Methods for Analysis
of Spatial and Temporal Patterns

Prepared by
Dr. Alan D. Jassby
Division of Environmental Studies
University of California, Davis
Davis, CA 95616

A Special Study of the
San Francisco Estuary Regional Monitoring Program
San Francisco Estuary Institute
1325 S. 46th Street
Richmond, CA 94804

September 1996

TABLE OF CONTENTS

LIST OF TABLES

(IN APPENDIX A)

LIST OF FIGURES

(IN APPENDIX B)

INTRODUCTION

Trace substance measurements have been made in the waters of San Francisco Bay by various groups now working under the auspices of the Regional Monitoring Program for Trace Substances (RMP), funded through the San Francisco Bay Regional Water Quality Control Board (SFBRWQCB) and under the management of SFEI. The water column data include near-surface measurements up to 3 times yearly since 1989 at up to 27 stations throughout the Bay and nearby portions of tributary rivers. Numerous trace elements and trace organics were measured, in addition to relevant environmental characteristics such as salinity and total suspended solids (TSS).

The purpose of this project is to suggest ways in which these data can be analyzed to describe spatial patterns and temporal trends. Four specific questions were used to guide and organize the data exploration:

1. How do we determine spatial boundaries within which the data can be summarized and between which comparisons should be made?

2. How can differences in space be detected?

3. How can significant relationships between trace substance concentrations and environmental characteristics be determined?

4. How can differences in time be detected?

Guided by these questions, we arrive at a set of recommendations for analyzing RMP data with an underlying goal of facilitating the upcoming five-year review of the data and program. Due to the size of the data set, the complexity of the issues, and the limited time available, we have not attempted to apply any of these techniques exhaustively to the data set for definitive answers. This is the task of the five-year review. We do, however, offer a specific example of each recommended approach.

We used water column total trace element data collected and provided by Russ Flegal and his group at the University of California, Santa Cruz. Although based on the trace element data, our recommendations should be applicable to other trace contaminants as well. The complete list of stations and sampling events (through mid 1995) are listed in Table 1 and Table 2, respectively. The RMP data collection effort officially began in 1993, so the earlier sites and years belong to a "pilot" program. In our analysis, we focus on the stations that have become part of the RMP program itself (Figure 1) and the sampling events that include these particular stations (events 6–15).

Unless otherwise stated, we used S-Plus (Statistical Sciences, 1995) to conduct the analyses and produce plots.

This project was supported by the San Francisco Estuary Institute (Project No. SFEI-135-95).

SPATIAL STRATIFICATION

Horizontal stratification of an estuary can be motivated by many different goals: (1) In the context of optimizing sampling design, stratified sampling refers to dividing a relatively heterogeneous estuary into more homogeneous subdomains and then carrying out either a random or systematic program of sampling independently within each subdomain (stratum). Insofar as the within-subdomain variability is reduced relative to the between-subdomain variability, stratification can lead to a more precise estimate of the mean than either simple random or systematic sampling (Cochran, 1977). The strong spatial correlation characteristic of estuaries (Powell *et al.,* 1986) suggests that stratification of sampling into spatially contiguous sub-regions might be appropriate. Stratification may be motivated, however, by other considerations as well. (2) Administrative convenience can be a valid reason when, for example, different sampling methods are required for different habitats of an estuary (e.g., shoals vs. channels). (3) Stratification may also proceed along political boundaries such as those between counties or states, particularly when the issue is one of compliance with government regulations.

(4) Division into subdomains can also be motivated by the need to understand underlying causal mechanisms, in which case one might want to stratify on the basis of covariability of different spatial locations in time.

Goal (1) is a relevant issue for the RMP, particularly in the context of trend detection. In order to study the efficacy of estuarine stratification in the context of this first goal, one must have a method for effecting a stratification and the data for evaluating it. Several methods are available.

Tree-based modeling or regression, which operates by successively splitting a dataset into increasingly homogeneous subsets or strata until some stopping rule comes into effect, is one of many approaches to the problem of grouping objects (in this case locations) into subgroups according to their similarity (Clark and Pregibon, 1992). Several features of tree-based regression attracted us originally. First, the criterion used for splitting a subdomain supports the goal of stratification for statistical estimation, namely decreasing the within-subdomain variance. Second, by operating through a binary recursive partitioning, it automatically preserves spatial contiguity within subdomains. Third, it can be applied to two- and higher-dimensional data.

Attempts to apply tree-based regression to datasets similar in size to the RMP dataset, however, have convinced us that more spatial locations are necessary for use of the technique. Under the auspices of this project, we have applied the technique to the USGS MIDAS datasets, which consist of approximately 5,000 data records per transect between Coyote Creek and Rio Vista. The details are described in a separate manuscript (Jassby *et al.,* 1996).

Another approach is through some kind of cluster analysis that can classify stations into subgroups that are similar within-groups but disparate between-groups. A number of clustering algorithms are available, differing in measures of similarity, in optimization criterion, and in search method. We examined an approach called model-based clustering, which is based on the assumption that the data are generated by a mixture of underlying probability distributions. Different clustering criteria can be chosen, depending on the assumed distribution of the cluster members (spherical or ellipsoidal) and whether the size of the clusters in the spherical case, and orientation, size and shape in the ellipsoidal case, is the same or not for all clusters. One of the advantages of model-based clustering is the availability of a supplementary technique for choosing the number of significant clusters using Bayesian analysis. A statistic known as the approximate weight of evidence for k clusters (AWE_k) is calculated for each k; the value of k that maximizes AWE_k is the number of clusters for which there is the most evidence. If all the AWE_k are negative, there is no evidence for any clustering. Given the "approximate" nature of this statistic, it functions best as a guide to the number of clusters rather than a dependable specification of that number.

We examined the behavior of model-based clustering using two separate clustering criteria. The first clustering criterion, which we refer to as SPHER (for "spherical"), assumes spherical clusters of different sizes determined by the data (Banfield and Raftery, 1992). The second criterion, which we refer to as UNCON (for "unconstrained"), assumes ellipsoidal clusters with different orientations, sizes, and exact shapes determined by the data (Scott and Symons, 1971). These two criterion are therefore near the opposite ends of the spectrum in terms of pre-specifying the nature of the clusters.

We applied each criterion to the trace element data set in two different ways:

1. In the first way, we attempted to cluster stations for each sampling event. For each event, we removed from the data matrix (stations x elements) any trace element for which one or more data points were missing. Data for each element were scaled by the maximum value of that element in that sampling event.

2. In the second way, we tried to cluster stations for each trace element, but all sampling events. For each element, we removed from the data matrix (stations x events) any event for which one or more data points were missing. Data for each event were scaled by the maximum value of that event for that element.

We used the "total" trace element data for sampling events 7–15, i.e., April 1992–April 1995 (Table 2). The combinations of events and trace elements available during this period are summarized in Table 3.

The results are summarized in Figure 2 and Figure 3. Three features stand out:

1. The results were often dependent on the exact model, i.e., whether the SPHER or UNCON criterion was used. This was true whether clusters were determined for single events (e.g., Figure 2, event 7) or single elements (e.g., Figure 3, Cr).

2. The AWE_k statistic indicates significant clustering in very few cases (Figure 2, events 10, 12 and 13; and Figure 3, Cr, Fe, Hg, Ni, Pb). In all of these cases, but Fe, the AWE_k statistic suggested only two significant clusters. In the case of Fe, the multiple clusters were based on a single event (event 7); Fe data were unavailable for subsequent events. Furthermore, significant clustering was implied only with the UNCON criterion.

3. The clusters are dependent on the sampling event and on the specific trace element. Even comparing only the analyses where two significant clusters were detected shows little constancy of clustering. BA40 and BC20, for example, occur in the same cluster for event 12 but different clusters for event 10 (Figure 2). Similarly, BA40 occurs separately from BA20 and BA30 in the case of Cr, but together with them in the case of Ni (Figure 3). Some of the trace element clusters are similar (compare Hg and Ni, or Cr and Zn), but these appear to be the exception rather than the rule.

These results effectively point out the shortcomings of cluster analysis and its inappropriateness for choosing strata with this particular data set. Clustering algorithms will always result in clusters of some sort and, given the complexity of most ecological phenomena, especially in estuaries, rationalization of the results is usually easy to accomplish. The three features pointed out here, however, each correspond to a problem that should make us proceed with caution.

First and foremost, we need to decide beforehand what algorithm should be used, which in turn depends on how we plan to use the clusters. In the case of stratifying to decrease the variance of the global mean, however, it is not clear which of the available models is most appropriate. The unconstrained method has the advantage of fewest assumptions among model-based techniques, and it alone produced significant clustering in the analyses presented here, but it does not cluster with the same criteria as that required for optimizing precision of the mean through stratified sampling. We are thus left without an objective way to choose a model.

Second, even if we knew which model to use, the behavior of the AWE_k statistic suggests that the data set is too small. The clusters that are determined are probably spurious in large manner, with no more status than the clusters we could obtain with a random data set. This problem will not go away unless the number of sampling stations per event increases substantially.

Thirdly, even if we have confidence in a particular model and have a sufficient number of stations, clusters are simply too evanescent in time to be dependable groupings. Moreover, clusters for one trace element do not necessarily correspond to those for another. This dependence on time and variable was clearly observed in our study of the MIDAS water quality data, which focused on salinity, suspended particulate matter (SPM) and chlorophyll *a* (Jassby *et al.,* 1996).

A further problem with clustering algorithms, which can be observed by inspection of the data, is the lack of respect for spatial contiguity. One would expect that components of an ecosystem that are closer in space are more likely to be under the same generating processes (Legendre, 1987). This expectation legitimizes an approach in which clusters are valid only when composed of spatially contiguous stations. Clusters containing noncontiguous stations are often just artifacts. There are ways to modify the conventional clustering algorithms to respect spatial contiguity, (Oliver and Webster, 1991) but this is yet another reason why conventional algorithms should be viewed with caution.

For all of these reasons, we recommend against the use of conventional cluster analysis as a final determinant of estuarine stratification with respect to the trace element data. There is no harm in using the method as an exploratory technique, but one should be advised that the data are probably too few to make it worthwhile.

Are there any alternatives? Although tree-based regression and other techniques are inherently spatially-constrained and also are based on criteria more in keeping with the requirements of an

effective stratification, there is no way to avoid the two other basic problems: not enough data and an estuary constantly in flux.

It is useful to remember that a stratification does not need to be optimal in any sense to be effective. In the case of the MIDAS data, for example, we combined all the tree-based regression results for all variables and cruises and used that as a guide to effecting a compromise stratification (Table 4). This stratification scheme turned out to be very effective in reducing the variability of estimates of the global mean, compared with simple random sampling. In principle, then, we can test any scheme for efficacy on an existing data set.

A fundamental problem with the RMP data set, however, is that the sampling design is not probability-based and there is no proper way to calculate global estuarine properties and their variance. The efficacy of any stratification for reducing variance is a moot point. The desirability of such global estimates needs to be considered.

If global estimates are desired, the stations should be laid out so that proper estimates can be made of global properties such as subembayment means. Given the experience in other systems with significant spatial autocorrelation, a systematic (regular) grid of stations is to be preferred over a random one. This "primary" set of stations should be supplemented with a "secondary" set located nearshore by effluents suspected to be important sources of one or more contaminants. The purpose of the primary set is to establish regional status and trends. The purpose of the secondary set is to provide important supplemental local information that could bear on causality.

The number of primary stations needs to be determined on the basis of a model for the design and the desired performance in terms of trend detection. An important requirement for this determination is inclusion of the spatial correlation structure. The primary grid of stations remains fixed through time, although the exact subset of these stations sampled each year may cycle in some way.

Secondary stations, on the other hand, are determined by an understanding of possible sources. The secondary set may change from year to year in a flexible way depending on the accumulated data and changes in activities within the watershed.

REGIONAL DIFFERENCES
Choosing Subregions for Comparison

Although there may be no "self-selecting" subregions for improving estimates of statistics like the global mean, many other reasons exist for selecting spatially contiguous groups of stations and comparing their means (see above). One particularly important reason is to establish the location of sources. If one region has higher trace element concentrations than another, the first is more likely a more important source or closer to the actual source. In this section, we explore how to do this comparison. Before doing so, however, we need to choose subregions of the Estuary with which to make the comparisons. We could do this arbitrarily, of course, but this seems like a good place to explore what subregions might be most useful to the RMP. In order to guide the discussion, we have mapped all the total trace element data in the RMP program (Figure 4). Concentrations are proportional to the areas of the squares, with the largest square being the same size in each map. As a result, square sizes can be compared only within individual maps. If the scale were the same for all maps for a given element, then many squares would be too small to distinguish among their sizes.

The maps each show two horizontal dashed lines. The dashed line at 4,160 km separates out all stations south of the San Bruno Shoal. The dashed line at 4,200 km separates out all stations in North Bay (San Pablo and Suisun Bays). Because the San Bruno Shoal is an important hydrological boundary (Powell *et al.,* 1986), it is arguably a better boundary for the South Bay than the more traditional Bay Bridge location. These boundaries, then, can be said to divide the Bay into South, Central, and North Bays. We do not want to make an argument that these boundaries are optimal in any way, only to point out that they have hydrological, physiographic, and historical significance.

In fact, these boundaries appear to have utility in describing trace element concentrations. In the case of silver, chromium, copper, mercury, nickel, lead, and zinc, a sharp demarcation often occurs in concentrations between the central and northern stations. For these same elements, values below the San Bruno Shoal are clearly higher than for the central stations during many RMP sampling events.

In the latter case, however, rather than a clear step in levels between embayments, we see rather a more gradual tapering off of concentrations from the south to north. As a result, the exact placement of a boundary between the southern and central stations is not well-defined and an argument could be made to move it at least one station south in many instances. The argument is particularly strong for the April 1995 data, as an example. We have kept the boundary just south of the San Bruno Shoal because of its hydrological significance and because a single best demarcation line does not exist, but a line further to the south would be equally acceptable and might give clearer differences at times between subregions means.

The easternmost of the northern stations, particularly the river stations and Honker Bay, often have lower concentrations as a group than the rest of the northern stations, suggesting an additional boundary at, or east, of Martinez. Such a boundary could be useful for refining hypotheses about sources to the northern stations, although the power of such tests is in danger of being too small because of the smaller number of stations.

Note how the boundaries suggested by the trace element data compare to those for the MIDAS data (Table 4). Because of the density of the MIDAS data, more boundaries can be identified. The only real discrepancy between the two sets occurs with the central stations. The MIDAS data suggest a boundary at Angel Island, while the trace element data suggest one at Point San Pablo. Otherwise, combining regions 1 and 2, and also regions 4 and 5, for the MIDAS data yields subregions similar to those suggested for the trace elements. No doubt the differences have to do with the effects of strong localized sources for some of the trace elements.

Incorporating Spatial Correlation into Analyses

Now we move on to the question of determining differences among these subregions. For independent samples, one can compare the means of groups by means of a conventional analysis of variance (ANOVA). For spatial data, however, the samples are not necessarily independent. In particular, because of the mixing processes in any water body, neighboring values tend to resemble one another and so the spatial changes are quite smooth compared to a random sample. Most classical statistics are not robust to the presence of positive spatial autocorrelation, i.e., the tendency for neighboring samples to be more similar than random samples. If spatial autocorrelation is not taken into account, models may end up being specified incorrectly, parameter estimates may be biased and inferences may be incorrect. Spatial autocorrelation is a potential problem in analyzing estuarine data, whether the issue is ANOVA, correlation, regression, or many other techniques (Griffith, 1987). For many of the RMP data, differences among subregions are obvious from a visual inspection and it may seem wasteful to go through the following calculations. Nonetheless, not all differences are conclusive from a visual inspection and in any case an "objective" approach is required to confirm subregion differences that might have important consequences in terms of pollution abatement.

In order to address spatial autocorrelation in data analysis, one must either show that it is not a problem or account for it. Spatial effects can enter in many forms, however, so in either case one must make some assumptions about the form, i.e., postulate some specific underlying statistical model. We chose to focus on a model called the *spatial lag model* (Anselin *et al.,* 1993):

$$y = \rho W y + X \beta + \varepsilon, \qquad (1)$$

where y is a N by 1 vector of observations on a trace element (N is the number of stations), X is a N by k matrix of observations on k explanatory variables at the N stations, β is a k by 1 vector of regression coefficients, and ε is a N by 1 vector of error terms. The difference between this and the ordinary-least-squares (OLS) regression model is the presence of the term $\rho W y,$ where W is a N by N matrix of spatial weights expressing the influence of all stations on each individual station and ρ is a coefficient.

This model is interesting because of its consistency with the advection-diffusion equations describing the movement of water and associated substances. The advection-diffusion equation when discretized results in a second-order autoregressive equation that is essentially a special case of equation 1.

An alternative model, the *spatial error model*, is similar to OLS:

$$y = X\beta + \varepsilon, \quad (2)$$

but the error terms are now correlated at neighboring locations. The spatial dependence of the error term can be formulated in two ways, as a *spatial autoregressive error model:*

$$\varepsilon = \lambda W \varepsilon + \mu, \quad (3)$$

or as a *spatial moving average error model:*

$$\varepsilon = \lambda W \mu + \mu, \quad (4)$$

> where λ is the autoregressive parameter and the μ are independent and identically-distributed error terms.

There are various ways to express how stations influence each other through the spatial weights matrix W. We have used the "gravity" model in which the influence is proportional to the inverse square of the distance separating the stations. Another more complicated spatial weights matrix may be more suitable for estuarine conditions. The influence due to tidal mixing, for example, probably falls off rapidly beyond 5–10 km, whereas advective influences may extend over the spatial scale of the Estuary. Moreover, both of these influences are anisotropic. The southern Bay may require a different spatial weights matrix than the northern Bay because of their different hydrodynamic regimes. Other forms for the spatial weights matrix therefore need to be carefully devised and examined. For our purposes of demonstrating the method, we will assume that this model is sufficient. Otherwise, the number of analyses to be done quickly spins out of control.

Spatial ANOVA can be treated as a special case of both equations 1 and 2 through the use of dummy variables. An indicator value is defined for each subregion, taking on the value of 1 for stations in that subregion and 0 otherwise.

Several tests are available for determining whether an ordinary-least-squares (OLS) regression is sufficient or whether spatial effects need to be incorporated. These tests can also indicate whether the spatial lag model is an appropriate alternative or whether some other formulation of the spatial effect should be used. We used a battery of so-called *Lagrange Multiplier* (LM) tests (Anselin, 1988; Anselin *et al.,* 1996): LM_{lag} tests for the presence of a spatially-lagged dependent variable, i.e., for the appropriateness of equation 1; LM_{err} tests for the presence of spatial error dependence, i.e., for the appropriateness of equations 2 and 3 or 4; and LM_{sarma} tests for the presence of a higher-order model involving both a spatial lag and a moving average error process. We also used robust versions of LM_{lag} and LM_{err} that are less likely to be distorted by spatial error and spatial lag dependence, respectively.

We used Maximum Likelihood estimation to solve these models. The algorithms were developed by Luc Anselin, formerly at the National Center for Geographic Information and Analysis at UC Santa Barbara and now at the Regional Research Institute of West Virginia University. The core algorithms are available free of charge and have been translated into several high-level languages (including GAUSS and S-PLUS; Anselin *et al.,* 1993). They are also part of a more comprehensive commercial package called *SpaceStat*. Both the core algorithms (S-PLUS version) and *SpaceStat* were used in our analyses.

One caveat regarding the detection of spatial differences is related to the large number of potential tests with the RMP data. For example, if we wanted to test the differences between all pairs of three subregions for each of 10 elements, both total and dissolved, for the eight sampling events of Figure 4, we would have to conduct 480 tests. If we used a significance level of 0.05 for these tests, then we would detect 0.05 x 480 = 24 differences by chance even if all the data were random. Depending on the reason for conducting the tests, one might want to be protected against falsely rejecting the null hypothesis of no difference by a more demanding significance level. For example, when comparing the three pairs of subregions for each trace element and

sampling event, we might want to use 0.05/3 = 0.017 as a level of significance (the "Bonferroni" correction).

Another way that affords some protection is to first test for the presence of any subregion differences though ANOVA and then proceed to pairwise comparisons only if differences are indicated. In accordance with this approach, we first show some examples of ANOVA results using all three subregions, then proceed to the pairwise differences. We do, however, investigate all pairwise differences regardless of the overall ANOVA result.

Examples Using the RMP Trace Element Data
All Subregions

As an example, we applied this form of analysis to the total trace element data for 1994 and the three subregions illustrated in Figure 4. A number of separate analyses are required to arrive at the final conclusion. The Lagrange Multiplier tests are based on OLS estimation of the standard regression model. The first step is therefore to estimate the OLS model and examine the diagnostics, including the error distribution. Sample diagnostic results are shown in Table 5. A non-normal error distribution may distort subsequent tests for heteroskedasticity and spatial dependence. SpaceStat uses the Kiefer-Salmon statistic to test for normality, an asymptotic test that may not be reliable for small data sets. We found non-normality in 22 out of the 30 cases (Table 6).

If the errors are not normal, a log transform of the dependent variable can often make them so. For these data, only a single non-normal error distribution remained after the log transform: January arsenic. Quite possibly other transforms such as a member of the Box-Cox family of transformations could induce normality in the arsenic data but we did not pursue the issue further.

Next we checked for heteroskedasticity, a situation in which different subregions have different variances. Heteroskedasticity can lead to changed significance levels for ANOVA statistics and hence incorrect inferences. SpaceStat uses the Koenker-Bassett test when the errors are normal and the data are few. For non-normal error distributions (of which we have one), SpaceStat uses the Breusch-Pagan test. We found heteroskedasticity in only 4 of the 30 cases: selenium in January and August, mercury in April, and zinc in August.

Finally, we checked for spatial dependence of the errors using the various LM statistics and their robust forms. We found spatial dependence in 9 of the 30 cases. In two of these cases (silver and cadmium in April), the LM_{sarma} statistic indicated the presence of a higher-order process. In the other seven cases, the LM_{lag} statistic or its robust form was significant, which tends to support our belief that the spatial dependence in the Estuary arises primarily from transport processes. In one of these cases (selenium in April), both the robust spatial lag and spatial error terms were significant. As each of these tends to protect against effects of the other kind of dependence, the results suggest that some higher-order process is at work involving both spatial lag and error processes.

These diagnostic tests occur in various combinations. In the 19 cases where errors were normal (after log transforming, if necessary) and there was neither heteroskedasticity nor spatial dependence, the OLS model could be used directly for assessing differences among subregions with the F-test (Table 5).

In the one case where errors were normal and there was no spatial dependence but heteroskedasticity was present (mercury in April), we used a robust form of the OLS model that compensates for heteroskedasticity (MacKinnon and White, 1985). In the other three cases with heteroskedasticity, spatial dependence was also present. As spatial dependence can masquerade as heteroskedasticity, we decided to treat these in the same way as the other spatially dependent cases.

We estimated a spatial lag model for the six spatially dependent cases which had a significant LM_{lag} statistic (robust or otherwise) (sample output in Table 7). The remaining three spatially dependent cases exhibited signs of a higher-order model, which currently cannot be estimated in SpaceStat. In two of the six spatial lag models, the heteroskedasticity that was present in the OLS model persisted.

For the 20 well-behaved OLS models (robust and otherwise), we detected significant spatial dependence in all but 2 cases: January cadmium and August arsenic. For the four well-behaved spatial lag models, we detected significant spatial dependence in all cases. To sum up, we were able to estimate well-behaved ANOVA models in 24 of the 30 cases, and we found evidence for distinct spatial subregions with 22 of the 24 models.

Pairwise Differences

The spatial ANOVA analyses were conducted as a dummy variable regression with no constant and an indicator variable for each region (with a value of 1 for stations in the region, 0 otherwise). The variance of a pairwise difference is (Snedecor and Cochran, 1967):

$$\mathrm{var}(\overline{X}_1 - \overline{X}_2) = \mathrm{var}(\overline{X}_1) + \mathrm{var}(\overline{X}_2) - 2\,\mathrm{cov}(\overline{X}_1\overline{X}_2), \quad (5)$$

where the $\overline{X}_i$ are subregion means. These variances and covariances are given by the coefficient variance-covariance matrix in the case of OLS, and the asymptotic variance-covariance matrix in the case of spatial lag models. For OLS, the significance of $(\overline{X}_1 - \overline{X}_2)\big/\sqrt{\mathrm{var}(\overline{X}_1 - \overline{X}_2)}$ is tested with the t-distribution; for spatial lag models, the significance is tested with the standard normal distribution.

Several patterns were seen (Table 8). The most common, which occurred in 18 of the 24 models, was a depression of Central Bay concentrations with respect to both South and North Bay concentrations. In two of these cases (event 14 chromium, event 12 mercury), North Bay concentrations were greater than those in South Bay. Cadmium and selenium had the most unusual distributions. Cadmium showed either no subregional differences (event 12) or a pattern in which South Bay values were elevated. Selenium values were also elevated for South Bay, at least compared to North Bay (event 12). Finally, note that a pairwise difference was detected for arsenic during event 14, despite the fact that the ANOVA was not significant. Some statisticians caution that special comparisons should not be considered significant if the overall test is not (Snedecor and Cochran, 1967).

CAUSAL MECHANISMS
Source Identification

As discussed above, mixing creates a spatial averaging effect where the values at neighboring locations tend toward each other. The further away the location, the less influence it will have. The spatial weights matrix for a given sampling event and variable explicitly describes how the effect of neighboring locations drops off with distance. Using the spatial weights matrix, we can predict in some sense the contribution of mixing to the concentration at a given location by the concentrations at all the others. Large deviations from this prediction then suggest the presence of important accumulations or deficits. In fact, the difference between the value expected on the basis of the spatial weights matrix and the actual value enables us to pinpoint the location of these anomalies in an efficient and objective way. Whether or not these anomalies are due to unusually large local sources, sinks, or both cannot be determined without additional information.

Our approach is graphical and makes use of a diagram called the *Moran scatterplot* (Anselin, 1994). Moran's I statistic describes the degree of linear association between a vector of observed values y and a weighted average of the neighboring values Wy, where W is the spatial weights matrix. Wy is sometimes called a *spatial lag*. If the y are standardized in deviations from their mean, then Moran's I is simply the slope of the regression of Wy on y (Figure 5). Points lying near the regression line are "typical" in terms of their relations to their neighbors. Points lying far below the line, however, have a higher y than one would expect. These are locations where a positive anomaly in concentration occurs, which may indicate a local source.

One of the problems with the I statistic is its susceptibility to single observations that can exert a large influence on the slope. Large sources may fall into this category, pulling the regression line toward them and diminishing our ability to detect them. In order to have a measure of the central tendency that was more robust to the presence of outliers, we also calculated a least trimmed

squares robust regression line. This line minimizes the sum of the smallest half of the squared residuals, as opposed to minimizing the sum of all of them. We then measured the residuals from this robust regression line and identified the three most negative on each graph by labeling them with their corresponding station code. These three stations usually included those stations that could be considered to lie unusually far from and below the robust regression line.

The stations with the most important positive anomaly for an individual trace element (total concentration) during sampling event 13 are listed in Table 9. Note that a positive anomaly suggests the location of a source, not its relative importance. For example, consider two point sources that contribute the same trace element flux to the Estuary, one with low concentrations and high discharges, the other opposite. The former will tend to be more similar to its neighbors and therefore less likely to stand out in a Moran scatterplot.

Tests of Association

In addition to proximity to sources, trace element distributions will be influenced by water quality variables such as total suspended solids (TSS) and chlorophyll a. Because of the limited amount of data, it is not possible to test simultaneously for the nature and importance of all the potential effects. With only 24 stations per sampling event, for example, one cannot expect to explore more than 2 explanatory variables at a time, at most, aside from the variable of spatial location. There is a way to take the results of individual tests and combine them into a larger causal picture, which we shall describe below. First, however, we have to decide exactly how to do the individual tests, keeping in mind that spatial autocorrelation precludes the use of conventional statistics.

One possibility is through spatial regression. The ANOVA examples discussed above are simply a special case of how spatial correlation can be included in studying the relationship between a dependent and one or more predictor variables. In the ANOVA case, the predictor variables are subregions. In other cases, we might be interested in other predictive variables, including continuous ones such as salinity and TSS. A complication arises because many of these other predictor variables themselves often have a spatial structure. One outcome is that the spatial error model or perhaps a more complicated one will be more appropriate than the spatial lag model.

Another approach called the *Mantel test* can be used to analyze complications due to the presence of spatial dependence (Mantel, 1967). The Mantel test basically examines for any (multivariate) variables X and Y how the *differences* between pairs of stations are correlated rather than the stations themselves. In order to calculate a (normalized) Mantel statistic, one first needs to decide on a measure of difference between stations. Here we use euclidean distance as a measure. For univariate variables, this is simply the absolute value of the difference between two stations. A distance matrix is constructed for each variable (each entry is the euclidean distance between the corresponding row and column variables) and then a correlation coefficient is calculated for the two matrices. We use the Pearson correlation coefficient for illustrative purposes, although in principle more robust choices can be made. The usual significance tests are not valid, but the significance of the statistic can be determined by randomly permuting one of the distance matrices and recalculating the statistic, at least 1,000 times. One major advantage of the Mantel test is that station location itself (e.g., UTM coordinates) can be used as one of the variables, in which case one has a simple straightforward test for the presence of spatial dependence in a variable.

The Mantel test can be extended in a very important way to take into account how spatial structure might be affecting the (and even causing a spurious) relation between two variables. In an estuary, most variables have a strong large-scale structure simply due to the presence of mixing. This mutual structure will tend to induce correlations between many variables, even though they have no real causal connection. The *partial Mantel statistic* provides a means for handling such situations (Smouse *et al.,* 1986). If we are interested in the relation between X and Y corrected for the influence of U, then we first construct distance matrices for each of the three variables. The residuals X_r are then calculated from the regression of the X distance matrix on the U distance matrix, and similarly for Y_r. Finally, the correlation between X_r and Y_r is calculated as above. The significance can be determined by random permutation of one of the residual matrices, holding the other one constant.

For illustration, we examined the associations between each of 10 trace elements and TSS for sampling event 12. Most of the associations are significant when we use the simple Mantel statistic (row 1 of Table 10). On the other hand, when we calculate the partial Mantel statistic, which removes the spatial effect by first regressing distance matrices for trace elements and TSS on interstation distances, only one of the eight (chromium) correlations remain significant. This example clearly illustrates how spurious spatially-generated correlations can arise in estuaries, as well as of the utility of the partial Mantel statistic in detecting and correcting for these.

ANALYZING CAUSALITY

The Mantel and partial Mantel statistics in principle can be used to build up a model of causality for trace element distributions. The procedure involves a listing of all possible models, followed by an analysis of the consistency between each model and the correlations and partial correlations among the components of the model (Legendre and Trousellier, 1988). As an example, we will look at a three-component model consisting of spatial location, a trace element total and a cognate variable. The corresponding distance matrices are SPACE, TE, and COG. The eight possible models, assuming that the cognate variable can determine trace element concentrations, but not vice versa, are listed in Table 11. Note that our set of models differs from those of Legendre and Trousellier by this assumption, as well as by including possibilities (models 5–8) where some or all connections may be absent.

Each model has a set of correlations and partial correlations that should be significant, a set that should not be significant, and arithmetic relations among the correlations and partial correlations that should hold. As an example, the following relationships should hold if model 4 is true:

1. COG·SPACE should be significant

2. TE·SPACE should be significant

3. (TE·SPACE)·COG should be significant

4. (COG·SPACE)·TE should be significant

5. (TE·COG)·SPACE should *not* be significant

6. (TE·SPACE)·COG ≤ TE·SPACE

7. (COG·SPACE)·TE ≤ COG·SPACE

8. (TE·SPACE) x (COG·SPACE) ≈ TE·COG

Each model needs to be checked for consistency with its corresponding relationships. As an example, we examined the trace element totals for sampling event 12 using TSS as the cognate variable. The relevant statistics are listed in Table 10. In this particular case, none of the models were found to be consistent with the data. For example, we can eliminate model 4 for all trace elements simply because the third condition above is violated in the case of every trace element.

One possible reason for the inability to select a model is that the power of the tests are too low. In other words, some of the relationships are true but were rejected by chance because the probability of a Type II error (rejecting a true hypothesis) is too large. One can address this problem in part by taking another approach, namely by eliminating *only* those models for which Mantel statistics that should *not* be significant *are* in fact significant. Here, the potential errors are equivalent to the probability of a Type I error (accepting a false hypothesis), which is set to only 5%. Let us examine chromium as an example. First, we can eliminate models 1, 6, 7, and 8 as these imply that COG·SPACE should not be significant, contradicting line 3 of Table 11 (in fact, this association eliminates these models for all trace elements). Next, we can eliminate model 5 because TE·SPACE is in fact significant. Further, we can eliminate model 4 because (TE·COG)·SPACE is significant for chromium, leaving only models 2 and 3 as possibilities. This is as far as we can go: neither of these models can be eliminated risking only a Type I error. Nonetheless, the result is of interest as it leaves only models in which TSS has a direct effect on chromium, supporting the contention that their association is not a spurious one.

Another possible reason that the causal analysis does not settle on a single model is the presence of other factors that can mediate a spatial effect on cognate variables or trace elements. One of

the models of Table 11 could then correctly express the causal relations among SPACE, COG, and TE, yet there would be additional unexpected correlations due to connections that cannot be seen from the narrow perspective of these three-variable models. This possibility of course is always lurking no matter what kind of statistical or dynamic modeling one undertakes.

There are two implications from the considerations of this section that should be emphasized. First, if statistical associations between trace elements and possible causal factors are going to be explored, then spatial effects must be taken into account and the relations must be explored in the context of a causal analysis framework. One can conclude very little from a single association, particularly in a spatially-structured estuary.

Second, because of the probably complex causal nexus and the possibility of Type II errors, a statistically-based causal analysis will at best be able to narrow down the set of possible models, which will be of some help but probably not sufficient. Additional stations can only help, but it is difficult to know how many are really necessary and there is a possibility of wasting any effort on more stations, at least in this particular context. Regardless of the success of a statistical analysis, understanding of these causal relationships must be founded also on general chemical and ecological understanding, as well as non-RMP data sets and experimental work. As the RMP encounters the limits of its baseline data collection program in assessing causality on a statistical basis, more attention should be given to how other kinds of field measurements and experiments can narrow down even further the set of possible models.

TEMPORAL TRENDS
Trend Testing for Individual Sites

The trace element totals are plotted as time series in Figure 6. Generally speaking, these plots with vertical lines at each data point are superior to plots in which successive data points are connected by a straight line when there are many data gaps (see plots for zinc). The eye tends to impute trends to the latter type of plot even when they are not warranted by the data frequency. One could, of course, simply plot data points and not connect them with lines, but many find these vertical-line plots less confusing when there is a lot of short-term variability. It is also common to refine the vertical-line plots so that the long-term mean of the data is first removed and the anomalies (departures from the long-term mean) are plotted. Trends can often be seen more clearly using anomalies, although they have the disadvantage of requiring an additional step (adding back the mean) in order to determine the value of an individual point, plus they are largely unfamiliar to lay people. One obvious disadvantage of the plots in Figure 6 stems from the fact that all scales are constrained to be the same: in some cases, data fall too close to the abscissa and it is difficult to discern temporal behavior. If these details must be seen, then the scales can allowed to float free from each other, but this has the disadvantage of requiring an explicit scale for each panel of the plot and a consequent decrease in size of the panels and increase in clutter (see plots for zinc).

The Mann-Kendall test is widely used for trend testing, particularly when many time series need to be evaluated at the same time. The Mann-Kendall test is basically an application of Kendall's tau correlation coefficient to the case in which one of the variables is time. It has the following important advantages (among others): no assumption of normality is required; it is resistant to outliers; and it admits censored data (as only ranks are used). These characteristics are big advantages when analyzing for trends in many time series (such as multiple trace elements at multiple stations), because they mean that it is not necessary to screen the data for normality and lack of outliers and the number of degrees of freedom can be increased by including censored data.

Accompanying the Mann-Kendall test is a method for estimating a robust trend line, called the Thiel slope. The Thiel slope is simply the median slope for the set of lines joining all possible pairs of points in a time series. The significance of the test for significance of the Thiel slope is identical to the significance of the Mann-Kendall test. Both the Mann-Kendall test and Thiel's robust estimate of the trend line have been shown to be superior to equivalent parametric tests with only slight departures from normality. Water quality data are usually skewed and often ill-behaved in other respects, which suggests that trends should be quantified and assessed using these more robust approaches.

One remaining requirement of these robust tests is that the serial correlation between successive sampling days is small. We believe that the intervals between sampling days are so large that serial correlation is not a problem. The samples are not taken during the same hydrological event such as a pulse flow, for example, which could induce such correlation. The data are still so few that a statistical test of this assumption has very little power, so we will assume it solely on scientific grounds for now.

We applied the Mann-Kendall test to the period 1991–1995. Values of Kendall's tau for the individual tests are shown in Table 12. Only 3 of the 240 tests are significant, even fewer than we might have expected on the basis of chance.

The trend results can be further explored by mapping them (Figure 7). Here we have relaxed the definition of a significant trend ($p<0.1$). Despite the lack of significance of most of the trends, their pattern in space is often nonrandom. Groups of contiguous uptrends or downtrends give more weight to the presence of a spatially localized trend than do the sites considered individually. In the case of copper, for example, there seem to be four clustered areas of homogenous trend. The group south of the San Bruno Shoal tends to be increasing, as does the group extending from northern Central Bay through western Suisun Bay. Note that the river and east Suisun Bay stations are decreasing, suggesting that the zone of increase in the northern Estuary is due to influences from local discharges and the Napa and Petaluma rivers, not from upstream in the Delta and beyond. Examination of the trend patterns, even when the individual trends are not significant, can lead to further hypotheses and understanding.

Combining Tests from Multiple Sites

The apparently nonrandom spatial distribution of up- and downtrends such as for copper, even when trends at individual stations are not significant, lead us to ask whether there is any way to consider contiguous sites simultaneously. The seasonal Kendall test for individual sites suggests a possible and novel approach to this problem. In a seasonal Kendall test, the Mann-Kendall test is applied to each season (e.g., month or quarter) separately and then the results are combined for an overall test (Hirsch *et al.,* 1982). Each season by itself may show a positive trend, none of which is significant, but the overall seasonal Kendall statistic can be quite significant. The test has all the advantages of the Mann-Kendall test, offering higher power because it removes short-term variability caused by seasonality that would otherwise appear as background noise in a Mann-Kendall test for the whole time series. When successive seasons are correlated, as they may be if seasons are defined by months or smaller intervals, a correction must be used based on the covariance among seasons (Hirsch and Slack, 1984).

In principle, the seasonal Kendall test can be applied to multiple variables or multiple sites instead of seasons. In our case, this means that the tests for trace element trends at individual contiguous sites can be combined into an overall test for the subregion consisting of all of these sites; we might call this a *spatial Kendall test*.

One difficulty with the seasonal Kendall test is that trends of opposite sign in different seasons may cancel each other out, giving the impression that no trends are present. Similarly, the spatial Kendall test may show no significant overall trend, even though trends are significant within contiguous subsets of the sites. Careful consideration of how sites are to be grouped is necessary. One can use contrast statistics to test for trend homogeneity (van Belle and Hughes, 1984; Lettenmaier, 1988), but graphical exploration of the data combined with a scientific understanding of the problem should be a good guide on how to group contiguous sites. In the case of copper, for example, the sites clearly fall into four subgroups.

As an example, we applied the spatial Kendall test to the copper data (Figure 7), including the covariance correction (Table 13). The seasonal Kendall statistic, its standard error and the large-sample approximation for the significance level are given. None of the four subgroups exhibit a significant trend as a group of stations.

Adjusting Data to Reduce Short-Term Variance

Exogenous variables. Most ecological variables will exhibit significant short-term variability that tends to disguise more long-term systematic changes, including trends, that we might be

interested in. In the case of stream or estuarine chemical concentrations, streamflow is often the generating mechanism behind this shorter-term variability. If we can remove this "noise", i.e., this variation due to "exogenous" variables such as streamflow, then we are more likely to determine if a trend is present. All tests become more powerful when the variance is reduced. Adjusting concentrations for streamflow, usually by regressing the concentrations on streamflow and working with the residuals, is a common procedure in trend analysis of stream chemistry.

One requirement is that the probability distribution of the exogenous variable does not change over the period of interest. If it does, then a trend in the residuals may not be due to a trend in the trace contaminant of interest. With the relatively small number of observation times in this data set, we cannot test for a changing probability distribution with any power and must assume an unchanging distribution in the exogenous variable.

As an example, consider the effect of Net Delta Outflow (NDO) on trace element concentrations. We will take a "mixed approach", in which the effects of NDO are removed by regression and a Mann-Kendall test is applied to the residuals. It is also possible to take a fully nonparametric approach in which the exogenous variable's effect is removed through some smoothing procedure such as LOWESS (locally-weighted scatterplot smoothing). One can also take a fully parametric approach using multivariate regression with the exogenous variable and time as explanatory variables. The more nonparametric the approach, the fewer the assumptions necessary; the more parametric, the higher the power...*if* the necessary assumptions are actually fulfilled. Careful graphical examination is required in the latter case.

A summary of the significance levels for regression of trace element totals on NDO is given in Table 14. For most elements, the number of significant regressions could be expected on the basis of chance alone. For arsenic, selenium and especially cadmium, however, NDO has a marked relation with water column concentrations. We examined the data for cadmium in more detail (Figure 8). Although we did not undertake a detailed examination of the residuals, the regression lines appear to do an adequate job of capturing the relationship. There appears to be more scatter at the low flows, but this may be due only to the higher data density at low flows. Note that the slopes are consistently negative, suggesting a dilution of point sources.

Time series plots of the residuals (Figure 9) can be compared with the uncorrected total cadmium levels (Figure 6). Of particular interest are the 1995 data points, which are no longer negligible and therefore do not exercise such a strong influence toward negative trends. The statistics from the Mann-Kendall test of trend are summarized in Table 15. The downtrend of BG30 is now eliminated (cf. Figure 7) and uptrends appear for BD40 and BD50 in the vicinity of the Napa River. Removal of NDO effects has therefore revealed that the BG30 trend is due largely to the effects of dilution from river flow, and that there may have been a local increase in cadmium in Napa River sources that was disguised by interannual variability in flow.

Choosing the exogenous variables for correction must be done with care, guided by a specific goal. If we are specifically interested in anthropogenic trends, the exogenous variable itself must be free of human interference. In the case of San Francisco Bay, for example, Net Delta Outflow would then not be a suitable exogenous variable. The principles on which Delta hydrology is managed has changed over the years and will probably continue to change in the future. By removing the effects of Delta outflow, we could well be removing a major human influence on trace contaminant concentrations. The same can be said for any index of salinity distribution (such as X_2), as the salinity field is so closely tied to Delta outflow. An index of discharge that is unaffected by human hydrological manipulations, such as for example the Four River Index, is more suitable as an exogenous variable. On the other hand, if we are trying to investigate local source variability independently of other anthropogenic influences such as flow through the Delta, then NDO is the appropriate correction to use.

Seasonality. Seasonal changes may also induce "background" variability that interferes with the detection of longer-term systematic changes. The seasonal influence may be mediated completely by stream discharge, in which case there is no need to act separately on this problem. But seasonality can remain even after streamflow effects have been removed. Trace element sources, for example, may have a distinct seasonal pattern. Several methods exist for dealing with seasonal effects, but they fall into two main categories: (1) treat season as a separate variable in a regression, either as dummy variables or with the use of periodic functions, or (2) conduct trend

tests only within corresponding seasons and combine the results for all seasons (e.g., seasonal Kendall test).

Given the present size of the data set, correcting for seasonality is premature. Unlike the earlier "pilot" years, the RMP program (beginning in 1993) has sampled consistently in the first three quarters (Table 16). If this consistency is maintained, it will eventually be possible to examine the efficacy of a seasonality correction. We do not recommend such an approach at this stage. For example, treating quarter of the year as a separate variable in a regression requires the use of three dummy variables, resulting in a large percentage decrease in the degrees of freedom. Under these circumstances, corrections can have only a large uncertainty, resulting in at best no advantage and at worst the introduction of some spurious and distorting correction factor. Given the regularity of sampling during the second quarter over the pilot and RMP program years (Table 16), it might be of interest to examine trends at available sites using the data for this quarter only.

Mixing diagrams. It is sometimes suggested that using the residuals determined by joining the end members in a mixing diagram can help remove noise due to upstream influences. To examine this further, consider two cases, the first in which outflow changes and the second in which outflow concentrations change. In the first case, the higher the freshwater inflow, the more diluted will be local point sources, and the smaller the trace element concentration at these local point sources. This will be true whether the concentration is expressed in absolute terms or is measured as a residual from a mixing diagram. The residuals therefore do tell us the locations of possible local sources and sinks but in no way do they eliminate the effects of varying discharge. In the second case, the residuals are in fact uninfluenced by the changing end member concentration. On the other hand, if local concentrations increase as a result of upstream source increases, then the increase should be considered part of a real trend in the Estuary and not simply a result of freshwater discharge fluctuations. We therefore do not want to remove this effect from the data when examining for trends. In summary, residuals from mixing diagrams will be useful for diagnostic purposes, but they should not be used as a "corrected" form of the data.

SUMMARY AND CONCLUSIONS

1. Horizontal stratification of the Estuary can be of value in reducing the variance of global estimates such as the mean. We investigated model-based clustering of the trace element data as a means for choosing the strata in the case of San Francisco Bay. Two problems were encountered. First, the data are sufficient to identify at most only two significant clusters for any sampling event. Second, the clusters change both with the trace element in question and the sampling event. The first problem is a consequence of the limited data. The second problem is an underlying feature of the Estuary. As a result, we believe that the use of clustering in this context is unlikely to be of help, and is prone to mislead unless cluster statistical significance is also assessed. Note that stratification of the Estuary does not have to be optimal in order to be effective in reducing variance so that clustering and other "objective" approaches are not actually necessary.

2. In any case, stratification to reduce variance is currently a moot point as the RMP stations are not a probability sample to begin with and cannot provide proper estimates of the Estuary's mean and variance. The desirability of such global estimates and a possible redesign of station siting needs to be considered carefully.

3. Stratification can also be pursued for other reasons, such as to identify local point sources. A visual examination of the total trace element spatial patterns leads to the choice of three or four subregions: (1) south of San Bruno Shoal; (2) San Bruno Shoal through Point San Pablo; (3) north of Point San Pablo. The latter stratum can be further subdivided between Honker and Grizzly bays, although this leaves only three stations in the upstream stratum.

4. Spatial autocorrelation in estuaries potentially precludes the use of classical ANOVA for assessing subregion differences. Spatial ANOVA, in which individual sites are

influenced not only by their location within subregions but also by the values at neighboring sites, provides the solution. We examined the data for ten trace elements during 3 sampling events. We were able to determine well-behaved models in 24 of the 30 cases; 4 of the 24 cases required incorporation of spatial autocorrelation effects. In 22 of the 24 cases, we found evidence for distinct spatial subregions. Further analyzing the pairwise differences, the most common pattern (18 of 24 cases) was a depression of "Central Bay" (stratum 2 above) concentrations with respect to both "South" and "North" bay levels; means of the latter two were not significantly different in these cases.

5. Anomalous stations for any trace element and sampling event can be identified using the Moran scatterplot. Using a robust fit to the Moran scatterplot, we identified the most important positive anomalies for each trace element during sampling event 13. Positive anomalies can be interpreted as important sources. The San Jose station was the most common anomaly, followed by the Petaluma River.

6. Spatial autocorrelation in estuaries will result in potentially spurious correlations between almost any two variables. The partial Mantel test can be used to examine correlations among variables that are also spatially autocorrelated. As an example, 9 of 10 trace elements were correlated with TSS during sampling event 12, but only one association remained after spatial autocorrelation was accounted for using the partial Mantel statistic. Note, however, that this is a conservative procedure in that a true relationship may exist between other trace elements and TSS, but one cannot use a correlation as evidence because of the posssible confounding effect of spatial structure.

7. Proper statistical testing of causal connection between two variables does not consist of a single association test, even if it incorporates a correction for spatial autocorrelation. A causal analysis includes an array of possible models, as well as the associations, lack of associations and arithmetic relationships among associations that accompany each model. The RMP data are very limited in their ability to support such a causal analysis, primarily because of low power. The data are sufficient to narrow down the range of possible models, however. An example using chromium supports a direct effect of TSS on chromium. In general, though, the RMP should not expect any definitive causal analysis resulting from statistical analysis of the RMP data set and should accordingly support other field and experimental approaches to determining underlying mechanisms.

8. The Mann-Kendall test is an appropriate way for determining trace element trends at individual sites. Individual site trends are by and large not significant. When trends are mapped in space, however, trends of the same sign tend to occur contiguously in apparently nonrandom clusters and suggest systematic changes for subregions of the Estuary.

9. The seasonal Kendall test can be adapted to test for an overall trend in groups of stations. The increase in the power is such that an overall trend may exist even when no trends can be detected for individual sites. Mapping of the individual trends can guide selection of station groupings. A correction must be made for the covariance among stations, similar to the correction for covariance among months in the conventional use of the seasonal Kendall test. An example is given with copper, but no overall trend was detected for any of the four subregions.

10. The power of trend tests can be increased by removing exogenous sources of short-term variance. Residuals are determined for a parametric (regression) or nonparametric (LOWESS) fit of the data and a Mann-Kendall test is applied to the residuals. An example is given using Net Delta Outflow (NDO) and cadmium. NDO has a negative effect on cadmium at all stations. Before accounting for NDO, only the San Joaquin station exhibited a (down)trend. After accounting for NDO, this station no longer had a significant trend while two stations near the Napa River showed significant uptrends. Selection of the exogenous variable depends on the exact question being asked and must be considered carefully.

11. Seasonality may also contribute to short-term variability, even after correcting for seasonal exogenous variables such as flow. The data set is too small at present to correct for seasonality using a dummy variable approach. At certain stations, data may be sufficient for trend tests using second-quarter data only, thus averting the issue of seasonality.

REFERENCES

Anselin, L. 1988. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis 20:1–17.

Anselin, L. 1994. SpaceStat Version 1.50: revision notes. Research Paper 9428. Regional Research Institute, West Virginia University, Morgantown, VI.

Anselin, L., A. Bera, R. Florax, and M. Yoon. 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics (in press).

Anselin, L., S. Hudak, and R. Dodson. 1993. Spatial data analysis and GIS: interfacing GIS and econometric software. National Center for Geographic Information and Analysis, Santa Barbara, CA.

Banfield, J. D. and A.E. Raftery. 1992. Model-based Gaussian and non-Gaussian clustering. Biometrics 49:803–22.

Clark, L.A. and D. Pregibon. 1992. Tree-based models. In Statistical models in S (Chambers, J.M. and T.J. Hastie, eds.). Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

Cochran, W.G. 1977. Sampling techniques, 3rd ed. John Wiley & Sons, NY.

Griffith, D.A. 1987. Spatial autocorrelation: a primer. Association of American Geographers, Washington, D.C.

Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20:727–32.

Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18:107–21.

Jassby, A.D., B.E. Cole, and J.E. Cloern. 1996. The design of sampling transects for characterizing water quality in estuaries. Estuarine, Coastal and Shelf Science (submitted).

Legendre, P. 1987. Constrained clustering. In Developments in numerical ecology. (Legendre, P. and L. Legendre, eds.). Springer-Verlag, Berlin.

Legendre, P. and M. Troussellier. 1988. Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. Limnology and Oceanography 33:1055–67.

Lettenmaier, D.P. 1988. Multivariate nonparametric tests for trend in water quality. Water Resources Bulletin 24:505–12.

MacKinnon, J. and H. White, H. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics 29:305–25.

Mantel, N.A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–20.

Oliver, M.A. and R. Webster. 1991. How geostatistics can help you. Soil Use and Management 7:206–17.

Powell, T.M., J.E. Cloern, and R.A. Walters. 1986. Phytoplankton spatial distribution in South San Francisco Bay: mesoscale and small-scale variability. In Estuarine variability (Wolfe, D.A., ed.). Academic Press, London.

Scott, A.J. and M.J. Symons. 1971. Clustering methods based on likelihood ratio criteria. Biometrics 27:387–97.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35:627–32.

Snedecor, G.W. and W.G. Cochran. 1967. Statistical methods. Iowa State University Press, Ames, Iowa.

Statistical Sciences. 1995. S-PLUS, Version 3.3 for Windows. Statistical Sciences, Seattle, WA.

van Belle, G. and J.P. Hughes. 1984. Nonparametric tests for trend in water quality. Water Resources Research 20:127–36.

APPENDIX A

TABLES

Table 1. RMP and pilot study sampling stations.

Station	Code	Lat			Long		
		deg	min	sec	deg	min	sec
Alameda	BB70	37	44	50	122	19	24
Alcatraz	10-AZ	NA	NA	NA	NA	NA	NA
Alcatraz	11-AZ	NA	NA	NA	NA	NA	NA
Angel I.	12-AI	NA	NA	NA	NA	NA	NA
Angel I. / Treasure I.	09-AI	NA	NA	NA	NA	NA	NA
Benicia Br.	19-BB	NA	NA	NA	NA	NA	NA
Benicia Br.	BE10	NA	NA	NA	NA	NA	NA
Berkeley Flats	15-BF	NA	NA	NA	NA	NA	NA
Berkeley Flats	BC40	NA	NA	NA	NA	NA	NA
Berkeley Flats	09-AI	NA	NA	NA	NA	NA	NA
Chipps I., nr Buoy #20	25-CI	NA	NA	NA	NA	NA	NA
Coyote Creek	BA10	37	28	11	122	3	50
Davis Pt.	BD40	38	3	7	122	16	37
Dumbarton Br.	02-DB	NA	NA	NA	NA	NA	NA
Dumbarton Br.	BA30	37	30	54	122	8	7
Extreme South Bay	01-xsb	NA	NA	NA	NA	NA	NA
Golden Gate	10-GG	NA	NA	NA	NA	NA	NA
Golden Gate	11-GG	NA	NA	NA	NA	NA	NA
Golden Gate	BC20	37	48	13	122	30	23
Grizzly Bay	21-GB	NA	NA	NA	NA	NA	NA
Grizzly Bay	BF20	38	6	58	122	2	19
Hayward Flats	05-HF	NA	NA	NA	NA	NA	NA
Honker Bay	23-HB	NA	NA	NA	NA	NA	NA
Honker Bay	22-HB	NA	NA	NA	NA	NA	NA
Honker Bay	BF40	38	4	2	121	55	56
Hunter's Pt. Channel	08-HP	NA	NA	NA	NA	NA	NA
Hunter's Pt.	BB40	NA	NA	NA	NA	NA	NA
Napa R.	BD50	38	5	47	122	15	37
New York Slough	BG10	NA	NA	NA	NA	NA	NA
New York Slough	27-NYS	NA	NA	NA	NA	NA	NA
New York Slough	26-NYS	NA	NA	NA	NA	NA	NA
Oyster Pt.	BB30	37	40	12	122	19	45
Pacheco Creek	20-PCK	NA	NA	NA	NA	NA	NA
Pacheco Creek	BF10	38	3	5	122	5	48
Petaluma R.	BD15	38	6	37	122	29	13
Petaluma R.	16-PR	NA	NA	NA	NA	NA	NA
Pinole Pt.	BD30	38	1	29	122	21	39
Pinole Shoal Channel	17-PSC	NA	NA	NA	NA	NA	NA
Pinole Shoal Nearshore	18-PSN	NA	NA	NA	NA	NA	NA
Pt. Isabel	BC41	37	53	2	122	20	33
Port Chicago	22-PTC	NA	NA	NA	NA	NA	NA
Port Chicago	23-PTC	NA	NA	NA	NA	NA	NA
Port Chicago	BF30	NA	NA	NA	NA	NA	NA
Red Rock	BC60	37	55	0	122	26	0
Redwood Creek	03-RC	NA	NA	NA	NA	NA	NA
Redwood Creek	BA40	37	33	40	122	12	34
Richardson Bay	BC30	37	51	49	122	28	40
Sacramento R.	26-SR	NA	NA	NA	NA	NA	NA
Sacramento R.	27-SR	NA	NA	NA	NA	NA	NA
Sacramento R.	BG20	38	3	34	121	48	35

Location	Code						
SF Airport	BB20	NA	NA	NA	NA	NA	NA
SF Airport (SF W. Flats)	06-SFO	NA	NA	NA	NA	NA	NA
San Bruno Shoal	04-SBS	NA	NA	NA	NA	NA	NA
San Bruno Shoal	BB15	37	37	1	122	17	0
San Joaquin R.	BG30	38	1	24	121	48	27
San Jose	C-3-0	37	27	43	121	58	32
San Leandro Channel	07-SLC	NA	NA	NA	NA	NA	NA
San Leandro Channel	7SLC	NA	NA	NA	NA	NA	NA
San Pablo Bay	BD20	38	2	55	122	25	11
San Pedro Pt.	BD10	NA	NA	NA	NA	NA	NA
San Pedro Pt.	14-SPP	NA	NA	NA	NA	NA	NA
San Pedro Pt.	15-SPP	NA	NA	NA	NA	NA	NA
San Rafael Br. Channel	14-SRBC	NA	NA	NA	NA	NA	NA
San Rafael Br. Channel	12-SRBC	NA	NA	NA	NA	NA	NA
San Rafael Br. Channel	BC65	NA	NA	NA	NA	NA	NA
San Rafael Br. Nrshore	13-SRBN	NA	NA	NA	NA	NA	NA
South Bay	BA20	37	29	41	122	5	20
Stake Pt., nr Buoy #20	24-SP	NA	NA	NA	NA	NA	NA
Stake Pt., nr Buoy #20	BF50	NA	NA	NA	NA	NA	NA
Sunnyvale	C-1-3	37	26	8	122	0	40
Yerba Buena I.	BC10	37	49	22	122	20	58

Table 2. RMP and pilot sampling events.

Event	Start	Finish
1	2 Apr 1989	2 Apr 1989
2	9 Aug 1989	11 Aug 1989
3	2 Dec 1989	2 Dec 1989
4	15 Jun 1990	15 Jun 1990
5	19 Sep 1990	21 Sep 1990
6	12 Jun 1991	14 Jun 1991
7	8 Apr 1992	11 Apr 1992
8	3 Mar 1993	6 Mar 1993
9	25 May 1993	28 May 1993
10	14 Sep 1993	17 Sep 1993
11	31 Jan 1994	9 Feb 1994
12	19 Apr 1994	29 Apr 1994
13	16 Aug 1994	25 Aug 1994
14	7 Feb 1995	16 Feb 1995
15	19 Apr 1995	28 Apr 1995

Table 3. Trace element data available from sampling events 7–15 for model-based clustering of sampling sites .

Event	Ag	As	Cd	Cr	Cu	Fe	Hg	Ni	Pb	Se	Zn
7	]		]		]	]	]	]	]		]
8	]	]	]	]	]		]	]	]	]	]
9	]	]	]	]	]		]	]	]	]	]
10	]	]	]	]			]	]		]	]
11		]	]	]			]	]	]	]	]
12	]	]	]	]	]		]	]	]	]	]
13	]	]	]	]	]		]	]	]	]	]
14	]	]	]	]	]		]	]	]	]	]
15	]	]	]	]	]		]	]	]	]	]

Table 4. Definition of a stratification scheme for the MIDAS transects in San Francisco Bay. Locations are specified in terms of UTM coordinates.

Stratum No.	Description	Size ±SD (km)	Northing (km)	Easting (km)
1	south of Dumbarton Br.	6.9 ±0.3	<151.4	
2	Dumbarton Br. to San Bruno Shoal	23.3 ±0.6	151.4–165.3	
3	San Bruno Shoal to Angel I.	28.7 ±0.7	165.3–188.8	
4	Angel I. to Mare I.	37.3 ±2.2	≥188.8	<564.6
5	Mare I. to Martinez	13.1 ±1.1	≥209.3	564.6–574.5
6	east of Martinez	51.8 ±1.7	≥209.3	≥574.5

Table 5. Sample SpaceStat output from an OLS ANOVA for ln-transformed chromium (event 13).

```
ORDINARY LEAST     SQUARES ESTIMATION
DATA SET EVENT13
DEPENDENT VARIABLE LNCRT OBS 24 VARS 3 DF 21

R2      0.6853 R2-adj  0.6553
LIK     -22.4857       AIC     50.9714        SC      54.5055
RSS     9.15211        F-test  50.7176        Prob    8.56E-10

SIG-SQ          0.435815        (0.660163) SIG-SQ(ML)      0.381338 (0.617526)

VARIABLE COEFF      S.D.    t-value Prob
REG_1 2.16004       0.26951         8.014686        0.0000
REG_2 0.106689      0.233403        0.457103        0.6523
REG_3 1.95511       0.208762        9.36529         0.0000

COEFFICIENT VARIANCE MATRIX
REG_1
0.072636        0       0
REG_2
0       0.054477        0
REG_3
0       0       0.043582

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 1

TEST ON     NORMALITY OF ERRORS
TEST        DF      VALUE PROB
Kiefer-Salmon 2     1.086929        0.580733

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHTS MATRIX INVD2STD (row-standardized weights)
TEST                MI/DF  VALUE PROB
Moran's        I        (error)  0.129686       2.0869 0.036894
Lagrange       Multiplier     (error)  1      0.8034 0.370092
Robust LM      (error)  1      0.2355 0.627473
Kelejian-Robinson       (error) 3       2.155585        0.5408
Lagrange       Multiplier     (lag)   1      1.4294 0.231868
Robust LM      (lag)   1      0.8615 0.353315
Lagrange       Multiplier     (SARMA)         2       1.6649 0.434989
```

Table 6. Summary of spatial ANOVA diagnostics.

Element	Event	Non-normal?	Non-normal after ln transform?	Heteroske-dasticity?	Spatial dependence? [1]
Ag	12	y	n	n	n
	13	y	n	n	LM-SARMA, LM-LAG, LM-ERR
	14	y	n	n	n
As	12	y	y	n	n
	13	n	-	n	LMR-LAG
	14	n	-	n	n
Cd	12	n	-	n	n
	13	n	-	n	LM-LAG, LM-SARMA, LM-ERR
	14	n	-	n	n
Cr	12	y	n	n	n
	13	y	n	n	n
	14	y	n	n	n
Cu	12	y	n	n	n
	13	y	n	n	n
	14	y	n	n	n
Hg	12	y	n	n	n
	13	n	-	y	n
	14	y	n	n	n
Ni	12	y	n	n	n
	13	y	n	n	n
	14	y	n	n	n
Pb	12	y	n	n	n
	13	y	n	n	LM-LAG
	14	y	n	n	n
Se	12	n	-	y	LMR-LAG
	13	y	n	n	LMR-ERR, LMR-LAG
	14	y	n	y	LMR-LAG
Zn	12	y	n	n	n
	13	y	n	n	LM-LAG
	14	n	-	y	LM-LAG

[1]LMR refers to the robust form of the LM test.

Table 7. Sample SpaceStat output from a spatial lag ANOVA for lead (event 13).

SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
DATA SET   EVENT13   SPATIAL WEIGHTS MATRIX INVD2STD
DEPENDENT VARIABLE LNPBT   OBS 24   VARS 4   DF 20
R2   0.669 Sq. Corr.   0.7275
LIK   -20.6559   AIC   49.3117   SC   54.024
SIG-SQ   0.29924 (0.547028)

VARIABLE	COEFF	S.D.	z-value	Prob
W_LNPBT	0.549267	0.183552	2.992433	0.002768
REG_1	0.3657	0.238328	1.53444	0.124922
REG_2	-0.82209	0.252982	-3.24961	0.001156
REG_3	0.295845	0.177063	1.670851	0.094751

COEFFICIENT VARIANCE MATRIX
W_LNPBT

0.033691	-0.01528	0.029934	-0.00693	-0.00384

REG_1

-0.01528	0.0568	-0.01357	0.003144	0.001743

REG_2

0.029934	-0.01357	0.064	-0.00616	-0.00342

REG_3

-0.00693	0.003144	-0.00616	0.031351	0.000791

SIGMA-SQ

-0.00384	0.001743	-0.00342	0.000791	0.007901

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHTS MATRIX INVD2STD (row-standardized weights)

TEST	DF	VALUE	PROB	
Likelihood	Ratio Test	1	4.688028	0.030373

LAGRANGE MULTIPLIER TEST ON SPATIAL ERROR DEPENDENCE

WEIGHT	STAND	ZERO	DF	VALUE	PROB
INVD2STD	yes	no	1	0.00163	0.967796

Table 8. Summary of spatial ANOVA results.

Element	Event	Model	Subregions Different?	Pairwise Differences
Ag	12	OLS	y	NB>CB, SB>CB
	13	-[1]	-	-
	14	OLS	y	NB>CB, SB>CB
As	12	OLS	-[2]	-
	13	LAG	y	NB>CB, SB>CB
	14	OLS	n	NB>CB
Cd	12	OLS	n	none
	13	-[1]	-	-
	14	OLS	y	SB>CB, SB>NB
Cr	12	OLS	y	NB>CB, SB>CB
	13	OLS	y	NB>CB, SB>CB
	14	OLS	y	NB>CB, SB>CB, NB>SB
Cu	12	OLS	y	NB>CB, SB>CB
	13	OLS	y	NB>CB, SB>CB
	14	OLS	y	NB>CB, SB>CB
Hg	12	OLS	y	NB>CB, SB>CB, NB>SB
	13	ROBUST	y	NB>CB, SB>CB
	14	OLS	y	NB>CB, SB>CB
Ni	12	OLS	y	NB>CB, SB>CB
	13	OLS	y	NB>CB, SB>CB
	14	OLS	y	NB>CB, SB>CB
Pb	12	OLS	y	NB>CB, SB>CB
	13	LAG	y	NB>CB, SB>CB
	14	OLS	y	NB>CB, SB>CB
Se	12	LAG	y	SB>NB
	13	-[1]	-	-
	14	LAG	-[3]	-
Zn	12	OLS	y	NB>CB, SB>CB
	13	LAG	y	NB>CB, SB>CB
	14	LAG	-[3]	-

[1]Diagnostics suggested higher-order process (Table).
[2]Non-normal.
[3]Heteroskedastic.

Table 9. Stations identified from Moran scatterplots as most important positive anomaly for each trace element during sampling event 13.

Trace Element (total)	Stations
Silver	San Jose
Arsenic	Petaluma R.
Cadmium	Petaluma R.
Chromium	San Jose
Copper	San Jose
Mercury	Dumbarton Br.
Nickel	San Jose
Lead	San Jose
Selenium	Sacramento R.
Zinc	San Jose

Table 10. Mantel tests of association between trace element totals, TSS and spatial position for sampling event 12. TE, COG and SPACE refer to the respective distance matrices for the trace element, TSS and spatial position. X·Y denotes the Mantel statistic for X and Y. (X·Y)·Z denotes the partial Mantel statistic for X and Y given Z. Statistics significant at the p=0.05 level are in bold.

Association	Ag	As	Cd	Cr	Cu	Hg	Ni	Pb	Se	Zn
TE·COG	**0.71**	**0.79**	**0.34**	**0.98**	**0.93**	**0.97**	**0.91**	**0.86**	-0.12	**0.88**
TE·SPACE	**0.15**	0.08	**0.33**	**0.15**	0.12	**0.16**	0.12	**0.14**	**0.59**	**0.13**
COG·SPACE	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**	**0.14**
(TE·COG)·SPACE	0.06	0.06	-0.39	**0.72**	0.41	0.56	0.34	0.41	-0.54	0.31
(COG·SPACE)·TE	-0.31	0.12	-0.18	-0.3	0.01	-0.66	-0.11	0.11	0.13	-0.26
(TE·SPACE)·COG	-0.09	-0.25	0.25	-0.49	-0.37	-0.4	-0.32	-0.18	0.61	-0.24

Table 11. Possible models of causal relationships among space, a cognate and a trace element, assuming that the trace element does not determine cognate distribution.

1) SPACE   COG
    ↓   ↙
    TE

2) SPACE → COG
      ↙
   TE

3) SPACE → COG
    ↓   ↙
    TE

4) SPACE → COG
     ↓
    TE

5) SPACE → COG

    TE

6) SPACE   COG
     ↓
    TE

7) SPACE   COG
      ↙
   TE

8) SPACE   COG

    TE
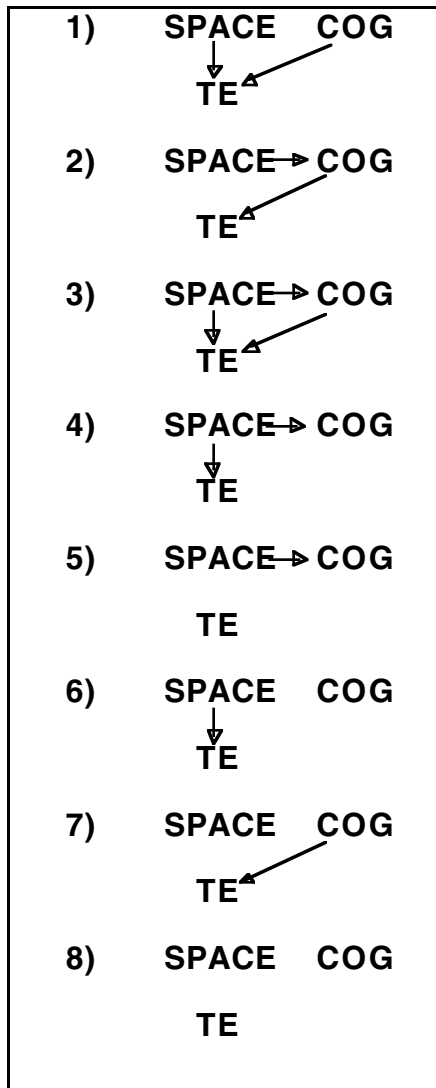
Table 12. Value of Kendall's tau for Mann-Kendall tests of trend during the period 1991–1995. Values in bold are significant (p<0.05).

Code	Ag	As	Cd	Cr	Cu	Hg	Ni	Pb	Se	Zn
BA10	-0.4	0.2	-0.2	0.2	0.4	0.2	0	0	0.4	-0.2
BA20	-0.2	0.22	-0.02	0.43	0.2	-0.16	0.02	-0.38	0.07	-0.16
BA30	0.24	0.22	0.02	0.29	0.51	0.47	0.56	0.02	0.54	0.47
BA40	-0.07	0.22	-0.11	0.5	0.02	0.38	0.07	-0.11	0	-0.02
BB15	-0.6	-0.4	-0.4	-0.8	-0.4	-0.2	-1	-0.6	0	-0.6
BB30	-0.11	0.22	0.02	-0.14	-0.02	-0.02	-0.07	-0.38	0.21	-0.07
BB70	-0.6	-1	-0.4	-0.6	-0.4	-0.4	-0.6	-0.6	-0.8	-0.6
BC10	-0.33	0.11	0.02	0.21	-0.02	-0.2	0.02	-0.24	-0.14	0.11
BC20	-0.42	0.06	-0.02	0.14	-0.22	**-0.69**	-0.29	-0.42	-0.36	-0.27
BC30	-0.38	0.06	-0.07	-0.14	-0.29	-0.29	0.02	-0.2	-0.11	-0.42
BC41	-0.29	-0.29	-0.07	-0.07	0.36	-0.21	0	-0.43	-0.29	-0.07
BC60	0.33	-0.4	-0.2	0.2	0.33	0.73	0.6	0.33	-0.6	0.33
BD15	0.4	0	-0.2	0.4	0.4	0.4	0.6	0.4	0	0.4
BD20	0.2	-0.07	0.07	0.14	0.24	0.2	0.29	-0.02	0.29	0.2
BD30	0.33	-0.07	-0.06	0.14	0.28	0.33	0.39	-0.22	-0.29	0.39
BD40	0.47	0.5	0.02	0.57	0.38	**0.64**	**0.64**	0.16	0.21	0.47
BD50	0.16	0.19	0.02	0.64	0.29	0.42	0.33	0	-0.29	0.2
BF10	-0.02	-0.33	-0.2	-0.21	0.02	-0.02	0.02	-0.16	0	0.09
BF20	-0.11	-0.22	-0.16	0.14	0.11	0.38	0.33	-0.11	-0.14	0.38
BF40	-0.05	-0.2	-0.43	0.4	-0.05	0.14	0.33	-0.05	-0.6	0.24
BG20	-0.28	0.07	-0.36	-0.07	-0.06	-0.02	0.07	-0.24	-0.5	-0.2
BG30	-0.38	0.07	-0.53	-0.29	-0.51	-0.33	-0.38	-0.38	-0.17	-0.56
C-1-3	-0.4	0	-0.4	0.2	0.4	0.4	0.2	0	0.8	-0.2
C-3-0	0.2	0.2	-0.2	0.6	0.4	0.6	0.4	0.2	0.4	0

Table 13. "Spatial Kendall test" statistics resulting from treating stations as seasons and sampling events as years. Stations were subdivided into four subgroups as suggested by the individual trend tests for copper (Figure 7).

Subregion	S_k	σ_{Sk}	Large-sample *p*
"SB"	45	79.4	0.580
"CB"	-31	91.5	0.743
"SP & east SU"	76	98.6	0.447
"rivers & west SU"	-26	28.9	0.387

Table 14. Significance (*p*-value) of slope for linear regression of trace element total on Net Delta Outflow.

Code	Ag	As	Cd	Cr	Cu	Hg	Ni	Pb	Se	Zn
BA10	0.27	0.239	**0.024**	0.848	0.948	0.551	0.74	0.794	**0.026**	0.929
BA20	0.602	0.34	**0.004**	0.865	0.689	0.767	0.785	0.972	0.053	0.791
BA30	0.803	0.365	**0.004**	0.964	0.835	0.679	0.517	0.851	0.472	0.598
BA40	0.763	0.335	**0.015**	0.599	0.269	0.894	0.681	0.802	0.155	0.99
BB15	**0.048**	0.067	0.06	0.484	**0.034**	0.227	0.389	0.498	0.073	0.621
BB30	0.466	0.271	**0.004**	0.24	0.122	0.322	0.372	0.613	**0.041**	0.307
BB70	0.055	0.094	0.063	0.069	0.269	**0.021**	0.163	**0.003**	0.124	0.076
BC10	0.399	0.242	**0.012**	0.611	0.683	0.389	0.882	0.878	**0.007**	0.437
BC20	0.613	0.087	**0.023**	0.12	0.573	0.407	0.723	0.993	**0.006**	0.523
BC30	0.726	**0.004**	**0.016**	0.284	0.806	0.337	0.981	0.81	0.287	0.404
BC41	0.743	0.053	**0.004**	0.556	0.243	0.623	0.223	0.569	**0.029**	0.42
BC60	0.799	**0.008**	0.054	0.092	0.075	0.149	**0.027**	0.19	**0.025**	0.061
BD15	0.898	0.851	**0.02**	0.472	0.734	0.689	0.305	0.496	0.681	0.812
BD20	0.345	0.224	**0.029**	0.785	0.645	0.744	0.256	0.868	0.87	0.764
BD30	0.4	0.067	**0.009**	0.858	0.965	0.582	0.499	0.932	0.161	0.998
BD40	0.622	0.698	**0.014**	**0.013**	**0.006**	**0.006**	**0.001**	0.565	0.446	**0.026**
BD50	0.852	0.217	**0.01**	0.478	0.736	0.934	0.444	0.624	0.4	0.963
BF10	0.24	**0.02**	**0.012**	0.723	0.773	0.646	0.866	0.945	0.754	0.899
BF20	0.669	**0.027**	**0.021**	0.949	0.661	0.745	0.758	0.843	0.162	0.783
BF40	0.497	0.311	**0.009**	0.882	0.687	0.835	0.553	0.493	0.423	0.983
BG20	0.995	**0.04**	**0**	0.462	0.995	0.981	0.28	0.794	0.327	0.613
BG30	0.971	**0.027**	**0.004**	0.882	0.324	0.432	0.984	0.695	0.503	0.289
C-1-3	0.959	0.359	**0.022**	0.678	0.932	0.843	0.359	0.483	**0.004**	0.684
C-3-0	0.555	0.22	0.07	0.864	0.675	0.99	0.582	0.7	**0.047**	0.964

Table 15. Kendall tau statistics for test of trend in total cadmium corrected for Net Delta Outflow.

Site	N	S	σ_S	tau	p
BA10	5	0	4.082	0	1
BA20	10	15	11.18	0.333	0.21
BA30	10	11	11.18	0.244	0.371
BA40	10	7	11.18	0.156	0.592
BB15	5	0	4.082	0	1
BB30	10	13	11.18	0.289	0.283
BB70	5	-2	4.082	-0.2	0.806
BC10	10	13	11.18	0.289	0.283
BC20	10	9	11.18	0.2	0.474
BC30	10	11	11.18	0.244	0.371
BC41	8	2	8.083	0.071	0.902
BC60	6	1	5.323	0.067	1
BD15	5	2	4.082	0.2	0.806
BD20	10	17	11.18	0.378	0.152
BD30	9	14	9.592	0.389	0.175
BD40	10	29	11.18	0.644	**0.012**
BD50	10	27	11.18	0.6	**0.02**
BF10	10	7	11.18	0.156	0.592
BF20	10	13	11.18	0.289	0.283
BF40	7	3	6.658	0.143	0.764
BG20	10	1	11.18	0.022	1
BG30	10	-15	11.18	-0.333	0.21
C-1-3	5	-4	4.082	-0.4	0.462
C-3-0	5	2	4.082	0.2	0.806

Table 16. Number of sampling events per calendar year quarter.

Year	Q1	Q2	Q3	Q4
1989	0	1	1	1
1990	0	1	1	0
1991	0	1	0	0
1992	0	1	0	0
1993	1	1	1	0
1994	1	1	1	0
1995	1	1		
1996				

APPENDIX B

FIGURES

Figure 1. San Francisco Bay and the Sacramento-San Joaquin Delta, showing the sampling sites for trace elements used in the RMP and prior pilot studies.

San Francisco Bay
Regional Monitoring Program

0   20 km
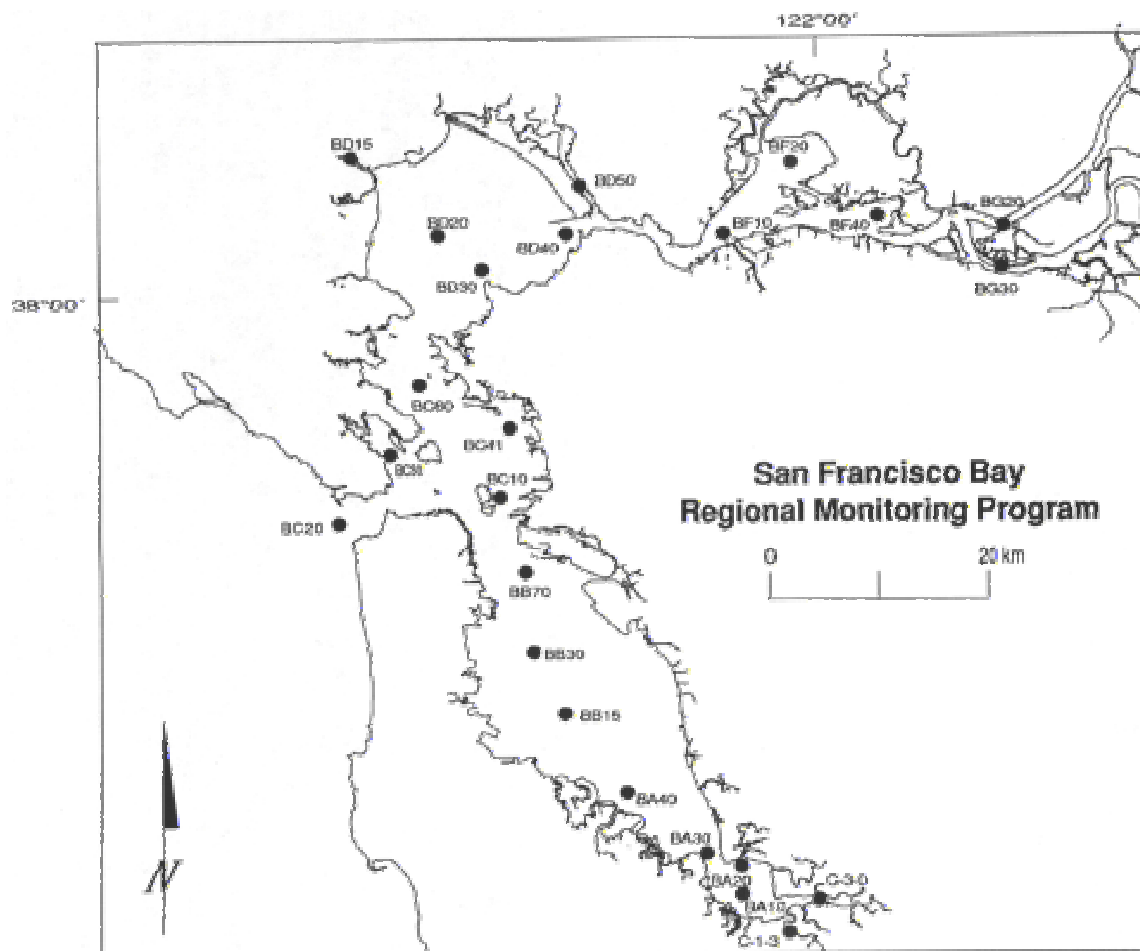
Figure 2. RMP station clusters for each sampling event based on all trace elements.

RMP Station Clusters
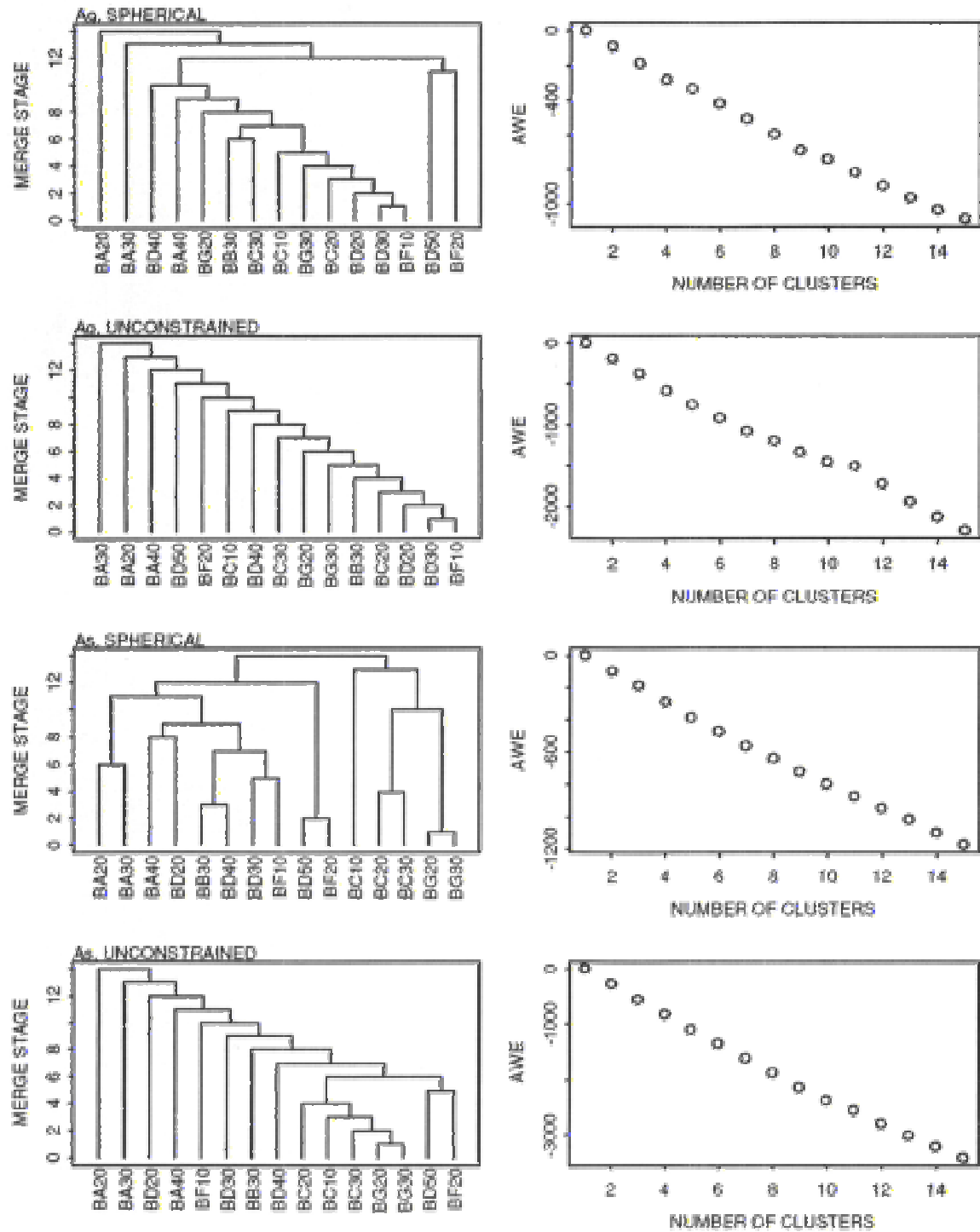
RMP Station Clusters

RMP Station Clusters

Cu, SPHERICAL

Cu, UNCONSTRAINED

Fe, SPHERICAL

Fe, UNCONSTRAINED

RMP Station Clusters

Hg, SPHERICAL

Hg, UNCONSTRAINED

Ni, SPHERICAL

Ni, UNCONSTRAINED
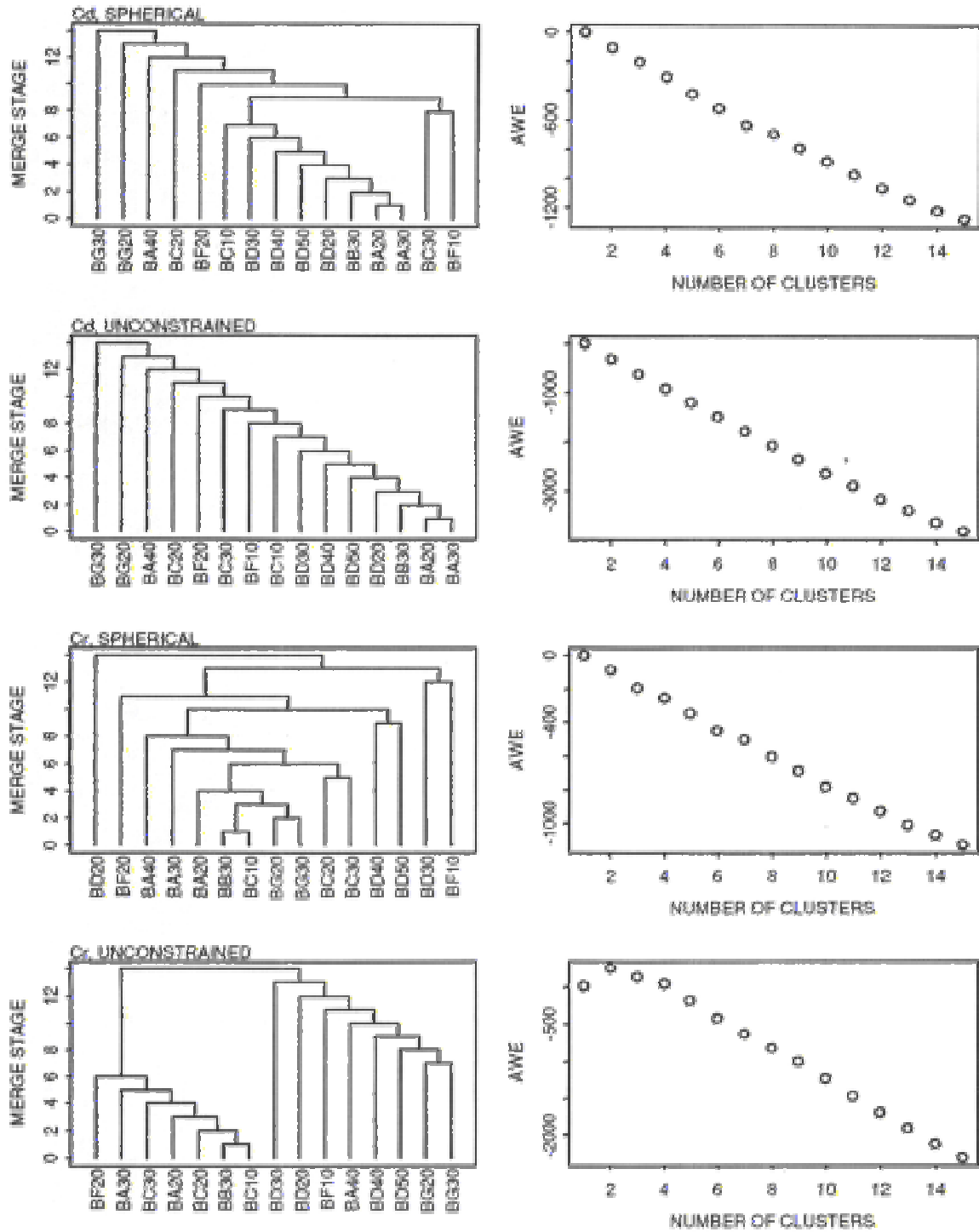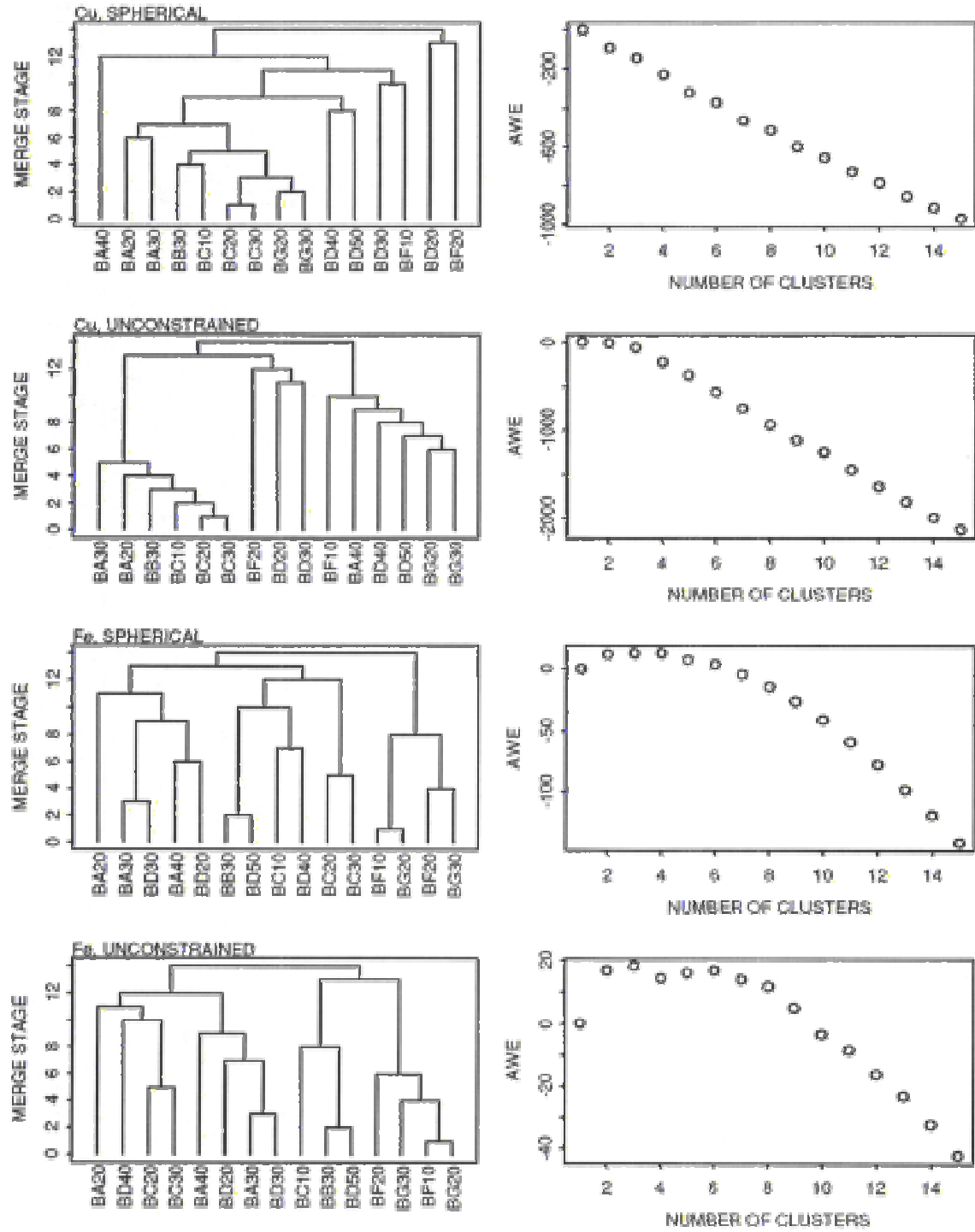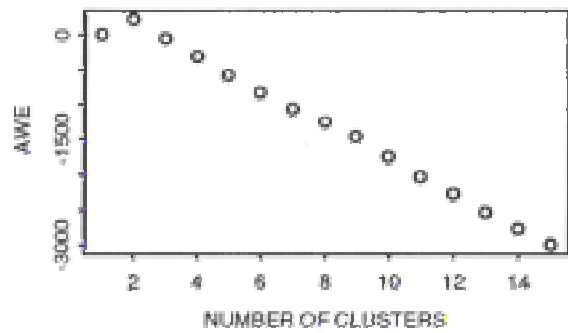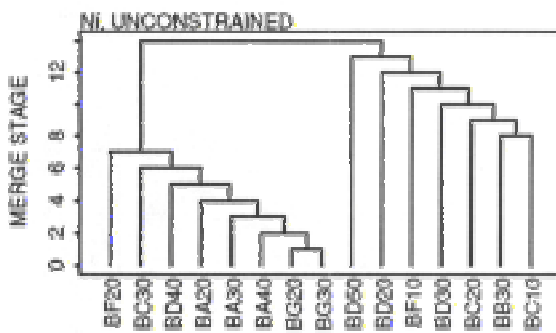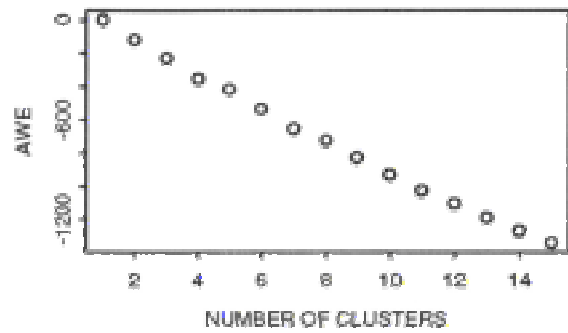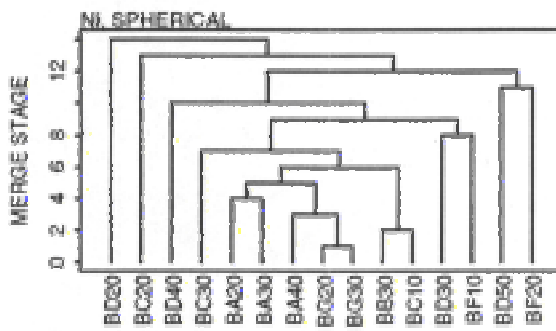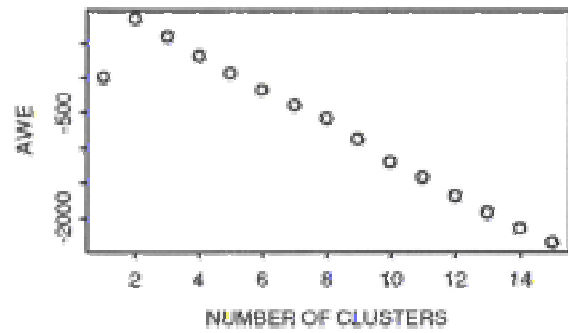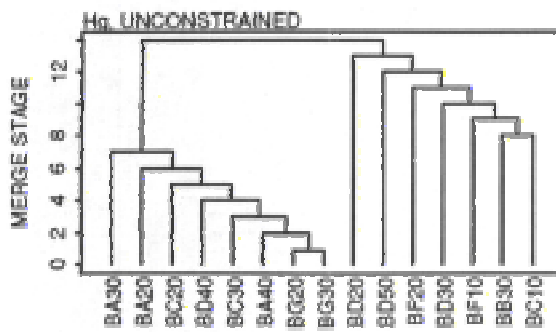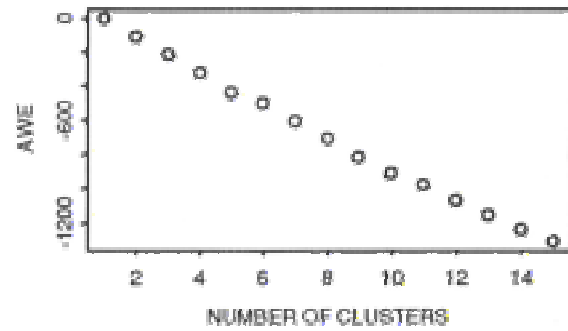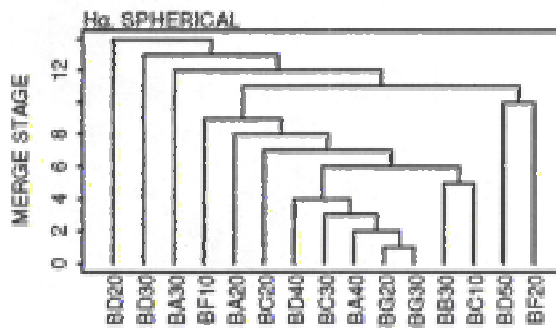
RMP Station Clusters
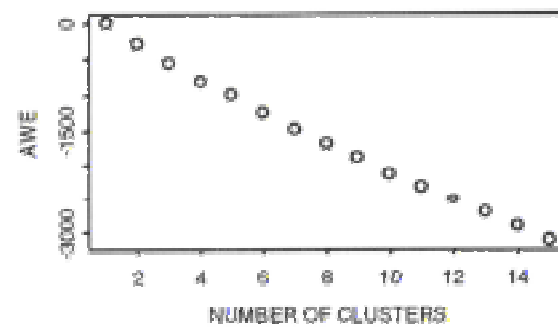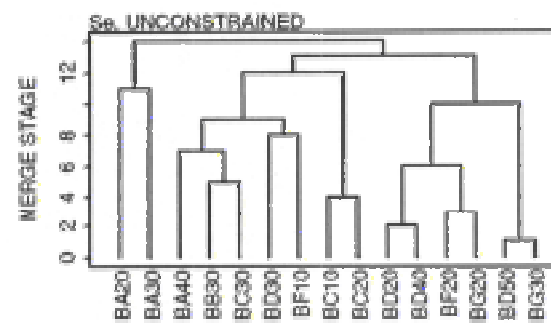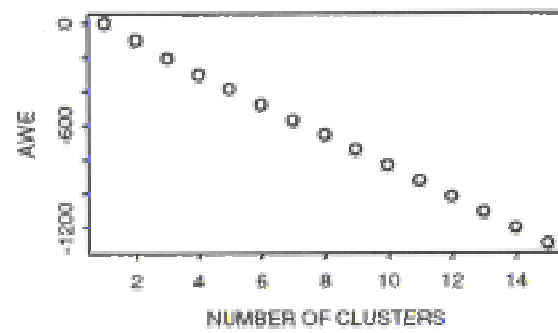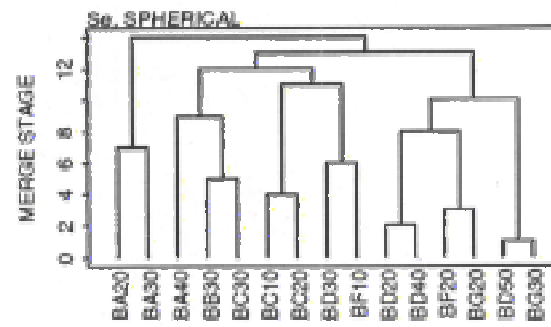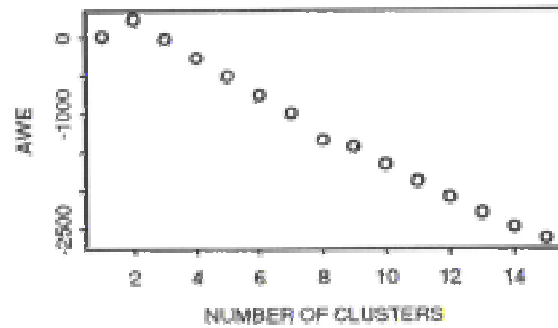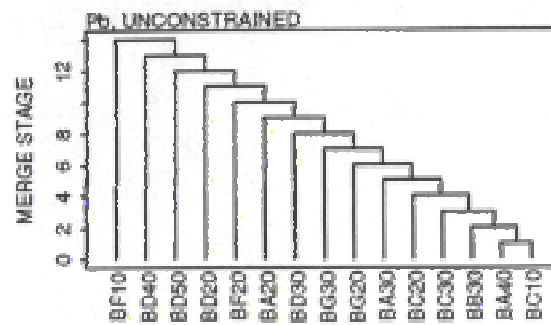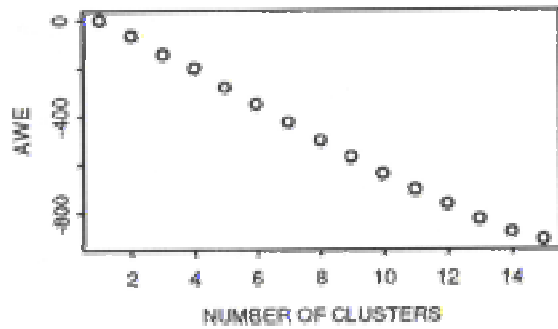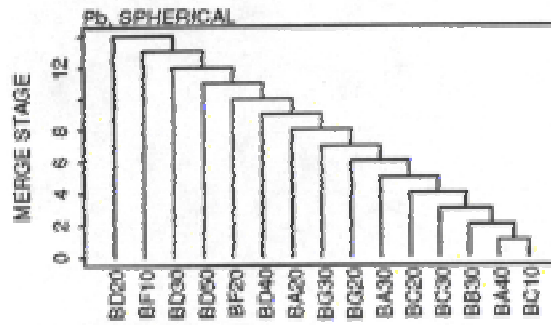
RMP Station Clusters
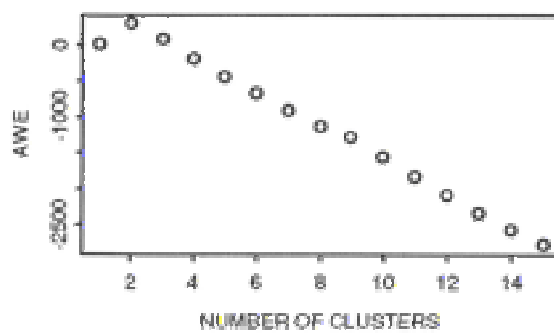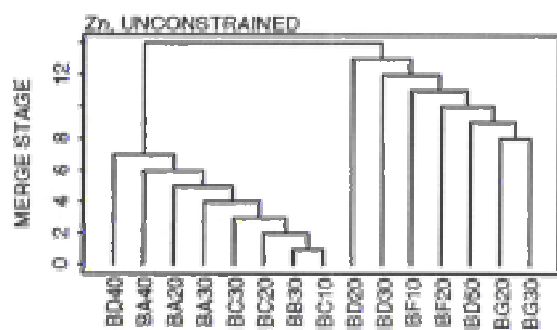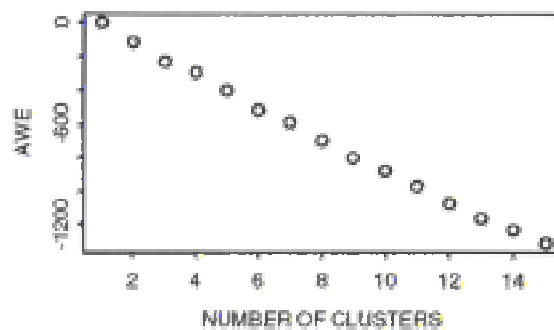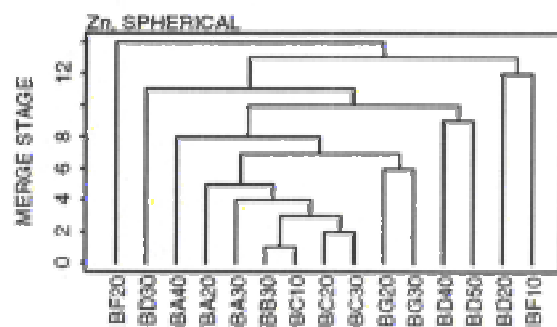
Figure 3. RMP station clusters for each trace element based on all sampling events 7-15.

RMP Station Clusters
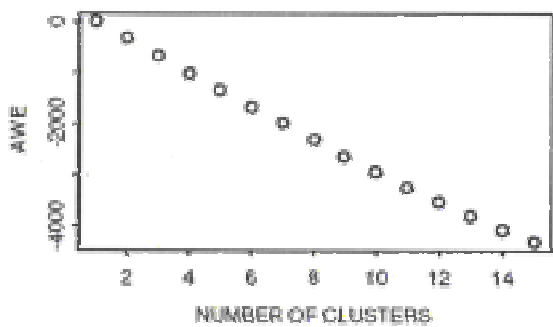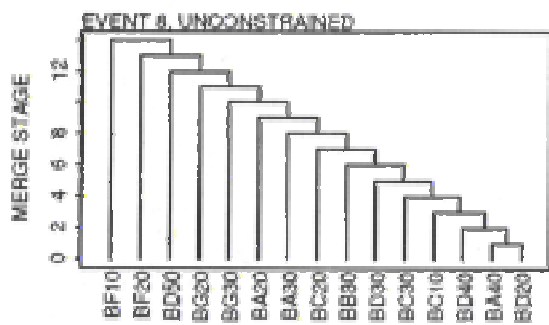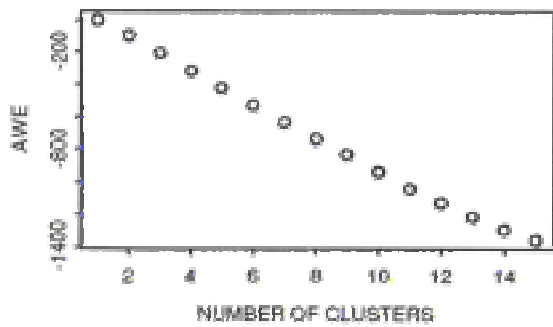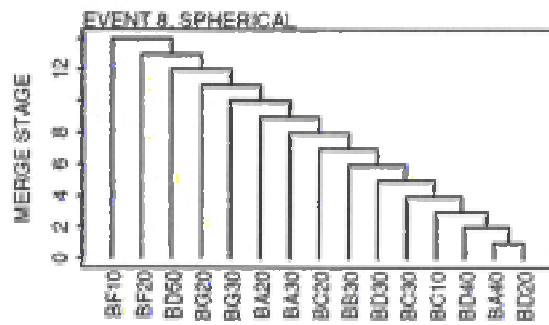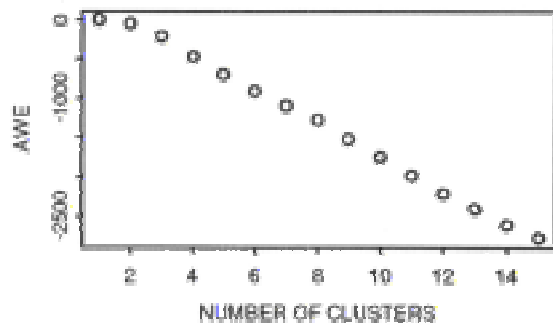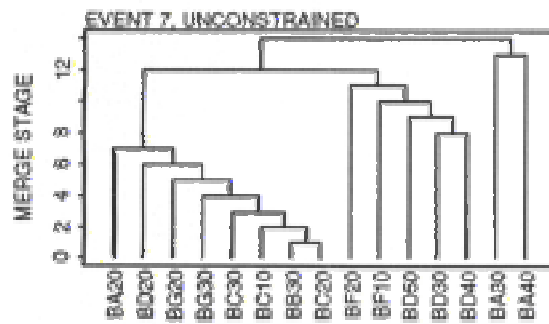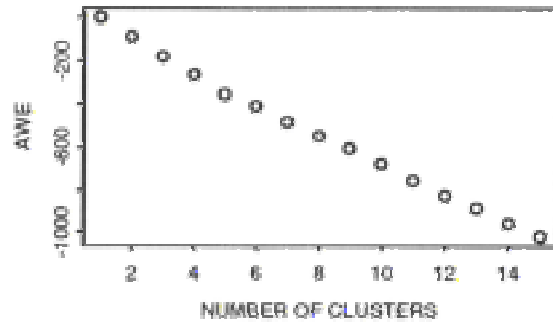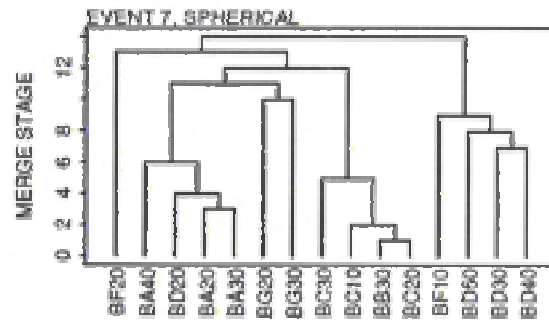
EVENT 7, SPHERICAL

EVENT 7, UNCONSTRAINED

EVENT 8, SPHERICAL

EVENT 8, UNCONSTRAINED

RMP Station Clusters

EVENT 9, SPHERICAL

EVENT 9, UNCONSTRAINED

EVENT 10, SPHERICAL

EVENT 10, UNCONSTRAINED

RMP Station Clusters

EVENT 11, SPHERICAL

EVENT 11, UNCONSTRAINED

EVENT 12, SPHERICAL

EVENT 12, UNCONSTRAINED

RMP Station Clusters

EVENT 13, SPHERICAL

EVENT 13, UNCONSTRAINED

EVENT 14, SPHERICAL

EVENT 14, UNCONSTRAINED

RMP Station Clusters

EVENT 15, SPHERICAL

EVENT 15, UNCONSTRAINED

EVENTS 7-15, SPHERICAL

EVENTS 7-15, UNCONSTRAINED
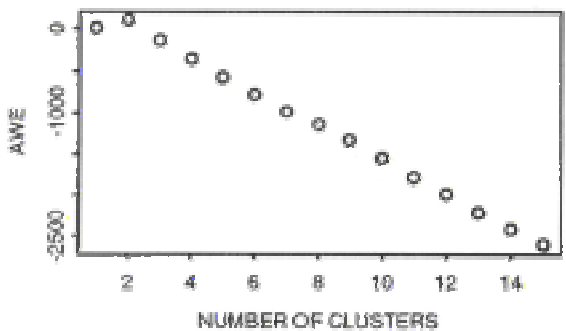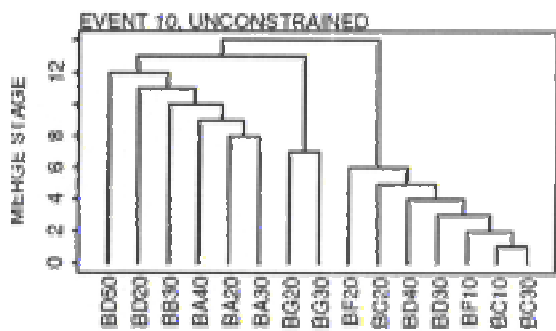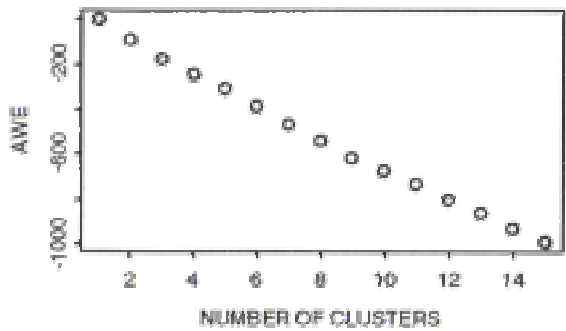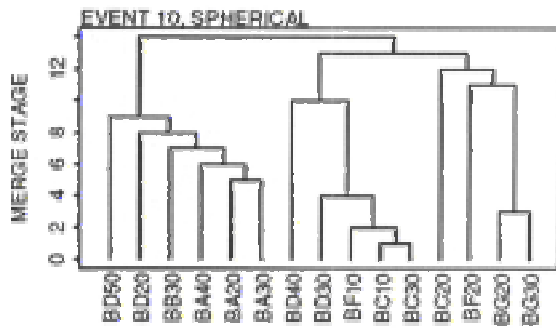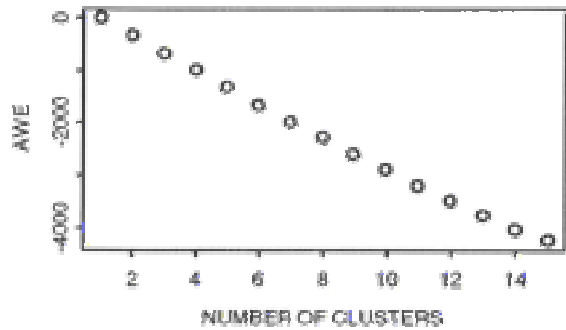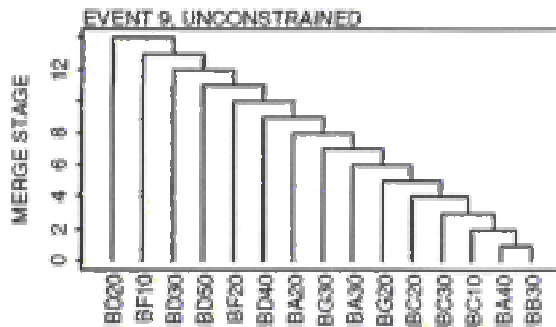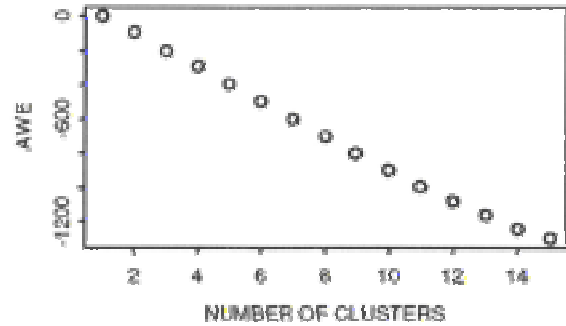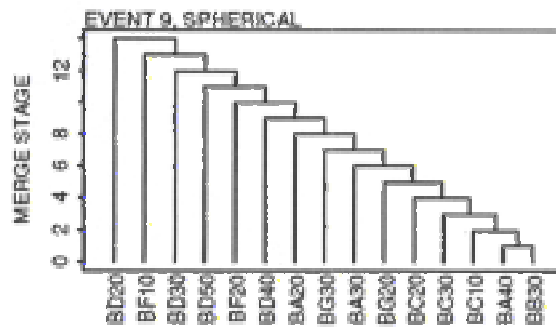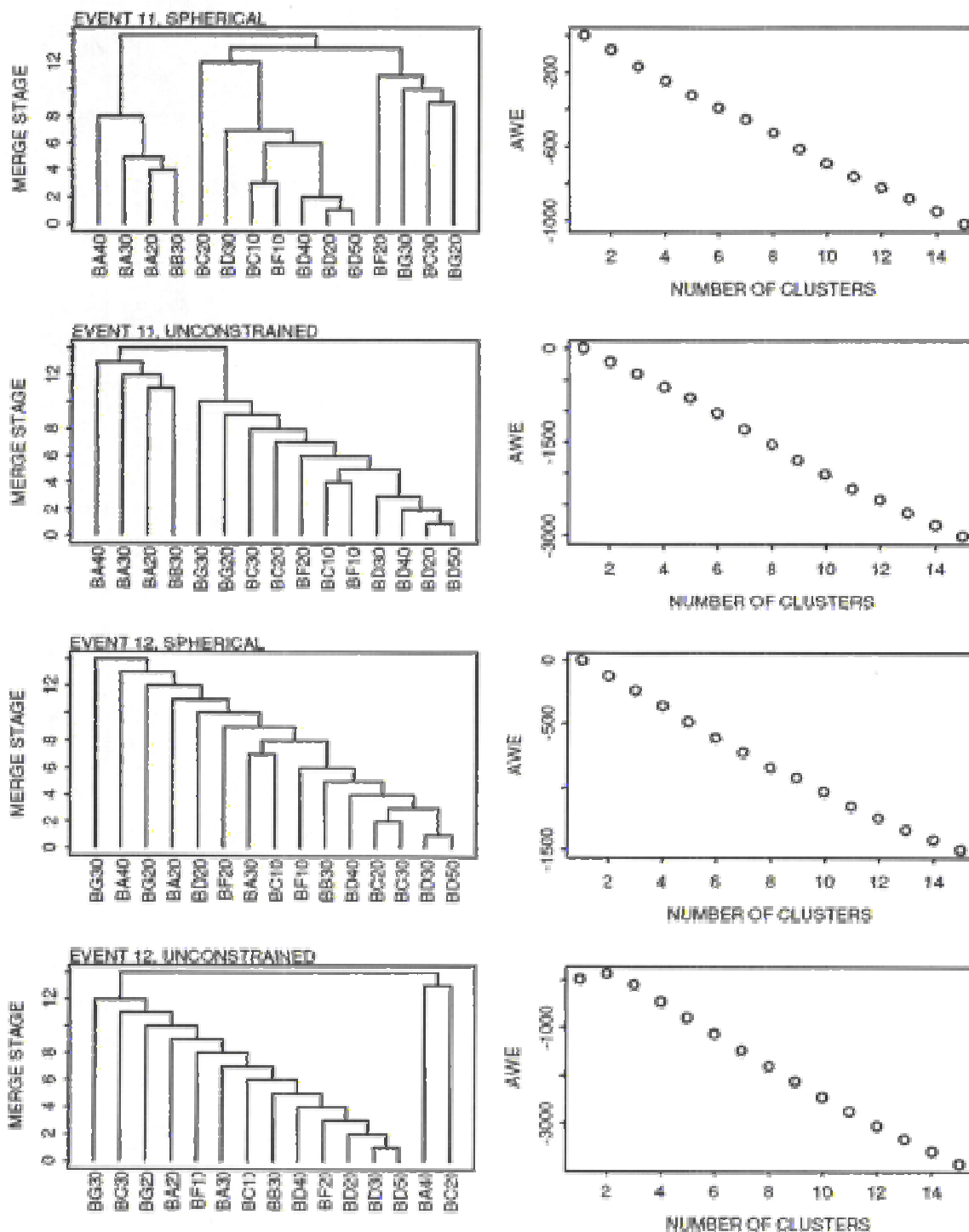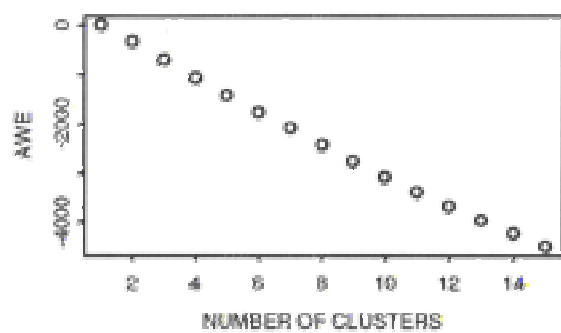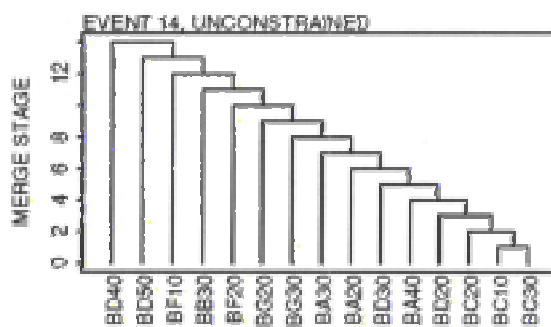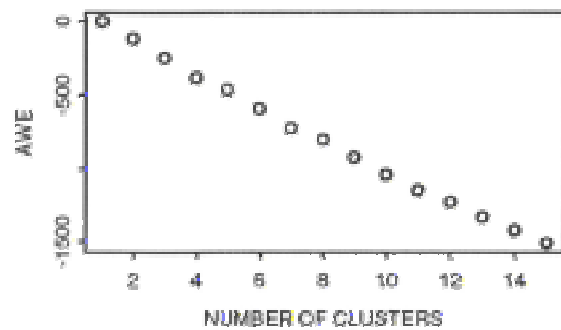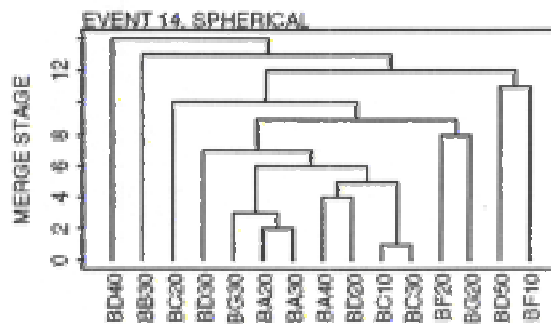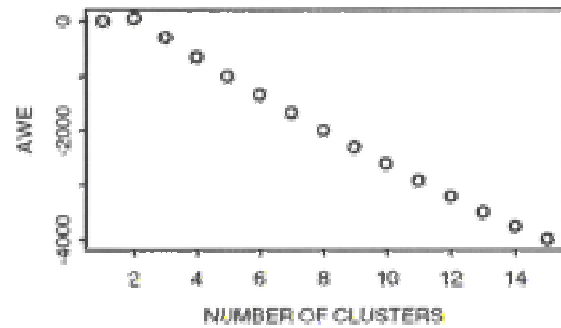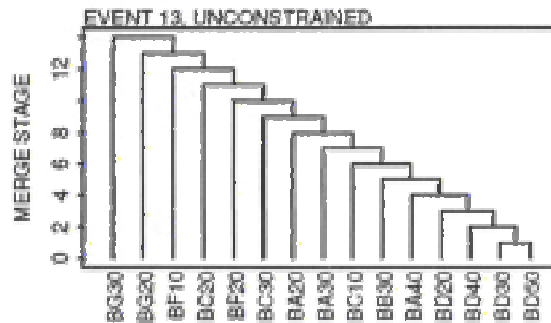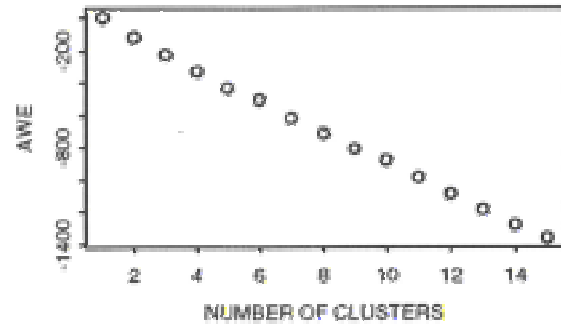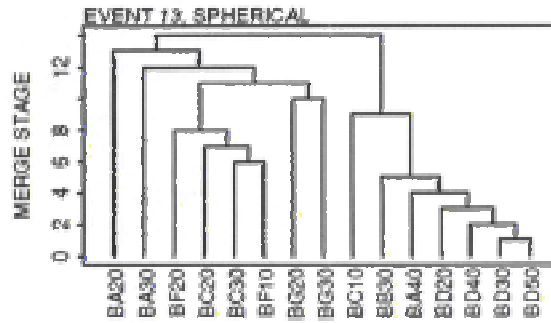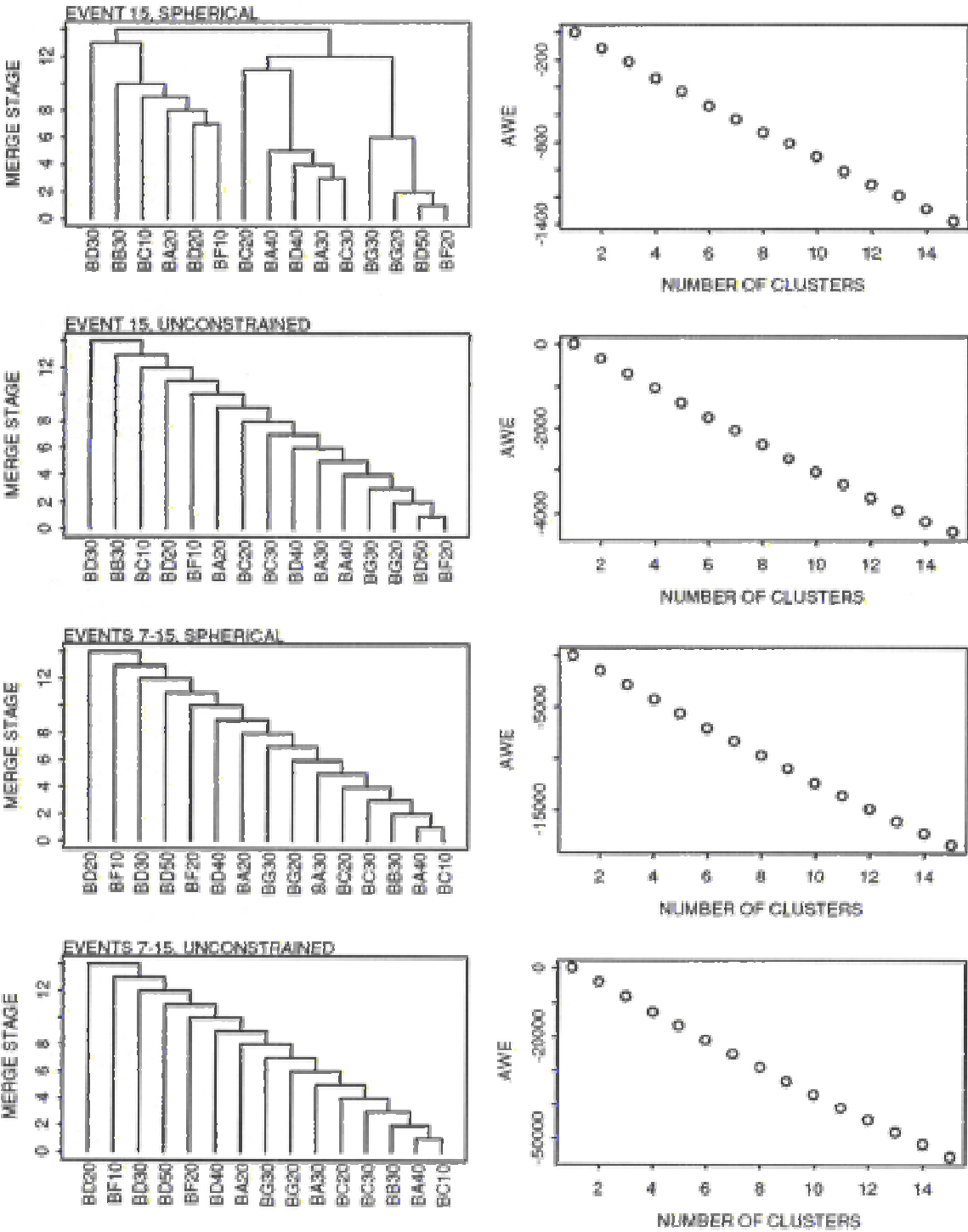
Figure 4. Total trace element distributions mapped in UTM coordinates. Each map corresponds to a single element and single RMP sampling event. The area of each square is proportional to the concentration, expressed as a fraction of the largest value found in each sampling event.

Total Silver
(scaled to sampling event maximum)
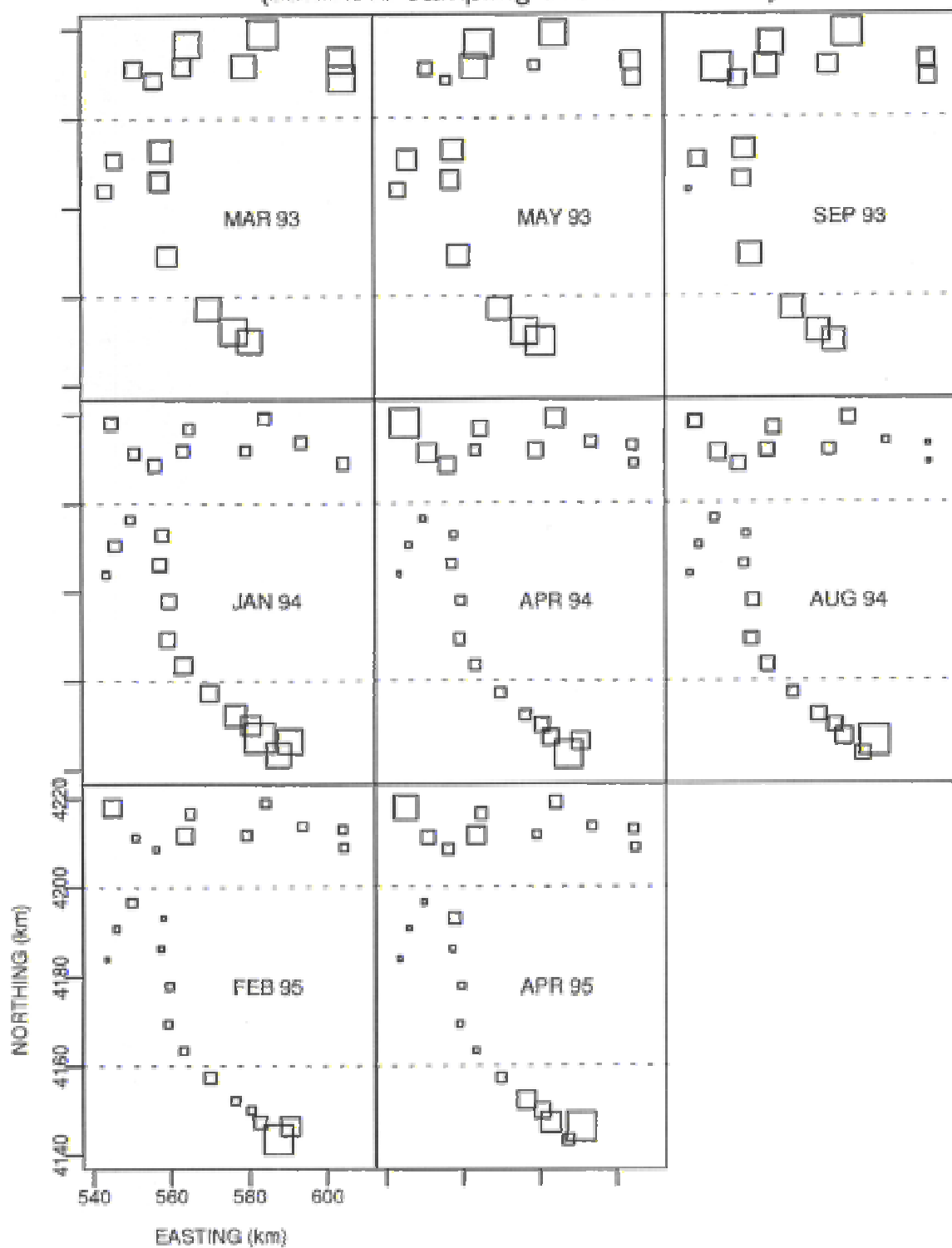
Total Arsenic
(scaled to sampling event maximum)

MAR 93

MAY 93

SEP 93

JAN 94

APR 94

AUG 94

FEB 95

APR 95

NORTHING (km)

4220
4200
4180
4160
4140

540   560   580   600

EASTING (km)
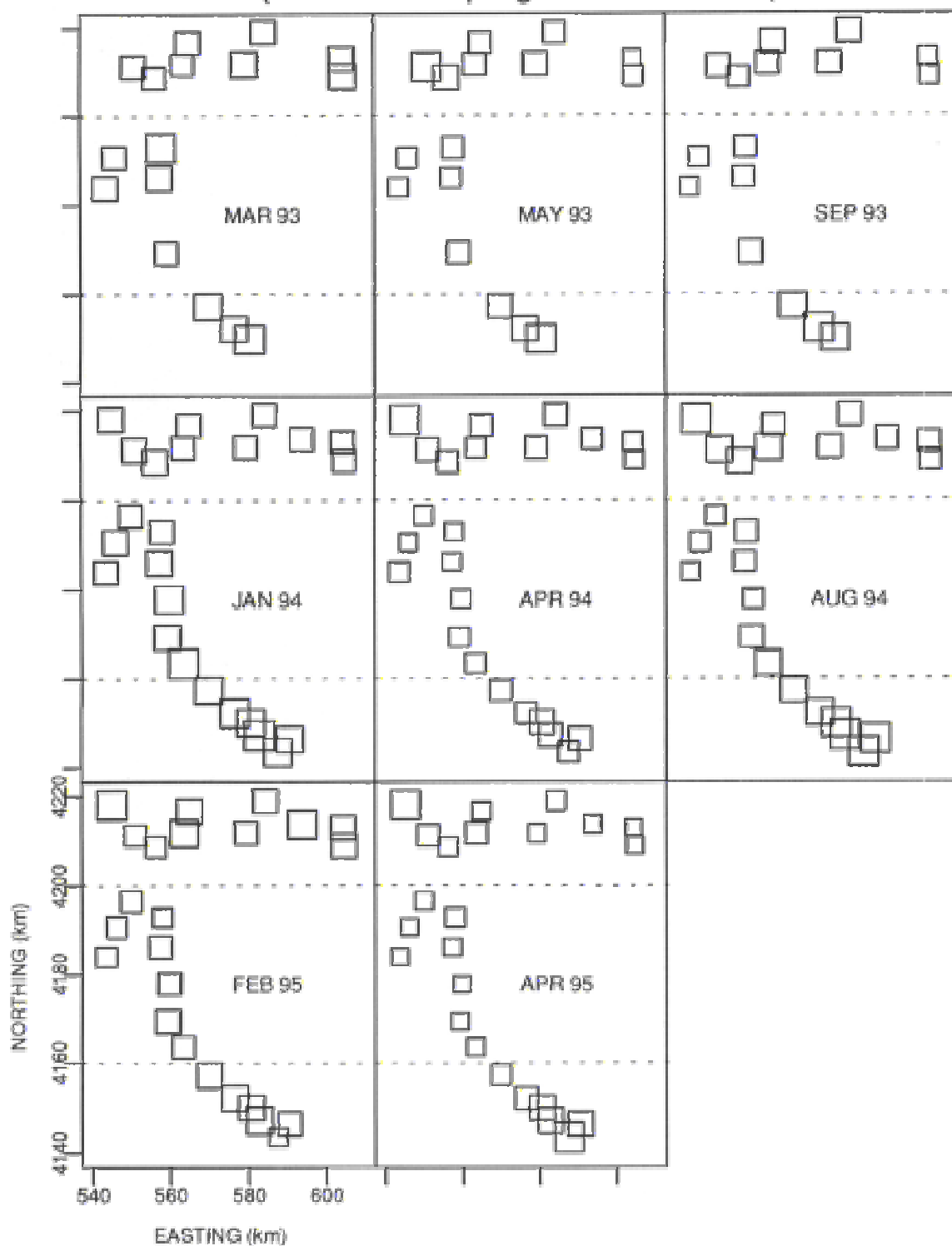
Total Cadmium
(scaled to sampling event maximum)
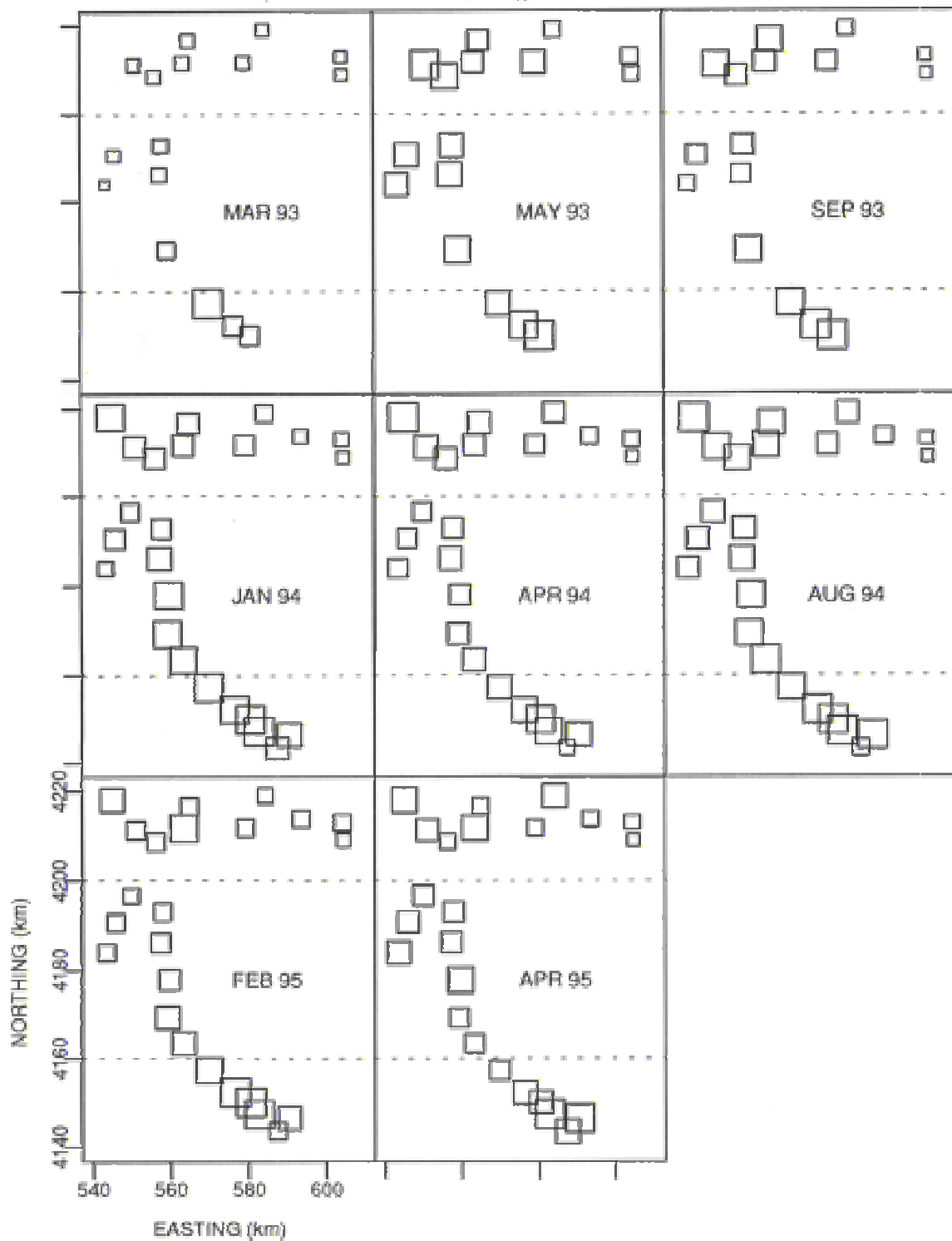
Total Chromium
(scaled to sampling event maximum)

MAR 93

MAY 93

SEP 93

JAN 94

APR 94

AUG 94

FEB 95

APR 95

NORTHING (km)

4220

4200

4180

4160

4140

540   560   580   600

EASTING (km)
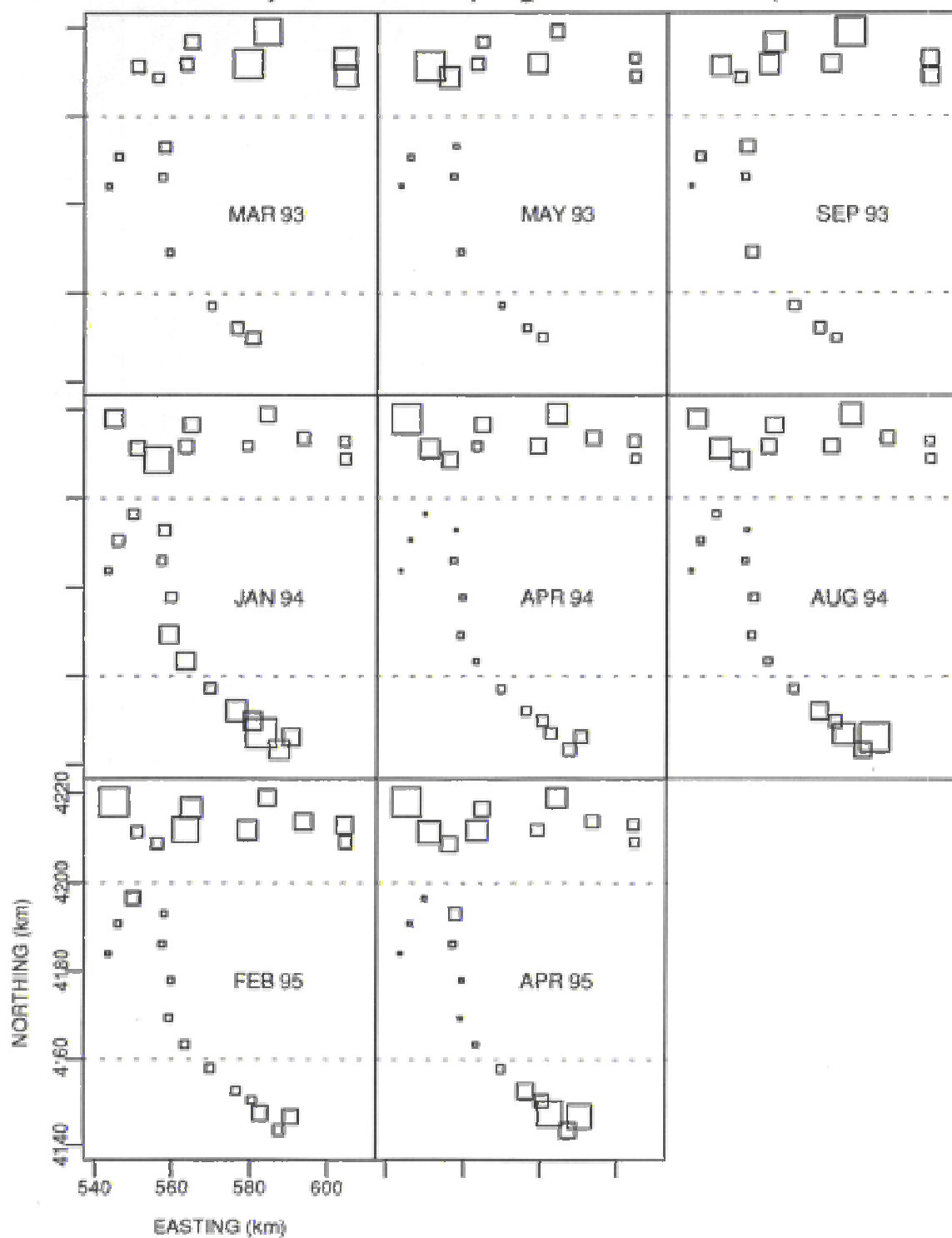
Total Copper
(scaled to sampling event maximum)
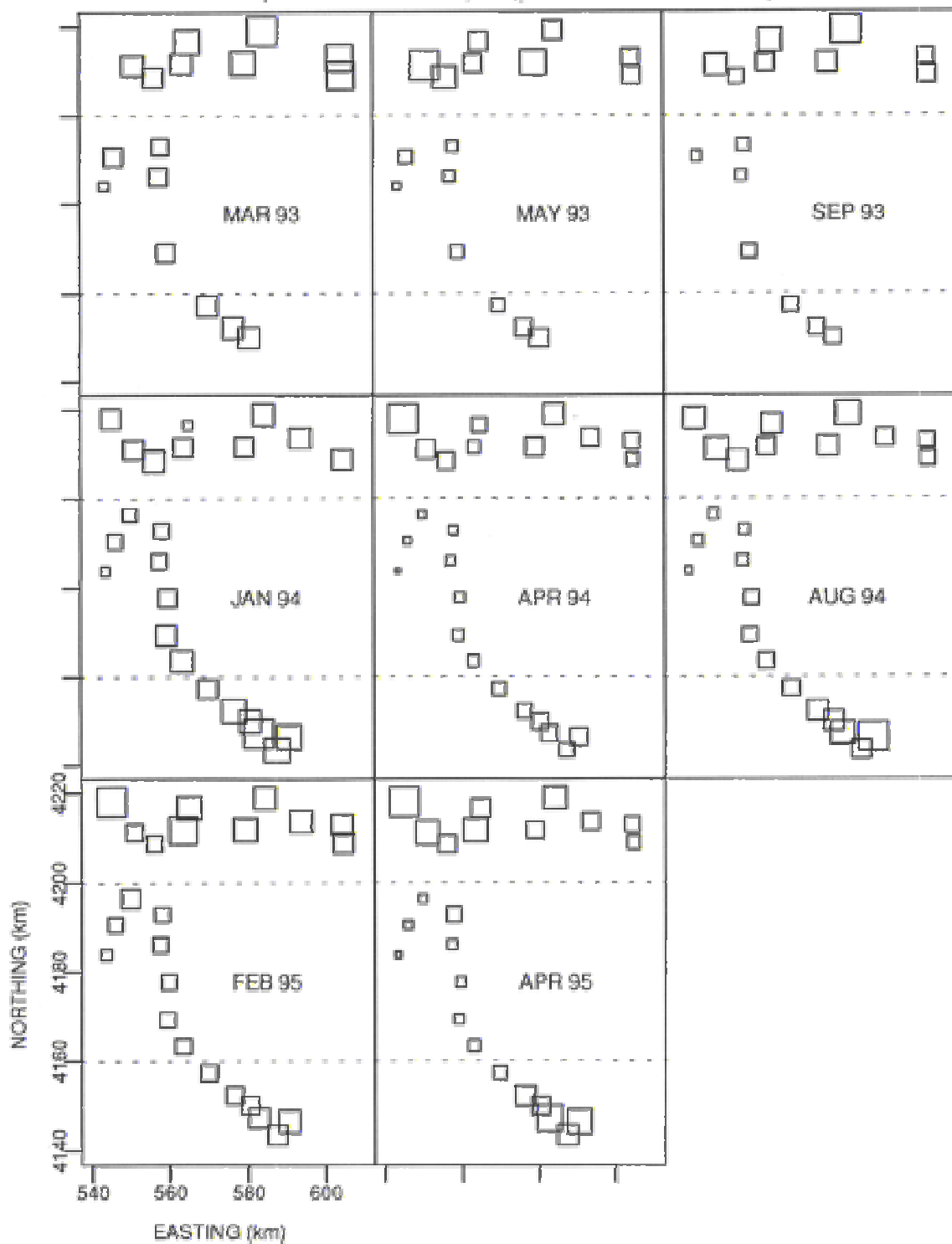
Total Mercury
(scaled to sampling event maximum)
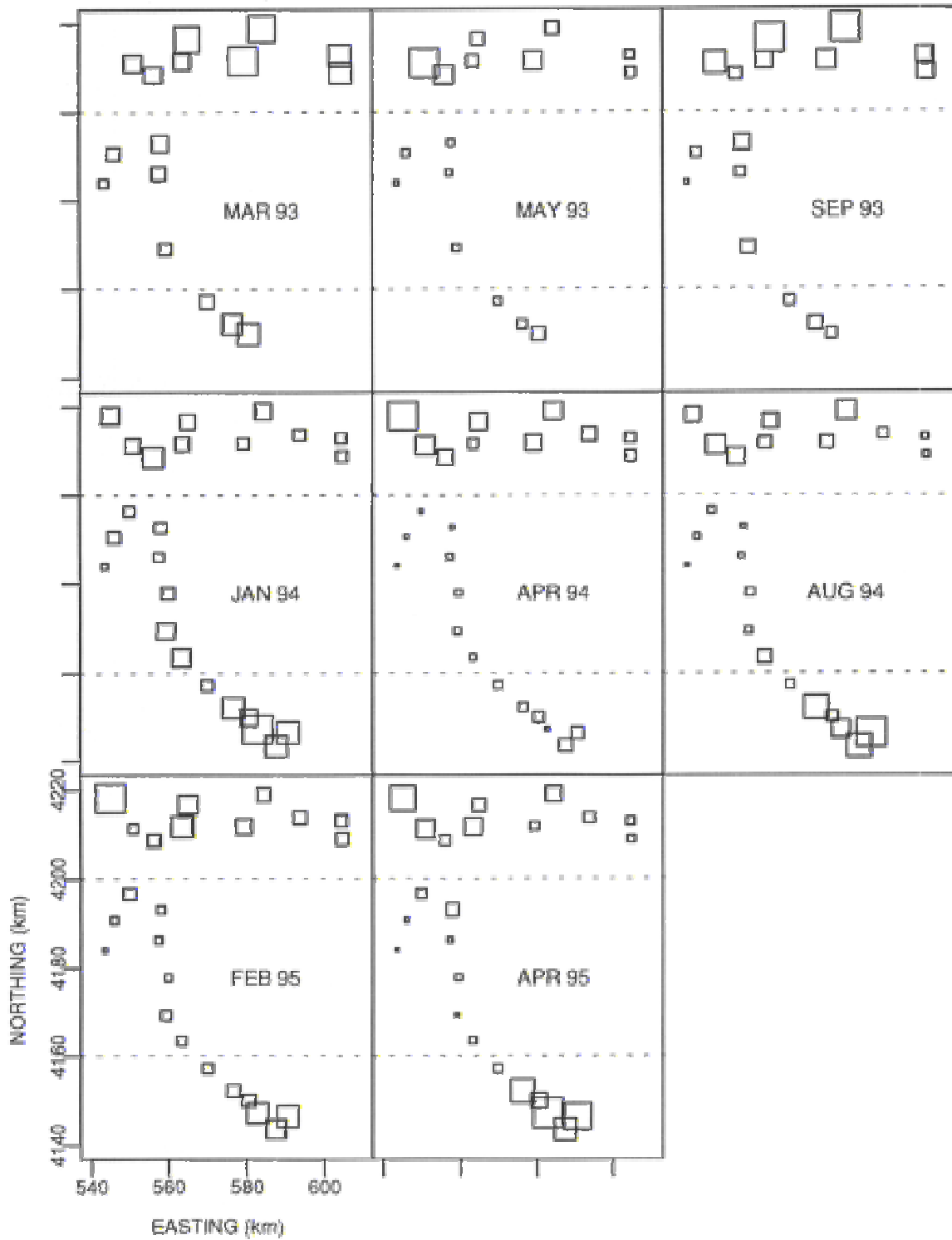
Total Nickel
(scaled to sampling event maximum)
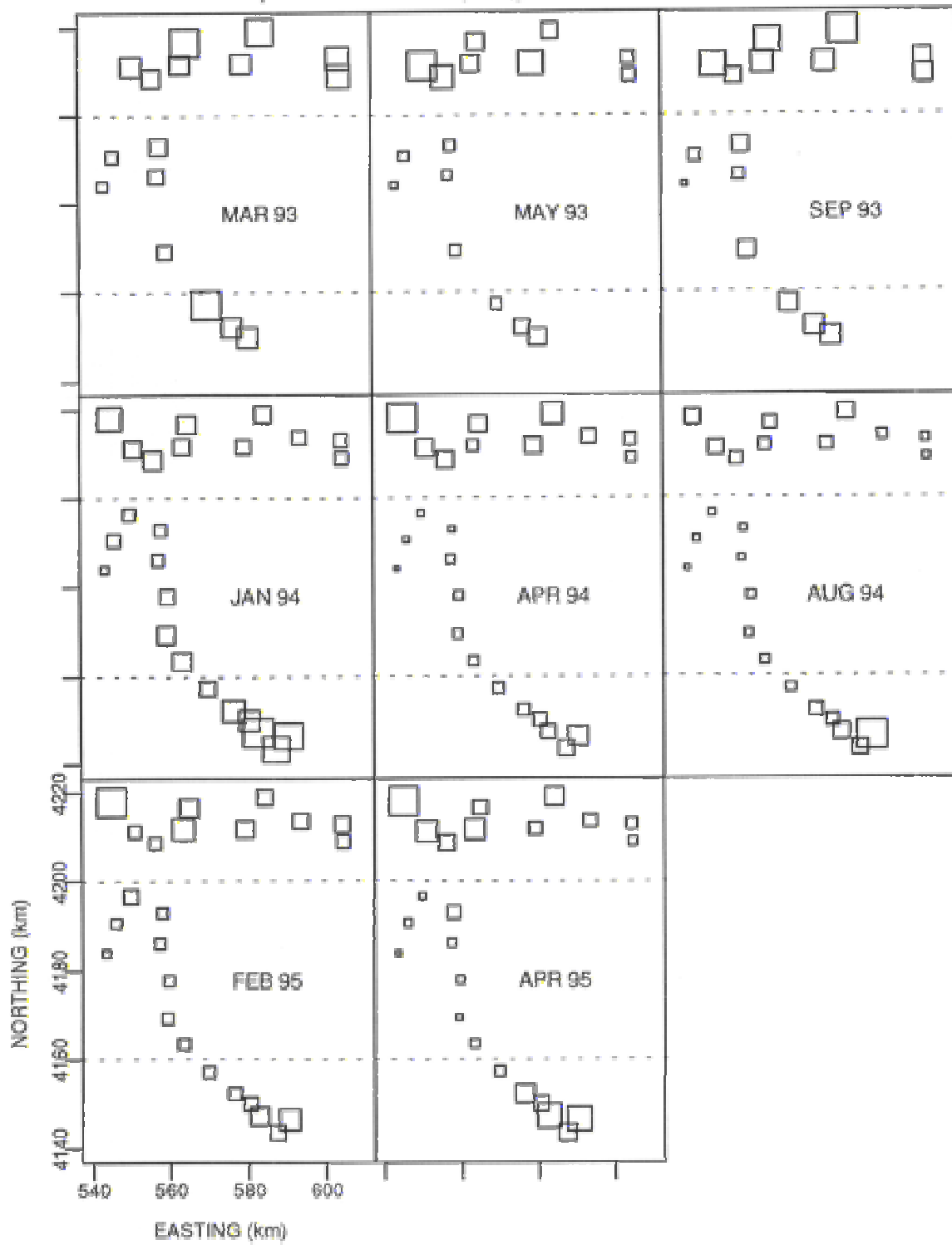
Total Lead
(scaled to sampling event maximum)

MAR 93

MAY 93

SEP 93

JAN 94

APR 94

AUG 94

FEB 95

APR 95

NORTHING (km)

4220
4200
4180
4160
4140

540   560   580   600

EASTING (km)
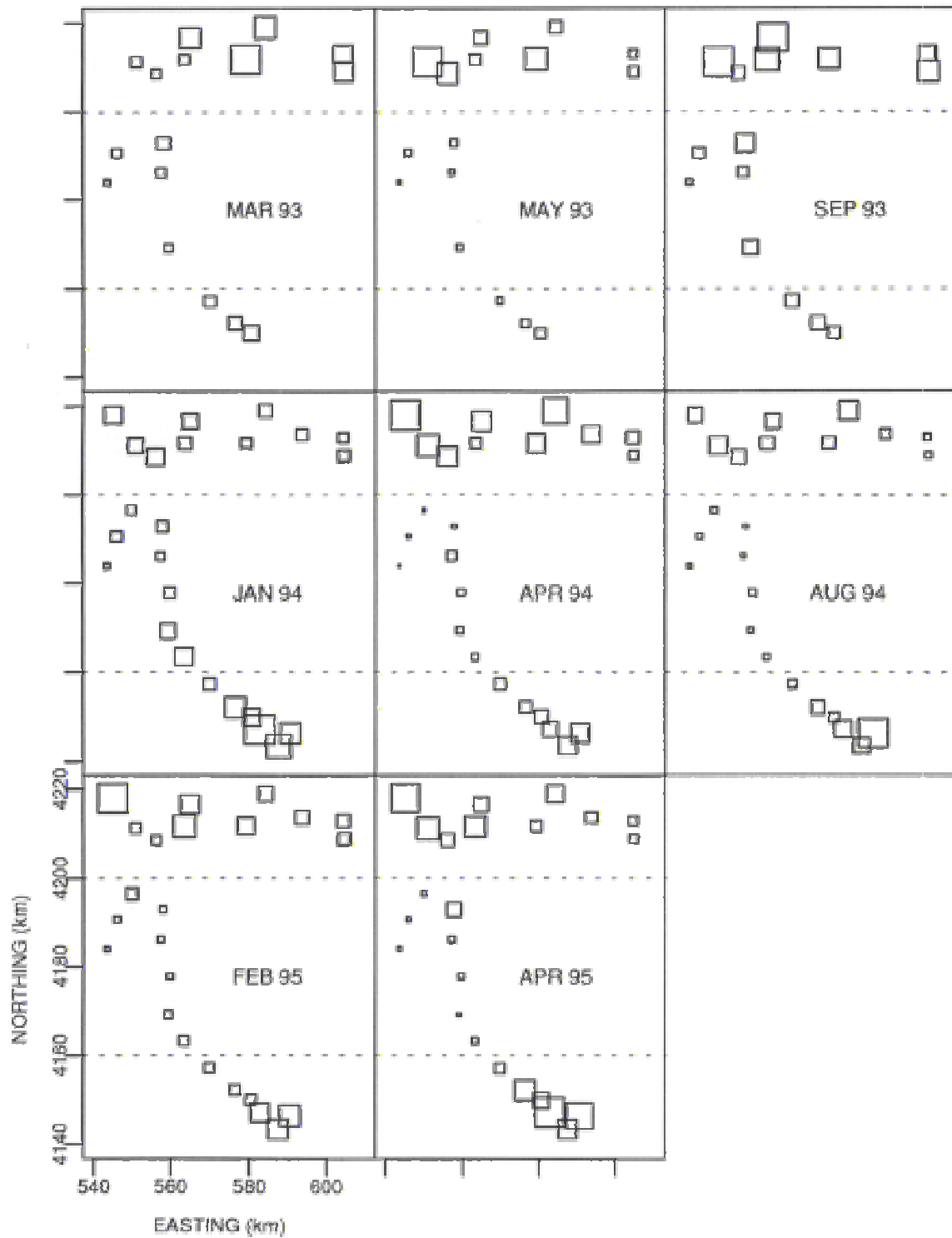
Total Selenium
(scaled to sampling event maximum)

MAR 93

MAY 93

SEP 93

JAN 94

APR 94

AUG 94

FEB 95

APR 95

NORTHING (km)

4220
4200
4180
4160
4140

540   560   580   600

EASTING (km)
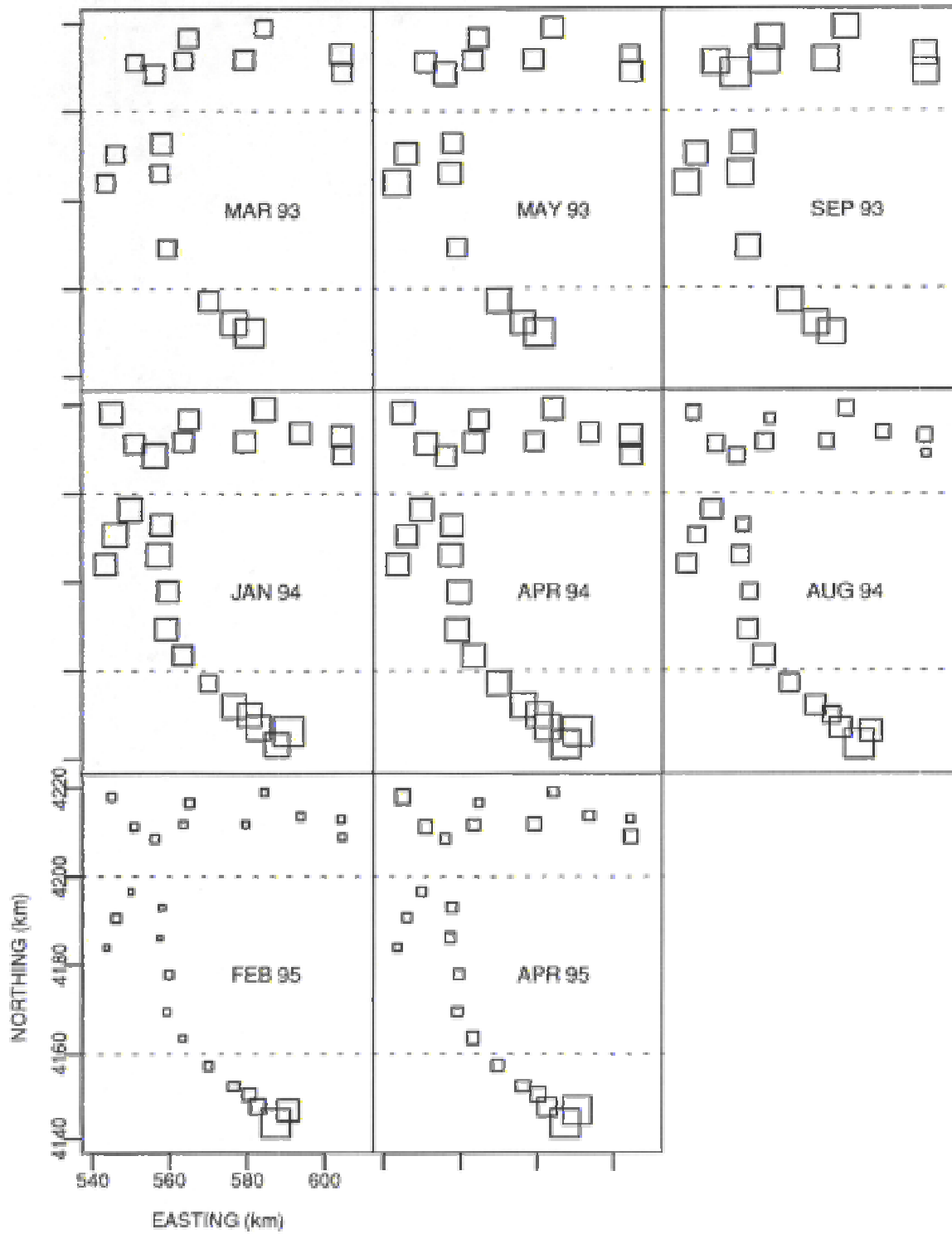
Total Zinc
(scaled to sampling event maximum)

MAR 93

MAY 93

SEP 93

JAN 94

APR 94

AUG 94

FEB 95

APR 95

NORTHING (km)

4220

4200

4180

4160

4140

540   560   580   600

EASTING (km)
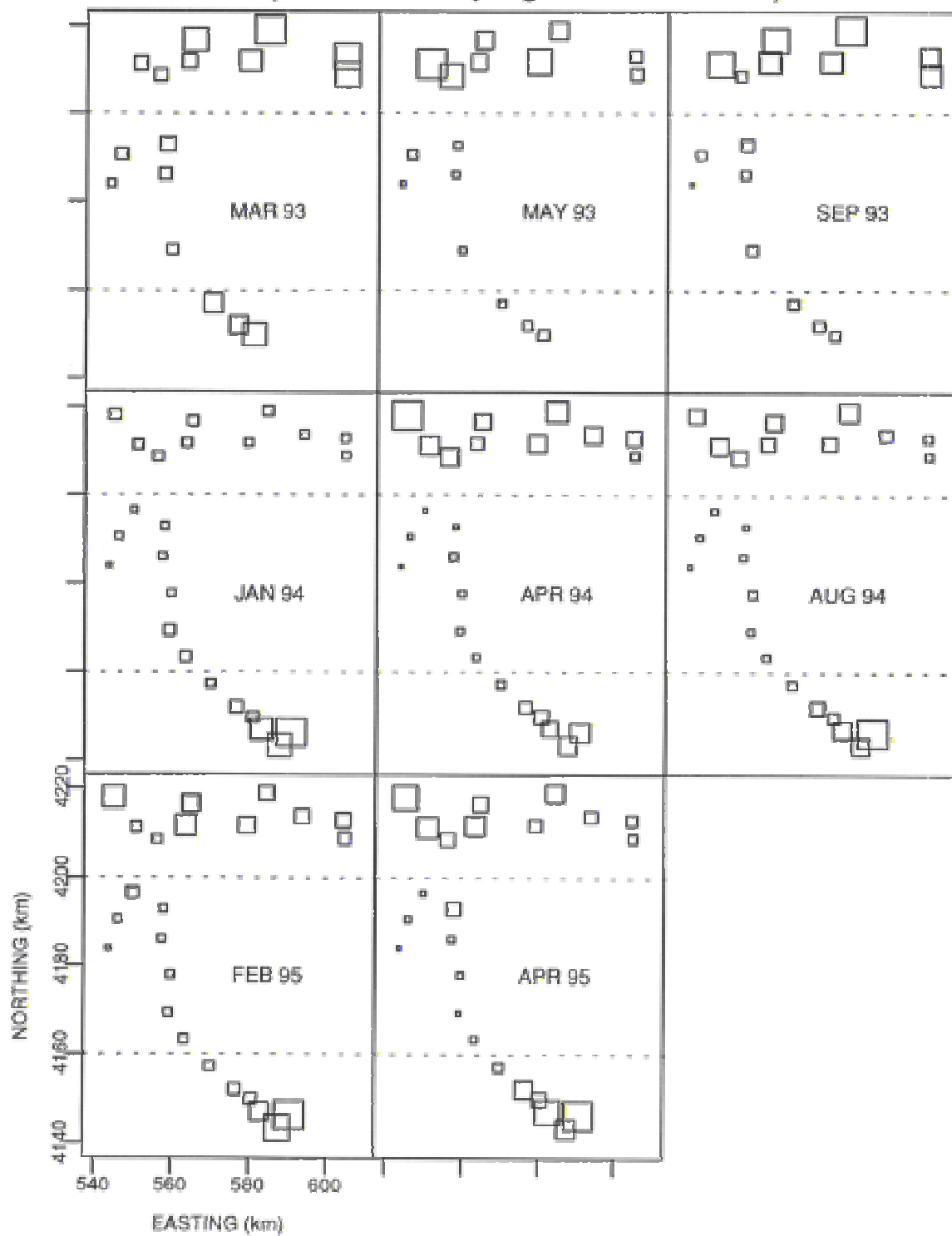
Figure 5. Moran scatterplots for each trace element during sampling event 13. The dashed line is the linear regression line. The solid line is the least trimmed squares regression line. Three points are designated by their station codes rather than by circles. These have the three most negative residuals with respect to the least trimmed squares line.

Moran Scatterplot
Total Silver
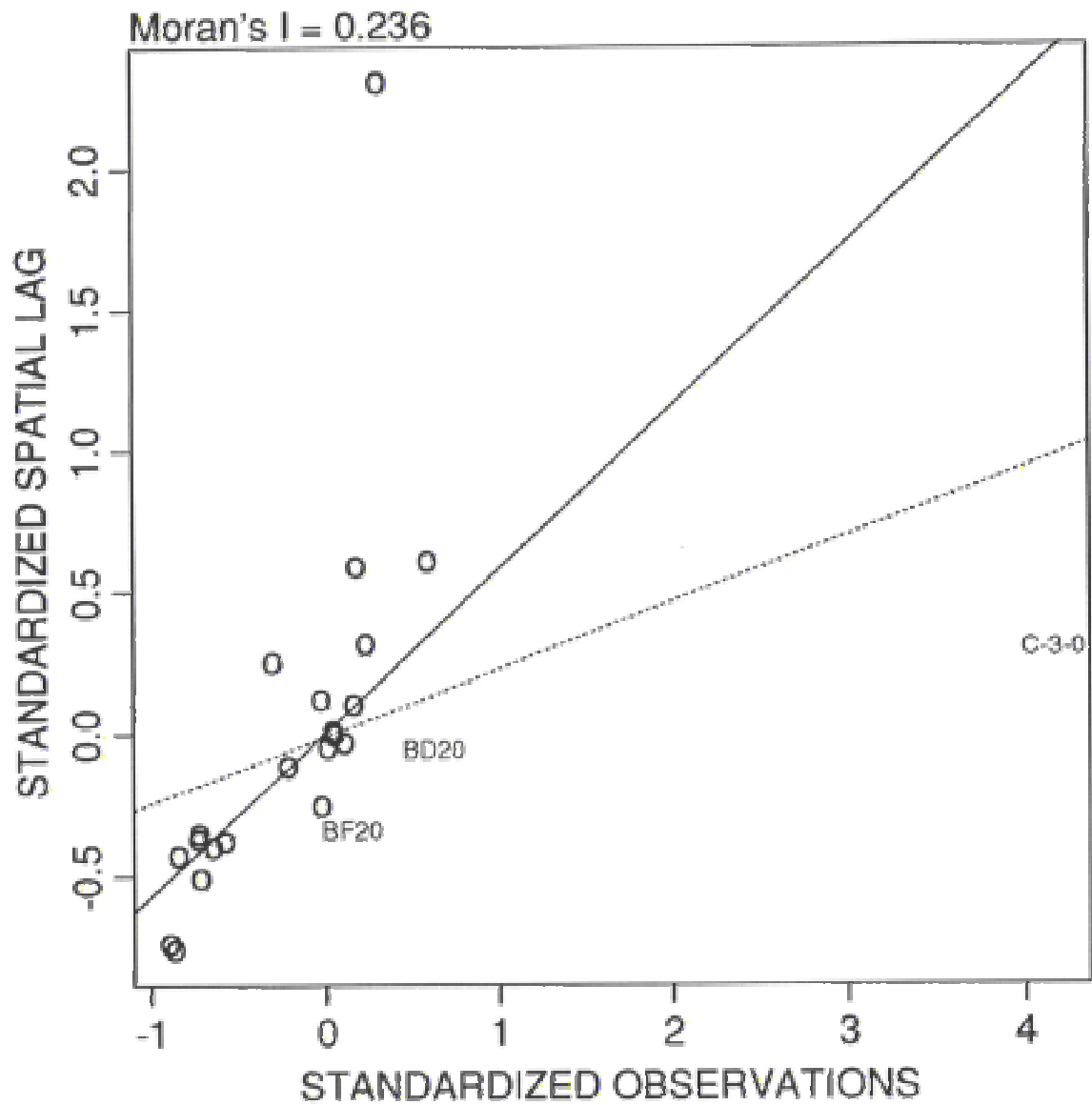
Moran's I = 0.236

Moran Scatterplot
Total Arsenic
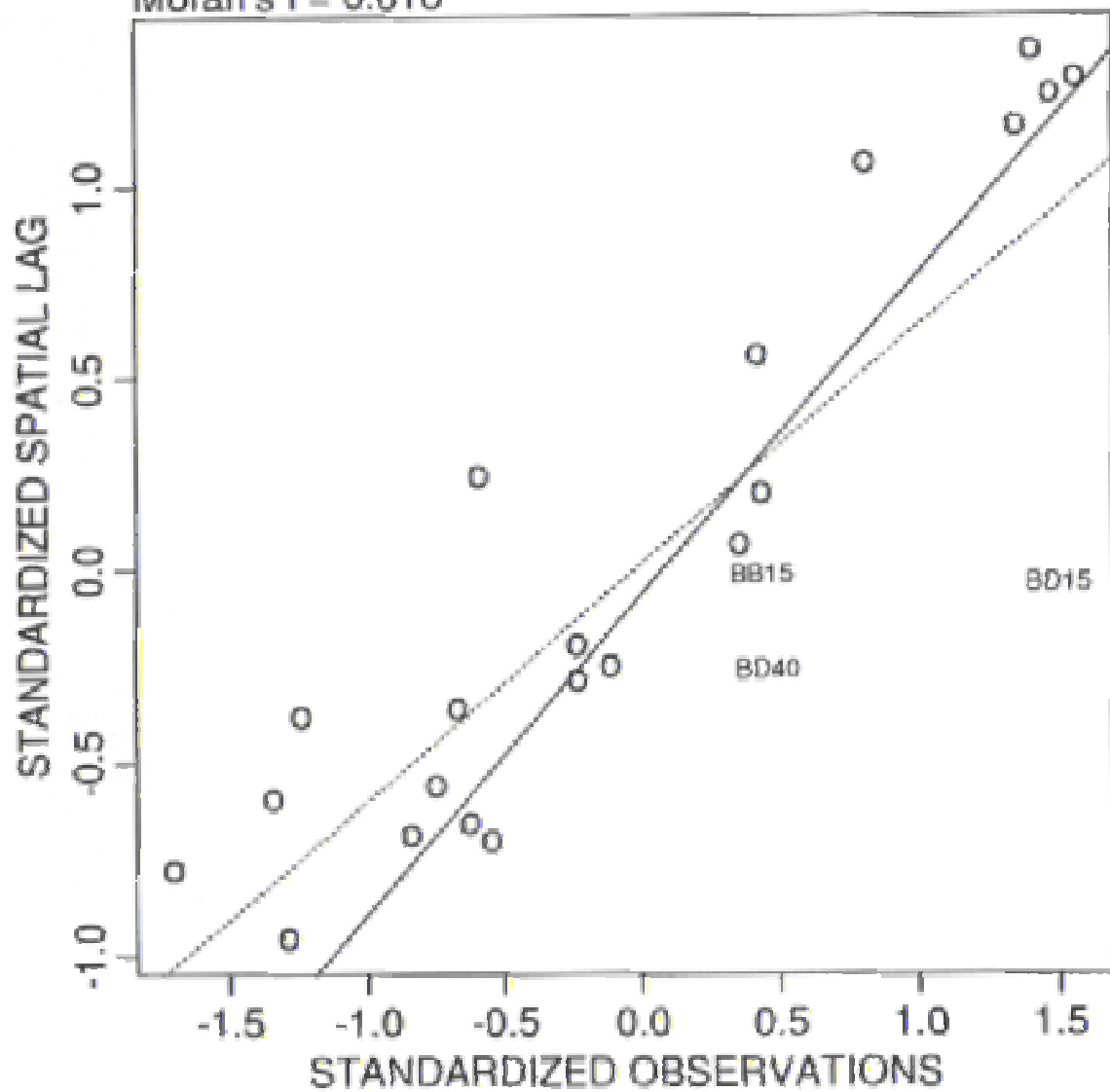
Moran Scatterplot
Total Cadmium

Moran's I = 0.547
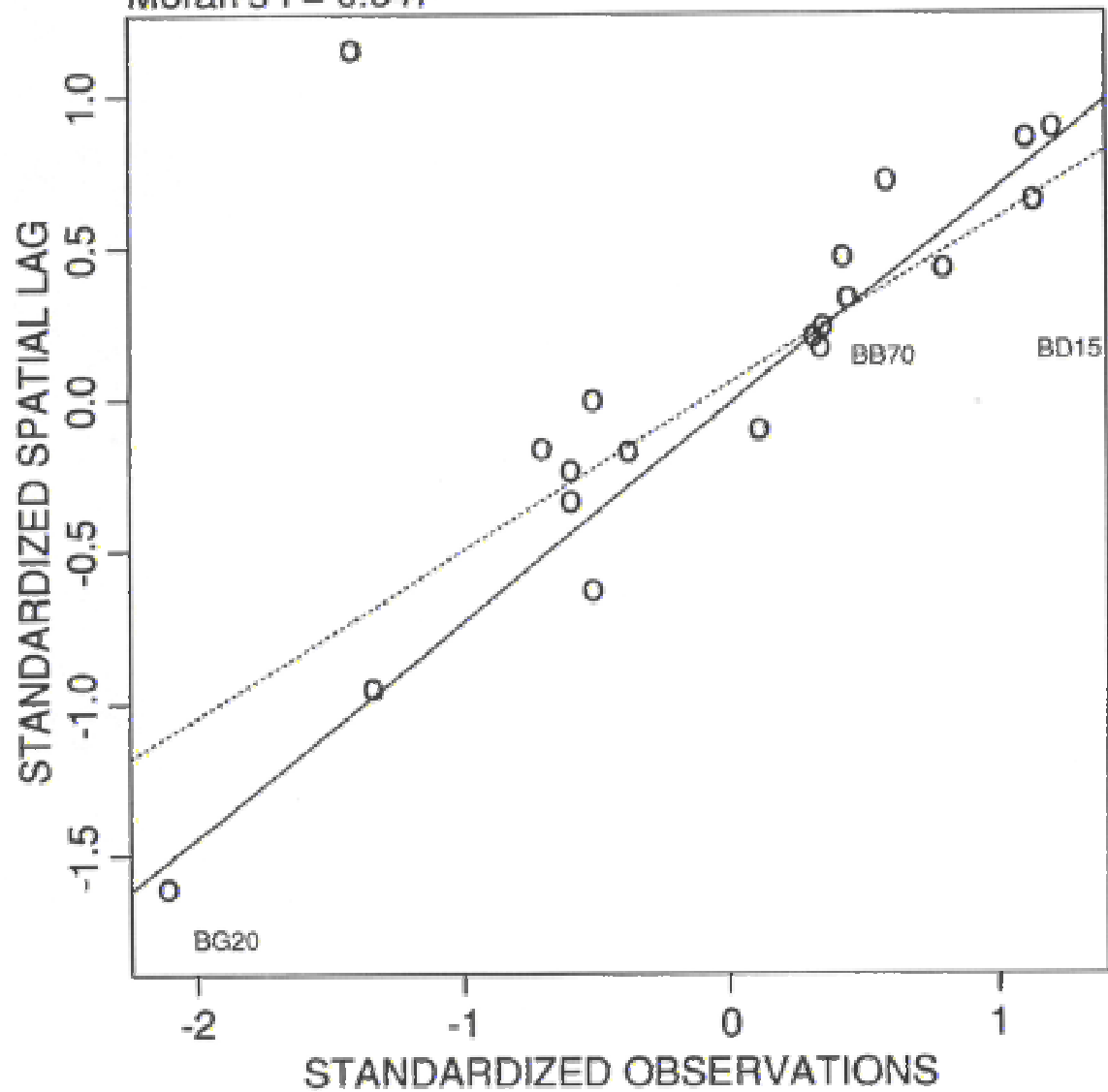
Moran Scatterplot
Total Chromium
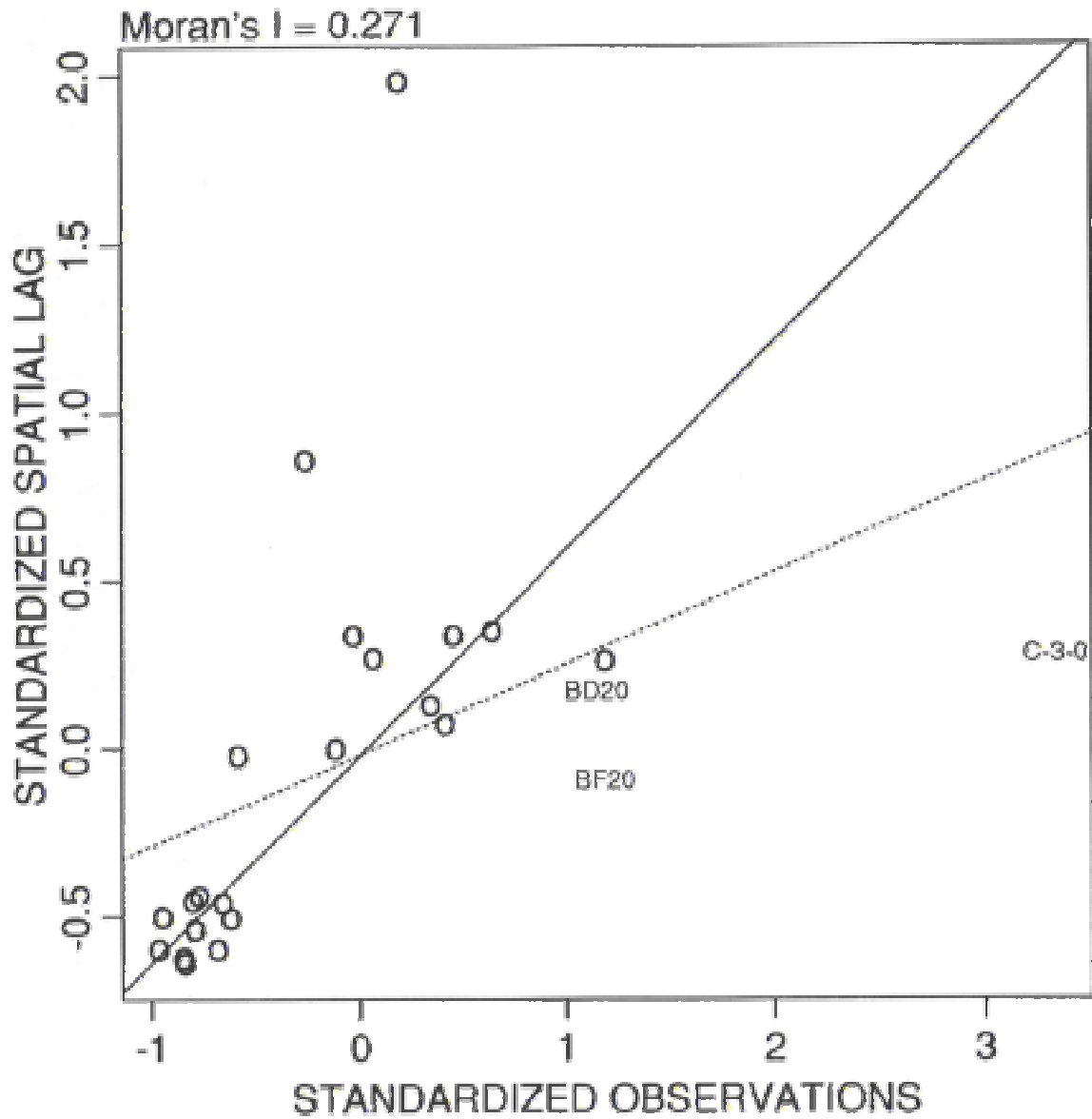
Moran's I = 0.271

Moran Scatterplot
Total Copper

Moran's I = 0.372

STANDARDIZED SPATIAL LAG

STANDARDIZED OBSERVATIONS

C-3-0

BD20   BF20
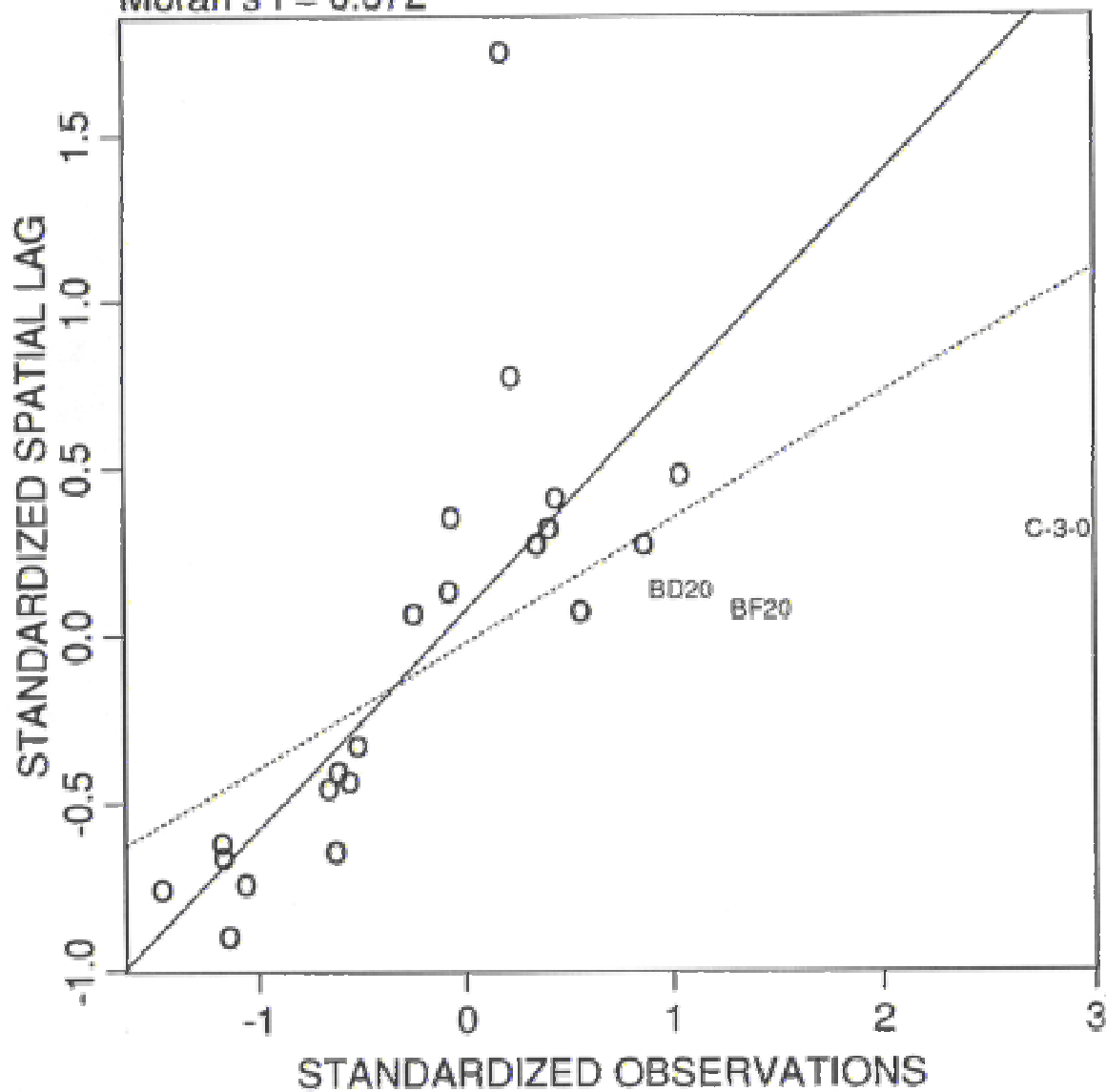
Moran Scatterplot
Total Mercury

Moran's I = 0.412
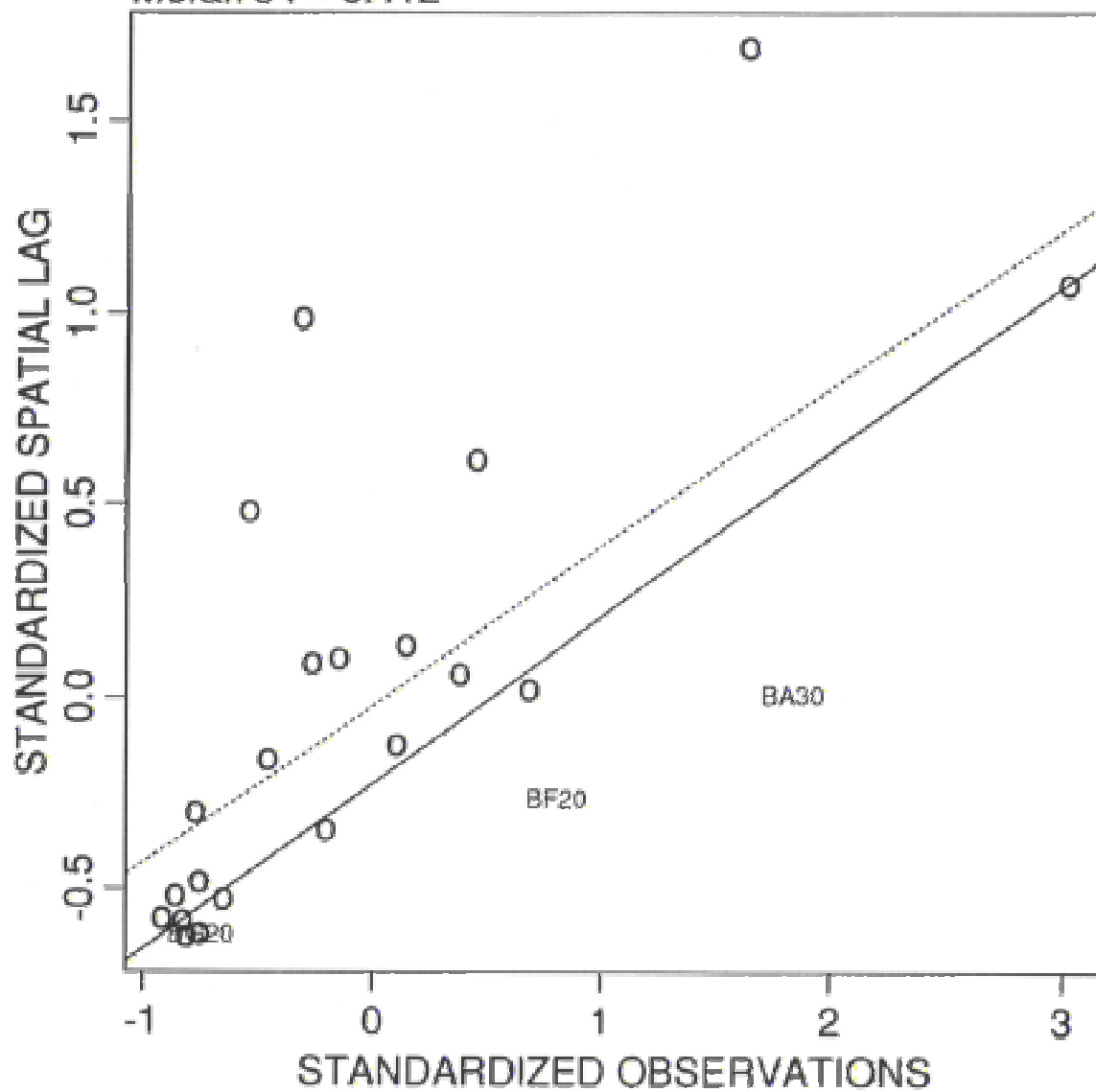
Moran Scatterplot
Total Nickel

Moran's I = 0.253
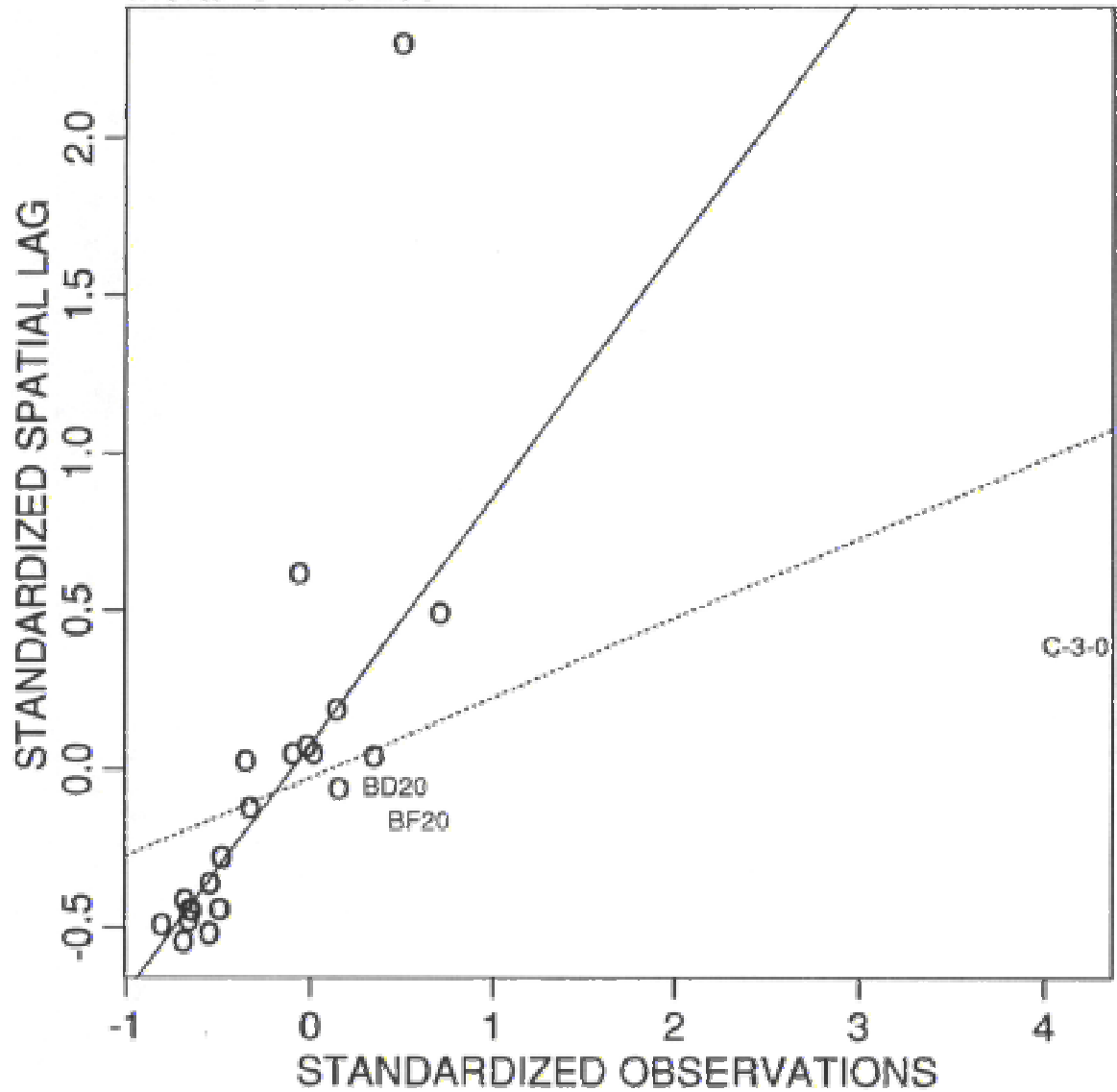
Moran Scatterplot
Total Lead
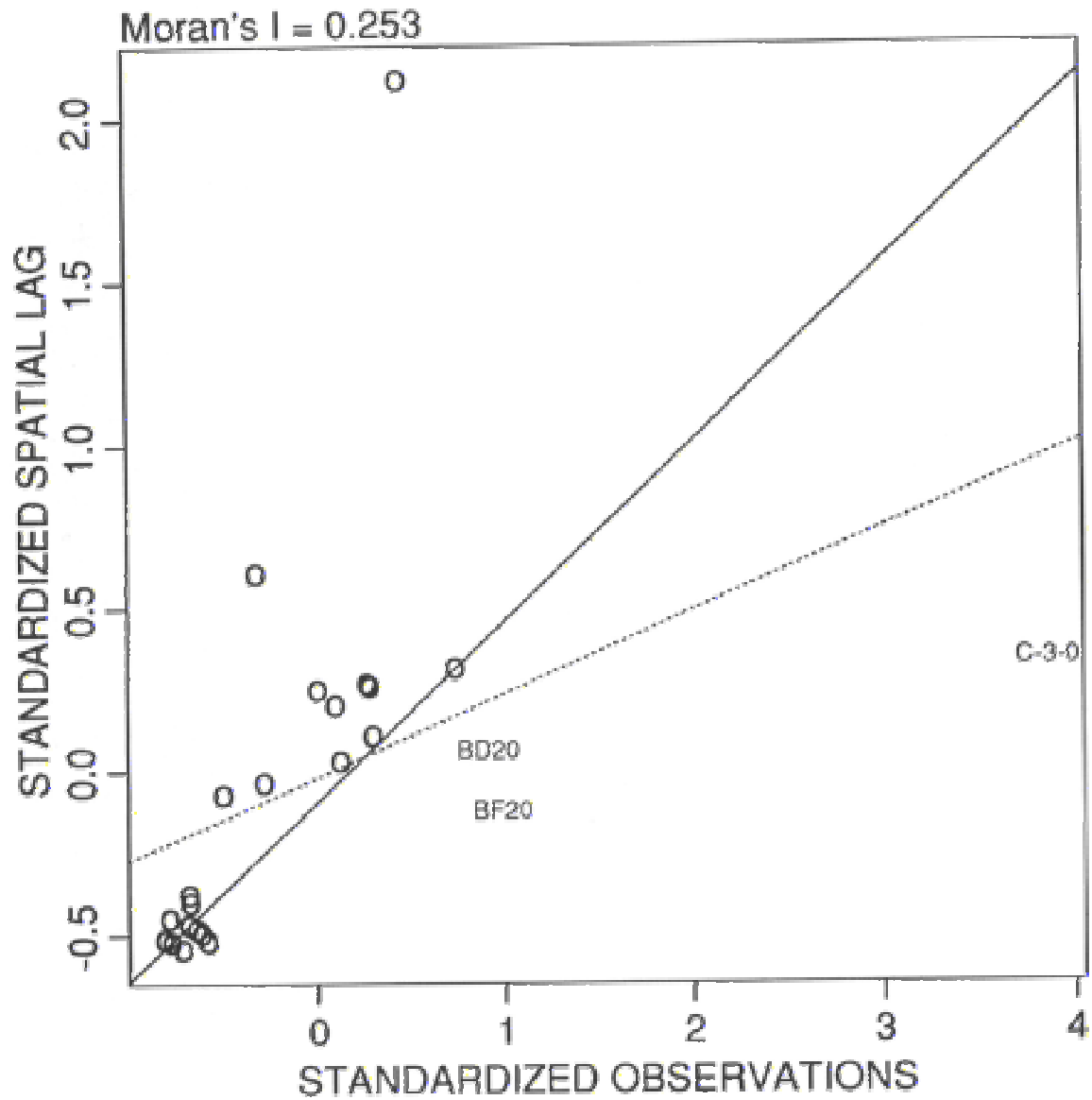
Moran's I = 0.253

Moran Scatterplot
Total Selenium
Moran's I = 0.441
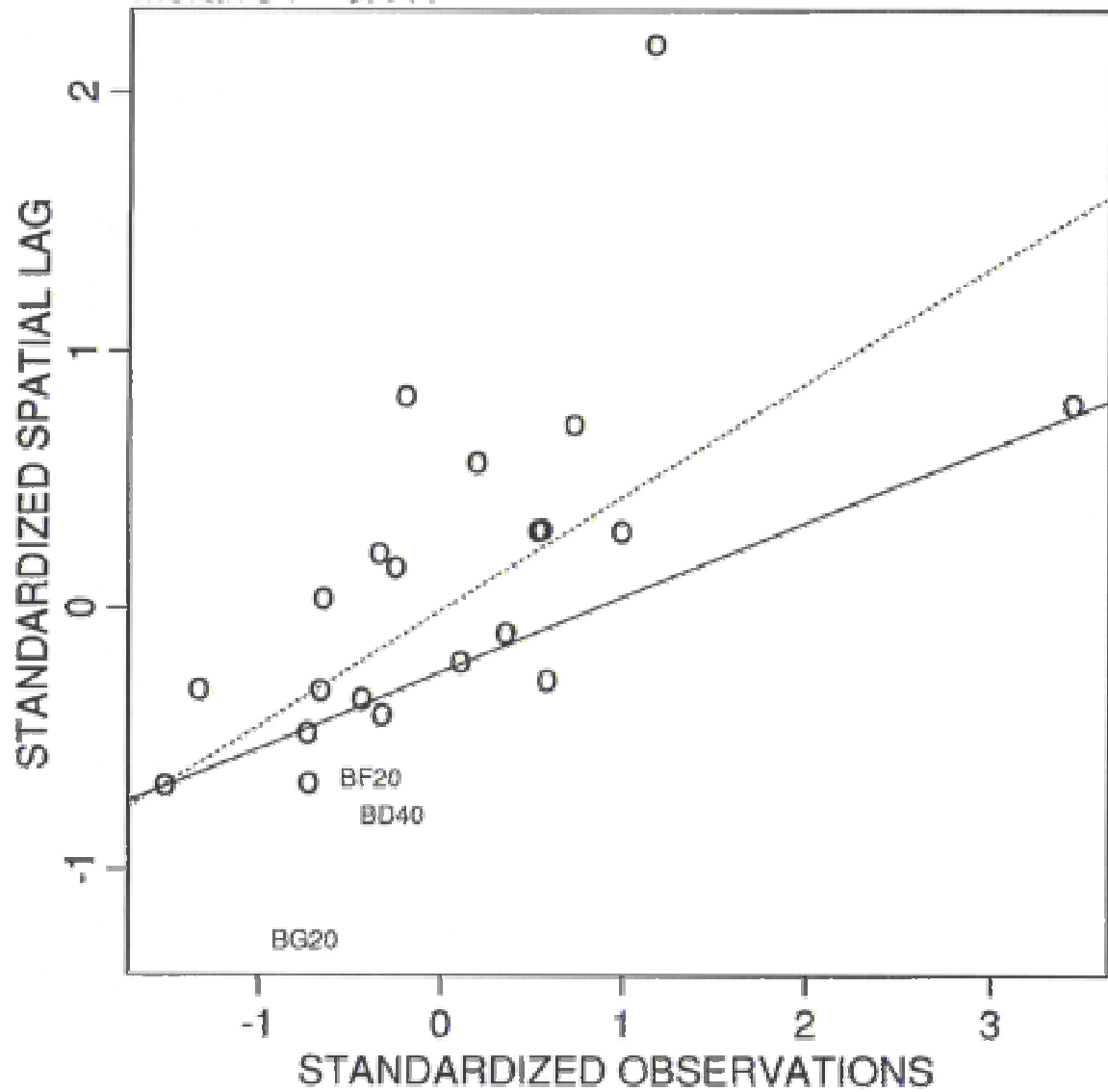
Moran Scatterplot
Total Zinc

Moran's I = 0.31
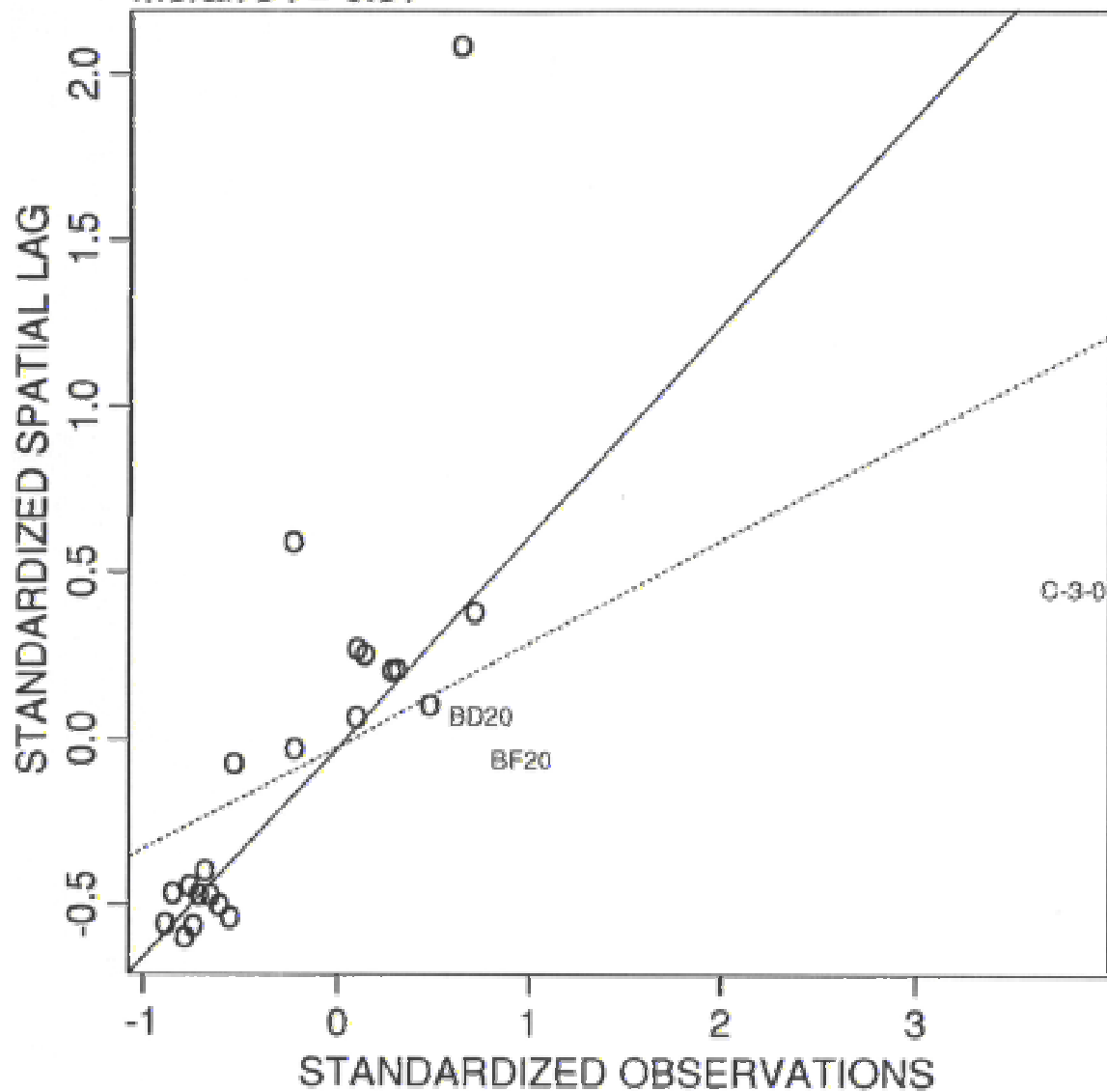
Figure 6. Time series of trace element totals for RMP stations during 1991-1995.

Total Silver

Total Arsenic

CONCENTRATION (ug/L)

YEAR

Total Cadmium
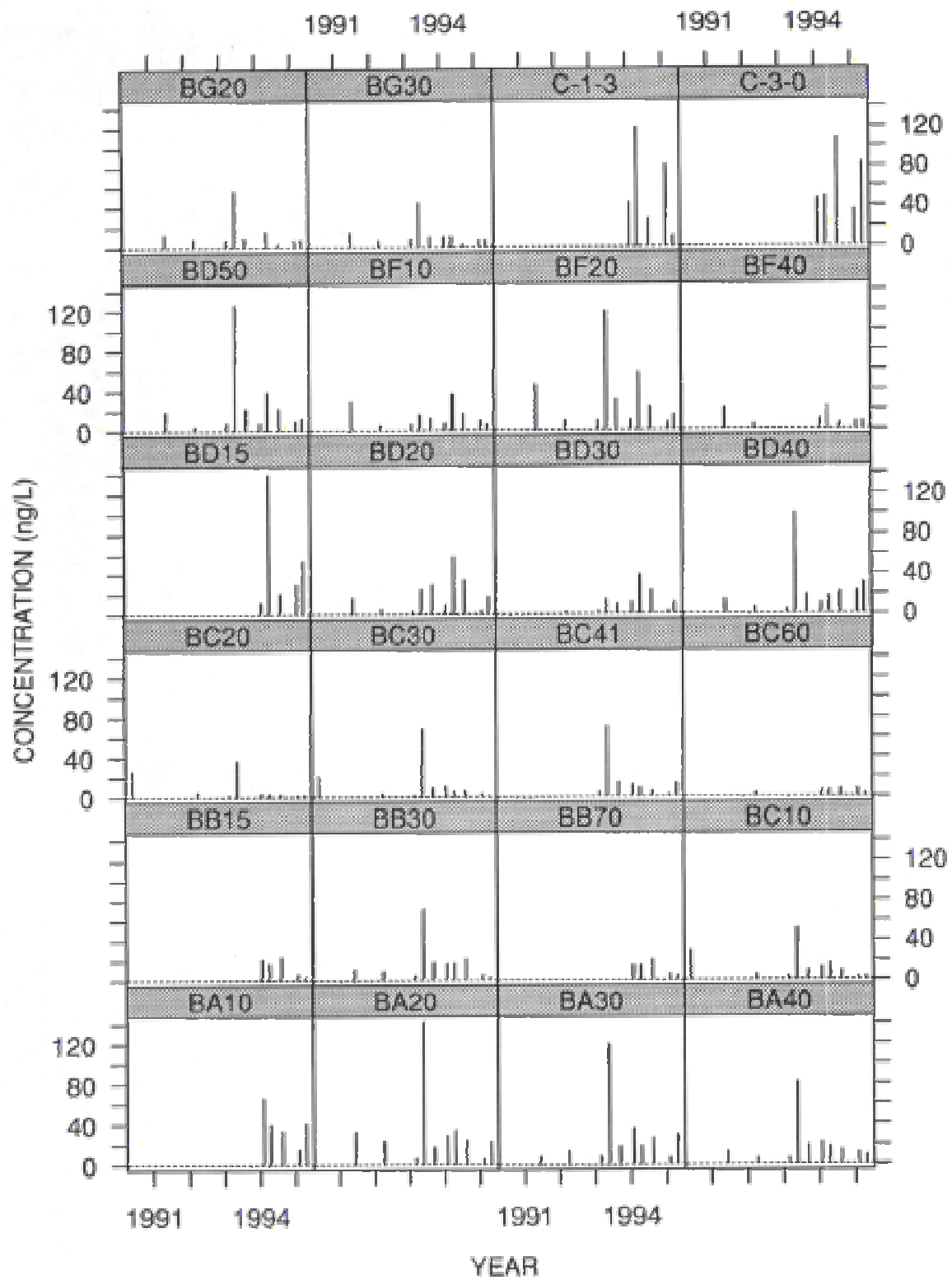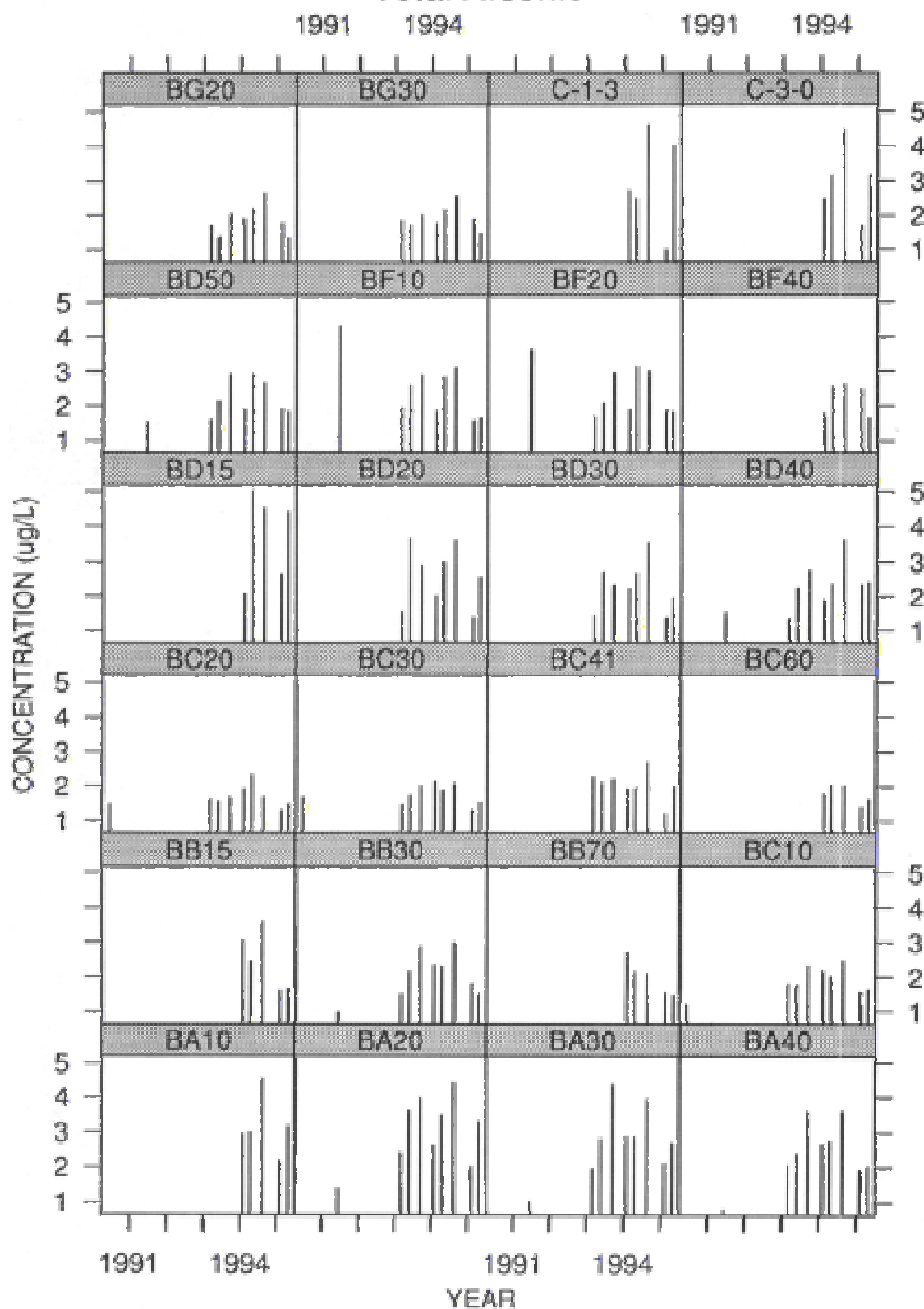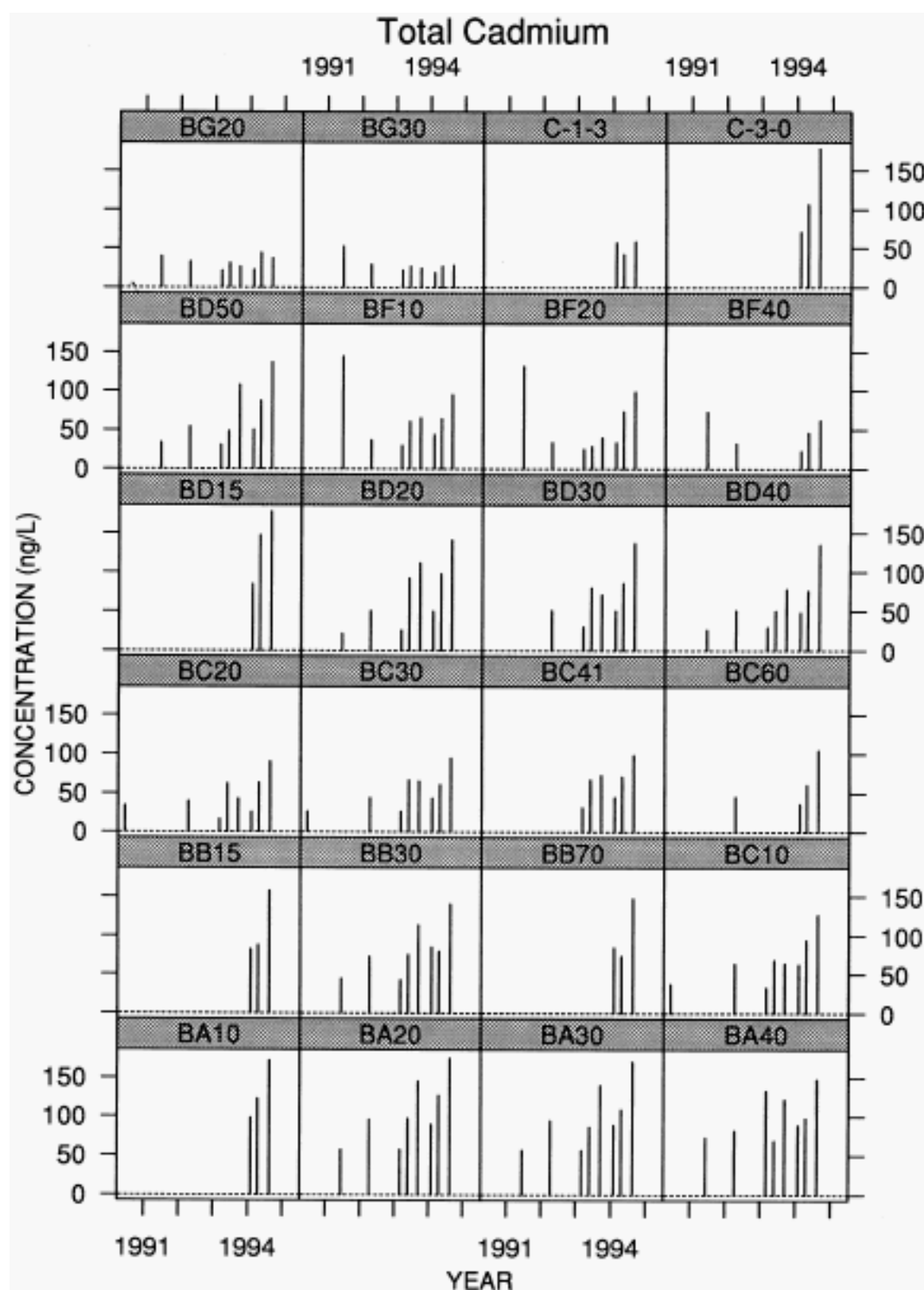
Total Chromium

Total Copper

CONCENTRATION (ug/L)
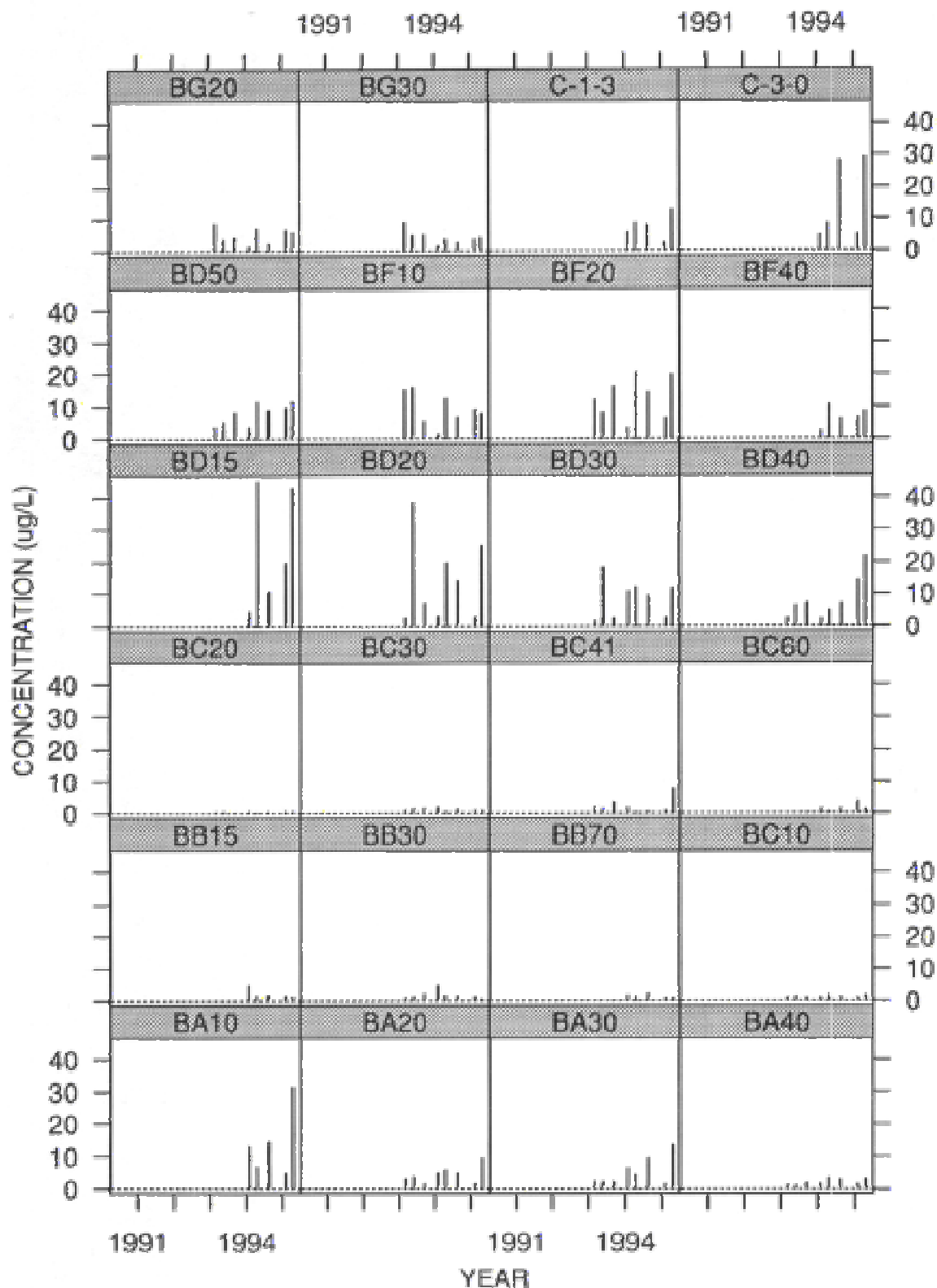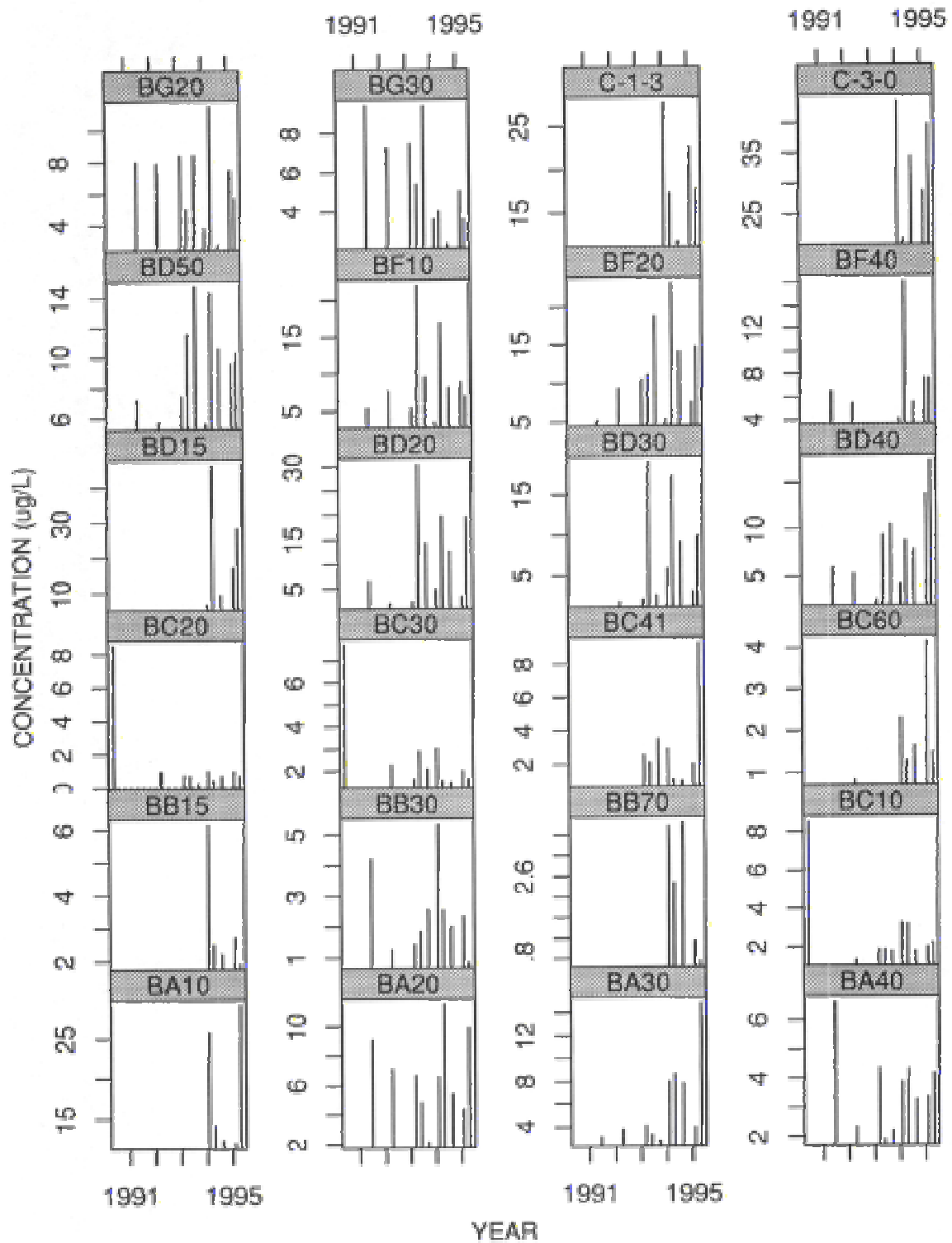
YEAR
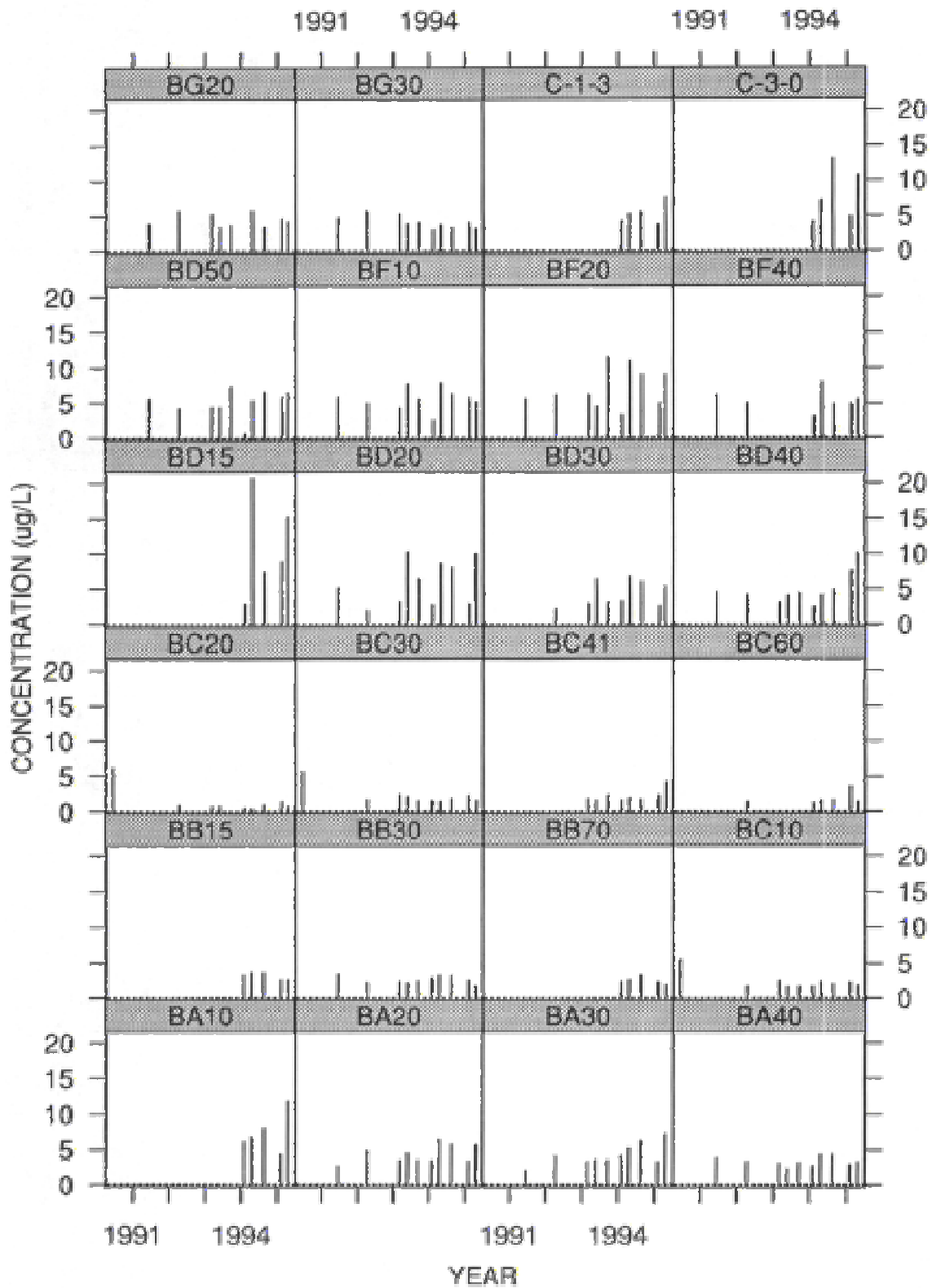
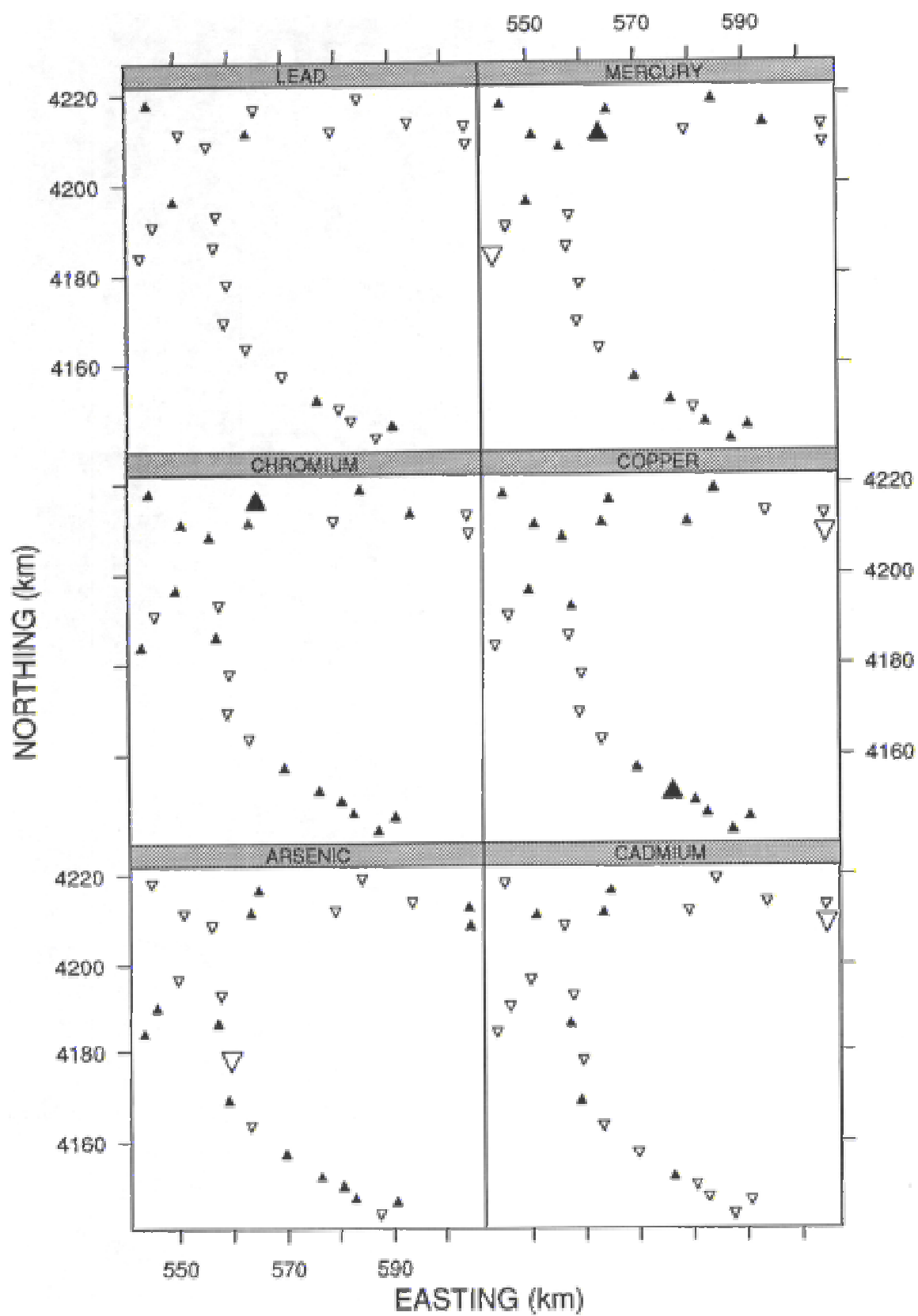Total Nickel

Total Lead
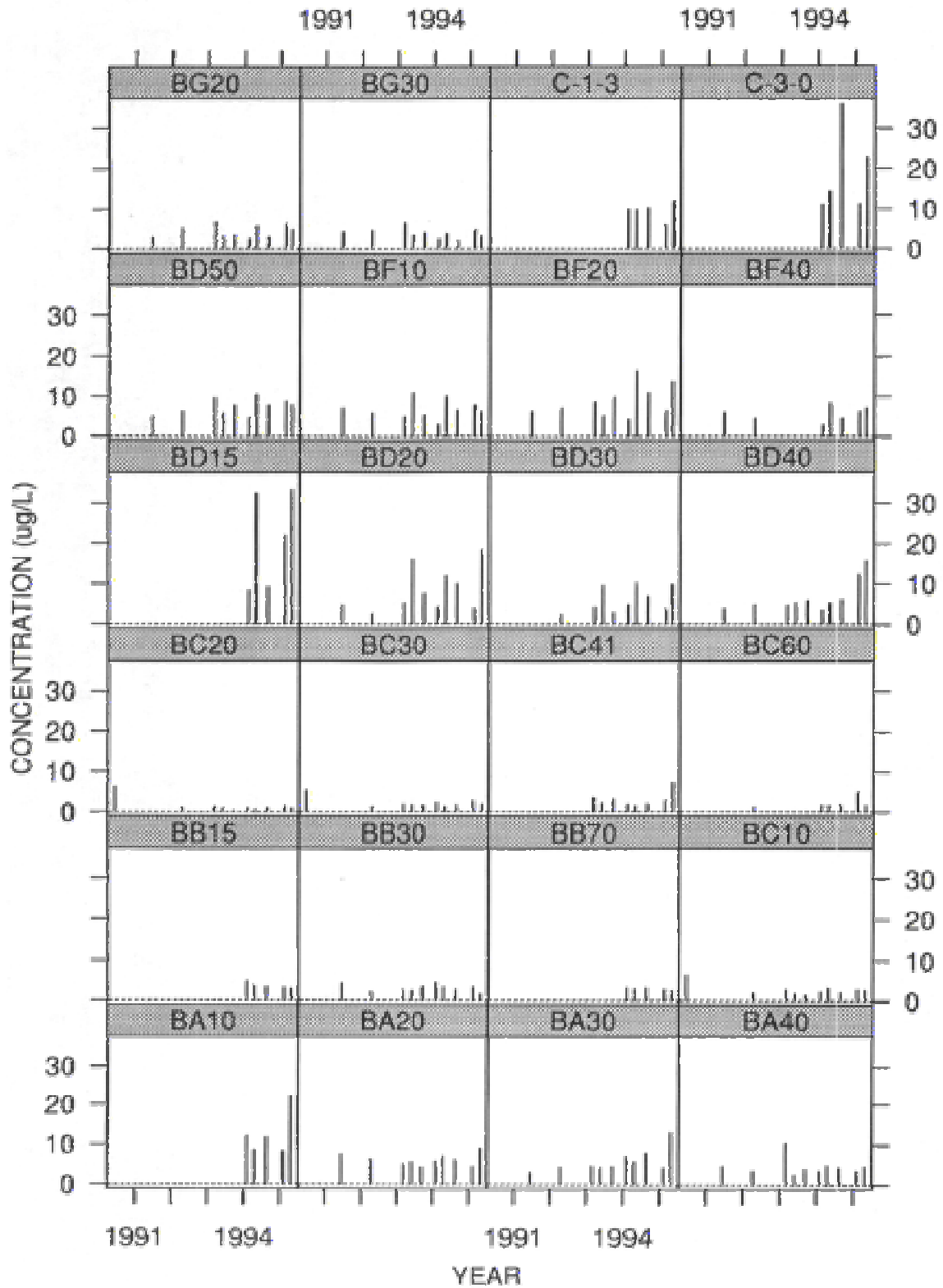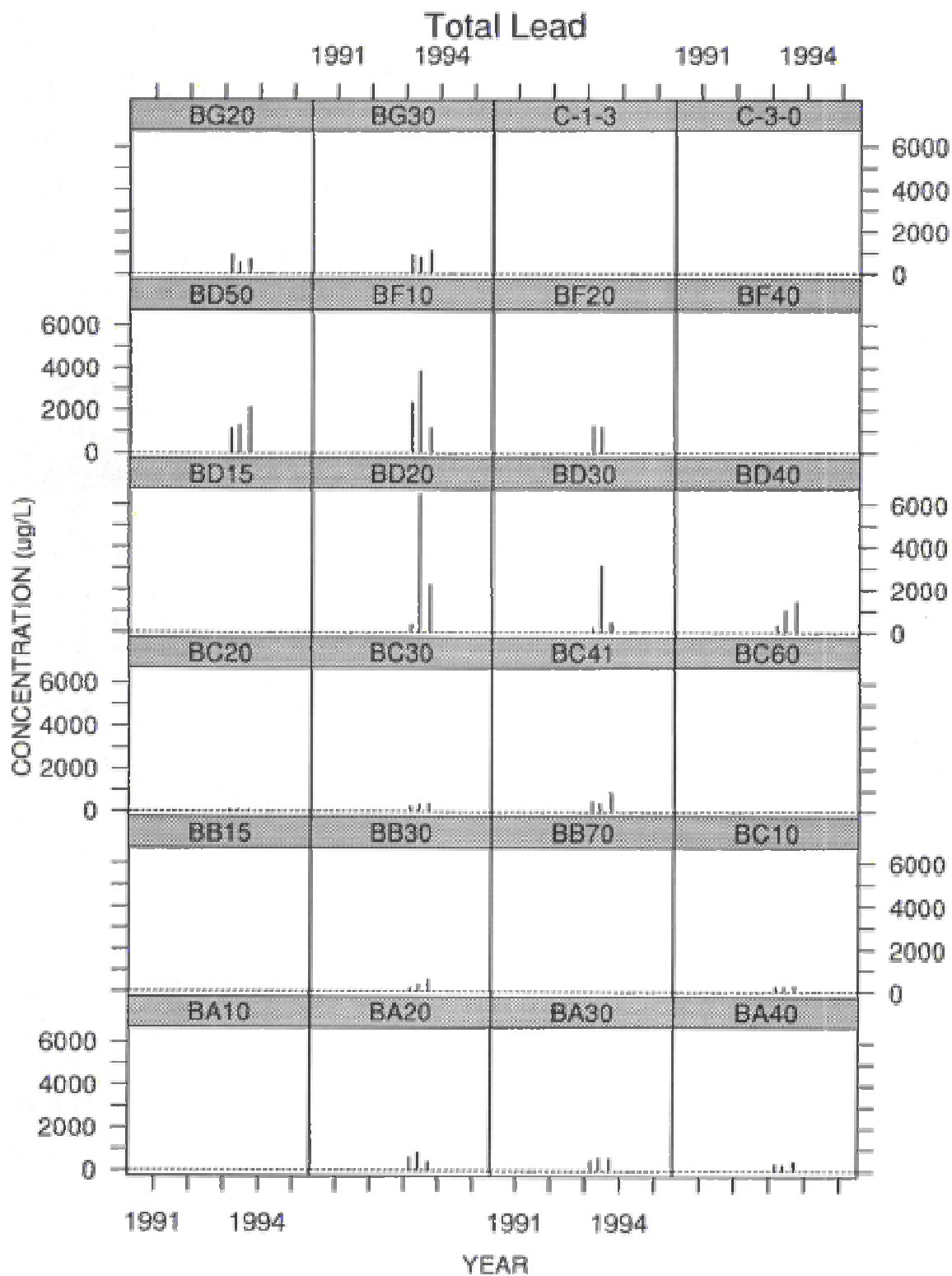
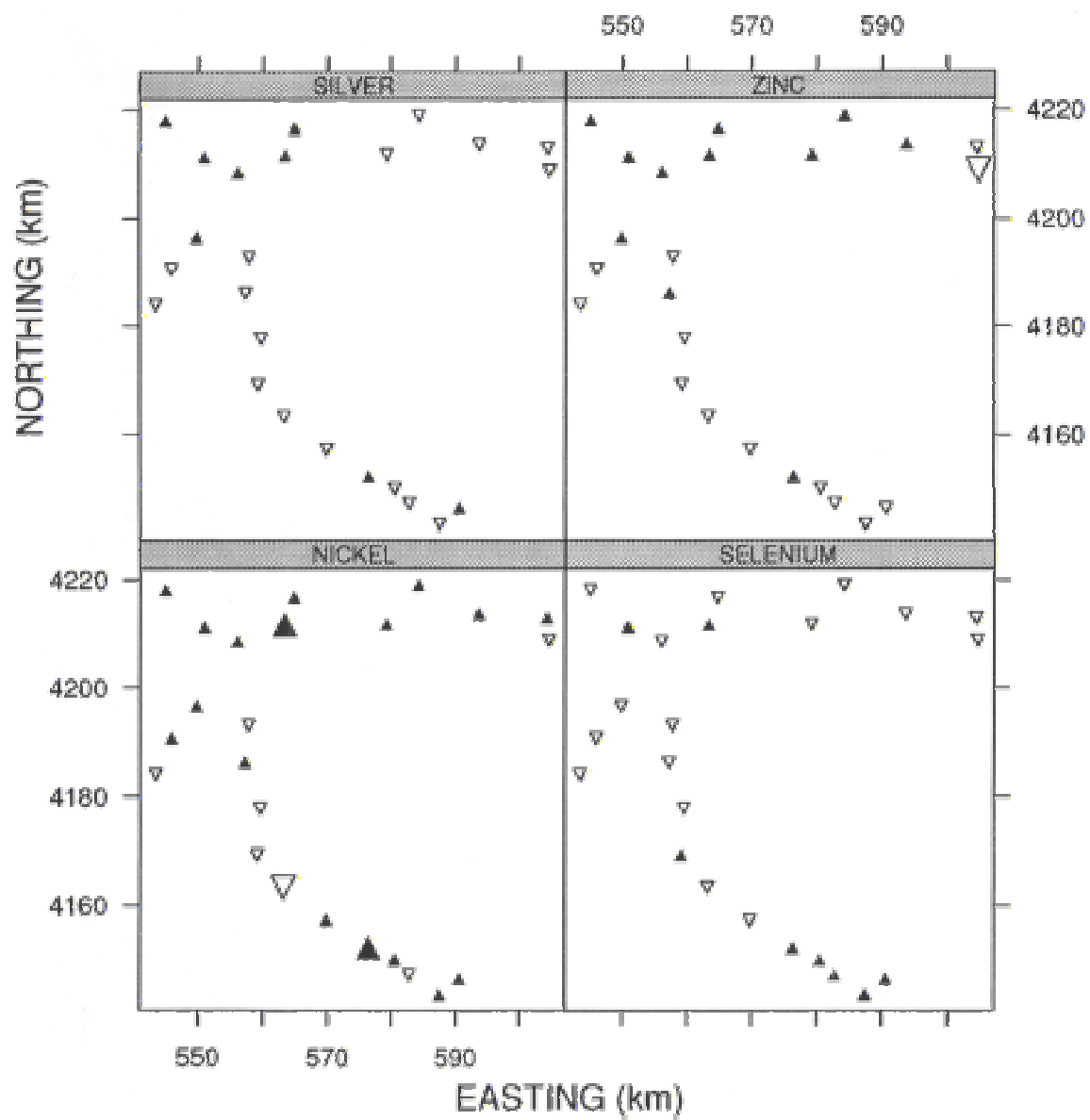# Total Selenium
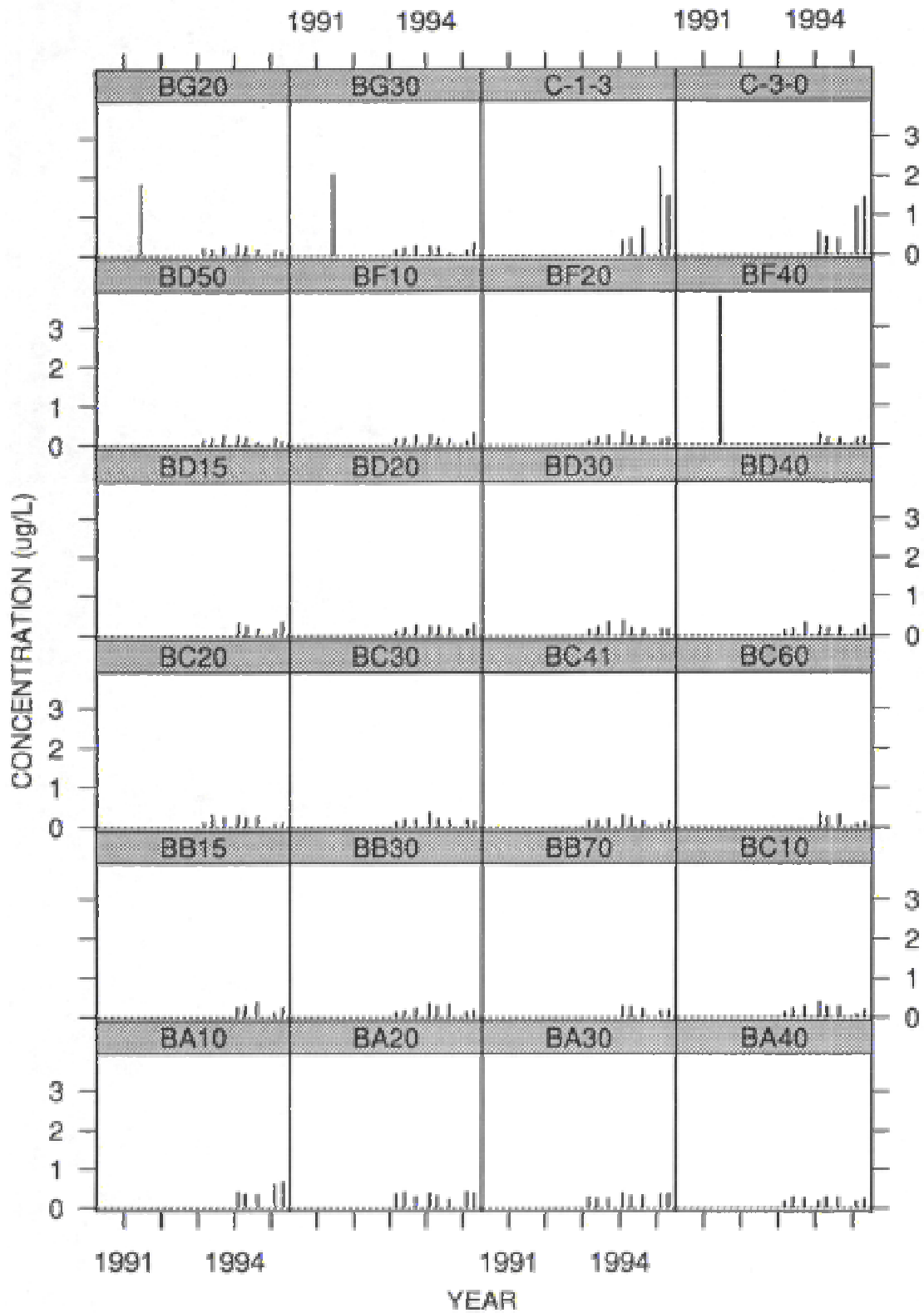
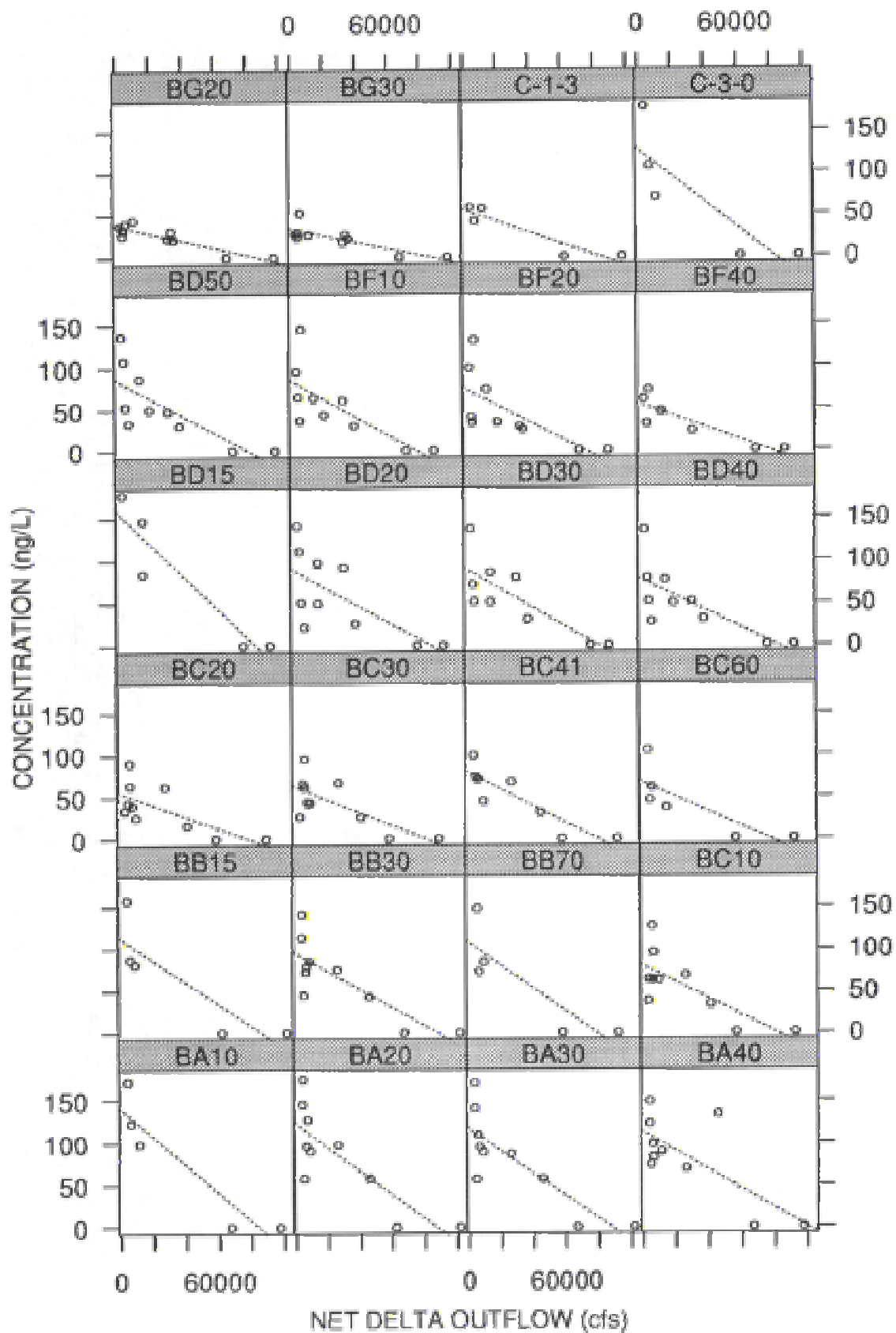Total Zinc
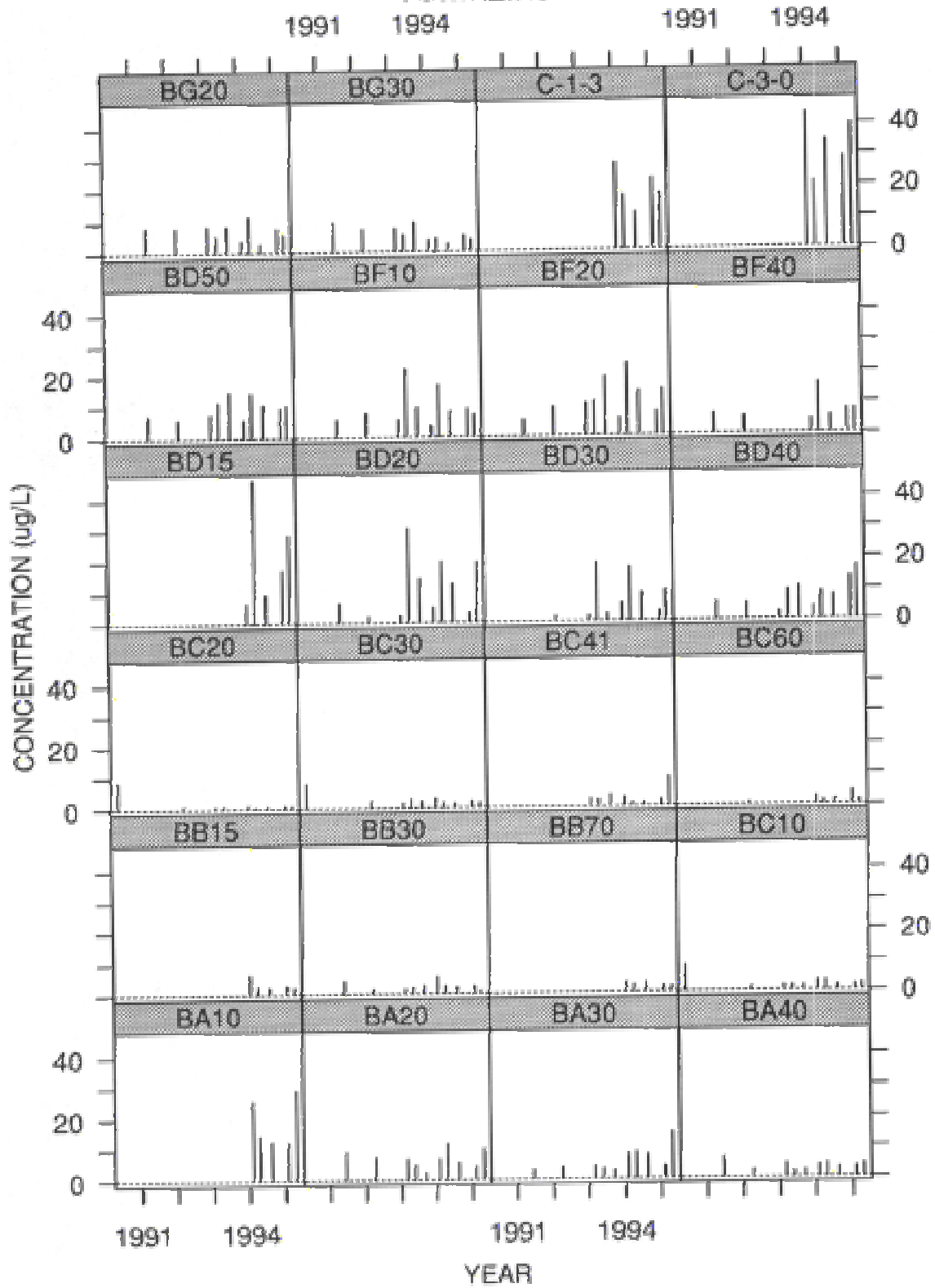
Total Zinc
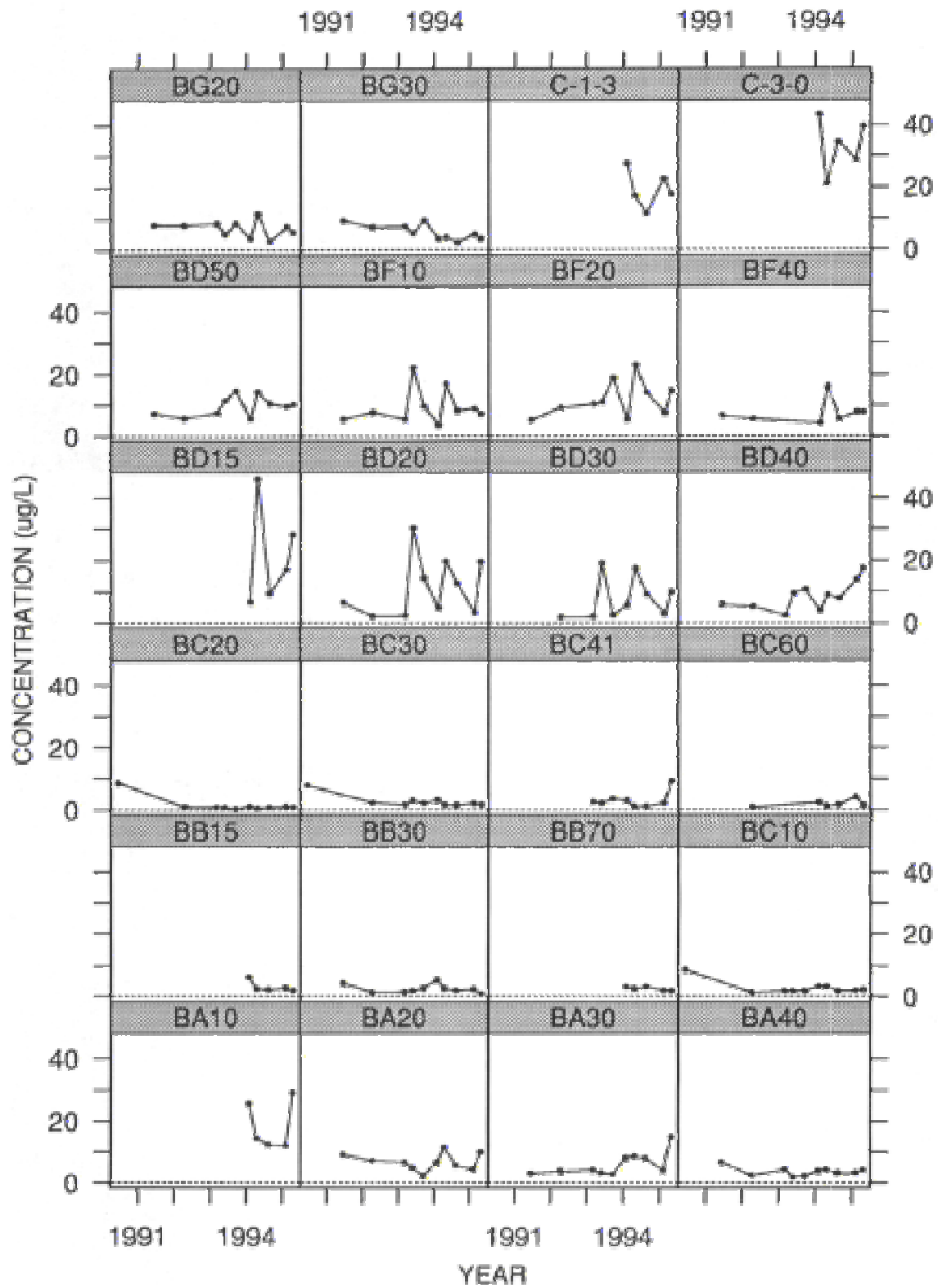
Total Zinc

Figure 7. Trends in trace element totals during 1991-1995. Upright triangles represent uptrends and inverted triangles represent downtrends. Weaker trends (p>0.1) are designated by small triangles, stronger trends (p≤0.1) by large triangles.

Figure 8. Total cadmium vs. Net Delta Outflow. The straight line in each panel is a linear regression fit.

Figure. CONCENTRATION (ng/L) versus NET DELTA OUTFLOW (cfs) for stations BG20, BG30, C-1-3, C-3-0, BD50, BF10, BF20, BF40, BD15, BD20, BD30, BD40, BC20, BC30, BC41, BC60, BB15, BB30, BB70, BC10, BA10, BA20, BA30, BA40.

Figure 9. Time series of total cadmium after removing the effects of Net Delta Outflow (NDO) through linear regression.