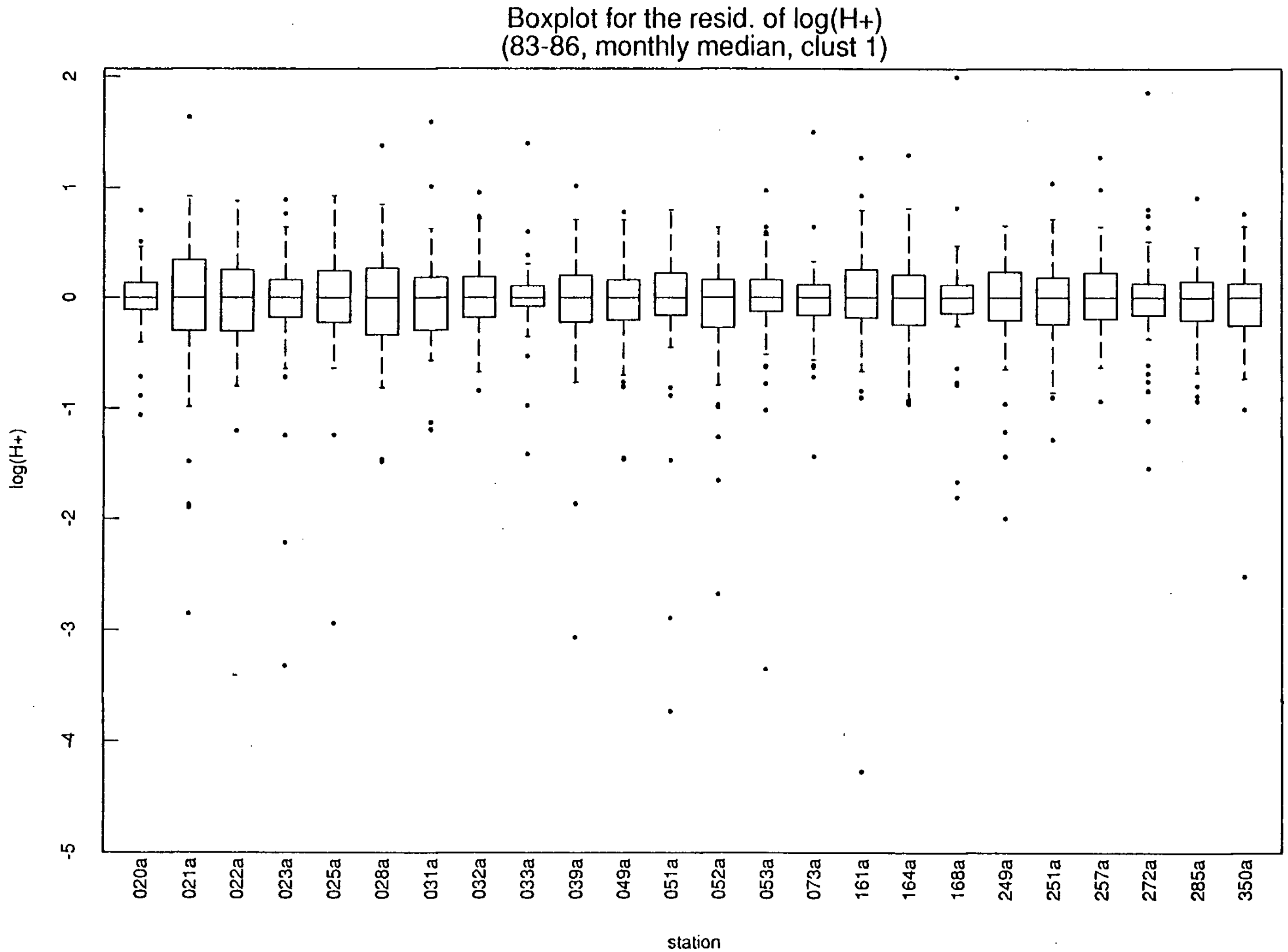


Figure 6.2.13(b)

Boxplot for the resid. of log(H+)
(83-86, monthly volume weighted mean, clust 2)

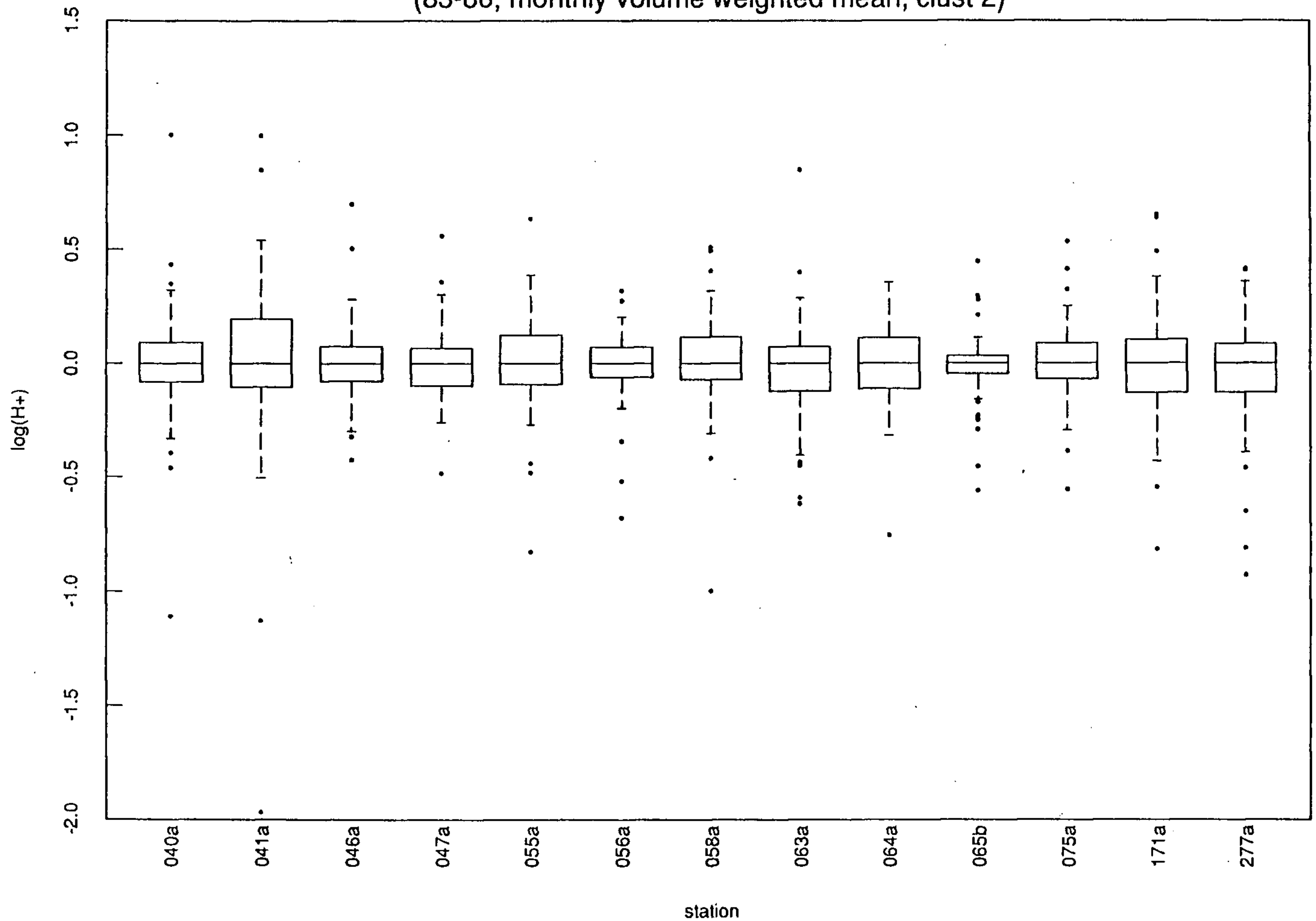


Figure 6.2.14(a)

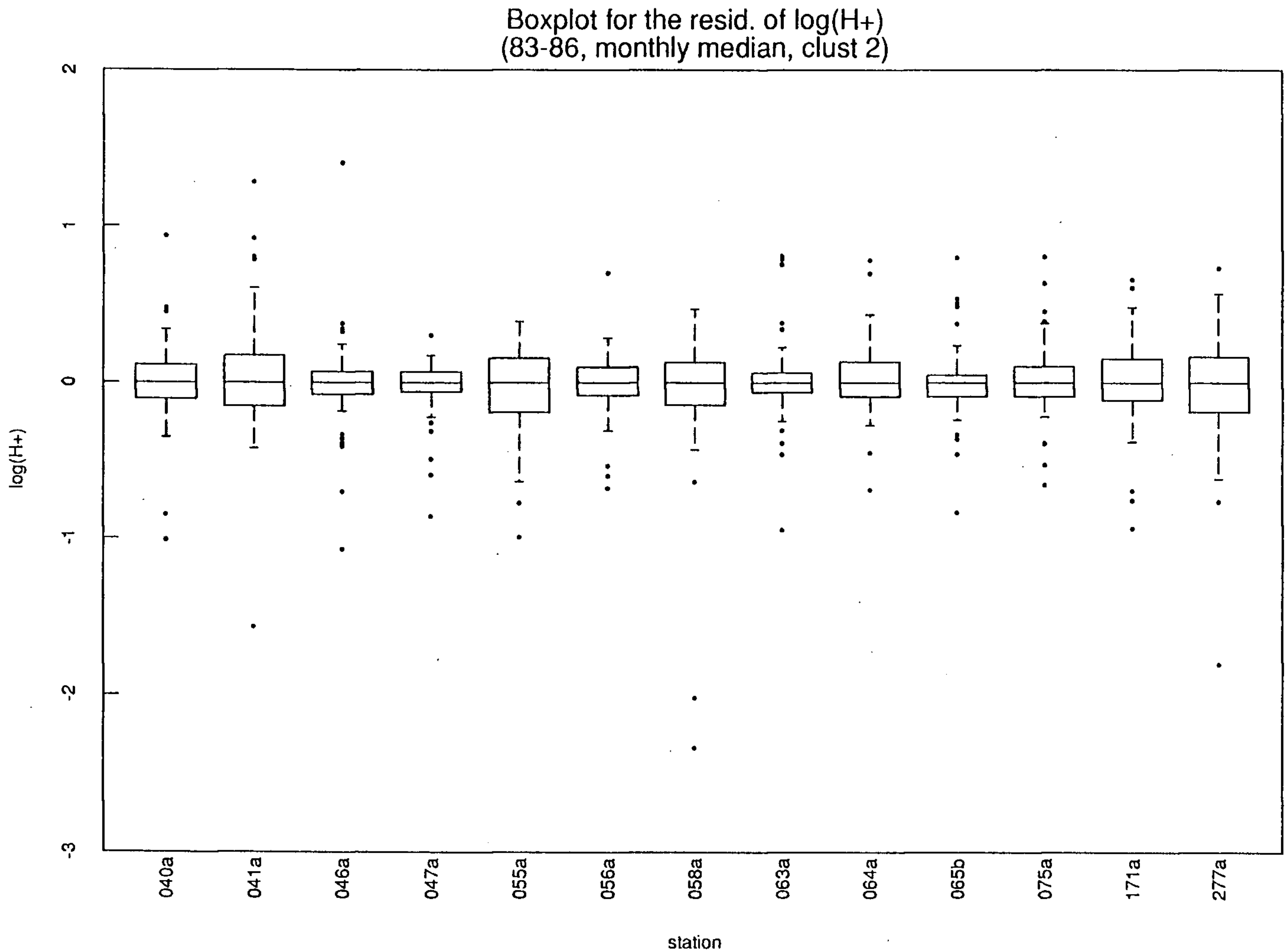


Figure 6.2.14(b)

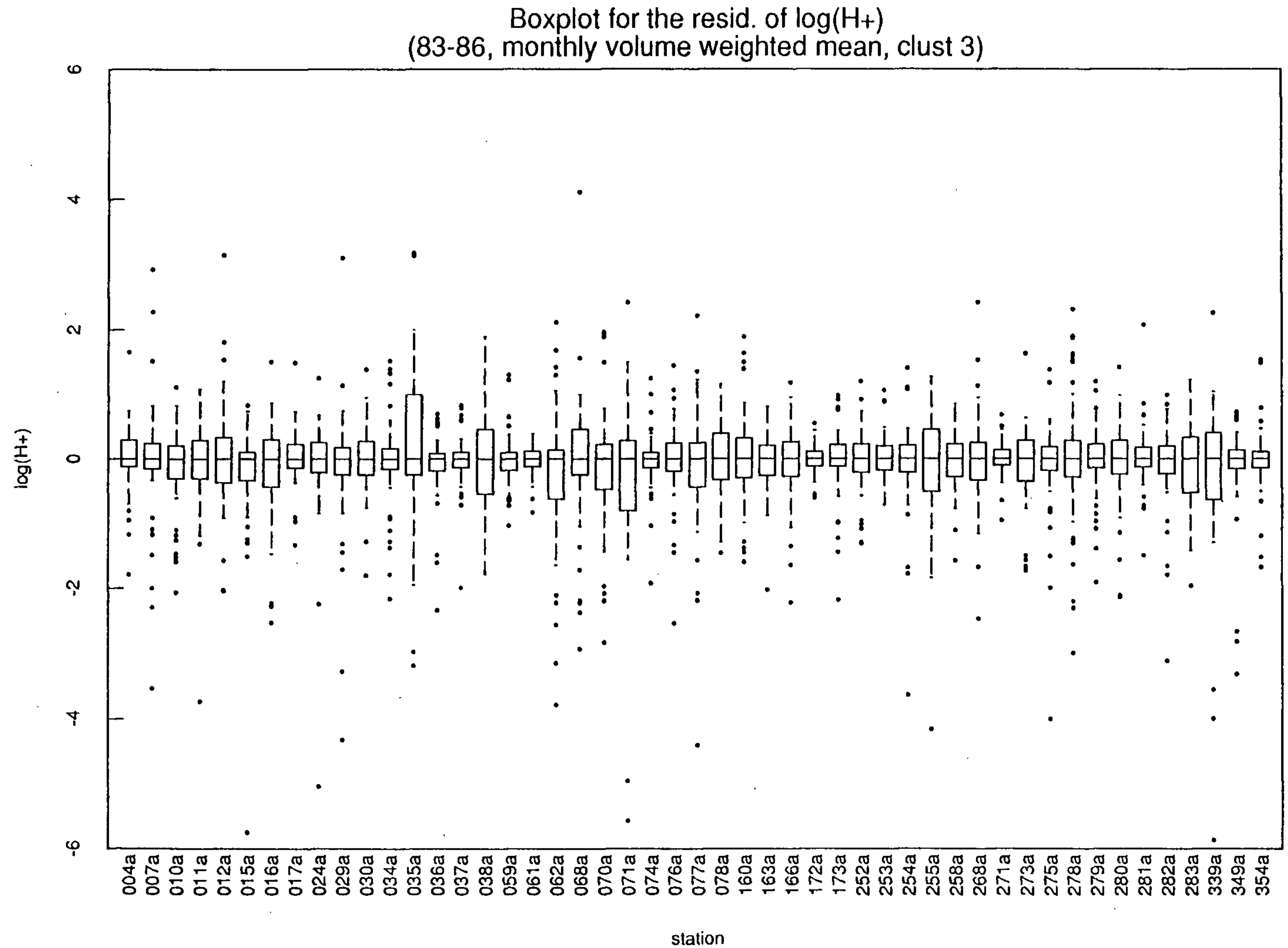


Figure 6.2.15(a)

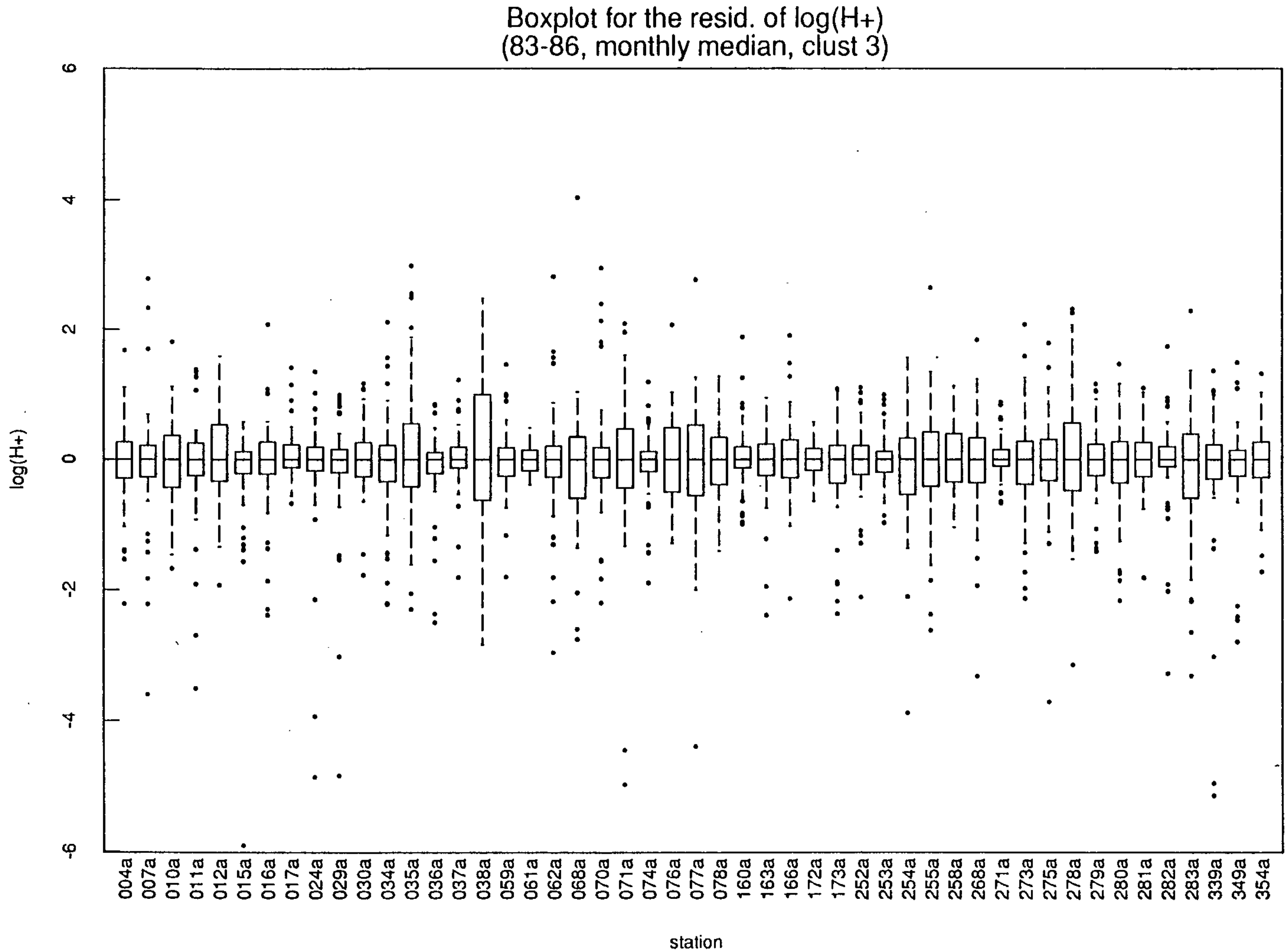


Figure 6.2.15(b)

Summary of the Effects and Residuals from Median Polish of log(H+)
(83-86,monthly volume weighted mean,clust 1)

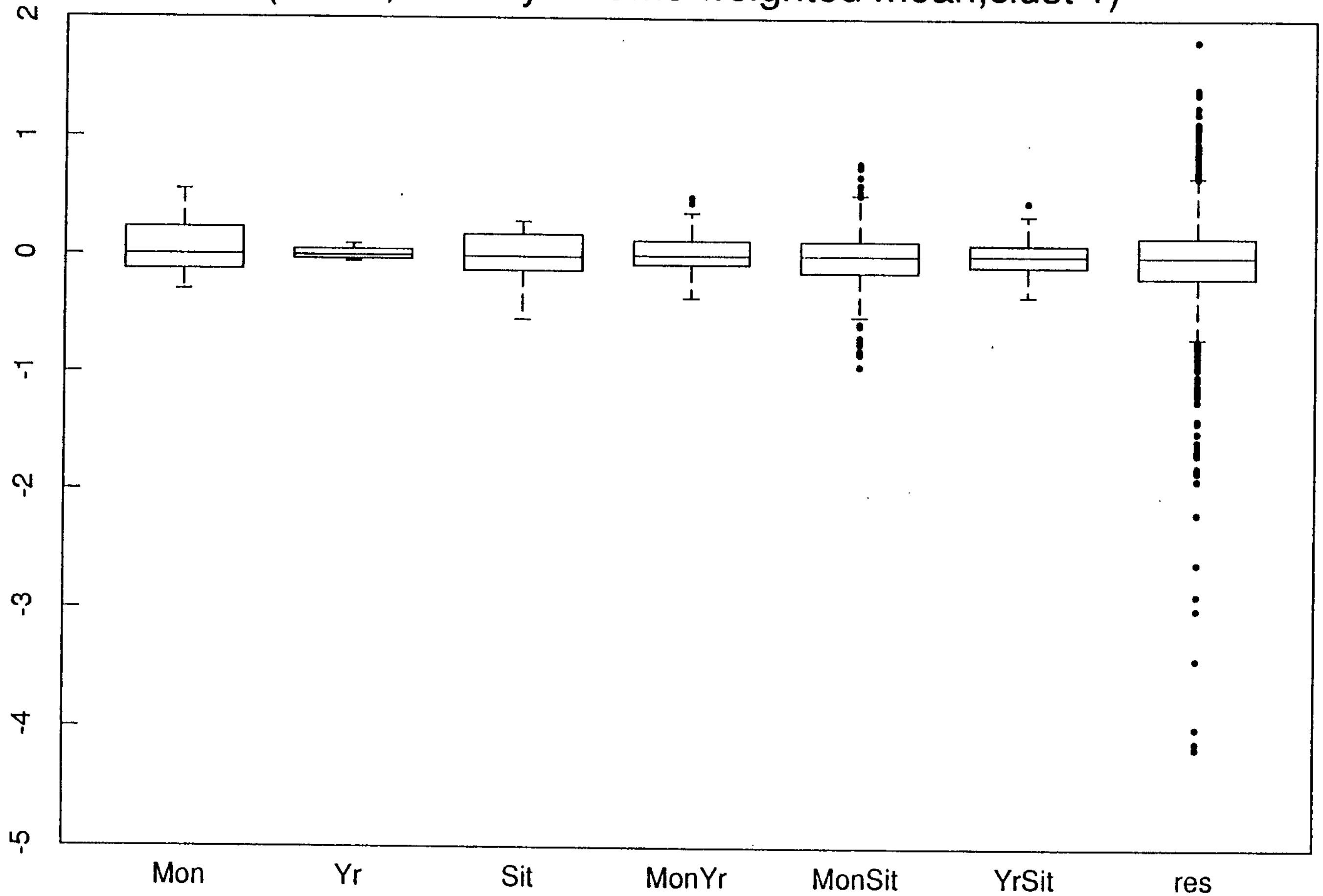


Figure 6.2.16(a)

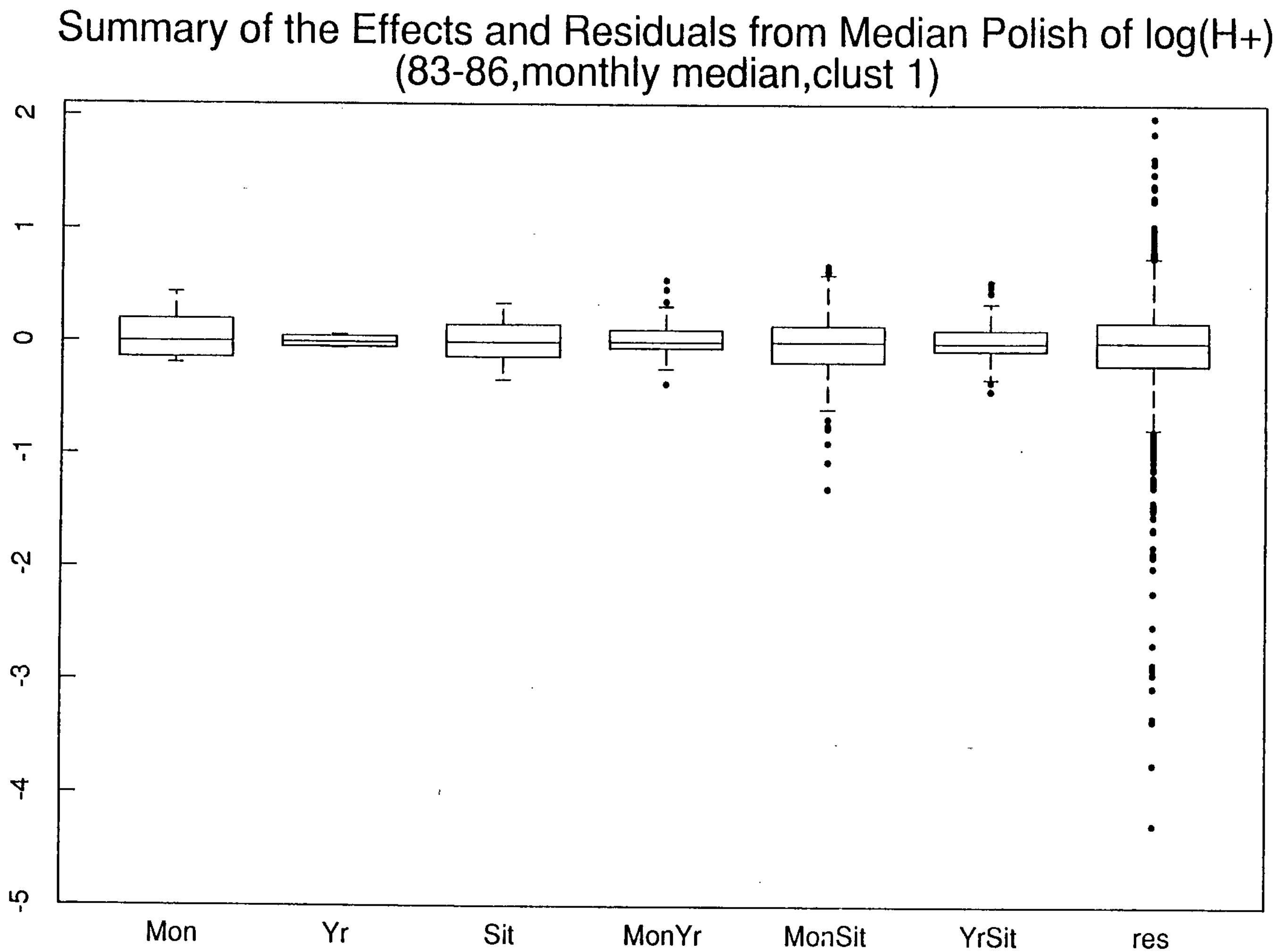


Figure 6.2.16(b)

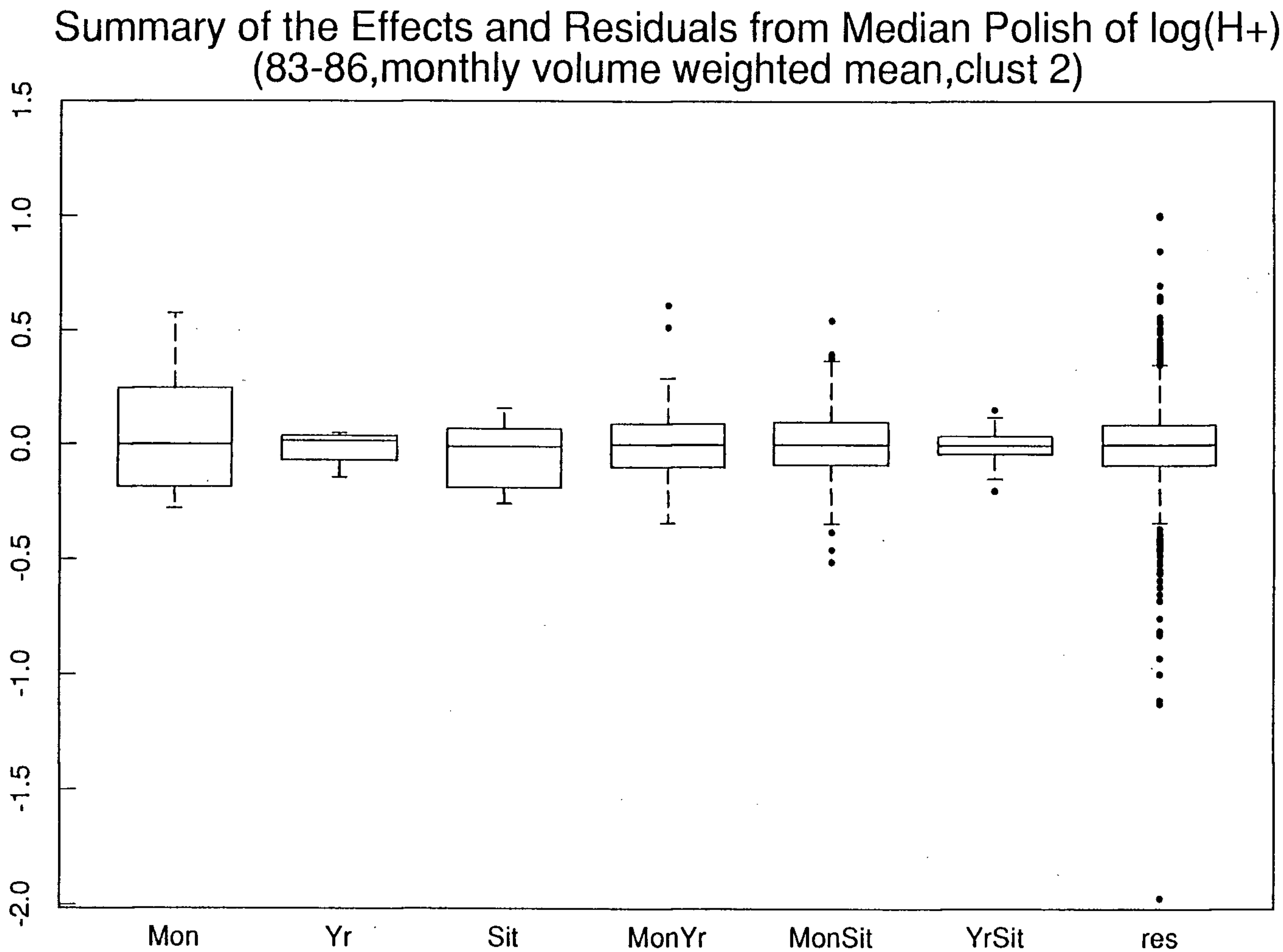


Figure 6.2.17(a)

Summary of the Effects and Residuals from Median Polish of log(H+)
(83-86,monthly median,clust 2)

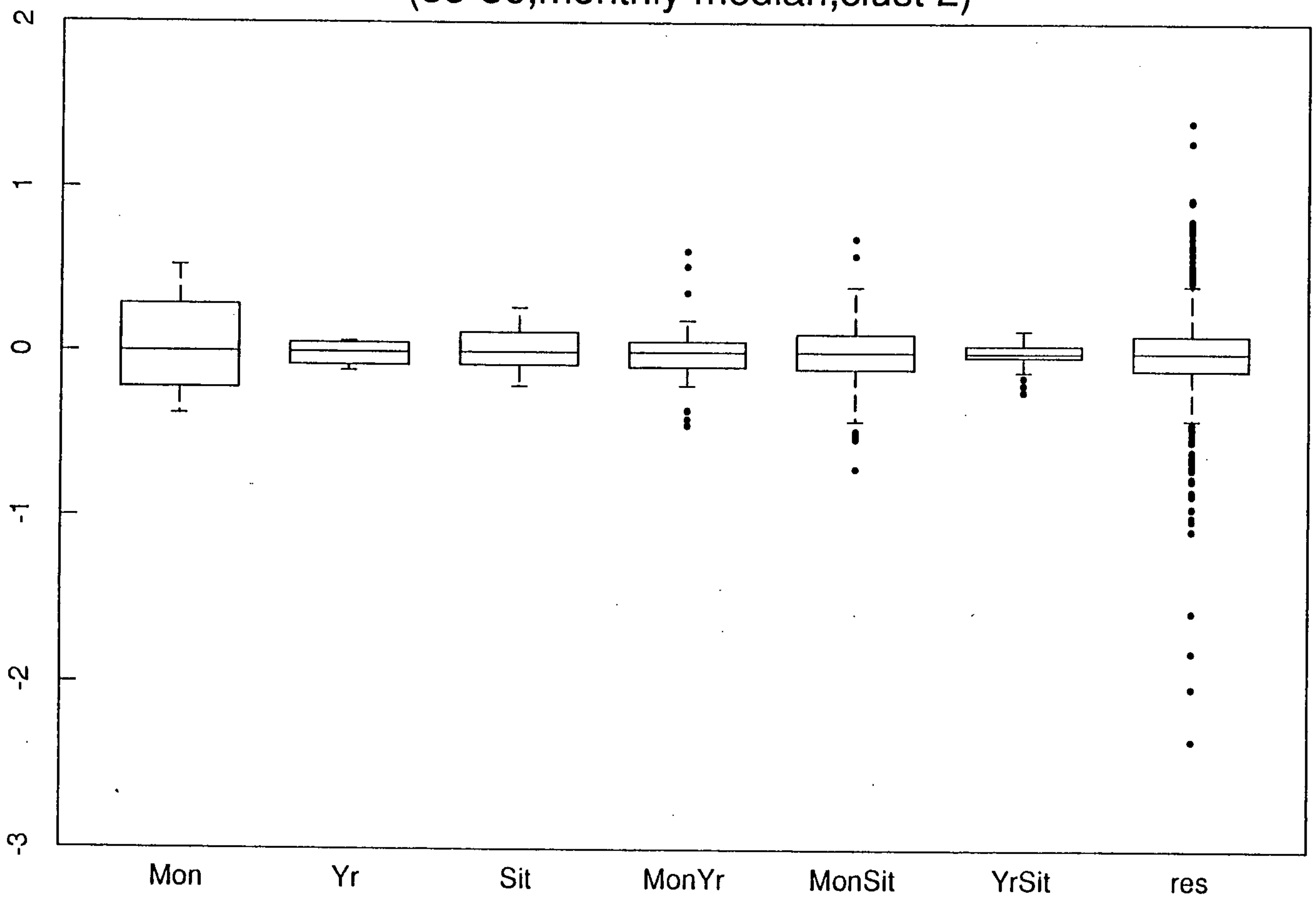


Figure 6.2.17(b)

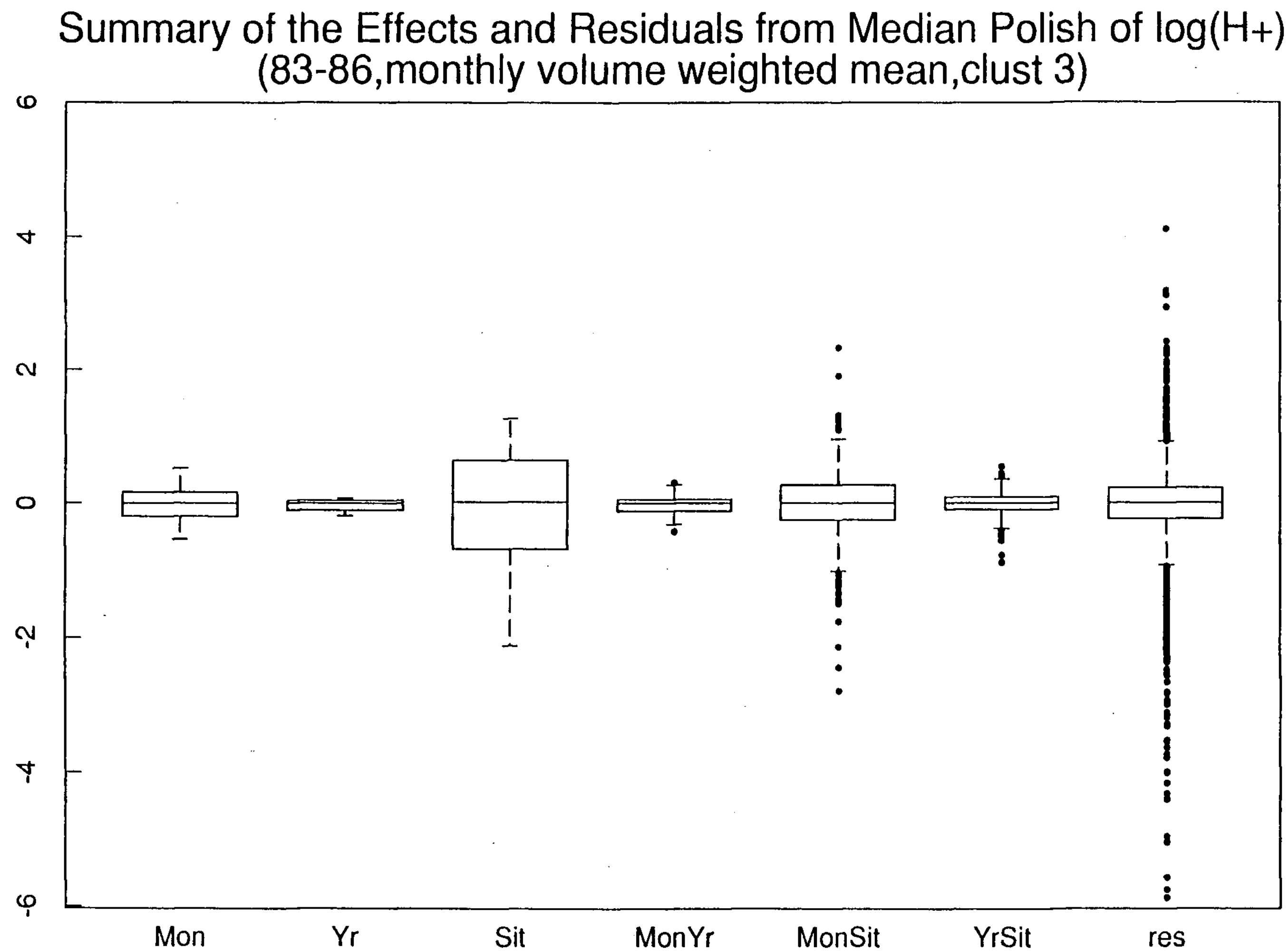


Figure 6.2.18(a)

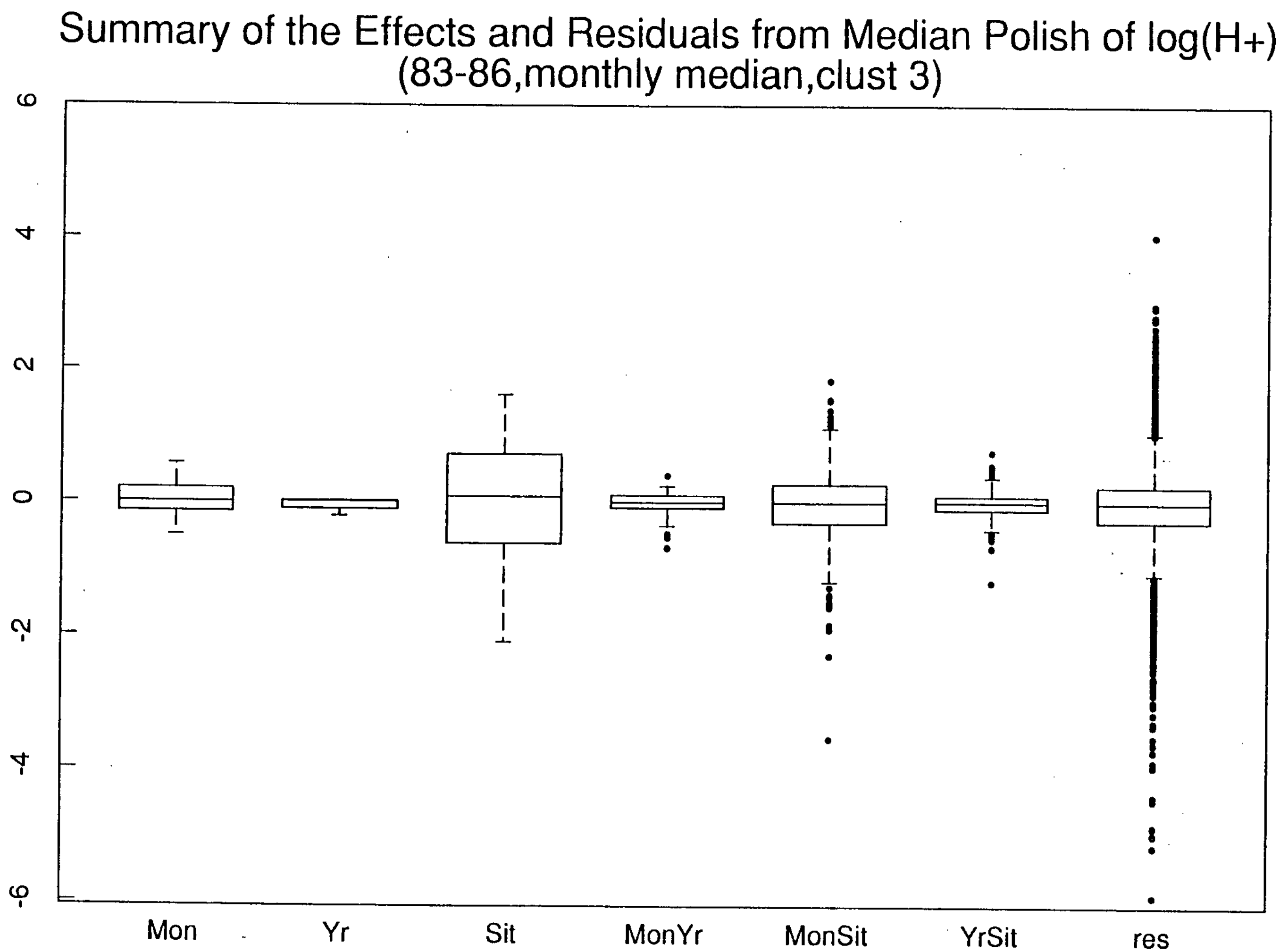


Figure 6.2.18(b)

Trend of log(H+) from 1983 to 1986 at the 86 Stations
(calculated by monthly volume weighted mean)

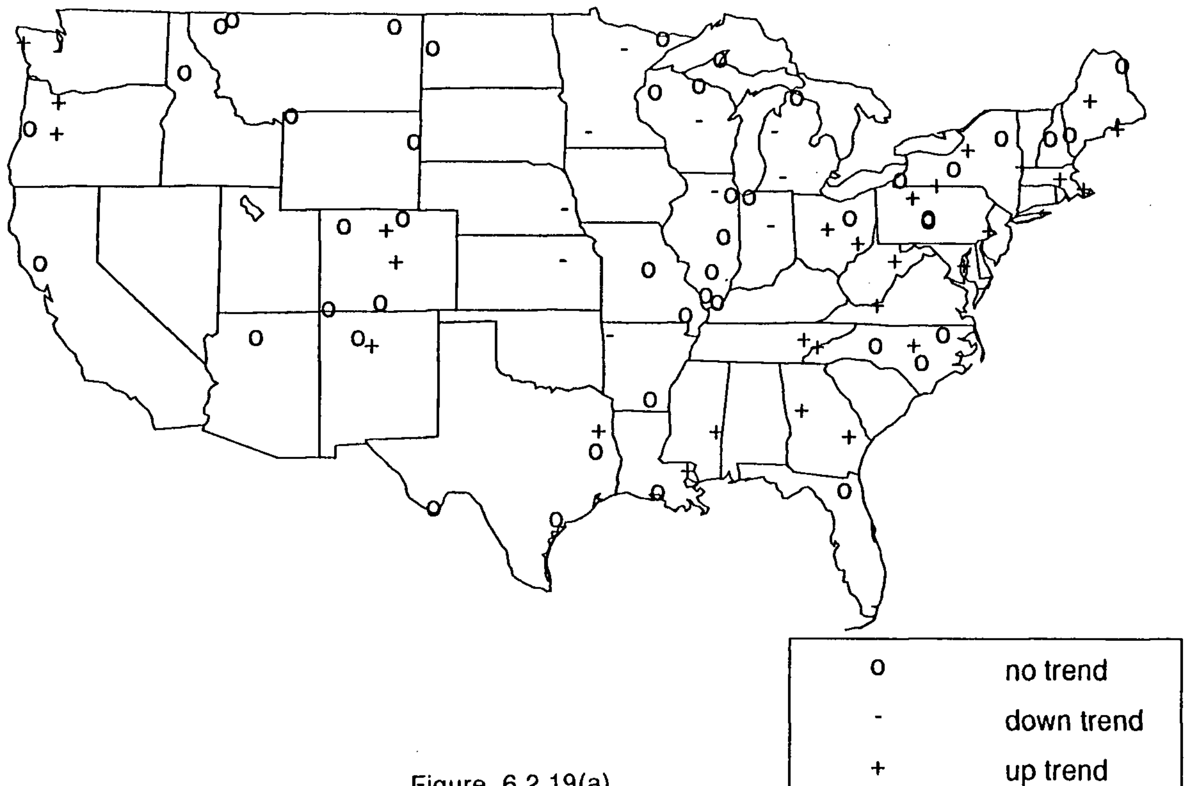


Figure 6.2.19(a)

Trend of log(H+) from 1983 to 1986 at the 86 Stations
(calculated by monthly median)

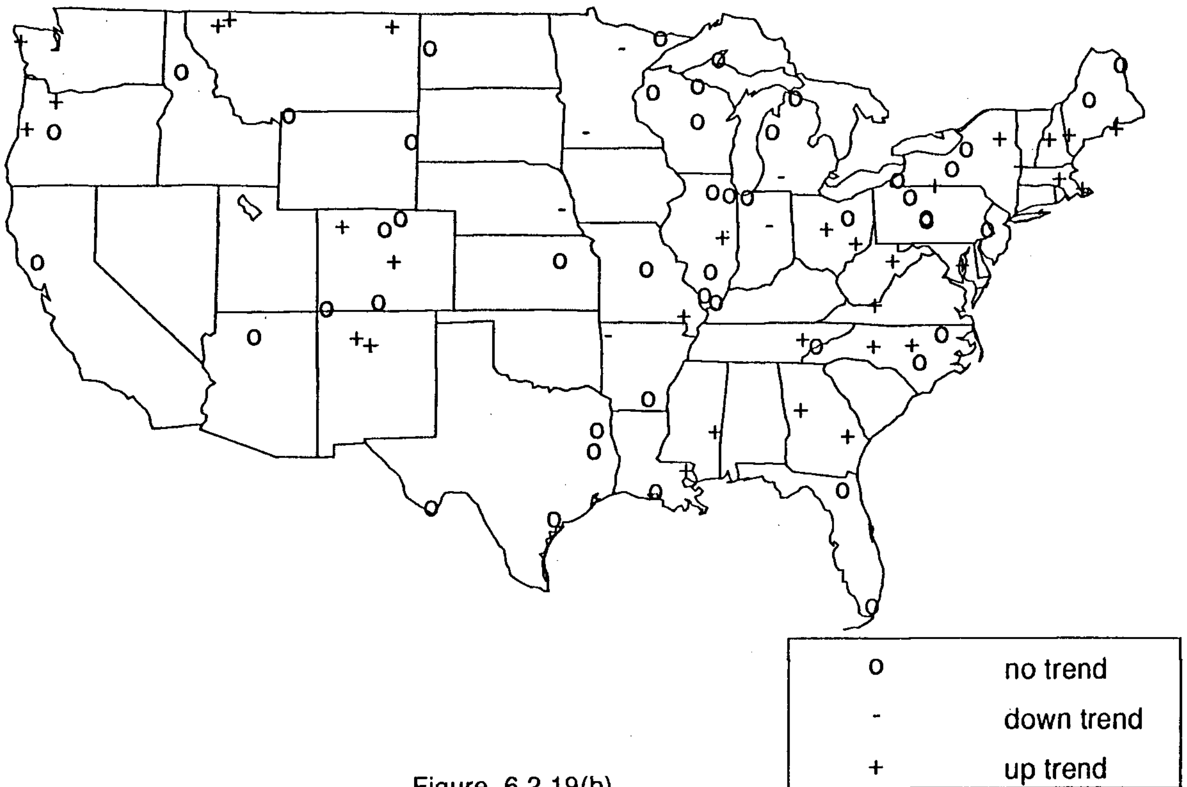


Figure 6.2.19(b)

Trend of log(H+) from 1983 to 1986 in the USA
(calculated by Kriging from monthly volume weighted mean)

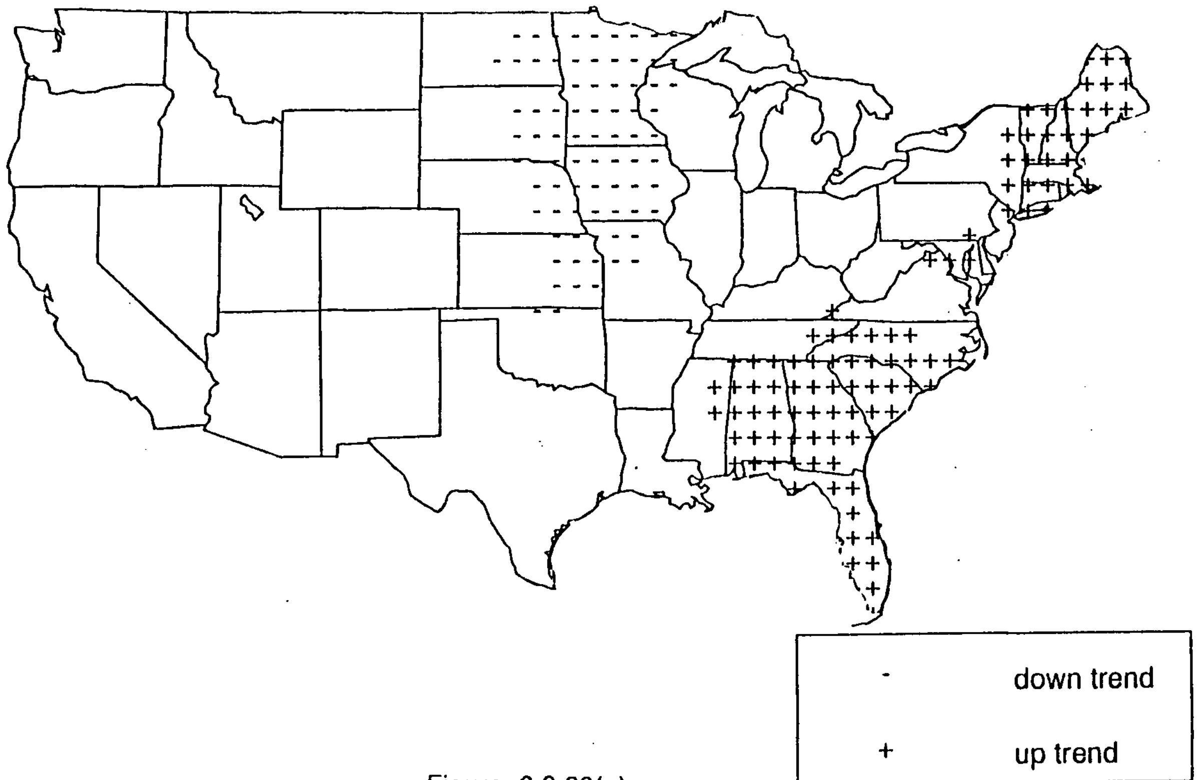


Figure 6.2.20(a)

Trend of log(H+) from 1983 to 1986 in the USA
(calculated by Kriging from monthly median)

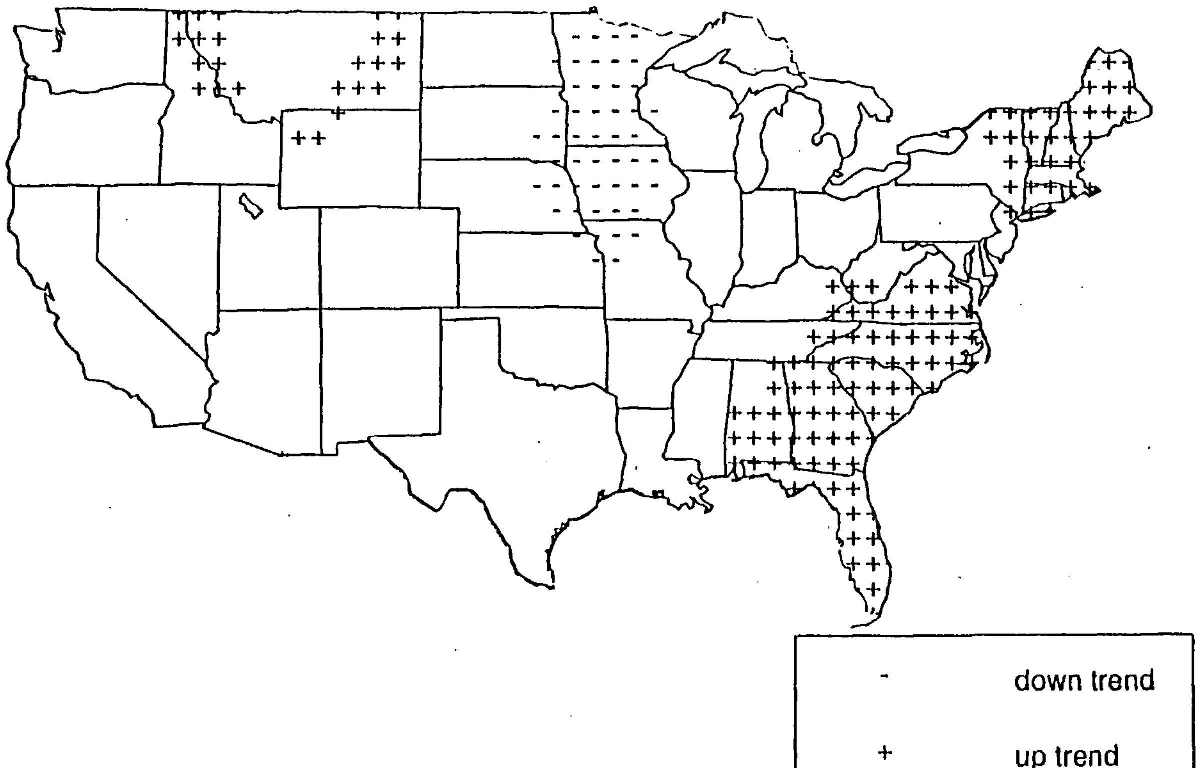


Figure 6.2.20(b)

TABLE 6.1.1

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF HYDROGEN ION
(Monthly Volume Weighted Mean, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.610	0.107	-0.004	-0.007	-0.001
010a	0.944	0.345	0.002	-0.001	0.007
011a	1.275	0.202	0.003	0.0	0.007
017a	0.961	0.336	0.002	0.0	0.005
020a	-1.768	0.077	-0.002	-0.004	-0.001
021a	-1.231	0.218	-0.002	-0.004	0.0
022a	-1.784	0.074	-0.002	-0.004	0.0
023a	-2.186	0.029	-0.003	-0.005	-0.001
030a	-0.877	0.381	-0.002	-0.005	0.001
031a	-1.115	0.265	-0.002	-0.004	0.0
032a	-0.169	0.866	0.0	-0.001	0.001
034a	-0.393	0.694	-0.001	-0.004	0.002
036a	0.769	0.442	0.001	-0.001	0.004
038a	-0.213	0.832	0.0	-0.007	0.005
039a	-1.432	0.152	-0.002	-0.005	0.0
040a	-0.654	0.513	-0.001	-0.003	0.001
041a	-0.724	0.469	-0.001	-0.003	0.001
049a	-0.731	0.465	-0.001	-0.004	0.001
051a	-2.477	0.013	-0.005	-0.009	-0.002
052a	-2.193	0.028	-0.004	-0.007	-0.001
053a	-1.106	0.269	-0.003	-0.006	0.0
055a	-1.753	0.080	-0.002	-0.003	0.0
056a	-0.387	0.699	0.0	-0.001	0.001
058a	-0.159	0.874	0.0	-0.001	0.001
059a	2.953	0.003	0.005	0.002	0.007
064a	0.066	0.948	0.0	-0.001	0.002
074a	1.385	0.166	0.002	0.0	0.004
075a	0.337	0.736	0.0	-0.001	0.002
076a	-1.288	0.198	-0.002	-0.006	0.0
168a	-1.042	0.297	-0.002	-0.004	0.0
171a	-0.457	0.648	0.0	-0.003	0.001
173a	-0.469	0.639	-0.001	-0.004	0.002

TABLE 6.1.2

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF HYDROGEN ION
(Monthly Median, '80 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.462	0.144	-0.004	-0.008	0.0
010a	0.191	0.849	0.0	-0.004	0.007
011a	1.656	0.098	0.007	0.001	0.012
017a	1.639	0.101	0.003	0.001	0.006
020a	-0.410	0.682	0.0	-0.003	0.001
021a	-1.419	0.156	-0.002	-0.005	0.0
022a	-1.923	0.054	-0.003	-0.006	-0.001
023a	-2.012	0.044	-0.003	-0.005	-0.001
030a	-0.466	0.641	-0.001	-0.005	0.001
031a	-1.127	0.260	-0.002	-0.004	0.0
032a	0.733	0.464	0.001	-0.001	0.003
034a	0.261	0.794	0.0	-0.003	0.004
036a	1.605	0.109	0.004	0.001	0.006
038a	-0.240	0.810	0.0	-0.008	0.005
039a	-0.259	0.795	0.0	-0.003	0.002
040a	-1.021	0.307	-0.001	-0.004	0.0
041a	-0.894	0.372	-0.001	-0.004	0.001
049a	-0.917	0.359	-0.001	-0.004	0.0
051a	-1.670	0.095	-0.004	-0.007	0.0
052a	-1.595	0.111	-0.003	-0.006	0.0
053a	-0.323	0.747	0.0	-0.004	0.003
055a	-0.205	0.837	0.0	-0.001	0.001
056a	0.282	0.778	0.0	-0.001	0.002
058a	-0.569	0.570	0.0	-0.002	0.001
059a	4.155	0.0	0.009	0.006	0.012
064a	-0.604	0.546	0.0	-0.002	0.001
074a	-0.008	0.993	0.0	-0.001	0.002
075a	-0.553	0.580	0.0	-0.002	0.001
076a	-0.323	0.747	0.0	-0.005	0.002
168a	-1.271	0.204	-0.003	-0.005	0.0
171a	-0.629	0.529	-0.001	-0.003	0.001
173a	0.373	0.709	0.001	-0.002	0.005

TABLE 6.2.1

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES FOR THE CONCENTRATIONS OF HYDROGEN ION (Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.926	0.054	-0.011	-0.019	-0.004
007a	-0.324	0.746	-0.002	-0.011	0.004
010a	1.769	0.077	0.007	0.002	0.014
011a	1.609	0.108	0.009	0.003	0.016
012a	0.947	0.344	0.008	-0.004	0.017
015a	0.670	0.503	0.001	-0.001	0.007
016a	-1.246	0.213	-0.009	-0.019	0.0
017a	3.918	0.0	0.016	0.009	0.021
020a	0.101	0.920	0.0	-0.003	0.003
021a	0.204	0.838	0.0	-0.004	0.003
022a	0.436	0.663	0.001	-0.003	0.007
023a	0.169	0.866	0.001	-0.003	0.005
024a	-1.502	0.133	-0.008	-0.014	-0.002
025a	-0.240	0.810	-0.001	-0.005	0.002
028a	1.526	0.127	0.006	0.001	0.012
029a	-0.542	0.588	-0.002	-0.008	0.004
030a	1.559	0.119	0.006	0.002	0.012
031a	-0.367	0.714	-0.002	-0.006	0.002
032a	-1.577	0.115	-0.004	-0.007	-0.001
033a	-1.822	0.068	-0.005	-0.009	-0.002
034a	-2.708	0.007	-0.016	-0.023	-0.009
035a	-1.507	0.132	-0.022	-0.036	-0.002
036a	2.729	0.006	0.012	0.006	0.019
037a	0.956	0.339	0.002	-0.001	0.006
038a	-2.568	0.010	-0.028	-0.042	-0.014
039a	0.898	0.369	0.005	-0.002	0.012
040a	1.271	0.204	0.005	0.0	0.009
041a	0.391	0.696	0.001	-0.003	0.005
046a	1.395	0.163	0.003	0.0	0.005
047a	1.841	0.066	0.006	0.002	0.009
049a	0.880	0.379	0.004	-0.002	0.011
051a	1.155	0.248	0.007	0.0	0.014
052a	-0.578	0.564	-0.003	-0.009	0.003
053a	2.446	0.014	0.009	0.003	0.016
055a	1.307	0.191	0.004	0.0	0.007
056a	2.195	0.028	0.006	0.003	0.010
058a	0.613	0.540	0.002	-0.002	0.005
059a	1.155	0.248	0.003	0.0	0.007
061a	2.301	0.021	0.005	0.002	0.008
062a	-0.335	0.737	-0.003	-0.016	0.008
063a	2.146	0.032	0.005	0.002	0.008
064a	1.253	0.210	0.003	0.0	0.006
065b	-0.147	0.883	0.0	-0.003	0.003
068a	-0.294	0.769	-0.002	-0.014	0.008
070a	-0.142	0.887	-0.001	-0.008	0.008
071a	-0.773	0.439	-0.007	-0.020	0.005
073a	1.683	0.092	0.007	0.002	0.013

TABLE 6.2.1 (continued)

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES FOR THE CONCENTRATIONS OF HYDROGEN ION (Monthly Volume Weighted Mean, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	3.136	0.002	0.008	0.005	0.012
075a	3.920	0.0	0.009	0.006	0.012
076a	1.027	0.304	0.005	-0.002	0.011
077a	-1.155	0.248	-0.010	-0.023	0.001
078a	-0.685	0.493	-0.004	-0.014	0.005
160a	1.096	0.273	0.007	-0.001	0.013
161a	-0.382	0.702	-0.001	-0.007	0.003
163a	0.871	0.384	0.003	-0.002	0.008
164a	0.161	0.872	0.001	-0.007	0.008
166a	0.624	0.533	0.004	-0.003	0.010
168a	0.994	0.320	0.005	-0.001	0.010
171a	2.755	0.006	0.009	0.005	0.014
172a	1.311	0.190	0.003	0.0	0.008
173a	0.578	0.563	0.003	-0.004	0.008
249a	1.448	0.148	0.008	0.0	0.015
251a	1.976	0.048	0.013	0.005	0.023
252a	-1.075	0.282	-0.007	-0.013	0.002
253a	0.862	0.389	0.003	-0.002	0.009
254a	-1.240	0.215	-0.005	-0.012	0.0
255a	0.453	0.650	0.004	-0.010	0.014
257a	1.922	0.055	0.009	0.003	0.016
258a	0.265	0.791	0.001	-0.004	0.007
268a	0.791	0.429	0.005	-0.003	0.012
271a	-0.792	0.428	-0.003	-0.008	0.002
272a	-1.733	0.083	-0.004	-0.008	-0.001
273a	-2.106	0.035	-0.011	-0.017	-0.005
275a	0.933	0.351	0.004	-0.002	0.010
277a	1.687	0.092	0.006	0.001	0.010
278a	0.744	0.457	0.009	-0.003	0.029
279a	2.160	0.031	0.011	0.005	0.017
280a	1.129	0.259	0.006	-0.001	0.013
281a	2.714	0.007	0.010	0.005	0.015
282a	1.698	0.090	0.007	0.002	0.014
283a	-1.467	0.143	-0.010	-0.021	-0.001
285a	1.316	0.188	0.007	0.0	0.015
339a	1.350	0.177	0.014	0.0	0.027
349a	2.036	0.042	0.011	0.005	0.016
350a	2.759	0.006	0.012	0.006	0.017
354a	0.450	0.653	0.002	-0.004	0.008

TABLE 6.2.2

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF HYDROGEN ION
(Monthly Median, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
004a	-1.825	0.068	-0.014	-0.022	-0.004
007a	0.0	1.000	0.0	-0.011	0.008
010a	0.382	0.702	0.003	-0.008	0.013
011a	1.324	0.185	0.010	0.0	0.019
012a	0.379	0.705	0.005	-0.010	0.018
015a	0.859	0.390	0.005	-0.002	0.011
016a	-0.743	0.457	-0.004	-0.012	0.003
017a	3.163	0.002	0.012	0.007	0.018
020a	1.770	0.077	0.004	0.001	0.007
021a	-0.356	0.722	-0.002	-0.011	0.005
022a	-0.791	0.429	-0.003	-0.008	0.002
023a	0.009	0.993	0.0	-0.005	0.004
024a	-0.178	0.859	-0.001	-0.007	0.005
025a	-1.093	0.274	-0.005	-0.009	0.0
028a	1.174	0.240	0.008	0.0	0.017
029a	-0.595	0.552	-0.003	-0.010	0.003
030a	1.247	0.212	0.007	-0.001	0.012
031a	-0.908	0.364	-0.004	-0.009	0.001
032a	-1.706	0.088	-0.006	-0.012	-0.002
033a	-0.613	0.540	-0.001	-0.005	0.002
034a	-1.875	0.061	-0.016	-0.026	-0.005
035a	-1.790	0.074	-0.022	-0.035	-0.009
036a	1.946	0.052	0.009	0.003	0.015
037a	1.657	0.097	0.006	0.001	0.011
038a	-2.265	0.023	-0.029	-0.042	-0.013
039a	2.373	0.018	0.013	0.007	0.019
040a	1.040	0.298	0.004	-0.001	0.008
041a	1.067	0.286	0.003	-0.001	0.008
046a	1.102	0.270	0.002	-0.001	0.005
047a	1.740	0.082	0.004	0.001	0.008
049a	1.004	0.315	0.005	-0.001	0.011
051a	1.541	0.123	0.008	0.001	0.014
052a	1.402	0.161	0.007	0.0	0.012
053a	2.563	0.010	0.015	0.007	0.022
055a	1.698	0.090	0.005	0.001	0.008
056a	1.929	0.054	0.006	0.001	0.010
058a	0.631	0.528	0.002	-0.002	0.005
059a	1.954	0.051	0.008	0.003	0.013
061a	0.985	0.325	0.002	-0.001	0.006
062a	0.084	0.933	0.0	-0.013	0.014
063a	0.688	0.492	0.002	-0.001	0.005
064a	0.755	0.450	0.002	-0.001	0.005
065b	0.084	0.933	0.0	-0.002	0.002
068a	0.142	0.887	0.002	-0.011	0.012
070a	0.718	0.472	0.007	-0.006	0.020
071a	-0.418	0.676	-0.007	-0.019	0.006
073a	2.622	0.009	0.010	0.004	0.016

TABLE 6.2.2 (continued)

RESULTS OF MANN-KENDALL TESTS AND SLOPE ESTIMATES
FOR THE CONCENTRATIONS OF HYDROGEN ION
(Monthly Median, '83 - '86)

site ID	Z	p-value	est'd slope	80% L-bd	80% U-bd
074a	2.210	0.027	0.008	0.003	0.013
075a	3.013	0.003	0.008	0.005	0.012
076a	0.0	1.000	0.0	-0.011	0.012
077a	-0.807	0.420	-0.010	-0.021	0.005
078a	-0.558	0.577	-0.004	-0.016	0.003
160a	-0.724	0.469	-0.004	-0.013	0.004
161a	-0.613	0.540	-0.003	-0.009	0.003
163a	0.843	0.399	0.004	-0.003	0.011
164a	1.790	0.074	0.011	0.003	0.020
166a	0.990	0.322	0.005	-0.002	0.015
168a	1.846	0.065	0.008	0.003	0.015
171a	2.480	0.013	0.010	0.005	0.014
172a	1.820	0.069	0.004	0.001	0.007
173a	1.395	0.163	0.009	0.001	0.016
249a	2.486	0.013	0.011	0.004	0.021
251a	2.955	0.003	0.023	0.014	0.030
252a	-0.800	0.424	-0.003	-0.011	0.002
253a	1.752	0.080	0.007	0.002	0.013
254a	-0.123	0.902	-0.001	-0.010	0.011
255a	1.182	0.237	0.009	-0.001	0.019
257a	1.477	0.140	0.008	0.001	0.014
258a	1.004	0.316	0.006	-0.001	0.014
268a	1.155	0.248	0.007	-0.001	0.015
271a	-0.039	0.969	0.0	-0.005	0.005
272a	-1.555	0.120	-0.006	-0.012	-0.001
273a	-0.729	0.466	-0.007	-0.015	0.003
275a	0.907	0.365	0.006	-0.003	0.014
277a	1.834	0.067	0.007	0.002	0.012
278a	1.772	0.076	0.015	0.004	0.032
279a	1.431	0.152	0.008	0.001	0.019
280a	1.715	0.086	0.013	0.002	0.024
281a	2.421	0.015	0.012	0.006	0.019
282a	1.058	0.290	0.003	0.0	0.008
283a	-0.622	0.534	-0.007	-0.022	0.008
285a	1.013	0.311	0.005	-0.001	0.012
339a	2.397	0.017	0.020	0.009	0.031
349a	1.790	0.074	0.010	0.002	0.016
350a	2.407	0.016	0.011	0.005	0.016
354a	1.780	0.075	0.014	0.001	0.026

Chapter 7 SUMMARY AND FURTHER STUDIES

The primary purpose of this study is to detect and estimate the possible temporal trends in different levels of chemical constituents of acid deposition at different locations. Spatial patterns and seasonal patterns of the levels of the chemical concentrations are also of interests. Three ions, sulphate, nitrate and hydrogen ion are analyzed in this study.

Some characteristics of the chemical precipitation data, for example, nonnormality, existence of missing data and the limited number of observations available, create some difficulties in using the traditional parametric statistical methods. Some nonparametric statistical techniques are thus suggested and used here, namely, hierarchical clustering, median polishing, the Mann-Kendall Test for monotone trend, Sen's slope estimate and Kriging.

The data are clustered into two to three clusters according their temporal patterns, then transformed by the log transformation so that the transformed data are approximately symmetric. The analyses are then based on the transformed data. The result of median polish suggests that there is a V-shaped trend for all the three chemicals in the subregion of highest industrial concentration. That is, the concentrations of the chemicals decreased first and then increased (or remained constant in a few cases) during the period, 1980 - 1986, with the change point around 1983. Strong seasonality and spatial patterns are discovered as well where the seasonal patterns of sulphate and hydrogen ion are similar but different from that of nitrate.

The Mann-Kendall Test for monotone trend and Sen's slope estimation procedure are applied to the deseasonalized data for each site. The main result is that there is overall an up-trend in the East for all the three ions during the period of 1983 - 1986. This result is consistent with the results of median polish. Sen's nonparametric slope estimate is obtained for each site. Based on these estimates, the slope estimate is obtained by Kriging interpolation for each integer degree grid point of longitude and latitude across the 48 conterminous states in the United States.

For further study, the outliers found in the data need to be examined; the serial independence of the data may be examined by testing independence to assure the validity of the trend test; and a rank test for umbrella alternatives can be used to test the V-shaped trend for which the monotone trend test is not valid. In addition, that the spatial patterns of the estimated trends for sulphate and nitrate are different from that for hydrogen ion suggests the possibility of multivariate analysis to explore the relationship of the concentration levels of different chemicals.

BIBLIOGRAPHY

Altwicker, E. R. and Johannes, A. H. (1987). Spatial and Historical Trends in acidic deposition: A Graphical Intersite Comparison, *Atmospheric Environment,Vol. 21 No. 1, pp.129-137.*

Bilonick, R. A. (1985). The Space-Time Distribution of Sulphate Deposition in the Northeastern United States, *Atmospheric Environment,Vol. 19 No. 11, pp.1829-1845.*

Bilonick, R. A. (1987). Monthly Hydrogen Ion Deposition Maps for the Northeastern U.S. from July, 1982 to September, 1984, Consolidation Coal Company, Coal Plaza, Pittsburgh, PA 15241.

Bradley, J. V. (1968). *Distribution-Free Statistical Tests*, Prentice-Hall, Englewood Cliffs, N.J.

Dana, M. T. and Easter, R.C. (1987). Statistical Summary and Analyses of Event Precipitation Chemistry from the MAP3S Network, 1976-1983, *Atmospheric Environment,Vol. 21 No. 1, pp.113-128.*

Delfiner, P. (1965). Linear Estimation of Non Stationary Spatial Phenomena, *Advanced Geostatitics in the Mining Industry*, Reidel, Boston.

Devary, J. L. and Doctor, P.D. (1981). Geostatistical Modeling of Pore Velocity, Battelle-Pacific Northwest Laboratory, Report #PNL-3789.

Devary, J. L. and Rice W. A. (1982). Geostatistics Software User's Manual for the Geosciences Research and Engineering 11/70 Computer System, Battelle-Pacific Northwest Laboratory Report.

Dietz, E. J. and Killeen, T. J. (1981). A Nonparametric Multivariate Test for Monotone Trend With Pharmaceutical Applications, *Journal of the American Statistical Association, Vol. 76, No. 373, pp169-174*

Egbert, G. D. and Lettenmaier, D. P. (1986). Stochastic Modeling of the Space-Time Structure of Atmospheric Chemical Deposition, *Water Resources Research, Vol. 22, No. 2, pp. 165-179*

Eynon, B. P. and Switzer , P. (1983). The variability of rainfall acidity, *The Canada Journal of Statistics, Vol. 11, No. 1, pp. 25-49.*

Farrell, R. (1980). Methods for Classifying Changes in Environmental Conditions, Tech. report VRF-EPA7. 4-FR80-1. Vector Research Inc., Ann Arbor, Mich.

Finkelstein, P. L. (1984). The spatial analysis of acid precipitation data, *Journal of Climate and Applied Meteorology, Vol. 23, pp. 52-62.*

Gilbert, R. O. (1987). Statistical Methods for Environmental Pollution Monitoring, New York: Van Nostrand Reinhold.

Hirsch R. M. and Slack, J. R. (1984). A Nonparamatric Trend Test for Seasonal Data With Serial Dependence, *Water Resources Research, Vol. 20, No. 6, pp727-732.*

Hirsch R. M., Slack, J. R., and Smith R. A. (1982). Techniques of Trend Analysis for Monthly Water Quality Data, *Water Resources Research, Vol. 18, No. 1, pp. 107-121.*

Interagancy Task Force on Acid Precipitation. (1982). National Acid Precipitation Assessment Plan.

Jernigan, R. W. (1986). A Primer on Kriging, report for U. S. Environmental Protection Agency, Washington, D.C.

Johnson, R. A., Wichern, D.W. (1982). *Applied Multivariate Statistical Analysis.* Prentice-Hall, Englewood Cliffs, N.J.

Jonckheere, A. R. (1954). A distribution-free k sample test against ordered alternatives, *Biometrika, Vol. 41, pp. 135-145.*

Kendall, M. G. (1970). *Rank Correlation Methods*, London: Charles Griffin.

Kendall, M. G. (1975). *Rank Correlation Methods*, 4th ed. London: Charles Griffin.

Le, D. N. and Petkau, A. J. (1988), The variability of rainfall acidity revisited, *The Canadian Journal of Statistics, Vol. 16, No. 1, pp.15.*

Mann, H. B. (1945) Non-parametric tests against trend, *Econometrica, Vol. 13,pp. 245-259.*

Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Matheron, G. (1973). The Intrinsic Random Functions and Their Applications, *Advances in Applied Probability, Vol. 5, pp. 439-468.*

NADP. (1980). NADP Quality assurance Report, Central Analytical Laboratory, 1/1/79 to 12/31/79. Colorado State University, Ft., Collins, CO.

NADP. (1982). National Acid Deposition Program Instruction Manual, Site Operation. natural Resource Ecology Laboratory, Colorado State University, Ft., Collins, CO.

NADP. (1984a). Instruction Manual, NADP/NTN Site Selection and Installation. natural Resource Ecology Laboratory, Colorado State University, Ft., Collins, CO.

NADP. (1984b). NADP Quality assurance Plan, Deposition Monitoring. natural Resource Ecology Laboratory, Colorado State University, Ft., Collins, CO.

Neuman, S. H. and Jacobson, E.A. (1984). Analysis of Nonintrinsic Spatial Variability by Residual Kriging with Application to Regional Groundwater Levels, *Mathematical Geology, Vol. 16, No, 5, pp. 499-521.*

Olsen, A.R. and Slavich A.L. (1986). Acid Precipitation in North America: 1984 Annual Data Summary from Acid Deposition System Data Base. *U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA/600/4-86/033.*

Robertson, J. K. and Wilson, J. W. (1985). Design of the National Trends Netwok for Monitoring the Chemistry of Atmospheric Precipitation. U.S. Geological Survey Circular 964 U. S. Geological Survey ; Alexandria, VA.

Schertz, T. L. and Hirsch, R. M. (1985). Trend Analysis of weekly Acid Rain Data -- 1978-83, U.S. Geological Survey, *Water Resources Investigations Report 85--4211*

Seilkop, S. K. and Finkelstein, P. L. (1987). Acid Precipitation Patterns and Trends in Eastern North America, 1980-84, *Journal of Climate and Applied Meteorology, Vol. 26, pp. 980-994.*

Sen, P. K.(1968a). On a class of aligned rand order tests in two-way layouts, *Journal of the American Statistical Association, Vol. 39, pp. 1115-1124*

Sen, P. K.(1968b). Estimates of the regression coefficient based on Kendall's tau, *Journal of the American Statistical Association, Vol. 63, pp. 1379-1389.*

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

van Belle, G. and Hughes, J. P. (1984). Nonparametric Tests for Trend in Water Quality, *Water Resources Research, Vol. 20, No. 1, pp. 127-136.*