Graphical and statistical methods to assess the effect of landfills on groundwater quality. [DNR-014a] 1987

Goodman, Iris

Madison, Wisconsin: Wisconsin Department of Natural Resources, 1987

https://digital.library.wisc.edu/1711.dl/DB6Z464MDFMTW8N

Wisconsin Groundwater Management Practice Monitoring Project
No. 15

GROUNDWATER
Wisconsin's
buried treasure

Wisconsin Department of Natural Resources

GRAPHICAL AND STATISTICAL METHODS

TO ASSESS THE EFFECT OF LANDFILLS ON

GROUNDWATER QUALITY

by

Iris Goodman
Land Resources Program
University of Wisconsin-Madison
1987

ABSTRACT

Wisconsin's groundwater protection law requires establishment of numerical standards for potential groundwater contaminants at landfill sites. For contaminants identified as being of public health or welfare concern, the standard is based on the assessed risk of the contaminant to human health. The standards for groundwater quality indicator parameters are based on changes in these constituents relative to background water quality conditions. Present methods used to set standards for indicator parameters are based on parametric statistical analyses of the data. Statistical characteristics of groundwater quality data may violate assumptions of common parametric methods and lead to incorrect inference. To determine if alternative statistical analyses might be preferable, I tested existing landfill monitoring data for skewness, serial correlation and seasonality and worked to develop meaningful and applicable methods for evaluating inorganic and COD data typically available from landfill monitoring. I found that skewness was most commonly observed, followed by positive serial correlation, and that nonparametric tests for assessing these characteristics performed better than corresponding parametric tests. The recommended procedure for analyzing groundwater quality data was designed with the following objectives: to be applicable to historical data; to evaluate evidence of contamination as shown by constituent concentrations which exceed background conditions; to evaluate evidence that the landfill is the source of the

contamination; and to evaluate whether the contamination is of environmental concern. The approach has the following advantages: it requires few distributional assumptions; it is little affected by missing data and erratic observations; it uses graphical techniques to provide meaningful context for the spatial and temporal variability of groundwater quality constituents; and it does not require extensive hydrologic characterization of a site, for which the existing information is often unavailable or incomplete. The recommended procedure includes standardization of central location and variation using the landfill site median and median interquartile range for each constituent. Graphical representations of the standardized data (boxplots) illustrate distributional characteristics of the data and facilitate identification of potentially contaminated wells by emphasizing their relative spatial variability. Tests for trend (Mann-Kendall) and time-series plots can then identify wells showing temporal degradation of groundwater quality based on data for inorganic constituents and COD.

FIGURES AND TABLES

I. INTRODUCTION

Wisconsin's groundwater protection law requires the establishment of numerical standards for potential groundwater contaminants. The law creates two types of standards, an enforcement standard and a preventive action limit (PAL). The enforcement standard defines a violation, while the preventive action limit functions as an early warning of potential groundwater contamination.

There are two bases for setting the standards. For contaminants identified as being of public health or welfare concern, the enforcement standard is based on the assessed risk of the contaminant to human health. The PALs for these contaminants are statutorily defined as a percentage of the enforcement standard. Thus, the target values for health-based standards are fixed and externally defined.

In contrast, the law requires that PALs for indicator parameters be based on changes relative to background water quality conditions. Further, the law directs DNR to consider how changes in indicator parameters might reflect changes in constituents for which there are health-based standards. Indicator parameters refers to certain chemical constituents which are found in high concentrations in landfill leachate. The concentrations of indicator parameters are not of concern per se; rather they are primarily important in terms of indicating changes in groundwater quality. Many indicator parameters occur naturally in groundwater, such as alkalinity, conductivity and hardness. If changes in these parameters are to be used to determine potential contamination, changes in their concentration due to waste disposal must be distinquished from changes due to natural variability. Thus, the target values for indicator parameters are determined by local hydrogeologic conditions and may vary from site to site or well to well.

To implement these standards, the law requires the Wisconsin DNR to adopt rules for determining compliance. Staff of the Bureau of Solid Waste (BSW) reviewed groundwater quality data from fourteen landfills in order to develop a method for setting PALs for indicator parameters. The BSW concluded that, in general, well-specific standards for each indicator parameter were preferred to PALs that applied to the entire site.

DNR published rules requiring that PALs for indicator parameters be calculated by selecting a minimum of 8 data points from the historical record and computing their mean and standard deviation. The PAL is set at the mean value plus three standard deviations, or the mean plus an increment specified by the DNR. The larger of these two values is used as the PAL for that constituent at that well. The purpose of this thesis was to determine if alternative methods for analyzing groundwater quality data might be preferable.

When assessing the effect of landfills on groundwater quality, three key issues require investigation: (1) evidence of contamination, as shown by constituent concentrations which exceed background conditions, (2) evidence that the landfill is the source of contamination, and (3) evidence that the degree of contamination is of environmental concern. Actual groundwater quality cannot be known; however statistical methods can be used to draw inferences regarding actual groundwater quality conditions and to express the uncertainty of these inferences in terms of probability.

This paper addresses these issues and presents a heuristic procedure for reviewing groundwater quality data within a statistical framework appropriate to the data and to the regulatory issues.

A. CONCEPTUAL AND STATISTICAL FRAMEWORKS

The major difficulty in regulatory analysis of groundwater data lies in reconciling the complexity of the data interpretation problem with the need for regulatory simplicity. Key aspects of the interpretive problem are summarized below:

Policy considerations:

- -- defining the policy decision that would be made if we had perfect information about constituent concentrations and transport.
- -- defining what constitutes a signal of important change in groundwater quality.

Physical setting:

- -- assessing the concentrations of multiple constituents which vary in three dimensions and over time.
- -- estimating the spatial and temporal variability of these concentrations from point samples.

Statistical methods:

- -- assessing the importance of statistical characteristics of hydrologic data which can violate common statistical assumptions.

-- selecting methods applicable to the available data which contain gaps in records, unequal sample sizes, transcription and analytical errors.

-- recognizing the limitations of simple tests of hypotheses.

B. POLICY CONSIDERATIONS

Determining compliance with a numerical standard is straightfoward if the standard is literally taken to mean "not to exceed X mg/L," and if an allowance for measurement errors and variability have been included in the standard. In this case, determining compliance simply requires comparing a statistic which estimates the groundwater quality variable with the standard concentration (Berthouex and Hunter, 1983).

Determining compliance becomes more complicated if: (1) there is disagreement over the measures of, or allowances for variability; (2) if the standard requires a finding of statistical significance between the statistic and the standard; or (3) there is disagreement over the statistical measure used to define a signal of important change in groundwater quality.

The BSW's monitoring data contain both hydrologic variability and sampling variability. For the historical data, these two sources of variabilty cannot be distinquished because there is no information available on sampling or laboratory errors. Other research on major inorganic constituents of groundwater has shown that, in general, hydrologic variability predominates over laboratory and sampling errors (Doctor, et al, Feb. 1985 and EPRI, 1985). In any case, changes in groundwater quality which are statistically significant are those changes which are evident despite the variability, or noise, in the data. In contrast, sampling and laboratory error play a far larger role for most groundwater contaminants of health concern. This is because the health-based standards for many of these substances are set close to or at the analytical detection limits.

When used to describe a statistical result, the adjective "significant" means that the result is statistically d
given the magnitude of the effect to be detected, and the size and variability of the sample data. A statistically discernible effect may or may not be important from a technical or policy perspective. If, however, we were completely certain that the effects which we discern from the data were accurate, we would still require policy criteria to distinquish between effects of minor or major environmental significance.

Several criteria have been suggested for establishing environmental significance (Dunker, and Beanlands, 1986). With respect to the effects being assessed, these criteria include: magnitude, spatial extent, duration, probability of occurence, confidence that the estimated effect has occurred, and the existence of externally derived standards for protection of public health. For this work, I generally adopted DNR's policy that a change in magnitude equivalent to three standard deviations indicated an environmentally important change.

Defining the quantitative measure to be used as a signal of important change is a policy decision. If this decision is to be used as for statistical inference, it must be capable of being translated to a statistical hypothesis. Figures 1 - 3 illustrate three commonly used statistical measures of change or difference. Figure 1 shows a comparison of means; Figure 2 shows a test that a single observation is drawn from an identified population; and Figure 3 shows a test for trend. In each case, the conceptual hypothesis to be tested is that no contamination is evident. The three examples show three different methods of expressing this concept statistically. It is the statistical hypothesis which is tested with the data; the interpretation of the test result with respect to the conceptual hypothesis is affected by policy.

For conventional regulatory purposes, choosing to test for similar means indicates that we are primarily concerned with "average" quality conditions. Such a measure is useful for comparing spatial variability, but is, by definition, relatively insensitive to detecting changes for compliance monitoring purposes. Selecting the second approach indicates that we are concerned with evaluating single observations in terms of the probability of observing each, given that we have defined the population from which we are sampling. Choosing a test for trend over time indicates that we are primarily concerned with detecting temporal changes in groundwater quality conditions.

Figure 1

Comparing Means of Two Samples
(eg., assuming identical distributions, but having
unequal variances)

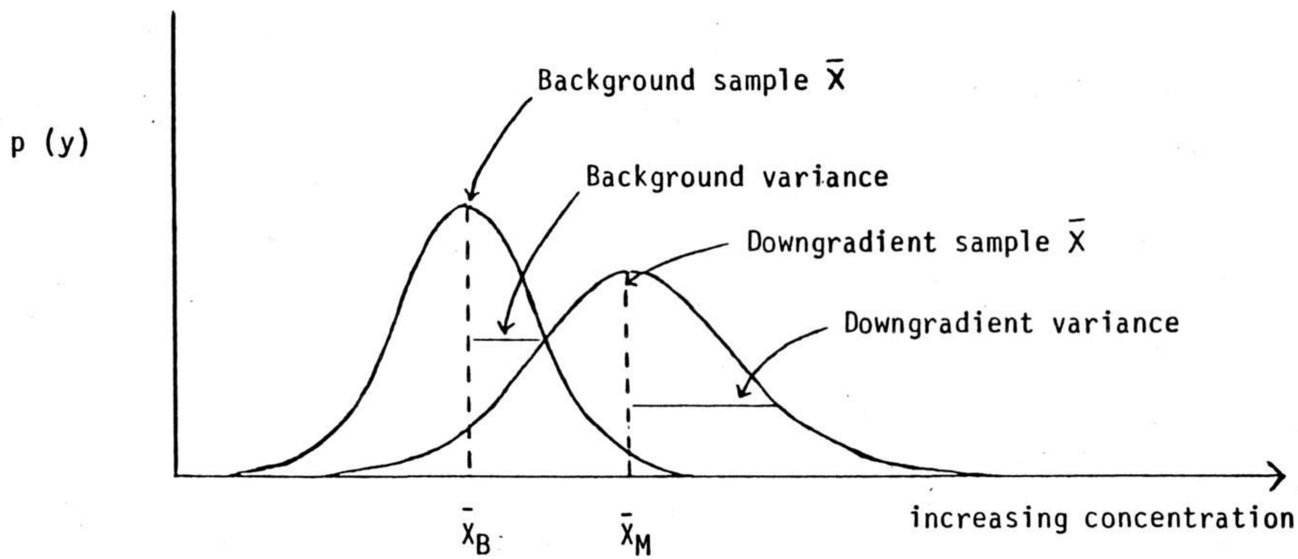
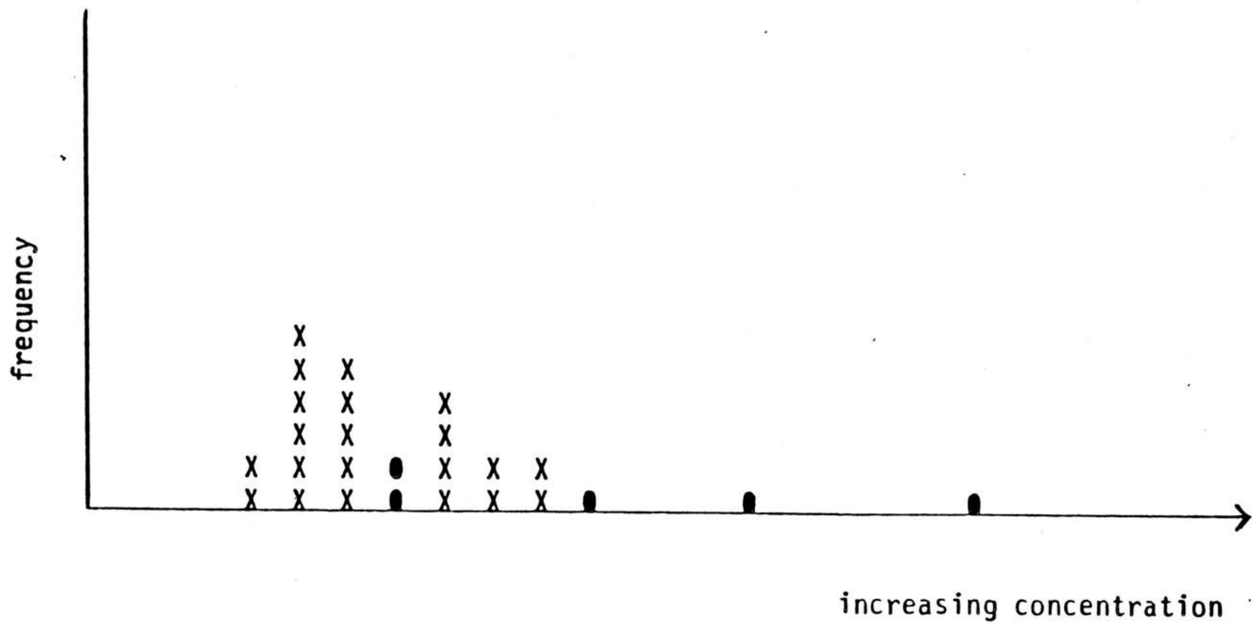


Figure 2

Testing that Single Observation
Comes From Common Population

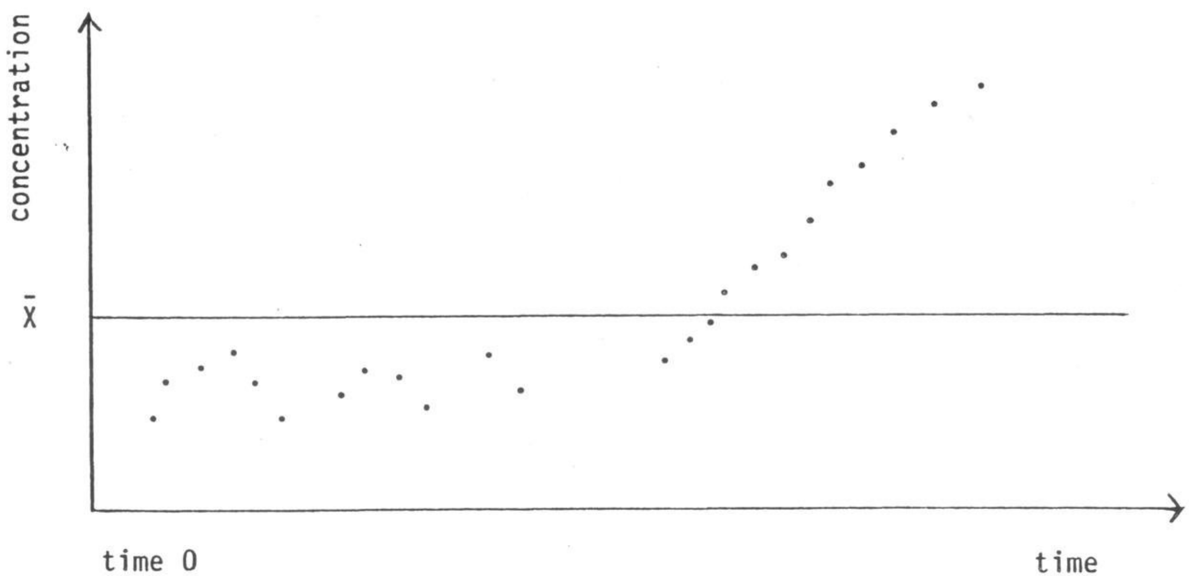


X represents single background observation

● represents single downgradient or "monitored" observation

Figure 3

Testing for Trend over Time



I used all three of these measures to review the historical data because of the strong spatial and temporal aspects of the groundwater quality interpretation problem.

C. LITERATURE REVIEW

Most of the literature regarding groundwater quality interpretation has been in response to the federal rules for the statistical analysis of indicator parameters from hazardous waste landfills (CFR 264.97 (h)). When EPA initially proposed these rules in 1980, they used a very simple conceptual model of groundwater quality as the basis for the statistical test. Because this test played such a dominant role in the early literature, it is briefly summarized below. In addition, the simplicity of the test illustrates several of the conceptual and statistical factors which must be considered in any proposed test.

Figures 4 and 5 illustrate the conceptual framework underlying the method EPA proposed in 1980 for analysis of groundwater quality data for indicator parameters at interim status RCRA facilities. (The minimum requirement for wells are shown; more wells could be included in the analysis). As shown in the figures, the test compares the mean concentration based on one quarter's sample data (and incorrectly assumes that the 4 aliquots, or replicate analyses, are independent observations) with the mean concentration obtained from four quarters of baseline sampling (again, assuming the 16 aliquots are independent observations). The test is repeated each quarter, for each

Figure 4

EPA's t-test for analyzing groundwater quality data

H_0: Downgradient mean $\leq$ upgradient mean

H_1: Downgradient mean $>$ upgradient mean

If the test results in rejecting the null hypothesis in favor of the alternative hypothesis, the regulatory policy requires investigation of the landfill as the cause of the increase in the downgradient well mean.

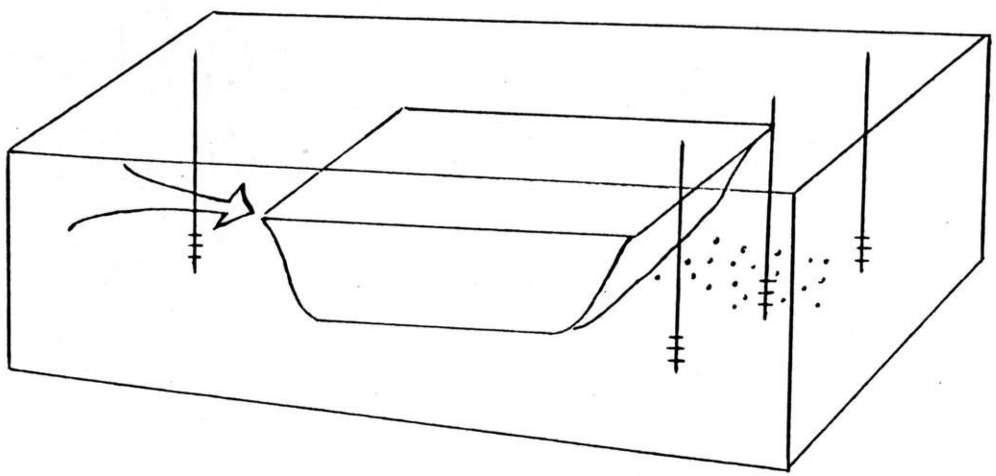
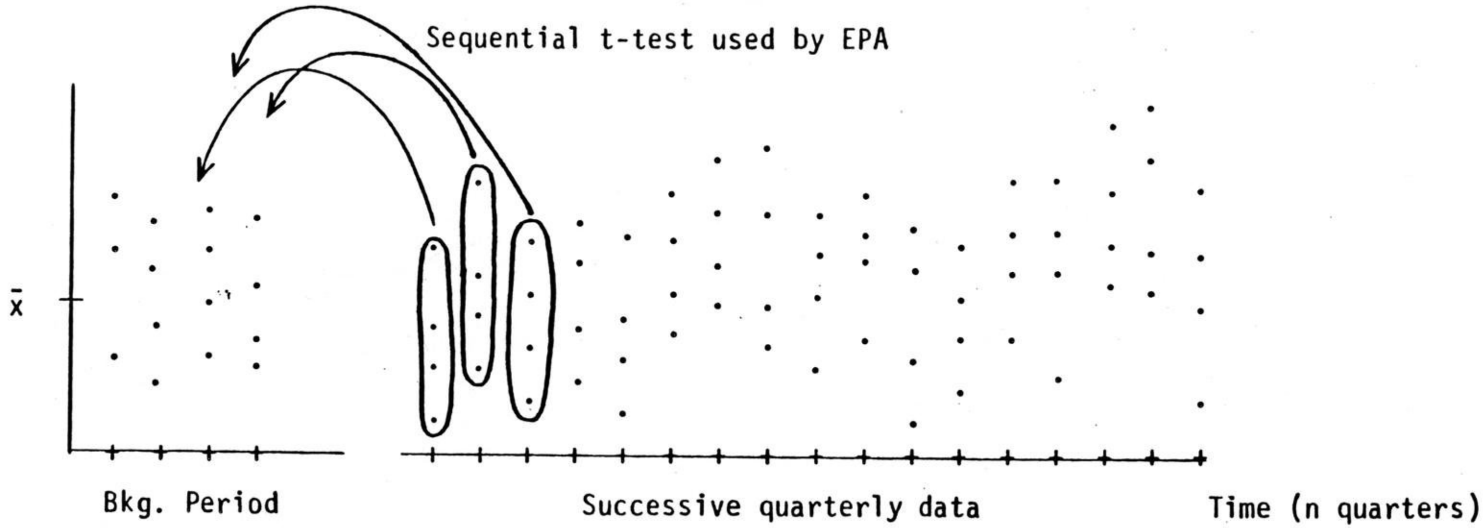


Figure 5

Sequential t-test used by EPA



Each new quarterly mean is compared to the original background mean. The test is repeated each successive quarter for each indicator parameter, for each well. The significance level for each individual test is .05%.

The test statistic, $t = \dfrac{\bar{x}_m - \bar{x}_b}{\sqrt{\dfrac{s_m^2}{m} \quad \dfrac{s_b^2}{n}}}$

where: $\bar{x}_m$ = UG well mean; $\bar{x}_b$ = DG well mean; m = #obs. DG well; n = #obs. UG well

s_m^2= variance DG well data; s_b^2= variance UG well data constituent, at each downgradient well. The individual downgradient well means are compared to the first year of baseline sampling for one or a few upgradient wells. The test has no provision for updating the original baseline estimate.

In proposing the use of a modified t-test , EPA set forth the concept that the appropriate measure of important changes in groundwater quality are those changes sufficient to significantly alter the mean concentrations.

Conceptually, the null hypothesis to be tested is that there is no evidence of groundwater contamination as shown by indicator parameters. For a one-sided test of means, the statistical hypothesis is:

H_o: downgradient mean is less than or equal to the upgradient mean

H_1: downgradient mean is greater than the upgradient mean.

or,

H_o: $u_{DG} \leq u_{UG}$

H_1: $u_{DG} > u_{UG}$

Snedecor-Cochran's approximation to the t-test is used for this comparison. The policy assumption is that a statistically significant difference is regarded as evidence of contamination. Establishing this difference, of course, means that the difference observed cannot be attributed to random variation in the groundwater constituents. Thus, estimates of constituent variability are included in the test statistic.

This application of the t-test has been widely criticized for a number of reasons: (1) potential violations of statistical assumptions of normality, independence and random sampling, (2) low power if assumptions are violated (Groenveld and Duval, 1985), (3) an ever-increasing false positive rate as the test is repeated over the duration of the landfill's operation (Unwin and Miner, 1985), and (4) incorrect comparisons of variability in space, over time, and of variability associated with chemical analyses (Silver, 1986).

The key conceptual flaws in this approach relate to inappropriate interpretations of spatial and temporal variability. These are (1) the interpretation that a significant difference between the means of wells in different locations is due to the landfill, rather than to inherent spatial variability or strata characteristics, and (2) the inappropriate comparison of analytical variability with temporal variability.

Alternative statistical methods were subsequently proposed. Virtually all the alternatives proposed have two features in common. First, the methods are prospective in nature; i.e., they are intended to be applied to a new landfill site with reliable baseline water quality data and well defined site hydrology. Second, the methods continue to assume the rigid upgradient/downgradient dichotomy first developed in the simple RCRA model (Doctor, et al, May 1985 and Plumb and Fitzsimmons, 1984). Under this assumption, only upgradient wells are used for estimating variability under conditions of no contamination as a reference for testing for the presence of contamination. A common result is that population characteristics are imprecise since they are based on few observations in time or space.

Finally, most of the literature is conceptual, in that there there is scant reporting on the statistical characteristics of actual data or of the application of the proposed methods to actual data.

D. RESEARCH METHODS

My research focused on reviewing existing data to clarify the conceptual framework for analysis and to assess the statistical characteristics of the data.

Since about 1974, the BSW has maintained a central computer tape for all the quarterly reports from landfills which require groundwater monitoring. The tape contained water quality data from 230 landfills and had over 440,000 records. The central data tape had never been comprehensively reviewed for research purposes.

The primary tasks were to: (1) select a subset of constituents for analysis; (2) select a subset of landfill sites; (3) devise a method for viewing the data in order to put its variability into context; (3) select appropriate statistical tests; and (4) compare the responses of indicator parameters and health and welfare constituents. Each of these tasks are described below.

1) selecting constituents: The tape contained no information regarding the constituents monitored at each site. I devised a method to screen the tape for constituents and found that seven constituents--chloride, COD, pH, alkalinity, hardness, conductivity and sulfate--

comprised over half of the observations recorded on the tape. Of these, chloride and sulfate are legally defined as "welfare" parameters and so would provide a basis for comparison with indicator parameters. Moreover, chloride and sulfate are mobile anions which are often used as tracers in groundwater studies (Kammerer, 1984, McFarlane, et al, 1983, and Dance and Reardon, 1983).

2) selecting landfills: The selection of landfills for analysis required three iterations. The tape had no information regarding the location of the wells relative to the flow field at the site. Consequently, I initially selected for analysis landfills for which DNR staff had already assembled this hydrologic information. A review of these sites showed that the data were too sparse for testing statistical characteristics or developing a test procedure. I then screened the tape for landfills which had at least several wells with 16 observations for 4 to 7 constituents. Sixteen landfills were identified and I began plotting and reviewing the data. An unintended result of this screening criteria was that it selected only older, usually unengineered sites which often had little supporting hydrologic information. I was thus unable to hydrologically confirm the cases of contamination evident by graphical analysis, nor could I study the behavior of groundwater at engineered sites. Finally, I selected a cross-section of landfill designs. I chose nine sites which had at least two years of data available for the selected indicator parameters, seven of which the DNR had sampled for volatile organic compounds (VOCs). These landfills are listed in Table 1.

TABLE 1
SUMMARY OF LANDFILLS REVIEWED

Name/Number	Design	Design Volume (m. cu.yds.)	Site Size (Acs.)	Waste Rec'd.	Date Filling Began	No. of Wells
Eau Claire 2821	Clay lined, leachate collection	1.2	24	MSW	1978	8
Ft. Howard 2372	Unlined	4.5	293	Paper Sludge	1964	25
Janesville 2822	Partial clay lined; partial unlined	0.7	18	MSW, IND	1961	17
Lacrosse 2637	Natural attenuation	1.38	55	MSW, IND	1976	12
Marathon 2892	Clay lined, leachate collection	1.5	10	MSW, IND	1980	6
Oakcreek (WEPCO) 2357	Unlined	4.0	130	Fly ash	1975	18
Sauk (old) 2051	Natural attenuation	1.0	14	MSW, foundry waste	1973	12
Verona 2680	Natural attenuation; partial leachate collection	2.0	49	MSW	1977	25
Wausau 2875	Mostly clay lined, leachate collection	0.25	4.8	Paper sludge	1981	5

Source: Feasibility studies and DNR file information

Note:
1) No. of wells refers to number of wells with sufficient data on tape to review.
2) IND refers to industrial waste.
3) MSW refers to municipal waste.

VOCs have been studied for use as a potential indicator parameter (Plumb and Pitchford, 1985). Thus, the VOC data provided another point of reference for determining the presence of contamination; in effect, I used the VOC data as an independent check on the inferences possible from the conventional indicator parameters.

3) Selecting appropriate statistical tests:

My statistical review of the data had two objectives: (a) to assess the statistical characteristics of the data, and (b) to assess the presence of contamination at a given well. These two objectives required different statistical methods.

The tests used to assess the presence of skewness, seasonality and serial correlation are described in the following section. The graphical techniques and trend tests used for assessing contamination at a given well are discussed in the section on Procedures.

4) Comparing responses of constituents: Comparisons between apparent contamination by VOCs and contamination evident based on analysis of COD and inorganic data are discussed in Conclusions.

II. STATISTICAL CHARACTERISTICS OF HYDROLOGIC DATA

A. Concepts

Hydrologic data often exhibit skewness, seasonal variation and serial correlation. The presence of these characteristics violates some of the assumptions necessary for strictly valid application of many commonly used statistical methods. In practice, all assumptions are rarely met. To the extent that several key assumptions are violated, the resulting inference is incorrect. Specifically, the Type I error probability level used to guage the strength of the statistical inference will be incorrect. One goal of my research was to assess the prevalence and extent of these statistical characteristics in the BSW's data base.

Skewness refers to how the shape of a distribution departs from symmetry about the mean value. The presence of skew in data violates the statistical assumption of normality. The general effect on a test of inference based on normality is that the probabilities associated with values in the tails of the distribution may be greatly in error; however, it is these tail probabilities where our intererst often lies. The specific effect of skewness on inference depends on the particular hypothesis and test statistic used. For example, when testing for similar means by use of the t-test, skewness inflates the variance estimate (relative to the assumed Gaussian, or normal, variance) used in the test statistic and increases the probability of false negatives.

Skewness in data is generally handled by transforming the data to make it more nearly symmetrical, choosing a probability distribution that more closely fits the observed data, or using non-parametric statistical methods which do not require the assumption of normality for establishing the critical region.

Seasonal variation in data refers to the presence of different underlying populations during different seasons of the year (Hirsch, et al, 1982). Possible hydrologic causes of seasonality include intermittent recharge of contaminants located near the ground surface or above the water table in hydrologic systems which are affected by direct infiltration. Seasonal changes in concentrations of constituents can also be due to changing conditions within the aquifer itself; eg., shifts in carbonate chemistry due to solution of limestone or dolomite by the seasonal influx of water high in carbon dioxide. The presence of seasonality violates the statistical assumption that groundwater observations at a single well are identically distributed. The general effect of seasonality on inference is that it confounds estimation of the random variation present, because it itself is a source of variation.

The confounding effects of seasonal variation can be addressed by (1) estimating the seasonal effects and removing them from the data or (2) by making comparisons only within a given season (i.e., within similar populations). Neither of these approaches are currently easily applied to typical groundwater quality data sets. The first approach is difficult because there are typically only several years worth of data available with which to estimate the seasonal effects. The second approach often results in a sharp decline in the discriminating ability of the test statistic to detect an actual change. This decline, or loss of statistical power, is because of the few number of observations per season in short data records (eg., data records of less than ten years) (Hirsch and Slack, 1984).

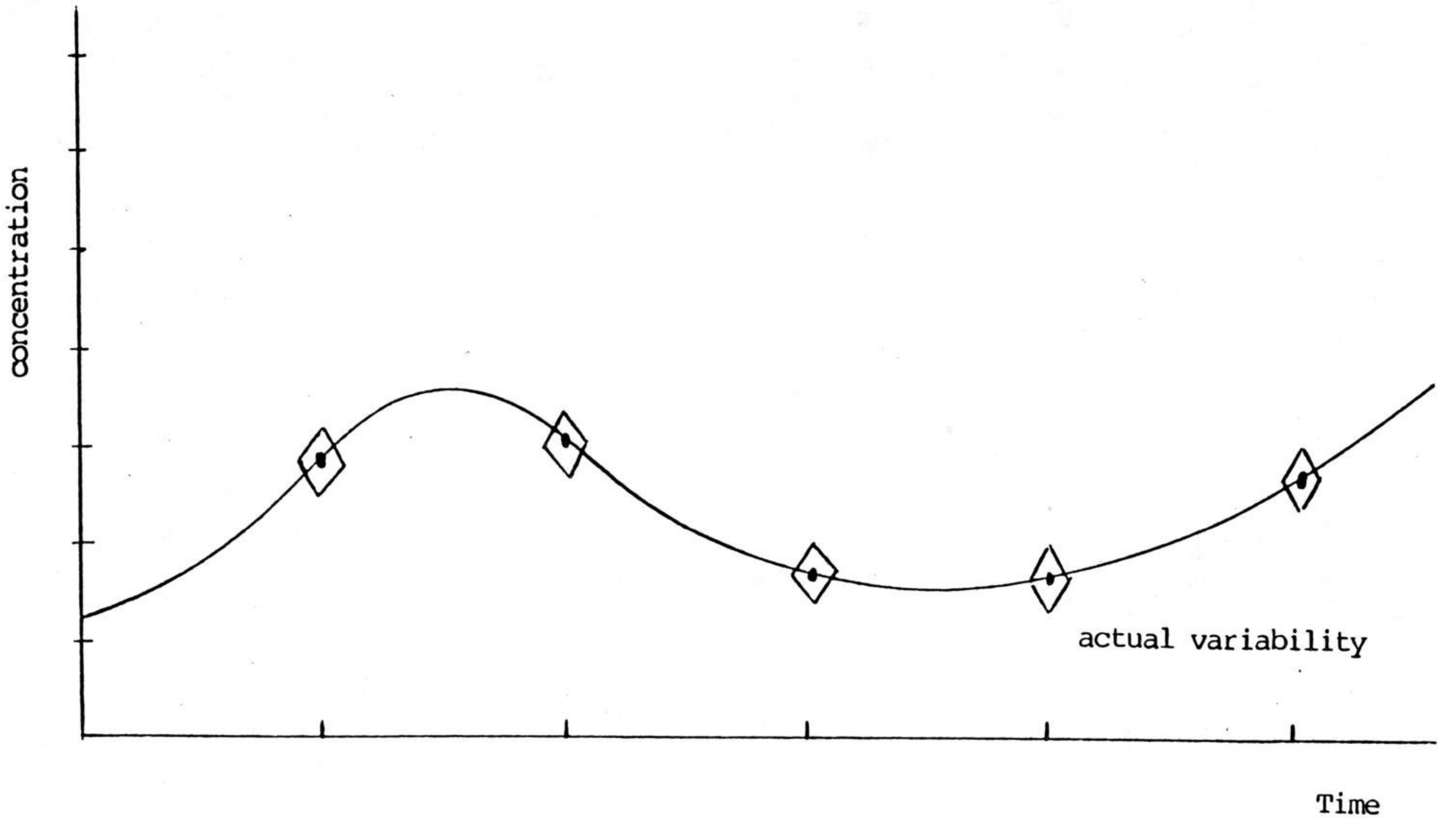
As a hydrologic concept, serial correlation refers to the "memory" of a system with respect to past influences acting upon it.

The low velocity of groundwater flow could be expected to act to integrate, or smooth out, sudden peak concentrations due to inputs of contaminants. As a result, high values may tend to follow high values, and low values to follow low values. The presence of serial correlation, as observed in the data, is a function of sampling frequency relative to the system characteristics. Serial correlation violates the statistical assumption of independence of the random variable, that is, the groundwater constituent observations.

When applicable, the assumption of independence greatly simplifies the derivation of the underlying probability models (Benjamin and Cornell, 1970). The presence of serial correlation has several effects on statistical inference. The general effect is that serial correlation obscures true variability because the sample we observe contains only a portion of the total variability present. This is illustrated in Figure 6. One result is that the sample size which is effectively available to estimate population parameters is reduced (Lettenmaier, 1976). Similarly, it is difficult to distinquish serial correlation from trend in the data for short periods of record (Hirsch and Slack, 1984). Simulations of tests for differences in mean values have shown that positive serial correlation inflates the false positive rate (Box, et al, 1978).

There are essentially two methods for accomodating serial correlation within statistical analyses. The first is to adjust the sampling frequency to remove correlation between observations (Loftis and Ward, 1980). Alternatively, conditional or joint probabilities

FIGURE 6
SERIAL CORRELATION AND SAMPLING FREQUENCY



Tick marks represent timing of samples
◊ represents data value obtained for each sample

can be calculated to correct the significance level of various tests. This approach generally requires additional assumptions and more sophisticated statistical methods (Hirsch and Slack, 1984 and Lettenmaier, 1976).

B. Methods:

Statistical inference can be used to test the existence of each of these hydrologic characteristics. We are primarily concerned with the presence of these characteristics in wells where contamination is absent, because the presence of contamination makes it difficult to separate artifact from genuine characteristics. For example, the presence of genuine trend results in "findings of skew"; erratic values can result in findings of "seasonality"; and serial correlation can appear as "trend."

As discussed in the section on Procedures, graphical and trend analysis provided the basis for partitioning the data into subgroups of clean and contaminated, thus enabling testing of the statistical characteristics of data at the clean wells. Since the data records at many of these wells are quite short (eg., 14-26 observations), it is difficult to statistically verify the presence of these characteristics. In order to improve the test inferences, I used both a parametric and nonparametric test to assess each characteristic.

1. Skewness:

a.) Skewness coefficient: I used the parametric skewness coefficient to determine lack of symmetry, or skew, in sample data at individual wells. The power of the skewness coefficient to detect skew is greater than alternative tests such as chi-square goodness of fit test (Shapiro, et al, 1968). In addition, statistical quality control techniques can be modified for compliance testing of nonnormal data if the skewness coefficient is known (Burr, 1967).

The sample estimate of the skewness coefficient is the ratio of the second and third (sample) moment about the mean. This coefficient is independent of scale, and is given by:

$$g_1 = m_3/(m_2)^{3/2}$$

where

$$m_3 = \sum (x - \bar{x})^3/n$$

$$m_2 = \sum (x - \bar{x})^2/n$$

(Snedecor and Cochran, 1980)

The sign of the skewness coefficient indicates the location of extreme deviations from the mean of the sample data; i.e., a negative coefficient indicates that extreme observations tend to be less than the mean, a positive coefficient indicates that extreme observations tend to be greater than the mean. For symmetric, Gaussian data, the coefficient is zero. Thus the size and sign of the coefficient indicates the direction and extent of skew in the data (Snedecor and Cochran, 1980).

The skewness coefficient calculated from the sample data is compared with tabulated values for the significance level of the test and sample size used. Values of skewness coefficients for small sample sizes typical of groundwater quality data sets have been tabulated by Harris, et al. (1987). I used a one-sided test for positive skew.

b.) Boxplots: As described in the section on procedures, boxplots are a visual method of informally assessing the skewness of sample data (Velleman and Hoaglin, 1981). This nonparametric graphical technique is based on order statistics from data at individual wells. Although the skewness suggested by boxplots cannot be quantified for summary purposes, these plots are useful for quickly assessing if extreme skew is present for a particular constituent at a particular well. I found that, in general, extreme skew as evidenced by boxplots is often due to trend in the data, or sample contamination.

2. Seasonality:

a.) Kruskal-Wallis multiple comparison: The nonparametric Kruskal-Wallis test is a multi-sample test used to compare K-populations for differences in mean population values. The objective of the Kruskal-Wallis multiple comparison is thus similar to the F-test used in the analysis of variance. The Kruskal-Wallis test, however, does not require normality of the populations as a background assumption. The relative efficiency of the Kruskal-Wallis test to the ANOVA F-test is not less than .864 when the assumption of normality is met, and can be much higher if the data are non-normal (Bradley, 1968). In addition, the Kruskal-Wallis test is one of the few nonparametric multi-sample tests which do not require equal sample sizes (Miller, 1972).

To adapt the test for assessing seasonality in quarterly groundwater quality data, observations taken during different quarters (or seasons) are used as the populations to be compared. The null hypothesis is that all four populations are identical. The alternative hypothesis is that the populations do not all have identical means (Conover, 1980). In order to obtain the most reliable inference, the analysis was generally limited to individual wells having at least four observations per season.

For comparisons of four groups (eg., four seasons), the distribution of the Kruskal-Wallis test statistic is approximately chi-square, with k-1 degrees of freedom. Tabulated probabilities are given in Lehmann (1975). (For the Ft. Howard site, the data for many of the constituents were biannual, thus representing only two seasons. For these constituents, we used the non-parametric Mann-Whitney test, which is the two sample equivalent of the Kruskal-Wallis test.)

b.) Autocorrelation functions: For data in a time series, the parametric autocorrelation function (ACF) measures the degree of linear dependence between successive observations. The sample ACF coefficient, r_k estimates the population coefficient, ρ_k and can be calculated for any time interval (also known as lag-k) between the observations. As discussed below, the sample ACF coefficient can be used to assess both seasonality and serial correlation in the time series data.

The sample ACF, r_k is defined as:

$$r_k = \frac{\sum_{t=1}^{N-K} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

where: N is the number of observations

K is the lag between observations

x_t is the observation at time t

$\bar{x}$ is the mean of the time series.

(Salas, et al, 1980).

For time-series data at a single well, each sample lag-k coefficient is a random variable. Values of the sample coefficient can vary from -1 to +1 for any k-lag. For a completely random process, the distribution of the sample ACF is approximately Gaussian with zero mean and variance of 1/n. Thus, the 95% confidence interval on the population ACF coeffficient is the value of the coefficient + $1.96/\sqrt{n}$ (Harris, et al, 1987). The sample ACF coefficient can be tested for significance by seeing if it falls within this interval.

If the observations at a single well are independent, there will not be any pattern to the coefficient values, nor will any of the coefficients be significant. Large positive coefficient values indicate a tendency for a large observation to follow a large observation, k - lags away. Conversely, large negative coefficient values indicate a tendency for alternating large and small values within the time-series (Harris, et al, 1987). To assess the presence of seasonality, I tested for significant lag-4 coefficients.

3. Serial dependence:

a.) Autocorrelation function: Using the same ACF coefficients described above, I tested for positive serial correlation in the data at each well using the lag-1 sample ACF coefficient. If significant, I concluded that the data were positively serially correlated.

b.) Runs test: The runs test is a simple nonparametric test for randomness of data in a time series. In this application, the test determines whether the sample data tends to "cluster" in a sequence rather than to vary randomly over time (Bhattacharyya, 1984).

For data in a time series, there are a finite number of ways in which the observations can vary above or below some criteria value. For the groundwater quality data, I used the median value at the clean wells to divide the time series into two segments: those values which cluster in a sequence above the median value and those which cluster below it. These sequences are known as "runs" of data observations. The number of runs observed in the data is a random variable.

The exact probability distribution of all possible combinations of these runs can be derived exactly or approximated for large samples using a Gaussian distribution. The probability associated with an observed number of runs indicates the probability of observing by chance such a sequence in the data. Tabulated significance levels for the runs test for small sample sizes are given in Draper, 1981. For larger sample sizes, the approximated significance level is computed and reported by the computer software package.

I used a one-sided runs test to test:

H_0: observations are random or too many runs present.

H_1: too few runs present.

"Too few" runs suggests positive serial correlation in the data, while "too many" runs suggests negative serial correlation.

C. Results:

Table 2 summarizes the statistical characteristics of the data for uncontaminated wells across all landfill sites. The results are aggregated by constituent, in order to assess possible differences in their statistical behavior. It should be noted, however, that the presence of any one of the characteristics may be site or well-specific. For example, almost half of the positive significant results for "runs" in conductivity data were from the Ft. Howard site; four other sites did not have any significant results for the runs test for conductivity.

As shown in the table, positive skewness is the most prevalent characteristic tested and was found in 129 of 316 tests. COD showed the greatest incidence of positive skewness (79%). This is likely related to the tendency for COD results to be affected by sample contamination. pH, in logarithmic scale units, showed the least positive skew (2%), while alkalinity showed positive skew in 15% of the tests.

Serial correlation is the next most prevalent characteristic

TABLE 2

STATISTICAL CHARACTERISTICS OF THE GROUNDWATER DATA:

RESULTS FOR UNCONTAMINATED WELLS

CONSTITUENT	SKEW	SERIAL CORRELATION ACF:	SERIAL CORRELATION RUNS TEST:	SEASONALITY ACF:	SEASONALITY KRUSKAL WALLIS:
CHLORIDE	27/53 (.51)	8/52 (.15)	18/53 (.34)	1/52 (.019)	2/49 (.041)
COD	42/53 (.79)	9/52 (.17)	19/53 (.36)	1/52 (.019)	0/49 (0)
pH	2/52 (.04)	5/53 (.09)	5/53 (.09)	0/53 (0)	0/49 (0)
ALKALINITY	4/27 (.15)	6/26 (.23)	6/28 (.21)	2/25 (.08)	2/25 (.08)
HARDNESS	18/44 (.41)	11/42 (.26)	16/44 (.36)	0/43 (0)	2/41 (.05)
CONDUCTIVITY	16/53 (.30)	12/50 (.24)	11/53 (.21)	0/51 (0)	0/51 (0)
SULFATE	20/34 (.59)	5/29 (.17)	10/34 (.29)	0/29 (0)	3/29 (.10)

Notes:

1) All tests are one-tailed, except for the Kruskal-Wallis test for seasonality which is two-tailed. Significance level for all tests is 0.05. Fraction shows number of significant results relative to total number of tests conducted.

2) Decimal values within parentheses show percentage of tests which were significant, for comparison with the .05 expected to be significant, based on choice of alpha.

tested, based on the results for autocorrelation functions and for the runs tests. On the basis of both tests, serial correlation is least evident for pH and alkalinity. It is interesting to note that the simpler and theoretically less efficient runs test generally resulted in more findings of serial correlation than did the parametric autocorrelation function. This may be due in part to missing data, which distorts the calculation of the sample autocorrelation coefficient.

Seasonality is the least evident characteristic of the sample data. The non-parametric Kruskal-Wallis test resulted in more findings of seasonality relative to that evident based on the lag-4 sample autocorrelation coefficient. Distortion of the coefficient due to missing data may again be a contributing factor. Based on the Kruskal-Wallis test results, seasonality might exist for chloride, alkalinity, hardness and sulfate. Chloride might be expected to fluctuate seasonally, if present in the leachate and if the well is vulnerable to seasonally high infiltrations. Alkalinity, hardness and sulfate are are non-conservative constituents and thus may be expected to show seasonal fluctuations.

III. PROCEDURE

We have seen that the conceptual null hypothesis to be tested is that of no contamination. There are however, several plausible reasons why the data may not conform to our expectations under this hypothesis. Each of these alternative reasons warrants consideration, since it should perhaps be considered more credible than the general alternative hypothesis of contamination. Deciding among these alternative possibilities is more difficult when the available data do not predate landfill operation and have been collected over different time periods.

In practice, we can frequently test the null hypothesis against only one alternative hypothesis at a time. It is the inability of a single statistic to differentiate among all possible alternative hypotheses that renders the result of a given test premliminary rather than conclusive. This fact poses a conflict between the need for regulatory simplicity and the complexity of the problem at hand.

In 1953, a quality control statistician at Bell Laboratories summarized a similar situation. He used the familiar analogy of drawing lots, or samples, from a bowl and using this sample information to address four statistical questions regarding description, selection of probability models, inference and quality control. Comments to provide context are added in brackets.

> "Given a set of observations about which we know everything [i.e., we can describe the sample data in terms of its central value and its dispersion]; was there a bowl? Let me say this another way: given what is supposed to be a sample from a single universe, could

> this reasonably be a sample from a single universe, or should it be treated as a set of subsamples from several universes with perhaps a few wild shots thrown in? . . .
>
> The working engineer . . . must tackle these problems in the reverse order [from that used to develop statistical theory]. He never knows what is in the bowl; he is never even sure that there was a bowl. He only has the numbers that he hopes form a sample. He must first try to determine whether or not they belong to a universe. If not all of them, then do most of them--and which ones? Only after he has answered this fourth question does the third [the question of statistical inference] make sense. And only after he has gotten the parameters can he hope to solve the second problem [of selecting a probability model] and make a prediction. He must of necessity start at the bottom and work up, whereas the mathematician starts at the top and works down." (Ferrell, 1953).

For similar reasons, I developed a procedure which combines quantitative and visual inference to first get a look at the universe of constituent data; to determine which constituent values likely do not belong and why; to determine if subgroups of constituent populations are necessary or helpful; and finally, to determine whether an individual well is contaminated. Implications for future work which might enable the use a very simple quality control technique for determining compliance at individual wells is also discussed.

A. OBJECTIVES IN DESIGNING PROCEDURE

Interpreting the data is made easier by the use of some simple graphical tools and tests for trend. Graphical methods are preferred because they better assist in the interpretation of test results than do tables of raw data. This is especially true when the data are meaningful primarily in the context of time and a physical hydrologic setting. I designed the following graphical procedure with the following objectives:

1) The procedure was designed for the analysis of existing landfill sites (those having at least 2 years of data available) rather than with a purely prospective orientation. This approach reflects the majority of landfill sites which DNR regulates and enables good use of previous investments in data collection.

2) The procedure was designed to create a framework for hierarchical review of the data. This framework assists DNR staff in getting the overall picture first and continues with guidelines for working down to whatever level of detail is required by the data; not necessarily the complexity of the site. A major advantage is that staff need not review all data by eye.

3) The procedure was designed to create a set of numerical and graphical summary statistics which are useful as a basis for discussing the data with site operators. An advantage of the procedure is that where the evidence of contamination is strong, it is unassailable. Where the evidence is uncertain, DNR has objective information with which to negotiate with site operators regarding the need for additional sampling or wells to clarify the uncertainty.

4) The procedure was designed so that the results are largely objective and reproducible. That is, while hydrogeologic judgement is needed to select the wells (and hence background conditions) to be compared, quantitative guidelines for selecting these wells are provided.

5) The methods are designed to be resistant to the erratic data values which are prevalent in historical groundwater quality data. In addition, the methods are robust against distributional assumptions, and thus provide assurance of relatively good performance across a broad range of distributions.

6) The procedure combines elements of the statistical concepts discussed in the literature. It is the combination of these techniques that strengthens the final conclusion regarding contamination against invalidation by violation of a single statistical assumption or other gap in available information.

B. DATA STANDARDIZATION

1) Purpose:

The first issue in assessing groundwater quality is to determine if there is evidence of contamination, as shown by constituent concentrations which are more extreme than we would expect under conditions of no contamination. We use as our frame of reference the variability we expect when no contamination is present. We need a picture of variability in constituent concentration over space and time so that we can estimate what is expected random variation for each constituent, for each well at each landfill.

A monitoring well represents a single sampling point. It is the sum of the water quality conditions at the well that indicates whether contamination is present. Since we use the response of several constituents to assess contamination, we need to have comparability across constituents, within the landfill site. Standardization is simply a technique for obtaining a common scale for comparison by "transforming" the raw data using a centering and scaling factor.

2) Theory: Finding a good measure of "center" and "scale" for groundwater quality data.

Conventional parametric estimates of mean and standard deviation are very good if the data are reasonably Gaussian and if they are fairly clean (i.e., they contain few transcription errors and few unreliable values). However, neither the mean or standard deviation is resistant to extreme values, nor is the standard deviation robust to departures from the Gaussian distribution (Bradley, 1968 and Hoaglin, et al, 1983).

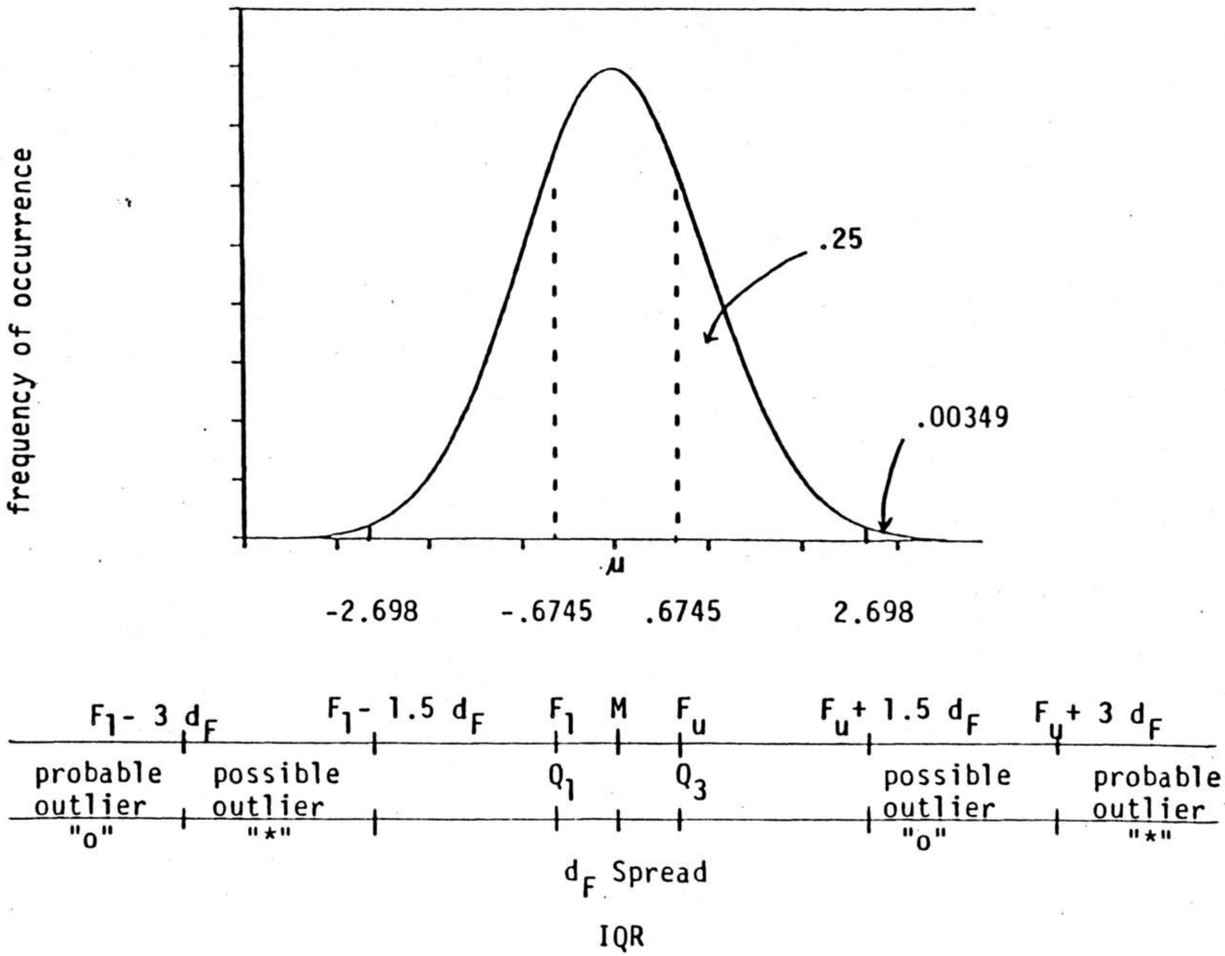
There are several non-parametric analogues to the mean and standard deviation; these also estimate the center and dispersion of the data. Two such estimators are the median and F-spread (which is essentially the same as the interquartile range, or IQR) They have the advantage of being relatively efficient estimators while also being resistant and robust estimators (Hoaglin, et al, 1983).

Because the median and IQR figure prominently in the recommended procedure, they are introduced here in the context of a Gaussian model. Figure 7 shows the relationship of the median and IQR relative to a Gaussian distribution. Thus we see that the upper quartile (Q_1) defines a spread from the mean equivalent to about +.6745 standard deviation units. Similarly, we see that the probability in each of

Figure 7

Estimates of Center and Scale:

Nonparametric Analogs to the Gaussian Model



the tail areas above and below the quartiles is .25. By convention, observations falling 1.5 IQRs above or below the quartiles are denoted as possible outliers. This corresponds to a standard deviation of +2.698, and an upper tail probability of .00349. (For finite samples, the average number of observations that fall beyond this cutoff can be approximated by .00698 + 0.4/n) (Hoaglin, et al, 1983).

As shown, the F-spread or IQR for the Gaussian distribution is 1.349 standard deviations. We can estimate a standard deviation unit based on this relationship. Thus, the F-psuedosigma is defined as:

$$\sigma_{F\text{-}ps} = F \text{ spread}/ 1.349$$

This analogue to the standard deviation is especially useful when the data may be near-Gaussian, except for a few erratic values. This situation appears to describe much of the groundwater quality data for uncontaminated wells (eg., the pronounced skew evident in COD data is due in part to apparent sample contamination incidents which cannot always be defensibly deleted from the data record).

3) Method:

Each sample observation observation is centered by the median concentration for that constituent at the site, and scaled by the median interquartile range (upper quartile - lower quartile) for that constituent at the site. Figure 8 shows the sequence of steps. The graphical analyses are conducted on these non-parametric standardized values which are given by:

$$\text{NP value} = (\text{obs.} - \text{median}_{\text{site}})/\text{IQR}_{\text{site median}}$$

Figure 8
Non-parametric Standardization of Data
for Graphical Analysis

Assume site has 21 wells and a total of 315 observations for chloride

Steps:

1) To find central location for whole site:

$$X_1\ X_2\ X_3\ .\ .\ .\ X_{315}$$

2) To find measure of scale for each well:
--rank from minimum to maximum value

--find lower and upper quartile

--take difference ⇒ measure of scale for each well

--find this scale measure for each well

⇒ 21 IQR scales

--take median IQR ⇒ 11th IQR scale

3) Standardize all chloride observations using site median and median interquartile range.

C. ADVANTAGES OF STANDARDIZATION

Use of these standardized values confers many advantages:

1) The NP value describes each observation relative to (i.e., above or below) a resistant estimate of the constituent's central location and variability for that site (i.e., the median value for that constituent for the whole site). Thus, an NP value of 0 represents a concentration equal to the site median for that constituent and probably represents uncontaminated conditions. An NP value of +10 means that the observation is more than 10 non-parametric scale "units" above the median value; -10 is more than 10 units below. As an approximation to the standard deviation, a one unit change in the NP value is equivalent to about a .74 unit change as measured in standard deviations. Thus, 5 NP units represent about the same spread as 3.7 standard deviations.

2) The NP value enables the use of a single number to describe the relative variability of <u>all constituents</u>. This NP value can easily be converted to absolute concentration values:

Absolute concentration = NP value x (IQR) + site median.

3) For a <u>given</u> constituent, the NP values are directly comparable between wells.

4) For <u>different</u> constituents, the NP values are comparable in that they reflect the relative variability of the different constituents at

the site (i.e., a +10 for chloride is the same as a +10 for COD). This is especially useful for comparing the response of each constituent at a single well location within the context of the variability for that constituent at that site.

5) NP values enable plotting as many constituents as desired on a single time series plot. Thus, the behavior of all the constituents at a single well can be viewed simultaneously despite differences in units or absolute magnitudes.

B. BOXPLOTS: an economical way of displaying data which shows the median value and its associated 95% confidence interval, the spread of the data, and which highlights extreme values.

1) Example:

```
         -----------------
    -----I    (  +  )    I----        *  *  *     0
         -----------------
```

2) Explanation of symbols:

--all symbols represent data values (i.e., it is an empirical, rather than an assumed, distribution)

--the + shows location of the sample median value

--the ( ) shows the approximate 95% confidence interval about the population median (the interval is estimated as +/- 1.58 x IQR/$\sqrt{n}$).

--the I I show the upper and lower quartile values and bracket the central 50% of the data

--the ----- extends to the most extreme data value, unless the most extreme data points lie a specified distance (defined by the variability of the data for a given well) away from the median

--if extreme values ar present,they are highlighted as * or 0 symbols (Velleman and Hoaglin, 1981).

3) Theory:

The components of a boxplot--as described above--are all descriptive statistics known as order statistics; i.e., their derivation depends only on the order of the observations when ranked from lowest to highest. As a result, these statistics are encumbered by few theoretical assumptions (eg., independence, or equal variances) (Hoaglin, et al, 1983)

The boxplot is economical in size, an important advantage when plotting many wells. The plot is also highly resistant to extreme values--up to 25% of the data values can be "wild" without distorting the central portion of the plot from which we draw inference regarding shifts in the median value of wells, as explained below.

4) Application:

Boxplots have several uses for interpreting data. These are listed below and discussed in later sections and by example in the case studies.

i) aid in editing data, while simultaneously reducing need to do so before drawing inference.

ii) boxplots often show that some wells can be grouped together based on similar median values (as identified by overlapping confidence intervals).

iii) boxplots often suggest an estimate of average background concentrations for wells which can be grouped.

E. EDITING DATA FOR ANALYTICAL PURPOSES

If objective criteria are available, it is desirable to edit data in order to improve the inference of subsequent statistical tests. Editing data to improve statistical inference is discussed below. Editing data to enhance the visual inference from graphical analyses is discussed in the next section.

An outlier is an observation which does not reflect the underlying behavior of the bulk of the data. An outlier may be a valid or an invalid observation. If invalid observations are left in the data, they can greatly influence the outcome of statistical tests. Because each well is potentially sampling from a different population, we need a well-specific criterion for identifying possible outliers. This criteria must itself be insensitive to extreme values.

As shown in Figure 7, exploratory data analysis theory suggests that a measure of spread equivalent to 2.698 standard deviation units is useful for defining extreme values. Using the interquartile range as the measure, rather than conventional standard deviation units, decreases the likelihood that this outlier criterion will be inflated by extreme values.

Figure 9 compares three visual representations of data: a Guassian distribution, a boxplot of the Gaussian data, and a time series plot of such data. This figure illustrates how a boxplot suggests important aspects of the data's distributional characteristics (eg., the plot for the Gaussian data is symmetrical and there are no outliers) and emphasizes that boxplots aggregate data over time in order to arrive at an economical display of variability over space.

Figure 10 shows three identical boxplots. All three show extreme values far from the bulk of the data for each well. By checking the time order of the observations, we can determine if it is possible to edit the data in a manner that is both reasonable and defensible. The time order of the extreme observations are noted for each plot (assume the plots are all based on 20 observations). By noting the time sequence for the extreme observations, we see quite different "pictures" of the pattern of variation at each well.

The boxplot to the left shows a consistent pattern of increasing trend; i.e., the extreme tail values are "pulling" the data upward.

The middle boxplot shows a pattern of decreasing trend; i.e., the initial sampling observations are extreme relative to the more stable bulk of the data. A plausible explanation is that the first several samples were contaminated by well-completion methods and are not

FIGURE 9

Three Alternative Visual Representations of Data

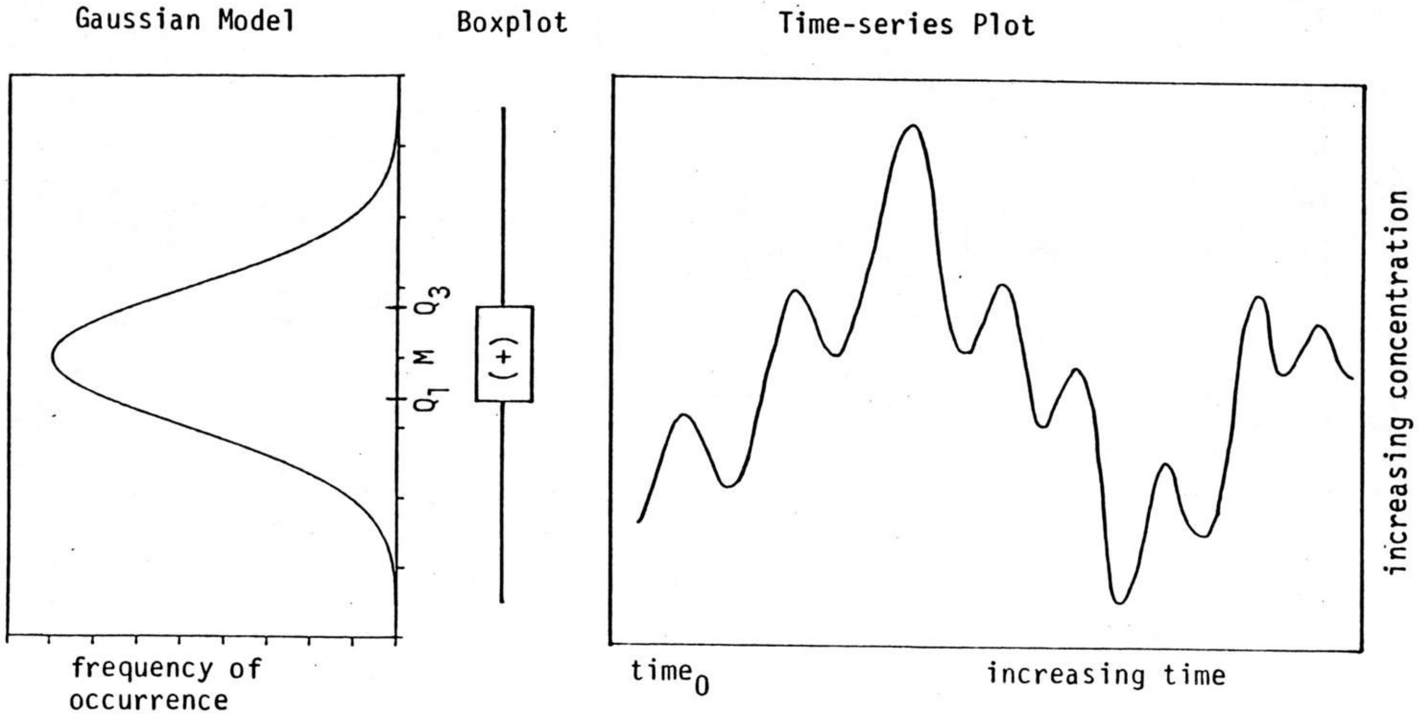
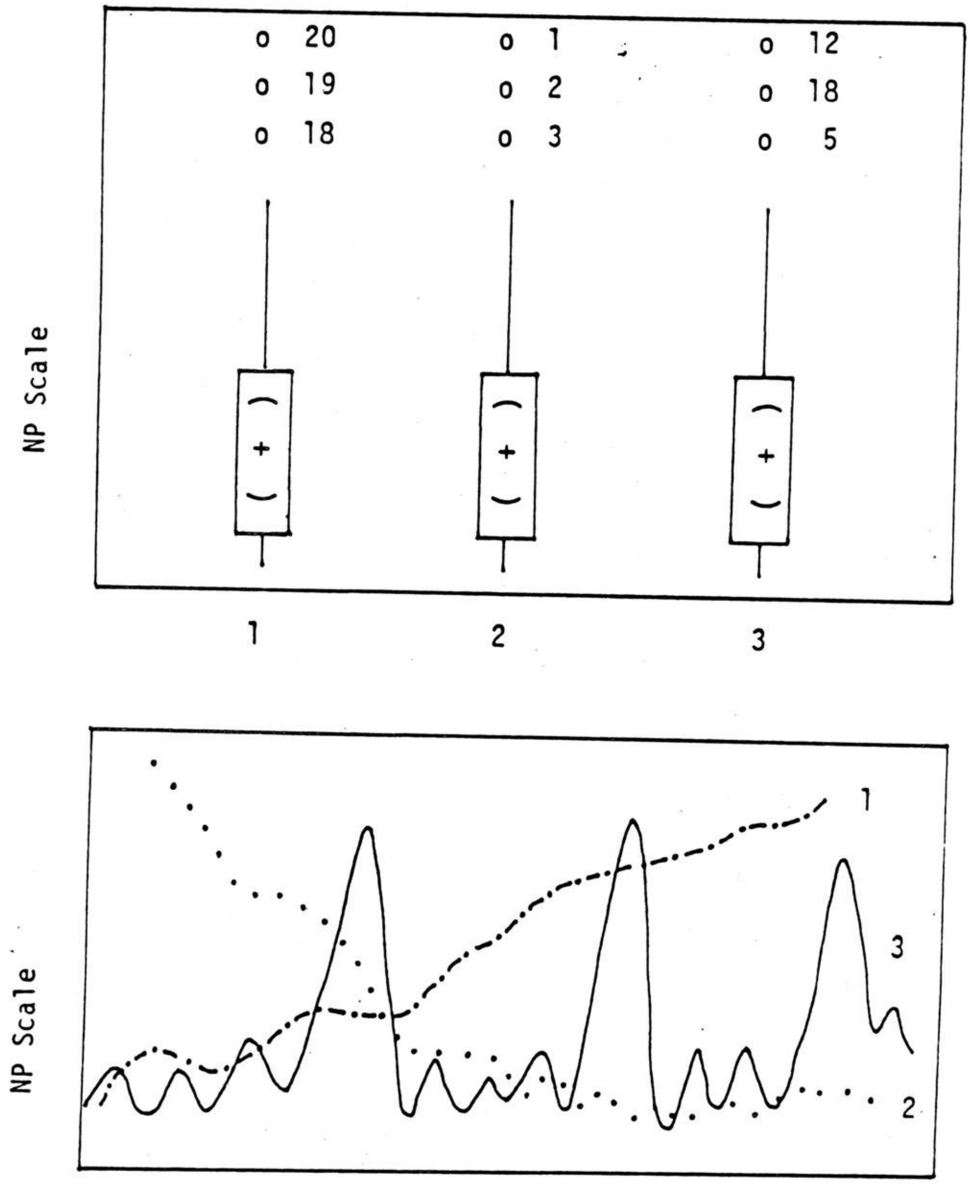


Figure 10

Comparison of Boxplots Having Different Time-series



The boxplots in the upper frame correspond to the time-series plots in the lower frame. The figure illustrates how annotated boxplots can suggest the behavior of data over time and can be used to quickly screen a site for unusual wells. The time-series plots for unusual wells should be checked to verify constituent behavior over time.

representative of the stable pattern in the subsequent samples. The boxplots provide statistically objective criteria for deleting these initial observations.

The boxplot to the right shows a pattern of random, albeit extreme variation. It is difficult to determine, in retrospect, whether these are valid or invalid observations. In this case, we choose to leave them in the data record, for several reasons: (1) their relative (not their absolute) magnitude can still provide information regarding a trend in the data (as discussed in the next section on trend); (2) because of the trend test chosen, their presence is unlikely to dominate the test results; and (3) selectively removing single values from the record is time-consuming and potentially subject to challenge.

F. EDITING TO ENHANCE VISUAL REPRESENTATION OF THE DATA

In practice, it is most useful to plot all the wells for a landfill on a single page. In such a case, it is important to note that the printed scale for each constituent's set of boxplots is determined by the most extreme value for that given constituent at the site. That is, the well with the highest concentrations will control the v

scale necessary to accomodate the highest values will compress the plots for the majority of the wells, rendering the plots far less informative. In this case, the extreme wells may be deleted for the purpose of improving the visual representation. It is apparent, of course, that any wells which so distort the plots are virtually certain to be contaminated; this should be noted, even if the well data is deleted from the plot.

G. TREND TESTS AND TIME-SERIES PLOTS

The objective of monitoring groundwater at landfills is to determine whether the landfill is degrading groundwater quality. It is apparent that finding a significant difference in means of constituents at two or more wells is inadequate to make this determination because it does not reveal why the means are different. Such a difference could be due to several factors: random spatial variability; the effect of geological strata; or perhaps even the effect of sample contamination in a down-gradient well which increases the mean constituent concentration.

In contrast, evidence of trend in concentration data is the single-most informative measure of a landfill's effect on water quality. This is because a trend test eliminates two major sources of confusion: inappropriate spatial comparisons cannot exist for a trend test for a single well, nor can inappropriate temporal comparisons, since time is one of the variables in the test. With respect to comparisons of mean values, however, it should be noted that extreme trends in the constituent concentrations can eventually result in significant differences in mean concentrations, relative to wells which do not show trends.

There are several nonparametric methods for determining trend in data. A method that has been widely used for surface water monitoring is Mann-Kendall's test for trend (Smith, et al, 1982). The null hypothesis tested is that constituent concentration is independent of time, versus the alternative hypothesis that concentration values are correlated with time.

The test statistic is based on the relative magnitudes of the concentration data within the time series. If higher observations occur more frequently <u>later</u> in the time series, the test reports a positive trend, and conversely for lower magnitude concentrations. The tau value reported by the test indicates whether this general pattern is positive or negative. The signficance of the tau value is established based on the consistency of the pattern toward greater or toward lower magnitude concentrations relative to the variance in the data. Because the test uses only information on the magnitude of each observation relative to the observations preceding it, high values (such as due to sample contamination) have less effect on the test result than for a linear regression of the same data. It should be noted that for any test of trend, the power of the test to correctly detect a small to moderate trend diminishes as the "noise" component of the data increases (Lettenmaier, 1976).

Another important feature of this test is that it can easily be modified to incorporate seasonal variability or serial correlation in the data (Hirsch and Slack, 1984). Seasonality is incorporated by testing for trend over years only within a single season, or quarter. I initially used the seasonal trend test for the analysis, but found the seasonal version was less powerful than the unmodified trend test when applied to the short data records available. The seasonal version requires at least 10 years of quarterly data for adequate power to detect trends. Serial correlation is incorporated by an adjustment to the test statistic which estimates the conditional covariance between seasons (Dietz and Killeen, 1981; Hirsch and Slack, 1984). This adjustment, however is invalidated by missing data.

I used the unmodified trend test for the final analyses based on the relative loss of power and the extensiveness of missing data and my finding that there was little statistical evidence of seasonality in uncontaminated wells. Serial correlation may overestimate the significance of trend results for some constituents at some wells. This, however, is not likely to appreciably affect the final inference since I use both highly significant trend results and relative magnitude of trend as criteria for determining contamination.

H. ANNOTATING BOXPLOTS AND TIME-SERIES PLOTS

The trend test results are reported as probability values or p-values. A p-value tells the probability of getting a result as or more extreme than the observed result; that is, it reports the most stringent significance level that could be used which would still reject the null hypothesis of no trend. Thus, a p-value tells the strength of the inference; eg., a p-value of 0.001 indicates a 1/1000 probability of observing by chance such a consistent pattern of increasing or decreasing concentrations, while a p-value of 0.50

indicates a progression of sample values which we would expect with a 50/50 probability. P-values are preferable to simply stating trend results as being significant or not significant at the 0.05 level. This is because there are many factors which might cause a p-value to be slightly over or underestimated by some unknown degree. A p-value enables the investigator to determine what is an important result in the context of the landfill site. For example, knowing that five constituents show marginally significant trends (eg., having p-values of .09) is more informative than knowing only that the trends are not significant at some conventional level. There are some caveats to these interpretations and examples of such situations are presented later.

By adding trend results and sampling dates to the boxplots, we can get a picture of how and why groundwater quality has changed at different wells while still retaining the spatial context of the data. This spatial context, or using the entire landfill as a frame of reference, is useful for assessing the importance of observed changes beyond purely statistical criteria.

I. EXAMPLES ILLUSTRATING IDEALIZED CONDITIONS:

Having seen the graphical tools used, it is useful to develop some intuitive feel for their meaning.

Figures 11 - 14 illustrate several plots from hypothetical landfills with idealized site conditions.

Figure 11

Boxplot Illustrating Idealized Landfill Conditions

(one constituent only)

Site 1: Assume homogeneous subsurface and assume sampling predates landfill operation

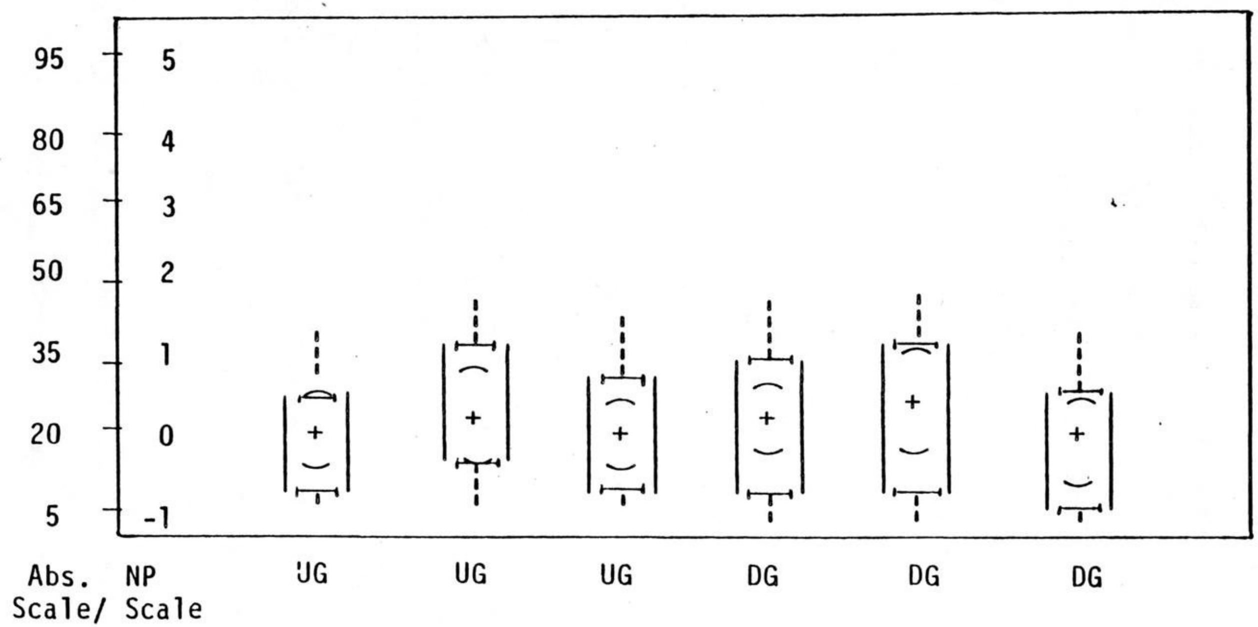


Figure 12

Boxplot Illustrating Idealized Landfill Conditions

(one constituent only)

Site 2: Assume homogeneous subsurface and assume sampling predates landfill operation

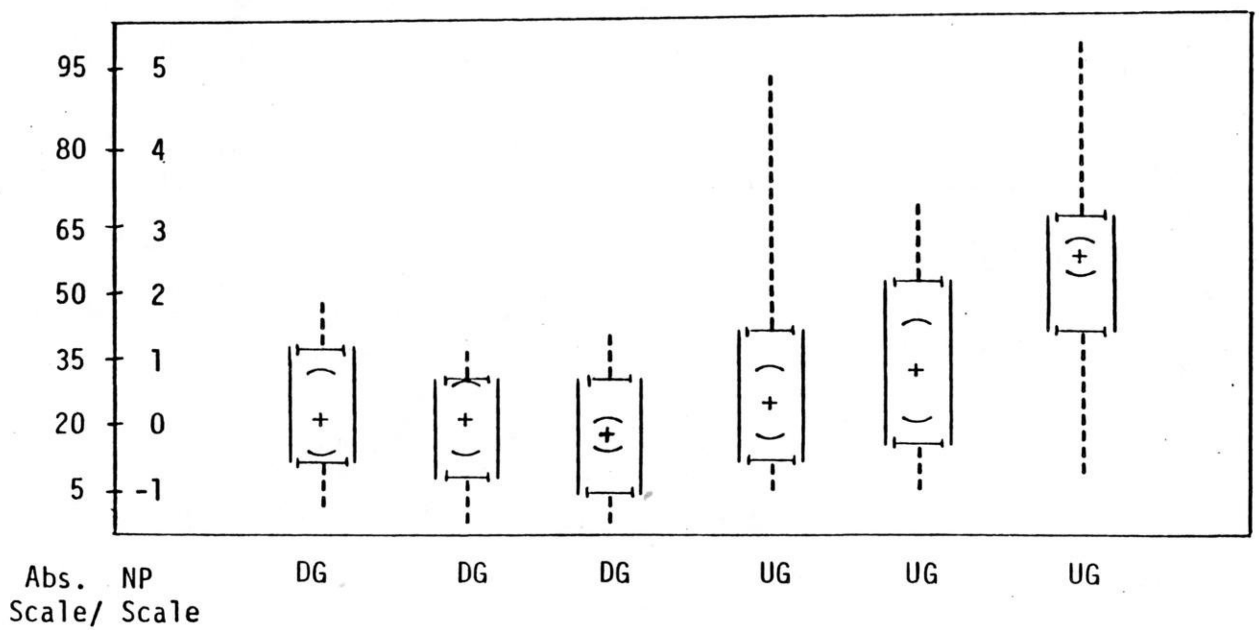


Figure 13

Boxplot Illustrating Idealized Landfill Conditions

(one constituent only)

Site 3: Assume heterogeneous subsurface and assume sampling predates landfill operation

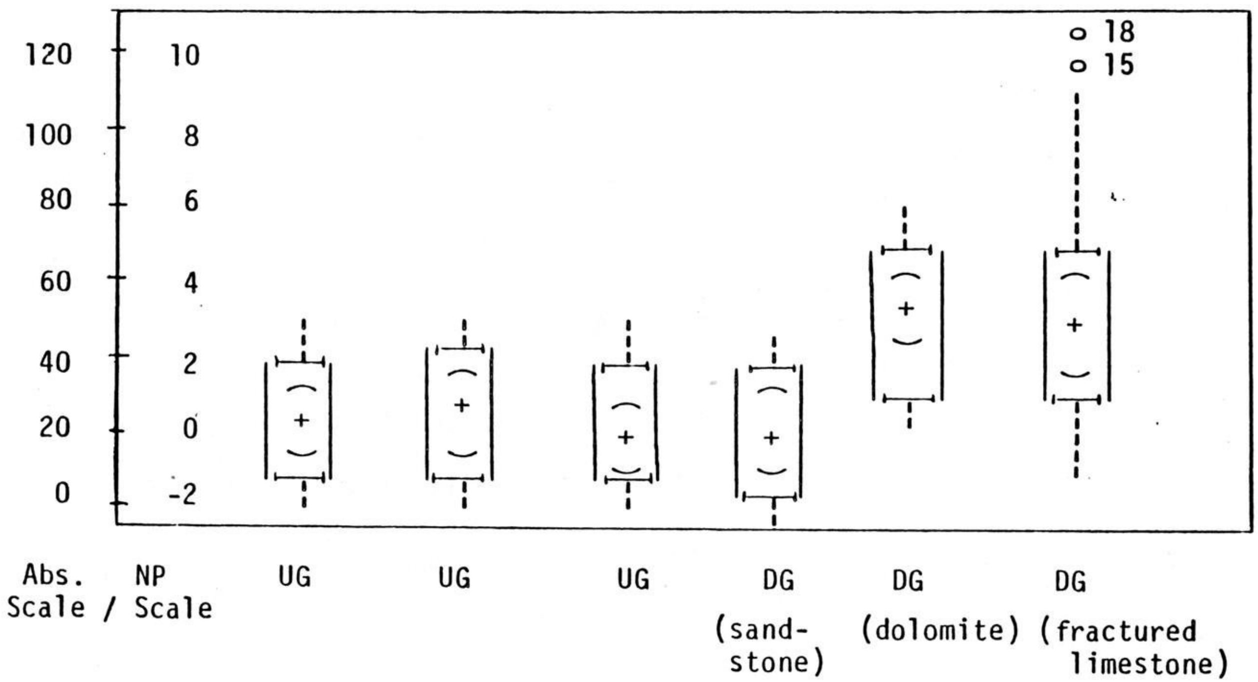
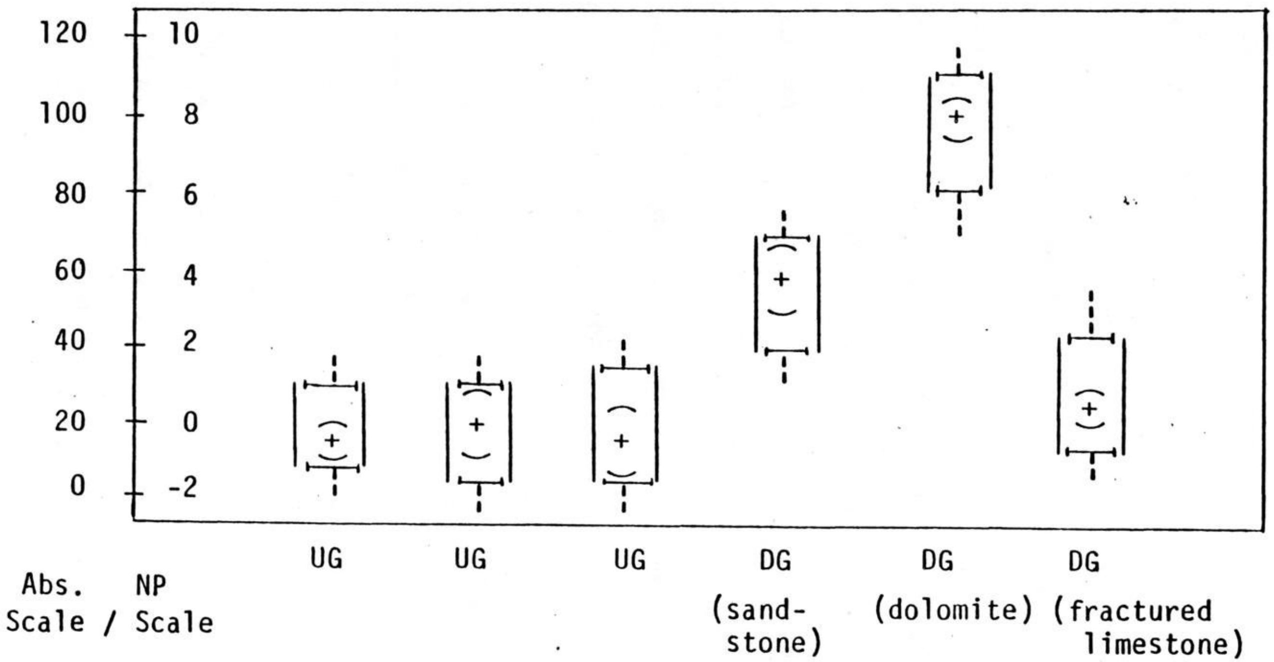


Figure 14

Boxplot Illustrating Idealized Landfill Conditions

(one constituent only)

Site 4: Assume heterogeneous subsurface and assume sampling began 10 years after landfill operations began



<u>Site 1:</u> At this site we assume a homogeneous subsurface: i.e., there is little physical reason to expect differences in groundwater quality <u>except</u> from the landfill. Further, we assume that we have genuine baseline data, taken before the site began operation, so that any changes in quality would be evident.

The graph shows boxplots for six wells, three up-gradient and three down-gradient. A glance at the plot tells us several things:

1) The site median value is 20 mg/L, which corresponds to the NP-scale value of 0. We can see from the relative location of the "box" edges and "tail" lengths of the plots that all of the wells show reasonable symmetry, i.e. low values occur about as often as high values and there is not any extreme skew to the data.

2) We can approximate a test for similar population means by comparing the location of all the confidence intervals about each individual well population median value. This individual confidence level is set at about 95%, equivalent to a significance level of about 5%.

To interpret these confidence levels, recall that we are using the sample median as an estimate of center for the population. Since we are using observed data, we know with certainty the value of the historical sample median value; its location is shown by an "+". The confidence interval shows us how certain we are about the location of the unknown population median value, given the number of sample data and its variation. The probable value of the population median value lies between the parantheses: "( )." In this case, the probability is about 95%.

We note that, in this example, the confidence intervals have roughly similar width and that all overlap with each other. Based on this, we infer that the population median values for the wells are statistically similar. This visual inference is an approximation. However, a t-test between two of these wells or a multiple comparison of well means would produce about the same result. The hydrologic inference is that there is no evidence of the landfill having changed the median concentration levels of this particular constituent. Finally, the dual concentration scale shows that both the NP scale values and the absolute concentrations are low. Thus, all wells at the site can be regarded as "background" or currently unaffected wells.

<u>Site 2:</u> We again assume a homogeneous subsurface and data which predates landfill operation.

1) A glance at this graph shows that it is different from Site 1. The upgradient wells are all centered near the site median value and show approximately equal spread. The down-gradient well plots look different: one shows a long tail toward higher concentration values, one shows a wider confidence interval, and the last shows far greater spread in the tails of the data.

Because the plots aggregate data over time, we cannot be certain what causes this variation in the data. By adding the trend results to the boxplots, we can distinquish random variation over time from a systematic trend in the data. Examples of this are presented in the case studies.

Site 3: At this site we relax our assumption of subsurface homogeneity and assume that up-gradient wells are screened in sandstone; down-gradient wells are screened in sandstone, dolomite and fractured limestone. We again assume that sample data predate landfill operation.

1) This graph shows differences between the up-gradient and two of the downgradient wells. The wells screened in sandstone all have comparable median values and variability. The wells screened in dolomite and limestone exhibit greater median values and variability for this particular constituent.

2) A visual test of medians leads us to infer that there are statistically significant differences in population median values. We need additional evidence before we can assess whether these differences might reasonably be attributed to the landfill or to differences in geologic strata. The plot for the downgradient well screened in limestone suggests a shift in median due to a trend in the data. This is suggested by the long tail extending toward higher concentrations and the extreme values near the end of the record. If the trend result and time-series plot for this well confirms the presence of trend in the data, we have evidence of degradation which is not likely due to inherent well differences.

Site 4: In this last example, we assume the same subsurface conditions as in site 3. In addition, we also assume that monitoring began a decade after the site had been in operation.

1) The graph again shows differences in the population median values. In some cases, the displacement of the median values is quite high, indicating that the median values are not only statistically different, but are different by the amount shown. Note that the double scale enables this amount to be easily interpreted in absolute concentration units or in terms of the constituent's variability at the site. This "internal scale" assists in evaluating the environmental importance of the observed differences in constituent concentrations between wells. Relative to Sites 1 and 2, we also note greater variability in individual well data (eg., the median of one down-gradient well is centered at 8 NP scale units above the site median).

2) As with Site 3, we infer that the median values are markedly different, but cannot yet infer the cause of this difference. All the plots suggest fairly stable water quality conditions, however, by virtue of their low spread about the individual well median values. Such a pattern suggests that the downgradient wells may have already been affected when water quality sampling began.

J. COMBINING INFERENCES: KEY FOR INTERPRETING DATA PLOTS

By combining the information contained within the annotated data plots, I grouped individual wells into four categories, based on the evidence that contamination is present or absent. The four categories are:

CATEGORY I : presumptive evidence that well is clean.

(eg., the well shows non-significant trends; low variability; and the location of the median values for all constituents is comparable to median values for wells in similar strata, or comparable to or less than the site median).

CATEGORY II: evidence that well is probably clean.

(eg., the well shows one or two significant, but low magnitude trends; the constituent medians are generally less than or comparable to site median).

CATEGORY III: evidence that well is probably contaminated .

(eg., the well shows several marginally significant trends; the median values for several constituent are greater than the site median; and the well shows high variability relative to other wells in similar formations).

CATEGORY IV: presumptive evidence that well is contaminated.

(eg., the well shows many significant, high magnitude trends; the median values for several constituents are displaced above-- or for pH and alkalinity, above <u>or</u> below-- the site median far beyond the amount needed to determine statistical significance).

Category I and II wells can be used for baseline information. Category IV wells require prompt regulatory attention, while category III wells should be watched for additional signs of degradation. Note that distinquishing between category II and III wells requires considerable judgement, yet it is for these wells which we need be most concerned about "false positives" and "false negatives." For the cases of category I and IV wells, we are unlikely to draw incorrect inferences, because the criteria are so clear.

K. COMPARISON OF SITE-WIDE AND CATEGORY I ESTIMATORS

As noted earlier, statistical theory shows that the median and IQR are resistant estimates of center and scale; that is, they are little affected by erratic values in the data. As a preliminary investigation of how well this strong theoretical advantage translates for use in an applied situation, I compared two estimates of medians and IQRs for each constituent reviewed at each landfill site. These estimates are shown in Figures 15 - 21. These figures enable (1) a visual comparison of constituent variability at clean wells across landfill sites having different design features, and (2) a comparison of location and scale estimates derived from the entire site data with similar estimates derived from the subset of clean wells. Good agreement between these two estimates would support the assumption that site-wide estimates approximate uncontaminated conditons.

Figure 15
Site-Wide Estimators vs. Category I Estimators: Chloride

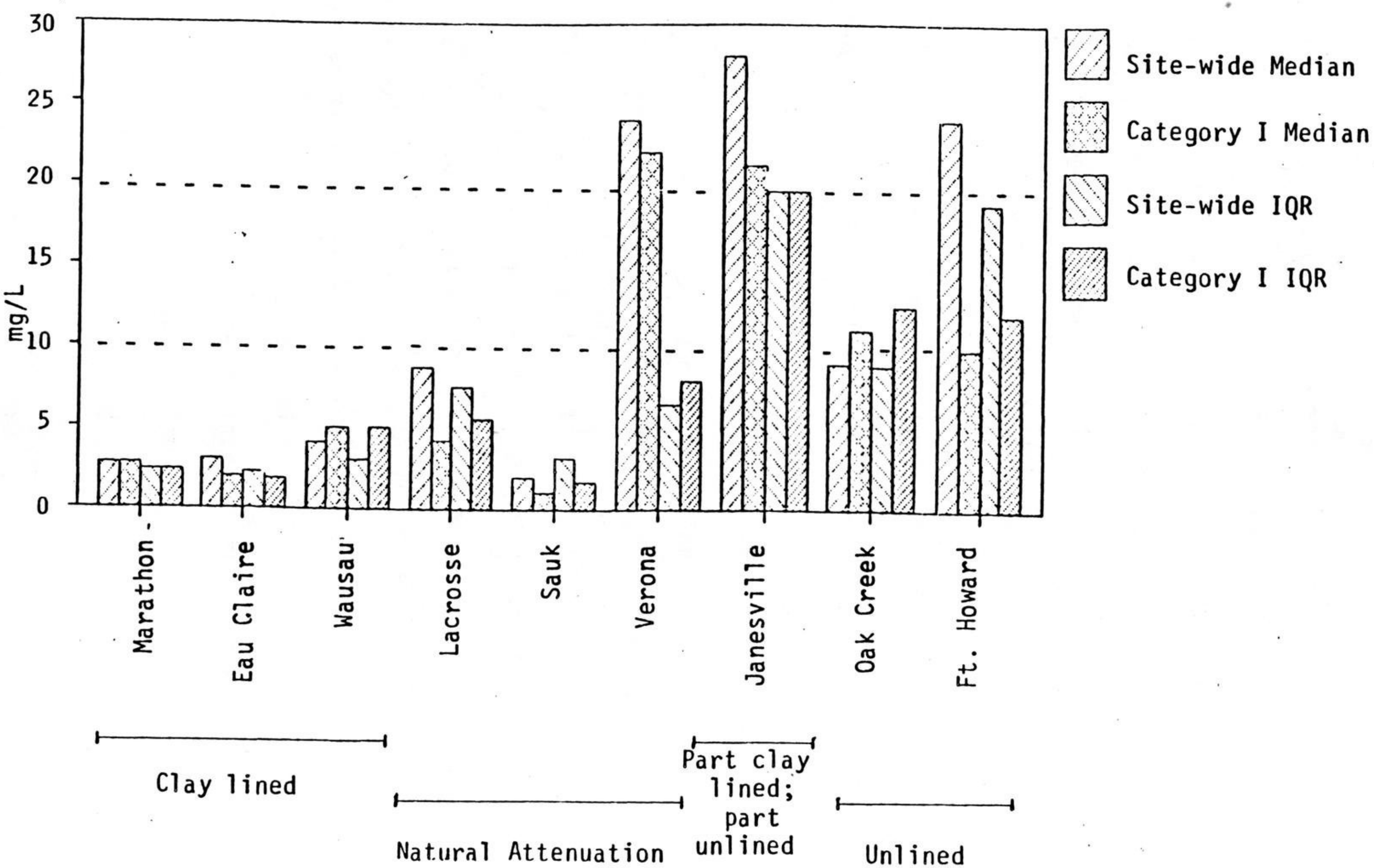


Figure 16
Site-Wide Estimators vs. Category I Estimators: Sulfate

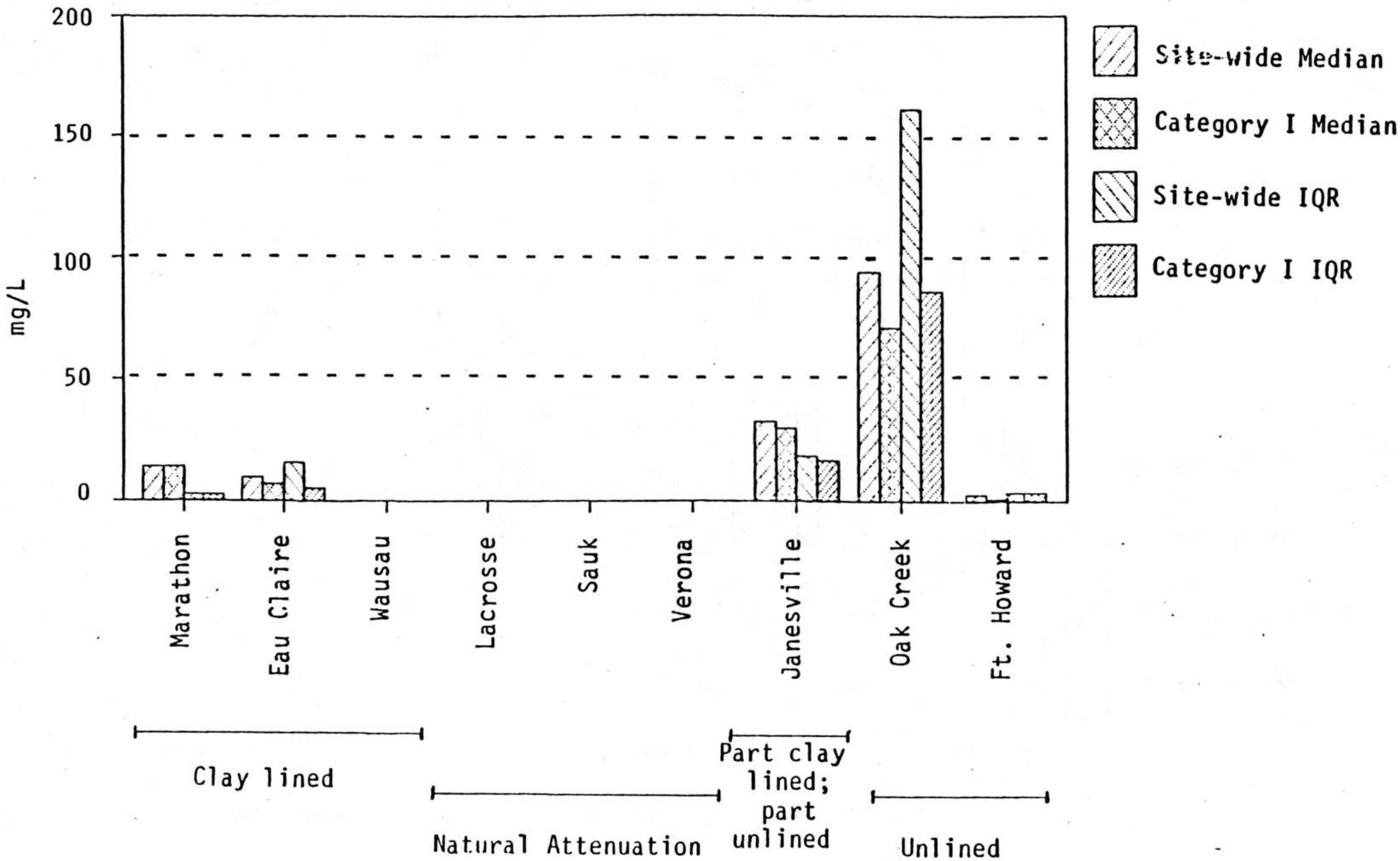


Figure 17
Site-Wide Estimators vs. Category I Estimators: COD

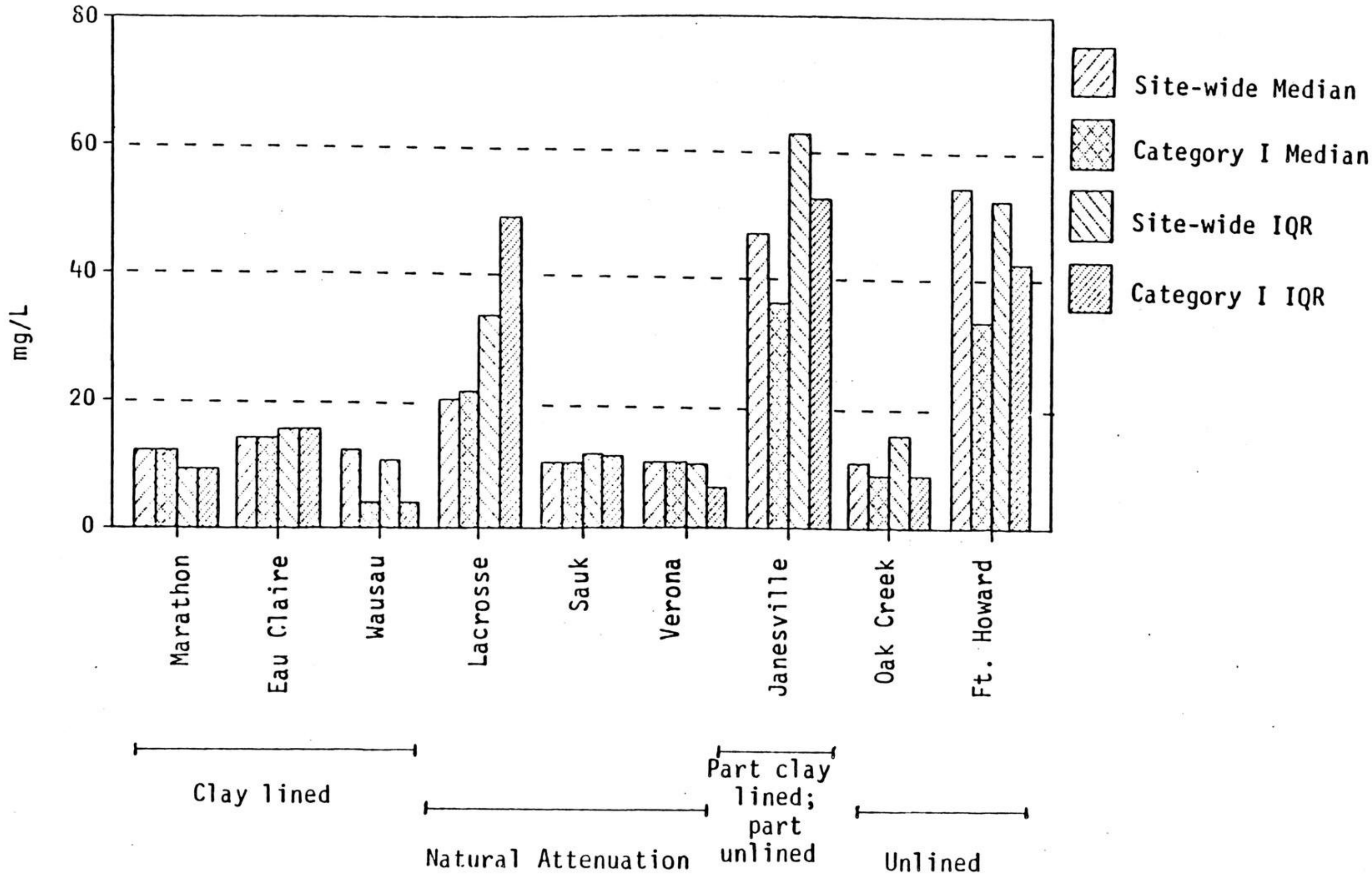


Figure 18
Site-Wide Estimators vs. Category I Estimators: pH

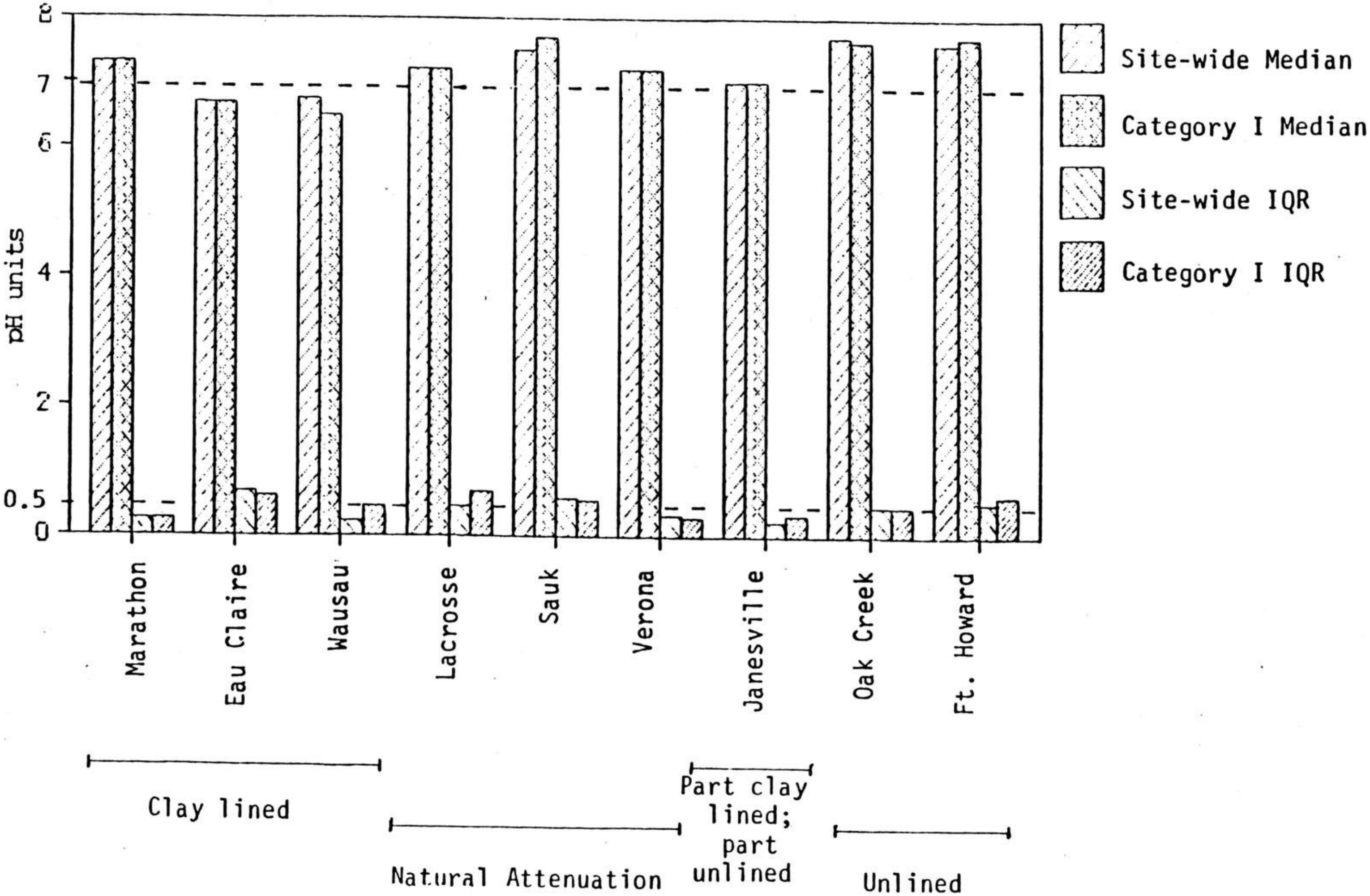


Figure 19
Site-Wide Estimators vs. Category I Estimators: Alkalinity

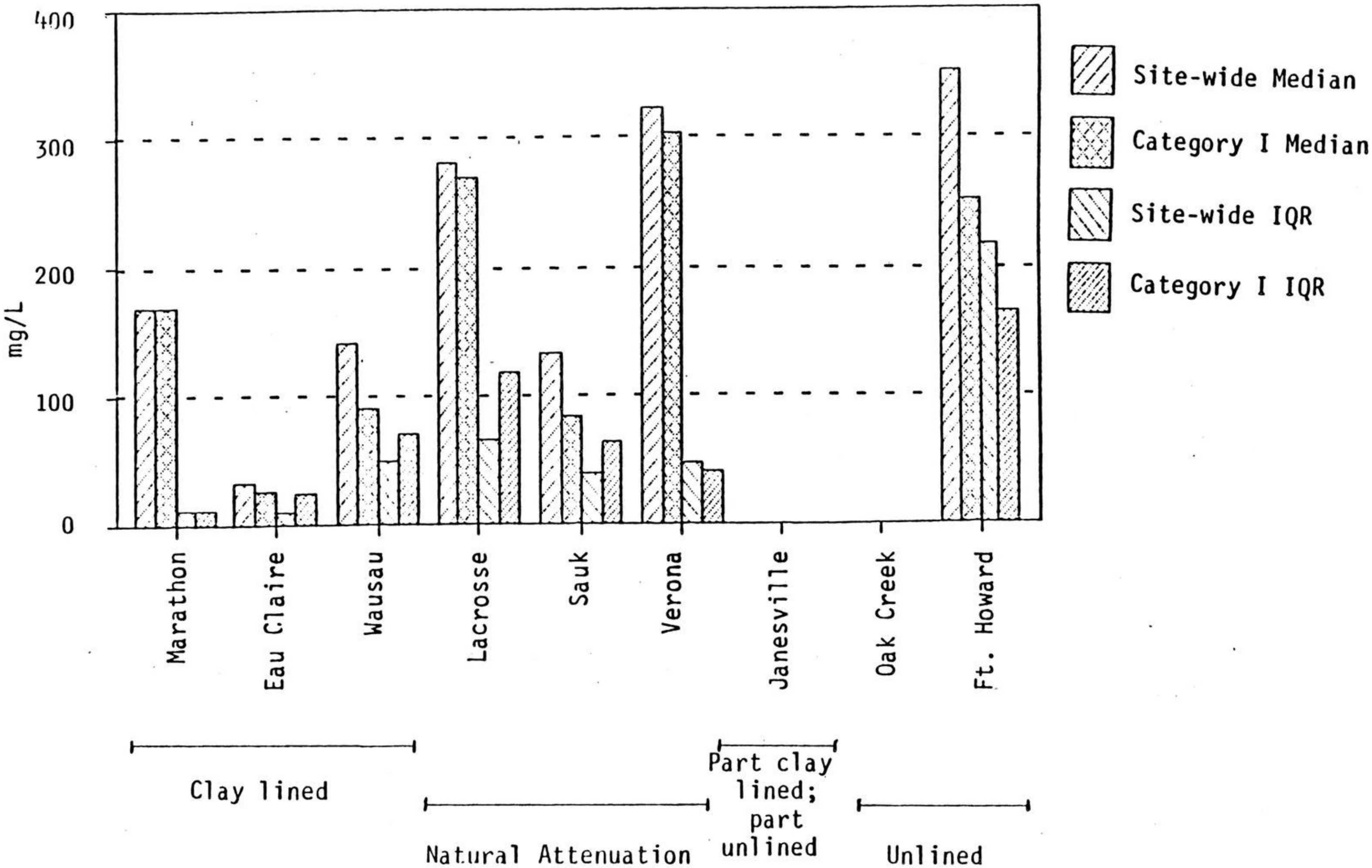


Figure 20
Site-Wide Estimators vs. Category I Estimators: Hardness

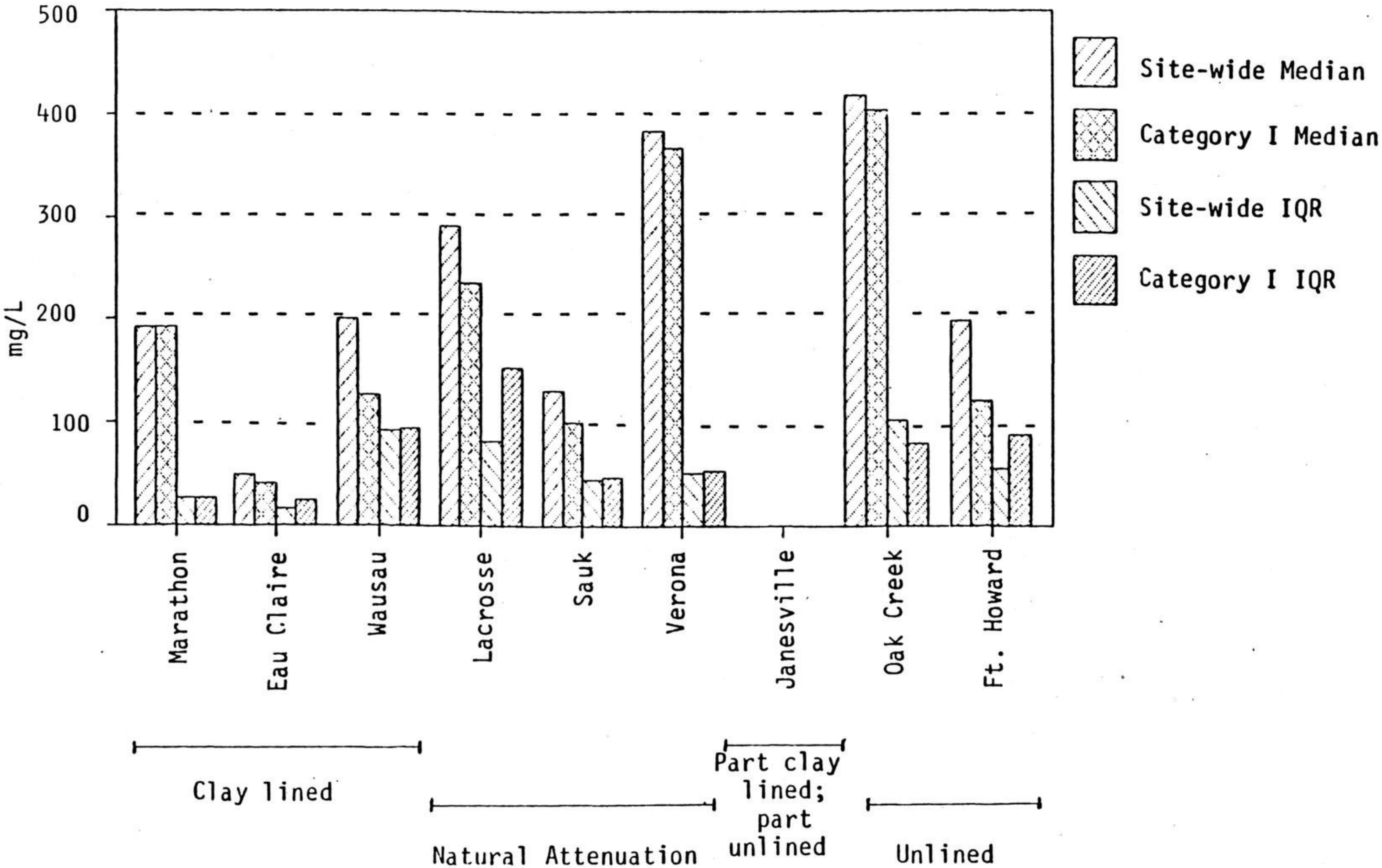
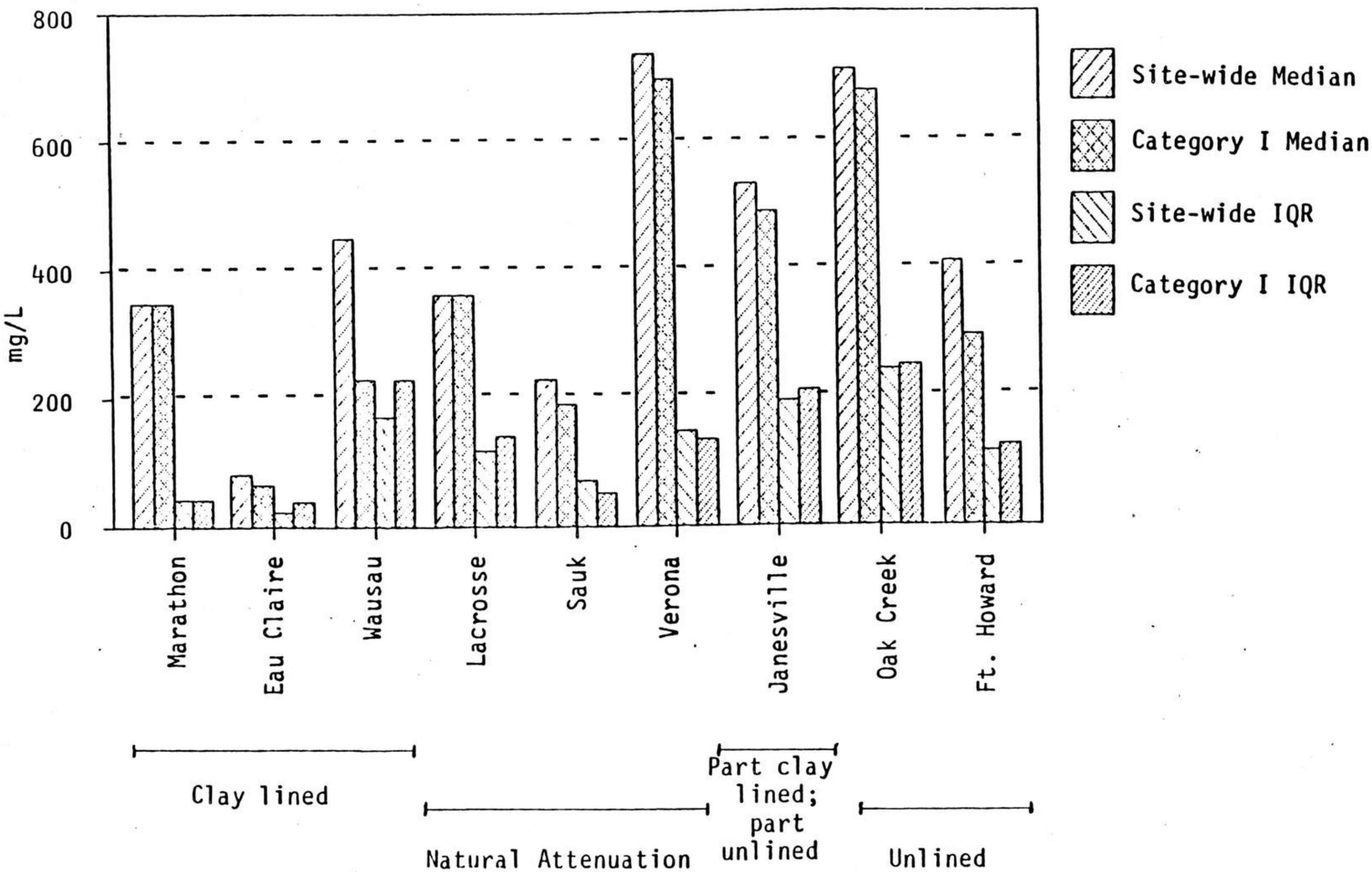


Figure 21
Site-Wide Estimators vs. Category I Estimators: Conductivity



1. Description of plotted statistics.

Two median estimates and two IQR estimates are plotted side-by-side for each landfill. These estimates were derived in very different ways. The "site-wide" median is the median value for a given constituent obtained from all the observations for that constituent at a site. These data were unedited, and thus contained not only transcription errors and erratic values due to sample contamination from well-completion, but also data for contaminated wells. Similarly, the "site-wide" median IQR is the median interquartile range obtained from all individual IQRs at the site. As with the median value, all wells and unedited data were used.

Similar estimates were obtained from all data contained within the subgroup of Category I wells, as determined by the described evaluation procedure. These data had been edited to remove the high "start-up" values and thus are the "cleanest" of the clean data. It is important to recall that the wells were categorized on the basis of the evidence from <u>all</u> constituents at a well, not on a constituent-by-constituent basis. Thus, neither the absolute concentrations nor variability for given constituent could be pre-determined or consciously biased as part of the analysis; i.e., a well which has extremely low chloride concentrations may also have very high conductivity concentrations.

Table 3 shows the number of wells used to obtain the site-wide estimates versus the number of wells in Category I used to obtain the "clean" estimates. Each median value, of course, represents only the

TABLE 3

Number of Wells used to Obtain
Site-wide vs. Category I Location and Scale Estimates

Site	Total No. of Wells Reviewed	No. of Wells in Category I
Marathon	6	6
Eauclaire	8	6
Wausau	5	3
Lacrosse	12	8
Sauk	12	6
Verona	25	10
Janesville	17	12
Oakcreek	18	9
Ft. Howard	25	11

single middle observation, just as the IQR median represents one IQR for an individual well.

2. Constituent variability

As shown in the figures, the landfills are grouped according to their design: clay-lined with leachate collection systems, natural attenuation, and unlined sites. The data from one location, the Janesville site, reflects water quality conditions obtained from wells located around the perimeter of several distinct disposal areas. Two such areas are unlined and waste disposal predates water quality monitoring by as much as 17 years. Adjacent to these is a disposal area which is clay lined and has had monitoring data collected simultaneously with waste disposal.

By comparing the estimates based on the Category I wells, it is evident that a concentrations for a given constituent may vary over two to three orders of magnitude across all landfill sites, even under conditions where contamination is likely absent. Except for chloride, there is not any consistent pattern of low concentrations being clearly associated with a type of landfill design.

3. Comparison of site-wide versus Category I estimators

For chloride, pH, hardness and conductivity, the agreement between these two sets of estimators is quite good for most of the sites. (It should be noted that all the wells at the Marathon landfill were placed in Category I, thus, by definition, these estimates are identical).

Table 4 lists the sites and constituents for which either of the site-wide estimate differed from the Category I estimate by more than 25%. As the table shows, discrepancies of this magnitude occurred in 23 out of the 56 comparisons. However, the table also shows the amount of the discrepancy in absolute concentration units. From this perspective, many of the discrepancies are negligible in terms of "environmental significance." For comparison, the table also lists for each constituent the statutory PAL or the minimum increase for an indicator parameter which DNR considers allowable. In only one case, for conductivity at Wausau, does the discrepancy in the estimates exceed the allowable increases.

These results are preliminary, yet they suggest that site-wide estimates obtained quickly from the site data may prove useful for initially screening wells at a landfill site for contamination. Further work might clarify which constituents appear to be most reliable and what multiple of the site-wide IQR estimate best distinquishes between contaminated and uncontaminated wells. At worst, such estimates could easily detect wells with extreme constituent concentrations; at best, they might closely approximate the conclusions based on the more labor-intensive analysis to determnine contamination for each individual well.

TABLE 4

Discrepancies between Site-wide and Category I Estimators

Constituent	Site	Estimator	Absolute Amount	Increment allowed by Policy or Law
Welfare Parameters				
Chloride:	Wausau	IQR	5 mg/L	PAL= 125 mg/L
	Lacrosse	M	5 " "	" "
	Ft.Howard	IQR	14 " "	" "
Sulfate:	Oakcreek	M	23 mg/L	PAL= 125 mg/L
		IQR	75 " "	" "
Indicator Parameters				
COD:	Wausau	M	8 mg/L	25 mg/L
		IQR	7 " "	" "
	Lacrosse	IQR	15 " "	" "
	Oakcreek	IQR	2 " "	" "
	Ft.Howard	M	21 " "	" "
Alkalinity:	Wausau	M	50 mg/L	100 mg/L
		IQR	20 " "	" "
	Lacrosse	IQR	53 " "	" "
	Sauk	M	49 " "	" "
		IQR	25 " "	" "
	Ft.Howard	M	100 " "	" "
Hardness:	Wausau	M	74 mg/L	100 mg/L
	Lacrosse	IQR	63 " "	" "
	Ft.Howard	M	100 " "	" "
Conductivity:	Wausau	M	220 umhos/cm	200 umhos/cm
	Ft.Howard	M	115 " "	" "

IV. INTRODUCTION TO CASE STUDIES

Three case studies are presented to illustrate the procedure for evaluating data to determine the presence of contamination. Three landfills are reviewed: Sauk County landfill, Marathon County landfill, and Dane County--Verona landfill.

The Sauk County landfill is particularly well suited for explaining the procedure. This is because the pattern and cause of contamination at the site are readily discerned from the data. Hydrologic information is not essential to the findings; if only water quality data were available, a strong case for landfill related contamination at specific wells could still be made.

The Marathon County landfill is an engineered landfill having a clay liner and leachate collection system. The case study is presented to illustrate how the data appear when there is no evidence of contamination. In addition, it illustrates that at such sites, statistically discernible degradation of groundwater quality may be evident long before the change is considered environmentally important in terms of policy criteria such as established by DNR.

Finally, the Dane County Verona landfill is presented as an example of a site having complex geology and somewhat ambiguous data. Even at this site, however, there are clearly contaminated and clearly uncontaminated wells.

The case studies summarize basic hydrologic and design aspects of the landfills. To assist in visualizing spatial variability in groundwater quality at a site, the boxplots for individual wells are grouped according to two criteria. First, wells screened in similar formations are grouped together. Second, wells screened in similar formations are plotted in order of their location within the apparent flow field at the site, going from upgradient, to cross-gradient to down gradient.

The boxplots are annotated as discussed in section III. H. For each well the p-value for the trend test is shown. In general, p-values less than 0.05 should be regarded as significant, and p-values from 0.051 to 0.10 as marginally significant. Th "+" and "-" following the p-value have two possible interpretations: (1) if a significant trend is present, they indicate its direction; or (2) if no trend is present, they simply indicate the general pattern toward greater or lower concentration values (i.e., they show the sign of the tau value used in the test statistic).

In addition, the case studies present the results of three comparisons:

1) <u>comparison with welfare parameters</u>: This section describes the extent to which data for chloride and sulfate materially added to the inference resulting from only the indicator parameters at the site.

2) <u>comparison with simple NP flag</u>: This section describes the results of using a site-wide flag to screen the site for contamination by COD

and inorganic constituents. Use of this flag is based on the concept that the site-wide estimates of median and interquartile range roughly approximate the corresponding statistics for wells in Category I (i.e., those wells which show no evidence of contamination). This characteristic makes the NP values useful for flagging potential contamination. An NP value of 4 is used for all constituents, primarily because it is equivalent to the median value plus approximately 3 F-psuedosigmas. Thus, this "flag" level should be close to, or perhaps slightly greater than, the amount of contamination DNR determined allowable in a preventive action limit for indicator parameters.

As a preliminary review of the efficacy of using a site-wide flag, each well was reviewed to see: (1) how many, if any, constituents exceeded an NP value of four; and (2) the behavior of the data preceding the flagged value. Since sample contamination could cause a false "flag," the presence of trend in the constituent data prior to reaching the flag threshold was used as an informal criteria for a flag that signals degraded groundwater quality. Only the data prior to the first flagged value was considered, so that the analysis would parallel the routine analysis of future monitoring data. Determining "trend" is a function of data record length and noise in the data. Consequently, the eight observations prior to the flagged observation were visually checked for trend. Eight observations were chosen as a compromise between the minimum number necessary to establish meaningful trend while still reflecting a time scale short enough to be sensitive to small but systematic trends.

Comments regarding the site-wide flags are provided in the summary tables for the landfills reviewed in the case studies and for the landfills discussed in Appendix B. Three types of comments are generally made: (1) there appears to be a trend present at the time the constituent is first flagged; (2) the flag comes too early in the data record to check for trend; or (3) the flagged observation shows an unusual "jump" (eg., is greater than the earlier observations by 4 NP values) and so may reflect sample contamination rather than actual changes in groundwater quality. Resampling or checking the next routine sampling report should clarify what has in fact occurred.

Figures 21a and 21b show examples for of the site-wide flag review. Figure 21a shows a well in which a single observation has triggered a flag. The preceding observations appear random and the presence of this single unusually high value does not indicate a consistent pattern of trend. Figure 21b shows a well for which three constituents were flagged. Checking the eight observations preceding the first flagged values, we see a strong trend in each constituent. This suggests that, for this well, the screening technique would have turned up a meaningful flag that warrants regulatory attention.

3) <u>comparison with VOC results</u>: This section presents the results of sampling for VOCs. Unlike sampling for inorganic constituents and COD, VOCs are not routinely sampled at monitoring wells. The available VOC data are difficult to analyze because the data are scant

Figure 21 a

Preliminary Review of Site-wide Flag

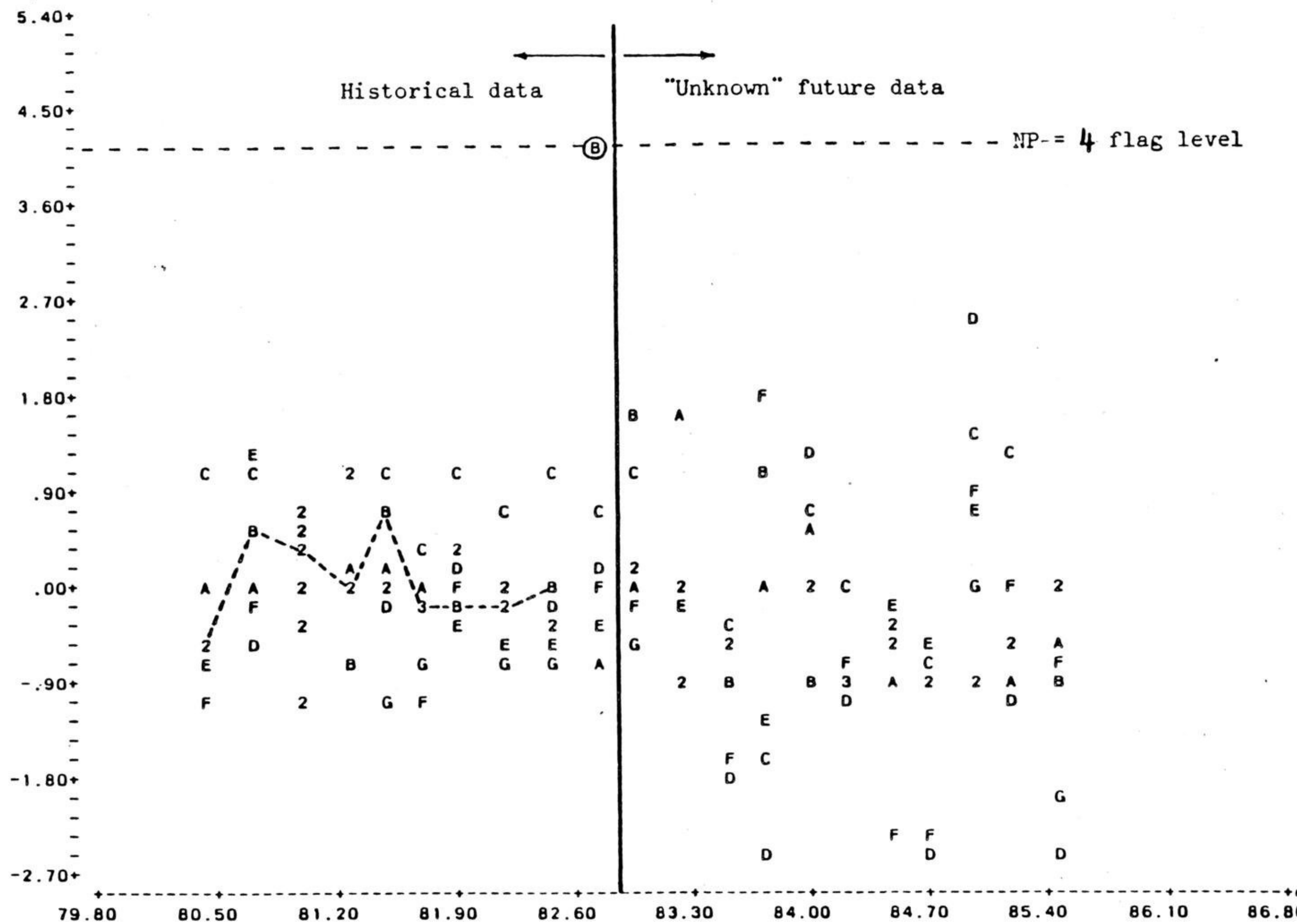
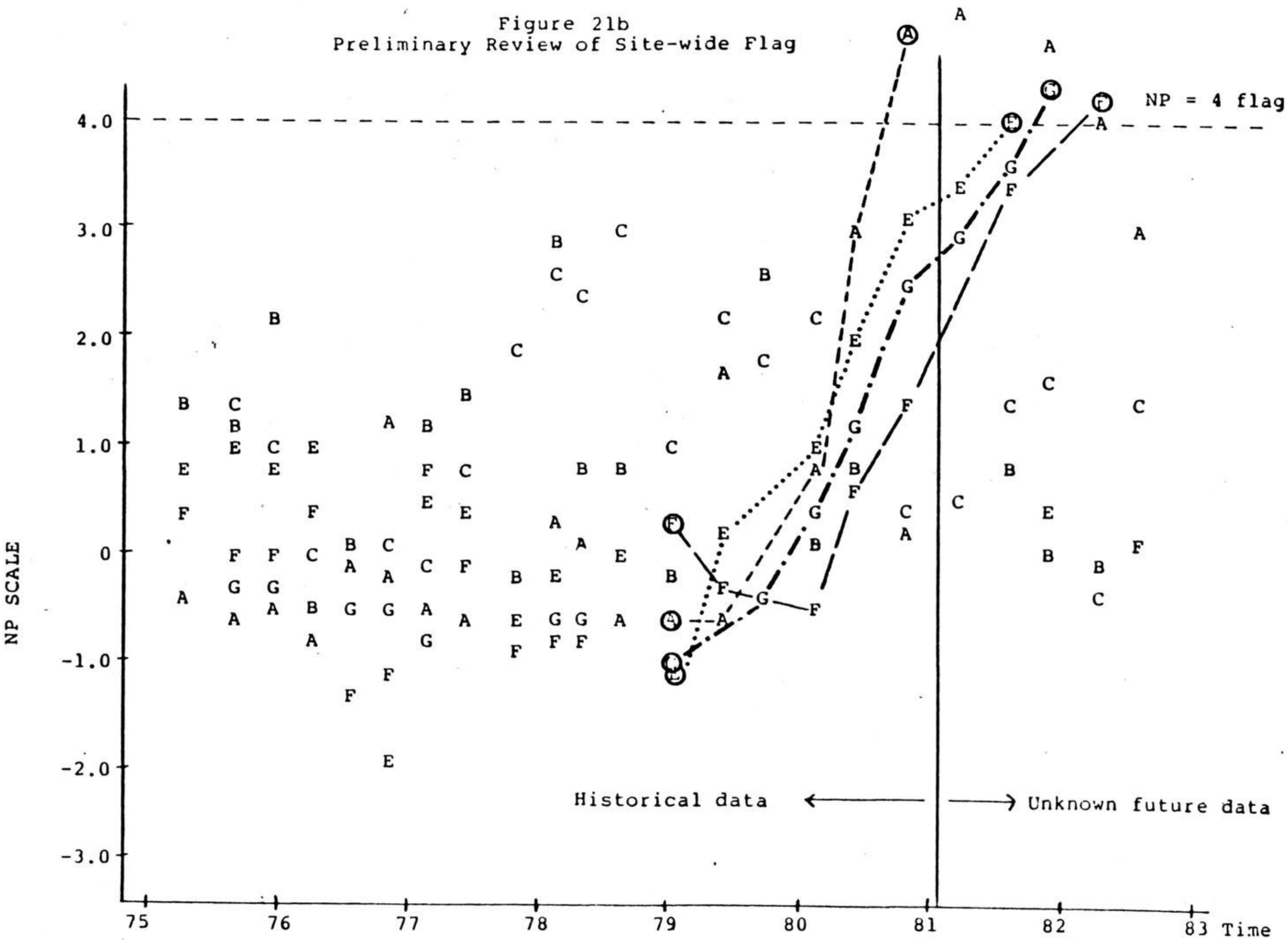


Figure 21b
Preliminary Review of Site-wide Flag



Key: A = chloride, B = COD, C = pH, E = hardness,
F = conductivity, G = sulfate

and highly variable, are of unequal sample size, include qualtitative and quantitative results, and often have no replicate measures of sampling error.

Given these limitations, a very simple measure of VOC response was devised and used as a basis for comparing with inorganic and COD data. This method is conceptually similar to previous research on indicator parameters and VOC data in that it focuses on frequency of detections rather than analysis of individual compounds (Plumb and Pitchford, 1985). VOC results are reported as the number of detections (qualitative or quantitative) over all sampling events, followed by a fraction indicating the number of times at least one VOC was detected when sampled for (eg., "23 Det., 2/4" indicates a total of 23 VOC detections over 4 sampling events; the fraction indicates that VOCs were detected at the well on only 2 sampling occasions. In essence, the number of detections simply reports the frequency of detection; the fraction indicates the consistency of detections over all samples for VOCs at a well).

A. CASE STUDY: (OLD) SAUK COUNTY LANDFILL SITE # 2051

1. SITE CHARACTERISTICS:

DESIGN: Natural attenuation design; no leachate collection. Landfill is 14 acres in size.

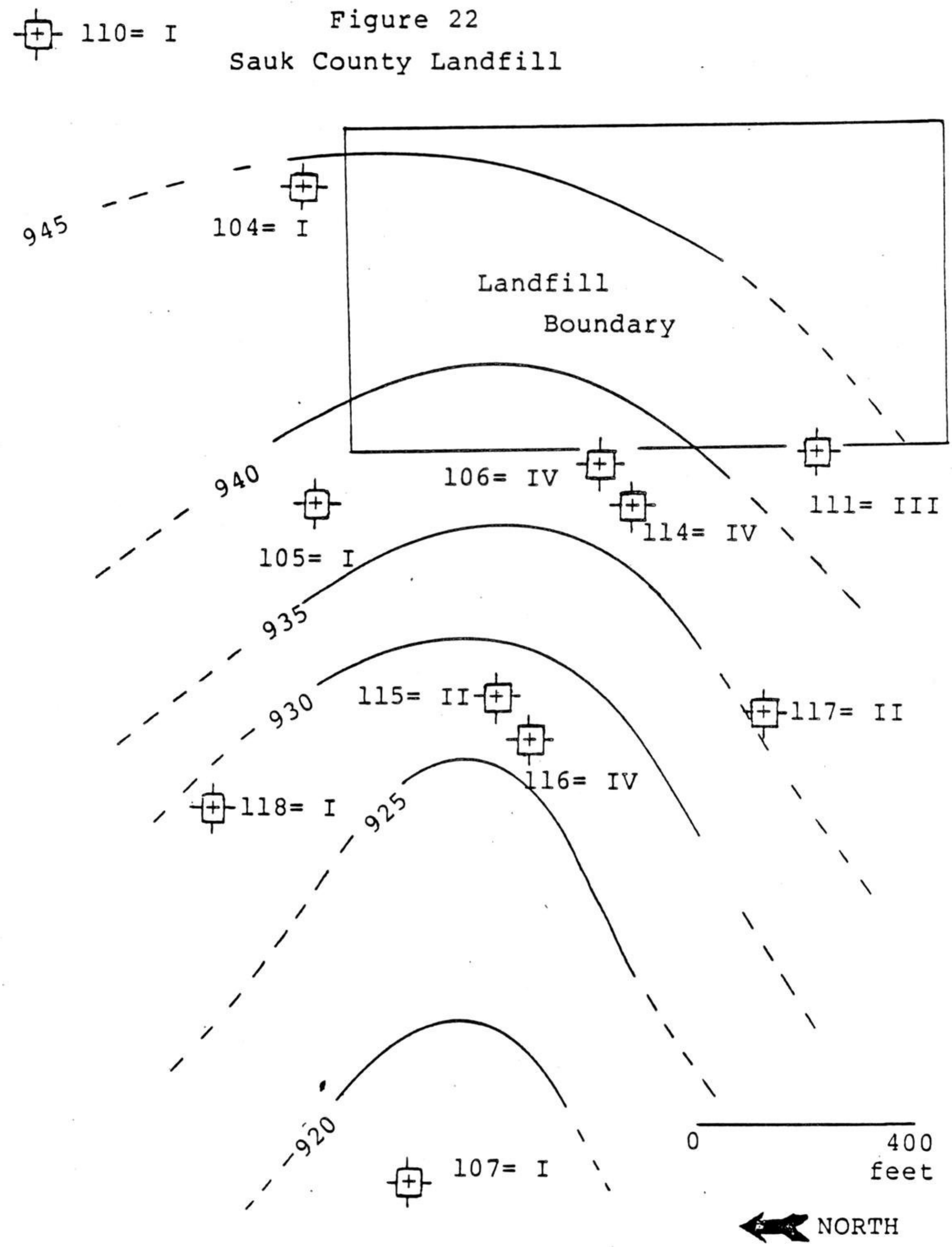
WASTES ACCEPTED: Municipal waste and foundry waste.

HYDROLOGIC SETTING: The landfill is located in an east-west trending valley between two sandstone ridges. Landfill is constructed in sand deposited by glacial lakes. Sand depths vary from 45-74 feet across site area, and is underlain by Cambrian bedrock sandstone.

FLOW FIELD INFORMATION: Groundwater generally flows from east to west or southwest in the site area. In 1981, horizontal gradients upgradient (east) of the site were estimated at 0.006 feet/foot; immediately downgradient of site, the horizontal gradient was estimated at 0.02 feet/foot. (See Figure 22 for a site map). Downward vertical gradients were calculated at two downgradient piezometers, indicating recharge conditions (well 115: 0.2 ft/ft; well 116: 0.4 ft/ft)

SAMPLING DATES: The landfill began operation in 1973 and is now being closed. Earliest sampling began in mid-1975 at wells 104, 105 and 106. Additional wells were installed and sampled from about 1979 to mid- 1982.

Figure 22
Sauk County Landfill



Groundwater elevations taken 7/29/81

2. ANALYSIS

TOTAL # WELLS REVIEWED: 12

NUMBER OF WELLS PER CATEGORY:

Category	#Wells
I	5 (104,105,107,110,118)
II	3 (115, 117,119)
III	1 (111)
IV	3 (106,114,116)

STRATIGRAPHY: Stratigraphic information is available for the eight most recently installed wells and was inferred for the earliest four wells. Wells depths vary from 40 - 90 feet, and are screened in sand, bedrock sandstone, or a combination. The sand is highly permeable; permeability estimates are 10^{-4} to 10^{-2} cm/sec. The bedrock sandstone aquifer is the source of private drinking supplies in the area.

The following table displays the results of the well categorization for the different strata. Only two divisions--clean and contaminated-- are used in order to maintain generality.

	I,II	III, IV
Formation type:		
Sand (uppermost)	4 (104,105,107,115)	1 (106)
Bedrock sandstone	2 (110,117)	2 (111,116)
1/2 sand; 1/2 fractured weathered sandstone	1 (118)	1 (114)
Sand, with some gravel, silt and clay.	1 (119)	

General comments:

We can regard as anomalies those wells which show signs of contamination despite their location apparently upgradient of, or beyond the hydraulic influence of the landfill. There are no anomalous wells at this site; all wells in categories III or IV are located downgradient of landfill. Moreover, the time-series plots support the finding of a vertical downward gradient at wells 115 and 116. Well 115 is shallower, and screened in sand, and only conductivity shows a significant trend. Well 116 is deeper and screened in sandstone; there are significant increasing trends for chloride, alkalinity, hardness and conductivity.

3. DESCRIPTION OF PROCEDURE

SPATIAL PATTERN ACROSS CONSTITUENTS: By comparing the boxplots for each constituent, we can learn whether there is a general pattern of spatial variability present. Figures 23 - 28 show the boxplots for the site; selected time-series plots are shown in Figures 29-32. We see immediately that well 106 and 114 are statistically (and dramatically) different from the other wells for all the constituents reviewed: chloride, COD, pH, alkalinity, hardness and conductivity. It is also clear that alkalinity, hardness and conductivity show similar patterns across all wells. This similarity corroborates the expected chemical relationships between these parameters: eg., hardness is measured as the sum of calcium and magnesium cations; alkalinity is measured as the equivalent concentration of bases; and conductivity is measured as the sum of ions in solution.

The boxplots also show a relationship between pH and alkalinity for wells 106 and 114: these wells have the lowest median pH and highest median alkalinity. The actual subsurface chemistry at these two wells is unclear, however, the plots suggest that the buffering capacity of these two wells may be exhausted at some future point, leading perhaps to further declines in pH. Note also that these differences in pH are statistically discernible from the rest of the wells at the site even though they are not one pH unit below the site median, which is the amount used by DNR to represent important changes in pH.

Although these relationships between constituent responses are elementary, they are nonetheless important. The observed chemistry of groundwater which has been affected by leachate depends greatly on leachate chemistry; if some constituents are not present it may

FIGURE 23

SAUK COUNTY LANDFILL
CHLORIDE

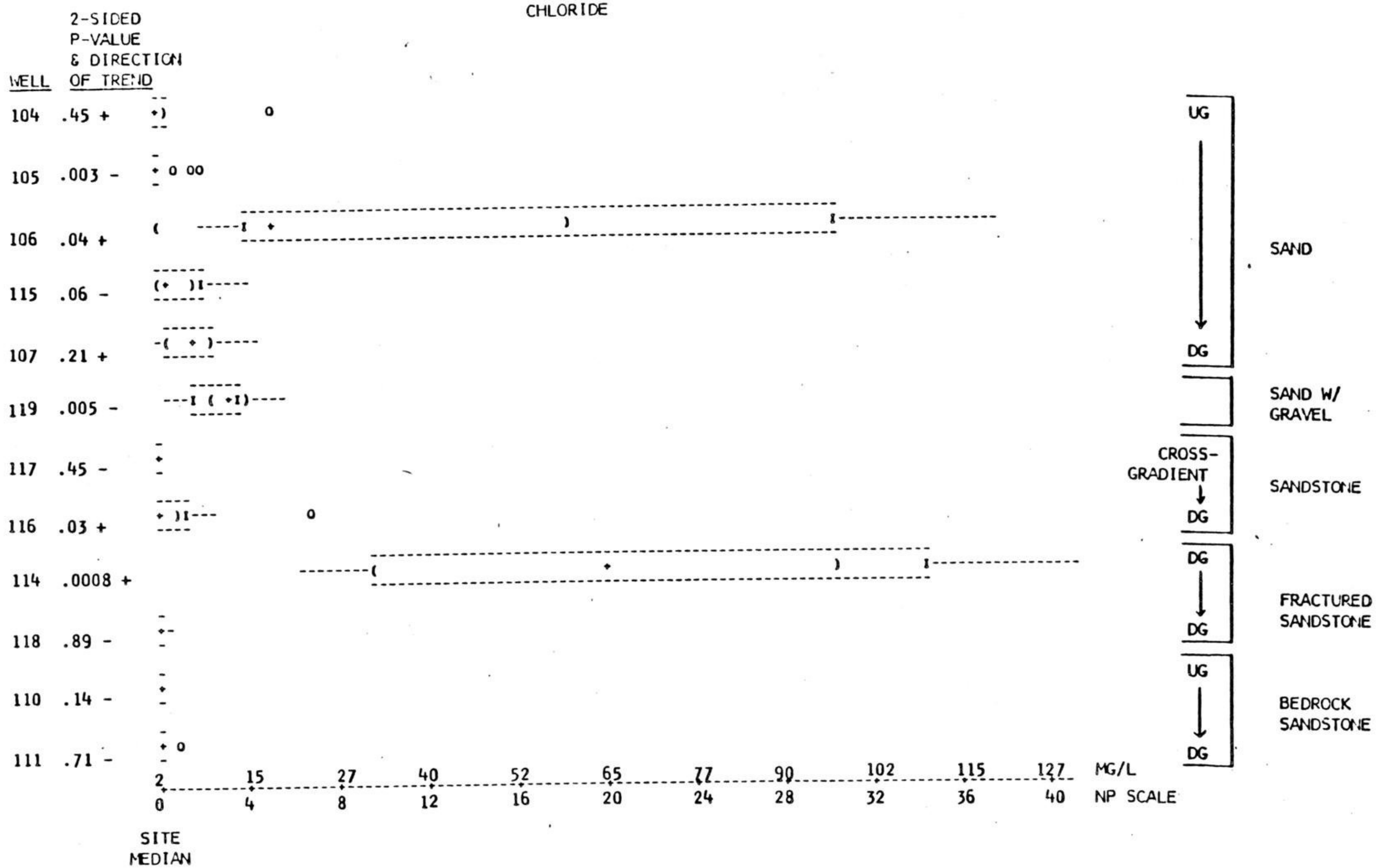


FIGURE 24

SAUK COUNTY LANDFILL

COD

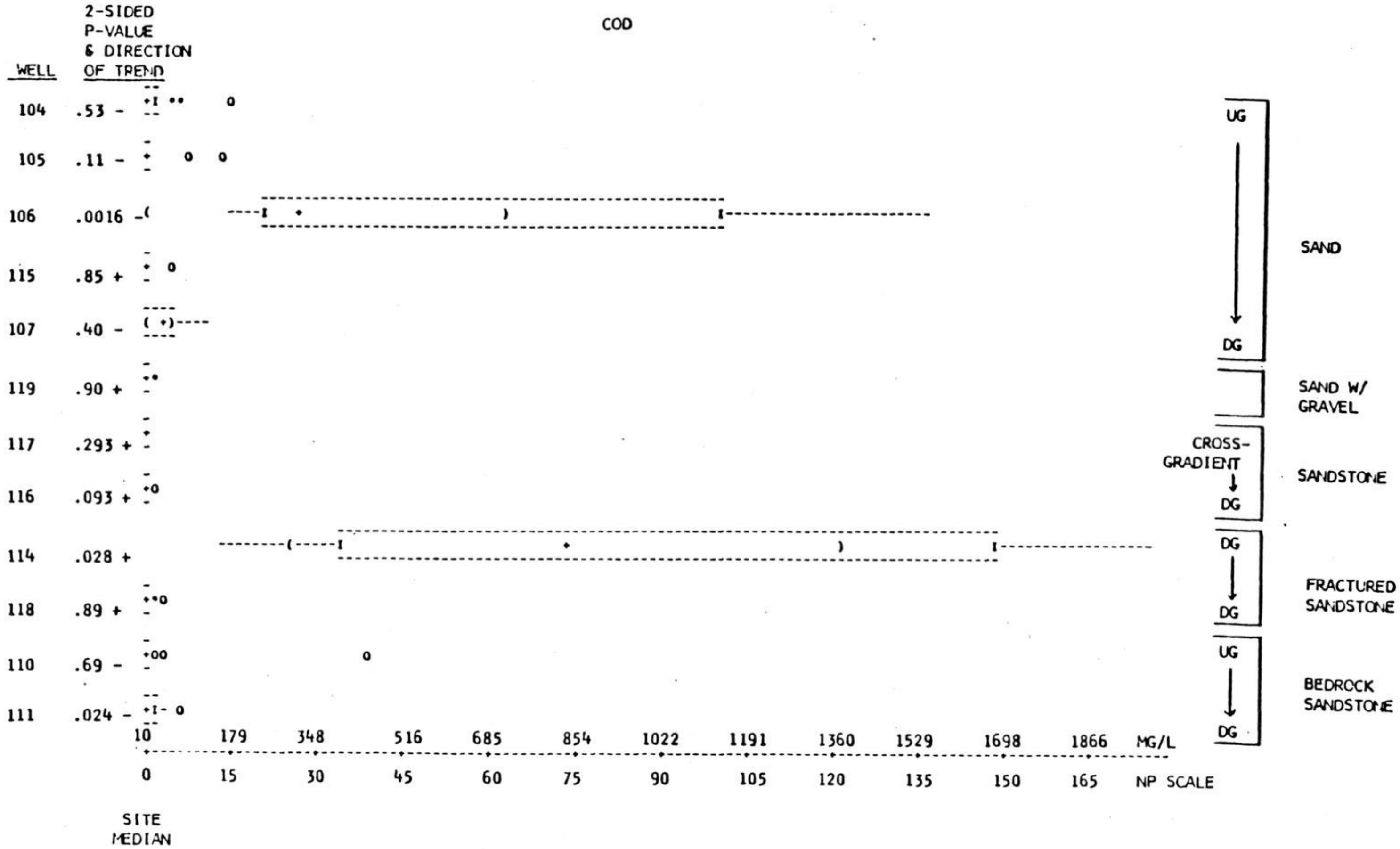


FIGURE 25

SAUK COUNTY LANDFILL

pH

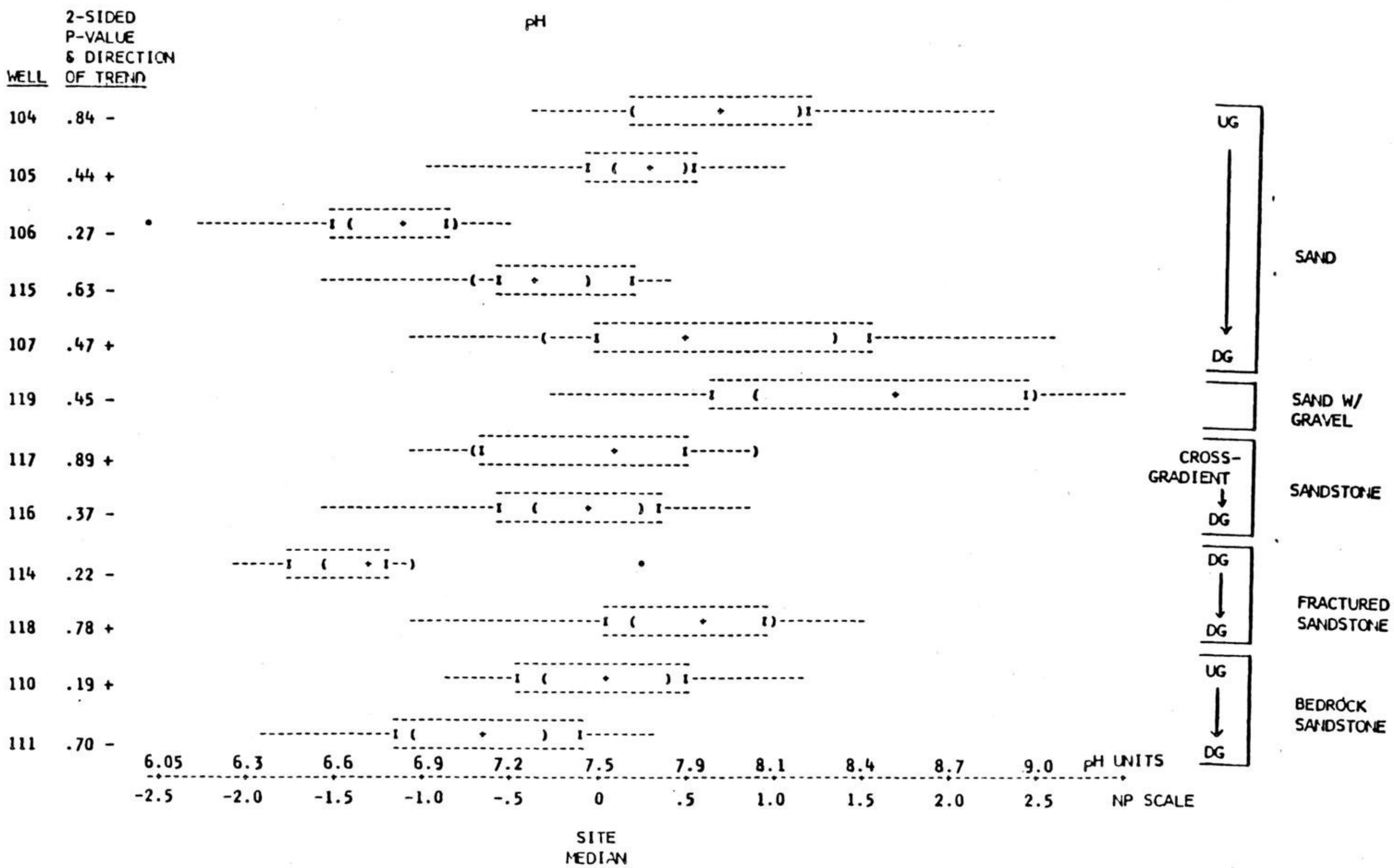


FIGURE 26

SAUK COUNTY LANDFILL

ALKALINITY

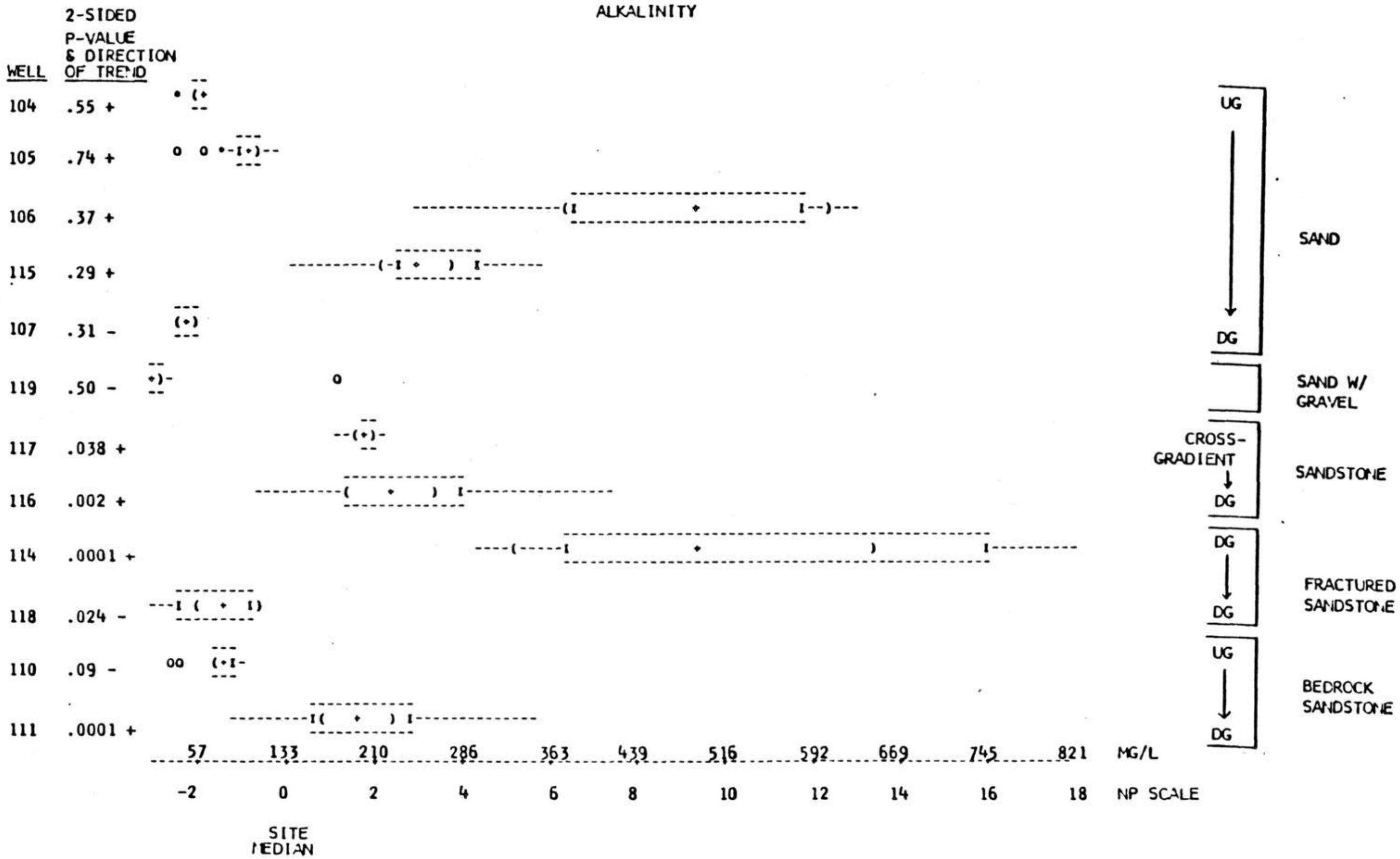


FIGURE 27

SAUK COUNTY LANDFILL

HARDNESS

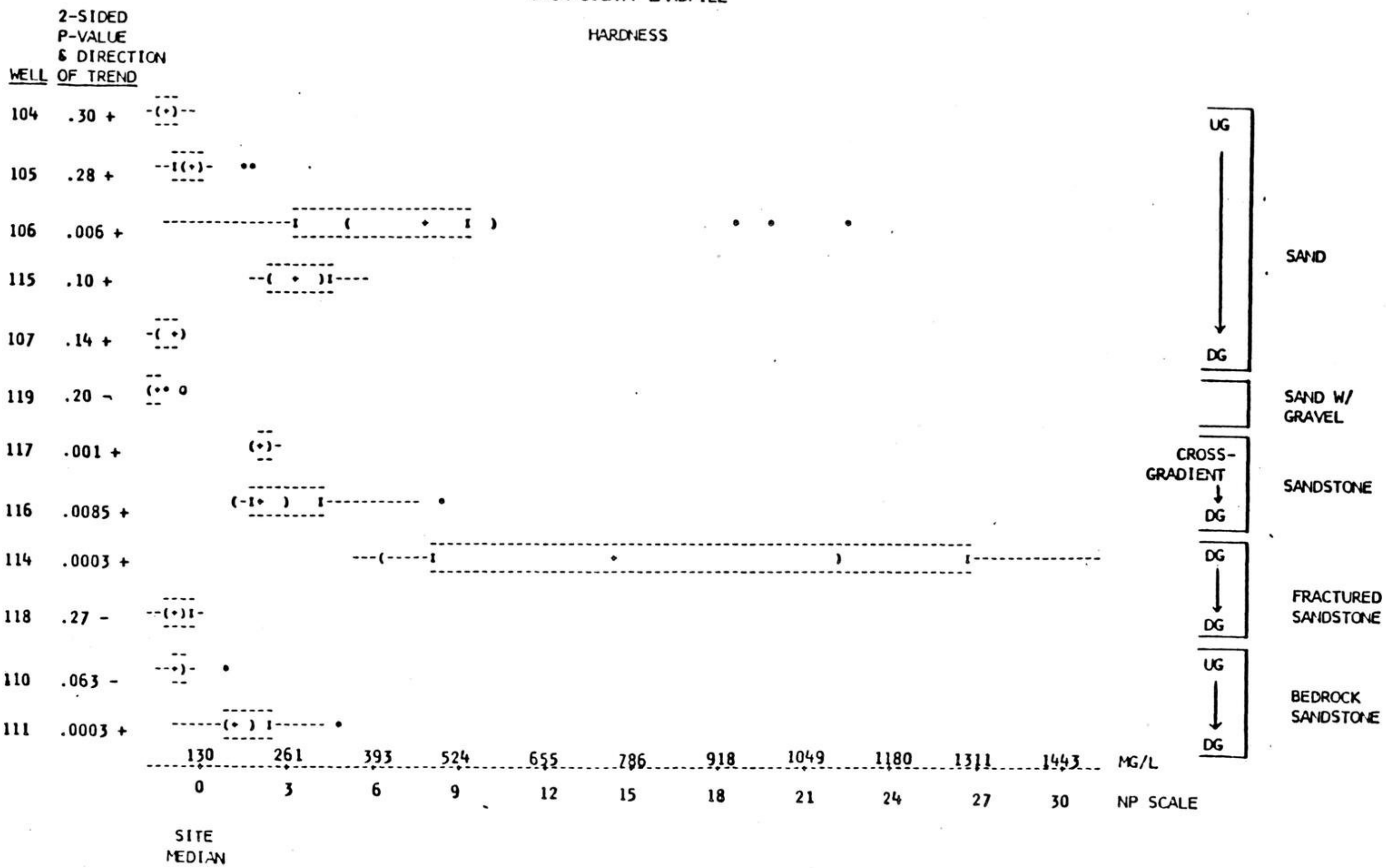


FIGURE 28

SAUK COUNTY LANDFILL

CONDUCTIVITY

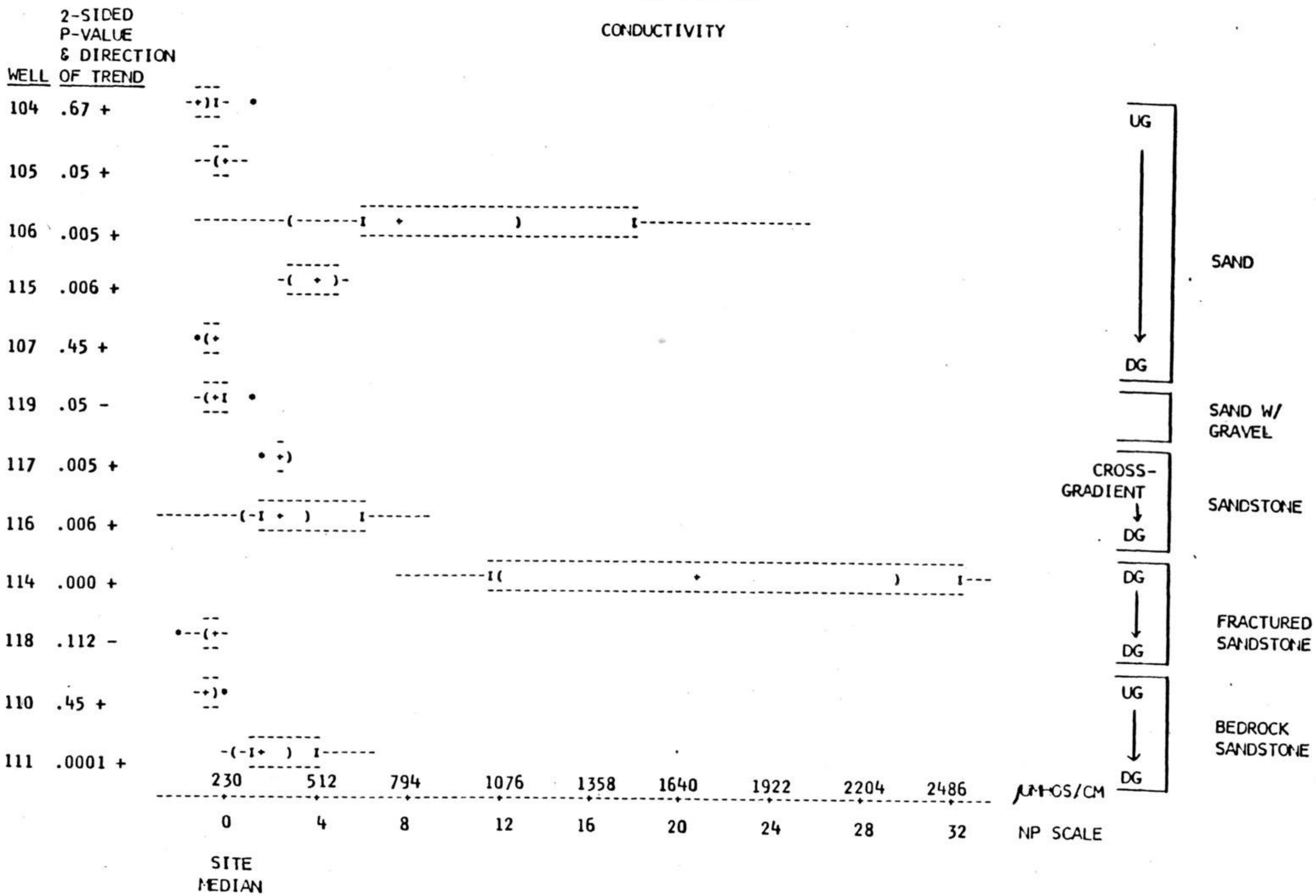


FIGURE 29
SAUK COUNTY LANDFILL
WELL 104 CATEGORY I SAND

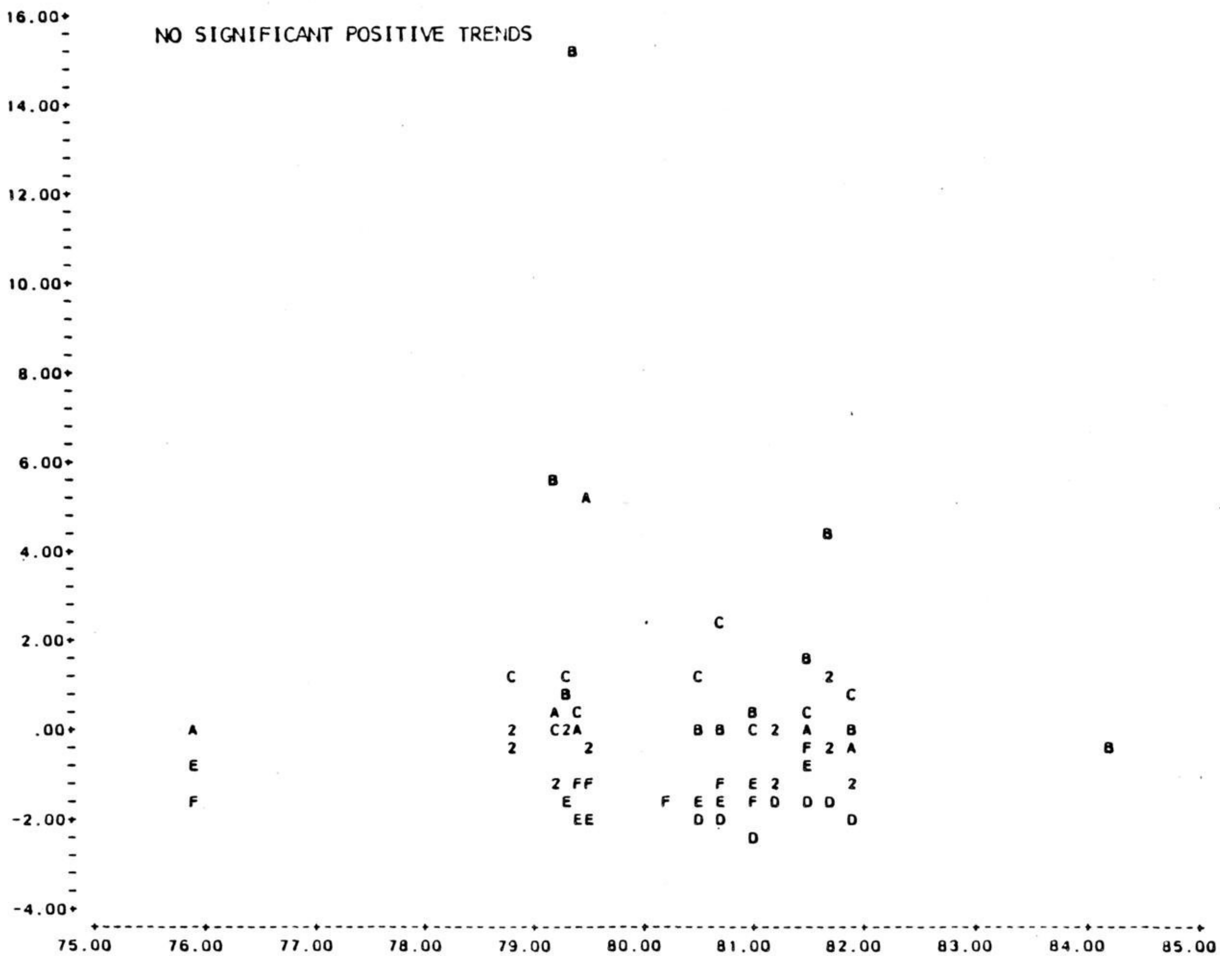


FIGURE 30
SAUK COUNTY LANDFILL
WELL 117 CATEGORY II SANDSTONE

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY
.038 + .0010 + .0005 +

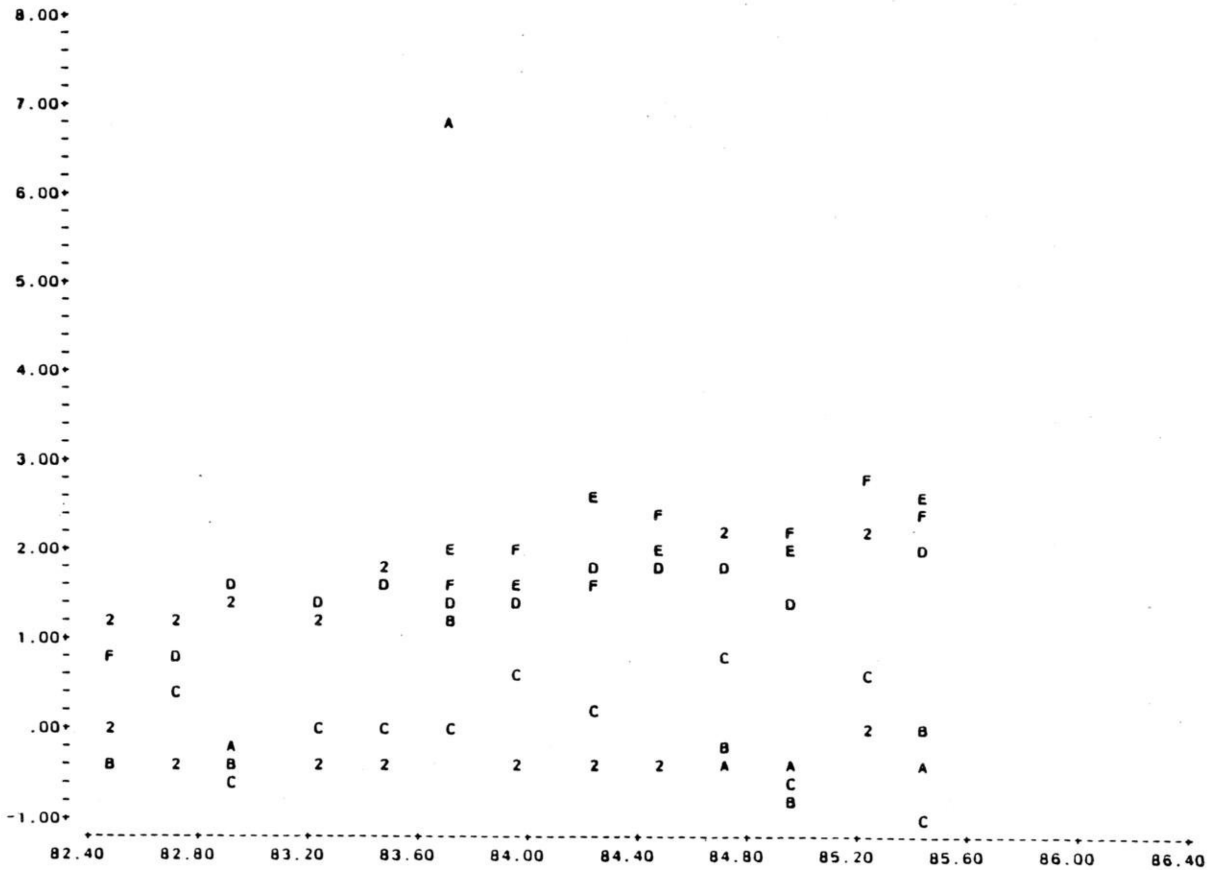


FIGURE 31
SAUK COUNTY LANDFILL
WELL 111 CATEGORY III BEDROCK SANDSTONE

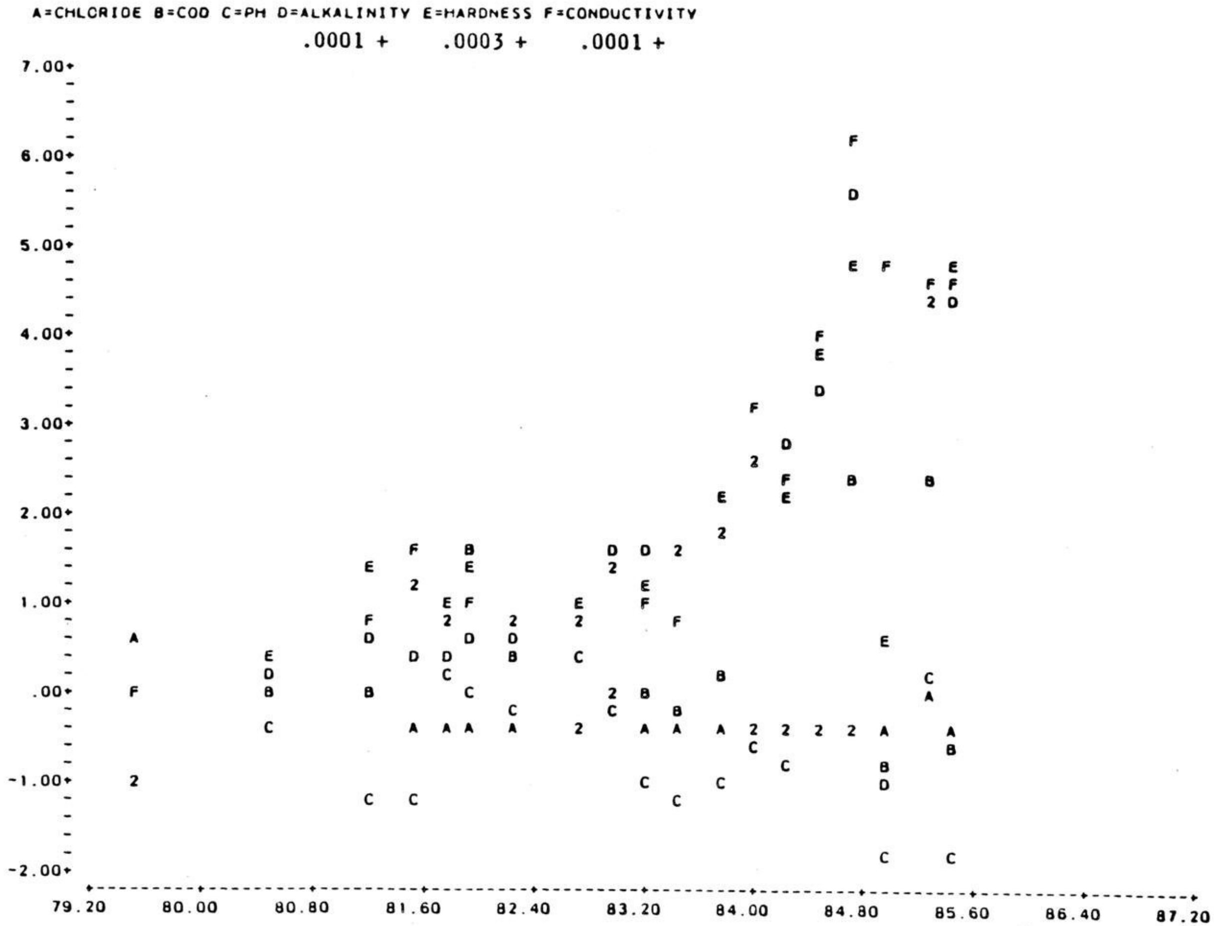
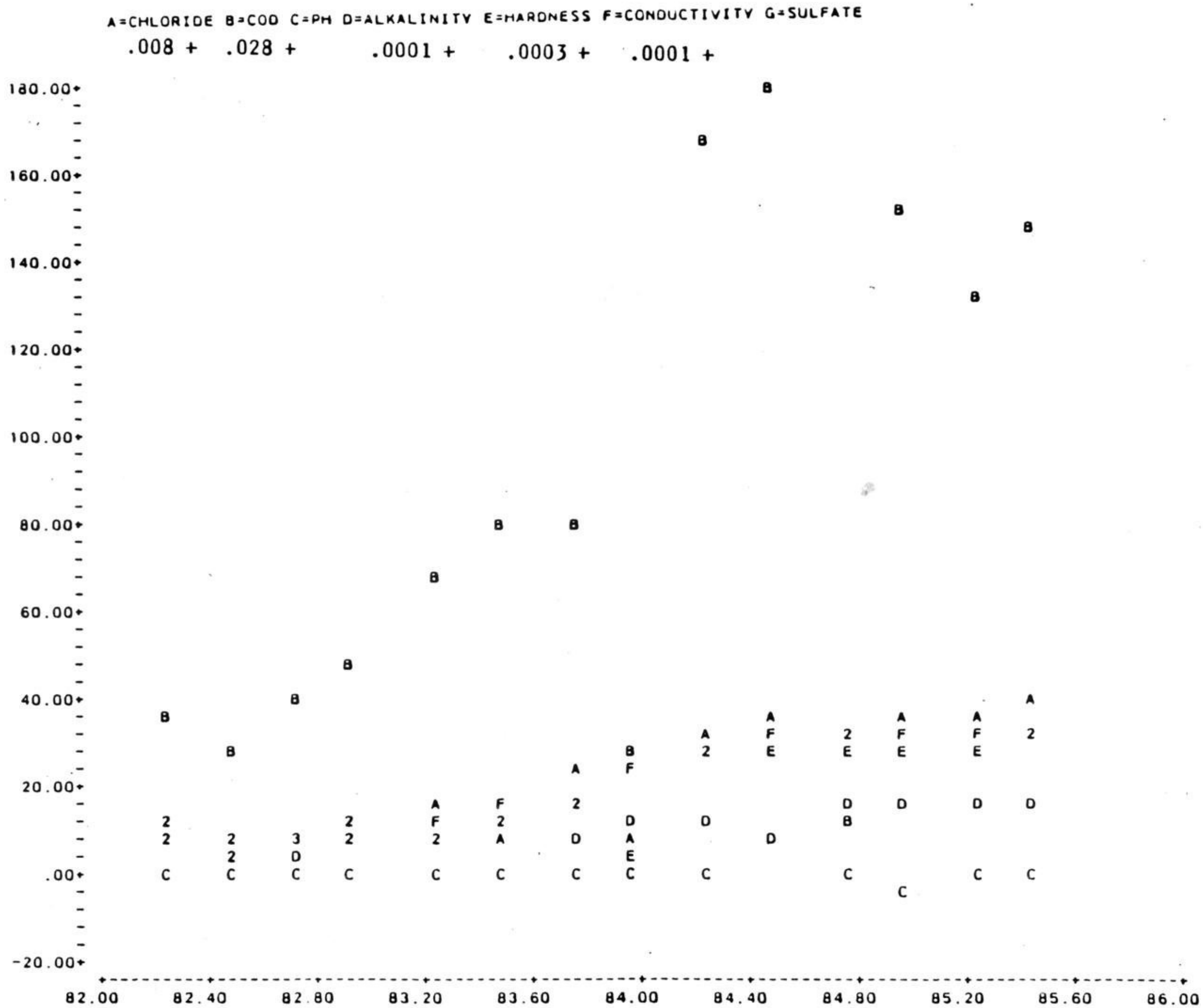


FIGURE 32
SAUK COUNTY LANDFILL
WELL 114 CATEGORY IV FRACTURED SANDSTONE



reflect only that they are not in the plume at that location. When the response of multiple constituents are corroborative, however, the evidence of contamination is clearly much stronger.

VISUAL TEST OF MEDIAN LOCATION: Having developed a visual image of spatial variability at the site, the next step is to review the confidence intervals about the individual well means for each constituent. The constituent hardness is reviewed as an example. By comparing the confidence intervals, we see that median hardness values for wells 106 and 114 are not just statistically different from the site median, but are different by a margin of about 5 to 6 NP values (or about 3.7 to 4.4 standard deviations). Similarly, median hardness values for wells 111, 115, 116, and 117 are statistically different from the site median value, by about 1 to 3 NP values. The median hardness values for the remaining wells are all centered slightly below the median value of about 130 mg/L. It is debatable, for example, whether the median for well 105 is statistically different from the median for well 107, but the spatial perspective provided by the boxplots indicates that a possible difference is not likely to be a policy concern.

TESTS FOR CHANGE IN QUALITY OVER TIME: From a hydrologic standpoint, it is important to determine whether these observed differences might be attributed to the landfill or to inherent strata characteristics. As described previously, a trend in quality at a single well is evidence of change unrelated to strata. It is often easier to quickly review trend results than to determine strata characteristics for large sites, especially if stratigraphic information is uncertain or unavailable. Thus, we can use trend results to focus on those wells for which stratigraphic information is especially important and conserve staff review time.

The Sauk County landfill is unusual in this respect. We see from the trend results that, for virtually every well which shows a significant difference in median values, this difference can be at least partially attributed to degraded water quality at each well during the period of the landfill's operation.

Returning to the plot for hardness, for example, we see that wells 106, 111, 114, 116, and 117 show highly significant trends; i.e., the greatest probability value for the trends at these wells is 0.0085, indicating an 8.5/1000 probability of observing such a pattern of increase by chance alone. It is important to remember that the Mann-Kendall test for trend is not based on absolute magnitudes of constituent concentrations, but rather on a pattern of increase or decrease. Thus, we note that the trend results for wells 111 and 114 are the same (0.0003). The boxplot, however, clearly shows that the magnitude of the trend for well 114 is far greater than for well 111. Note that the NP scale values on the time series plots also show this relative difference in magnitude between wells. Further, the sampling dates (as shown on the x-axis in Figures 31 and 32) show that this greater magnitude trend has occurred over a shorter period of time (i.e., data for well 111 is from 79.5 to 85.5; data for well 114 is from 82 to 85.5). Similar patterns are evident for conductivity and alkalinity.

From a statistical standpoint, we need go no further in our analysis, since we already have evidence of (1) a discernible trend in multiple constituents at several wells, and (2) different constituent populations, where these differences are plausibly explained by well location relative to the source of contamination.

The boxplots show that the lowest possible concentrations for the clearly contaminated wells are greater than the lowest concentrations for the unaffected wells. However, those wells having greater initial concentrations are located in vulnerable downgradient locations for which the earliest data was collected 6 to 9 years after landfilling began.

CATEGORIZING WELLS BY DEGREE OF CONTAMINATION: By drawing similar inferences on a well-by-well basis, it is possible to group the wells into four categories, related to the likelihood that contamination is present at each well. Examples of analysis and conclusions for each well category follow.

Category I, Well 104: This well is a water-table well which extends to a depth of 51 feet and likely ends in sand. Wells in similar formations are 105, 107, 115, and 119 (note that wells 104 and 119 are not located on the available site map).

A review of the trend results for each constituent shows that none of the five constituents had a significant increasing trend; in fact, the p-values are between .30 and .67, indicating a largely "random" pattern of variability. The plot for COD shows some outliers. The bulk of the data for COD, however, shows little variability, both at the well (varying between 0-45 mg/L), and relative to COD at all the other wells. The plot for chloride data also shows one outlier, but the rest of the data hardly varies. The plot for pH shows that the well median is within 1 NP of the site median. The plots for the remaining constituents are all on or below the site median. Finally, the time series plot for well 104 is checked. This confirms that the COD and chloride outliers are isolated incidents, perhaps sample contamination. Their presence does not change our overall conclusion that well 104 shows no sign of contamination.

Category II, Well 117: This well is screened at 65-75 feet in sandstone. Sampling began in mid-1982. Other wells screened in similar strata and thus useful for specific reference are wells 116, and perhaps well 110 and 111. For well 117, the individual median values for chloride, COD, and pH are all centered on the site median (as are all but one of the median values for the reference wells). The boxplots show that there is little variability about these low median values, and the double scale shows these median values are of low absolute magnitude. The trend results for chloride, COD and pH are not significant; these results are confirmed by the time-series plot for the well. Three constituents-- alkalinity, hardness and conductivity-- showed significant increasing trends. The total magnitude of each trend is less than two interquartile ranges. The median values for these constituents are significantly different from the site median, but again, by less than two interquartile ranges. Alkalinity, hardness and conductivity also show significant trends in two of the reference wells (111, 116), but these trends are of much greater magnitude than at well 117.

In summary, not all constituents show the same pattern and trend, and the magnitude of the effects which are evident are relatively small. We conclude that the well is likely not contaminated.

<u>Category III, well 111</u>: The boxplots show that chloride and COD concentrations are low and stable. The boxplot for pH shows that the well median is less than the site median value and statistically less than wells which do not show evidence of contamination (wells 104, 105, 107, 110). Hardness, alkalinity, and conductivity all show highly significant trends of moderate magnitude (i.e., about 5 IQR scale units). We conclude that well 111 is likely contaminated.

<u>Category IV, Well 114</u>: This well is so obviously contaminated, that only a brief summary of the evidence is necessary. As noted earlier, the boxplots for pH shows a significant displacement of the median pH below the values for all wells other than 106. Except for ph, all

constituents show highly significant, high magnitude trends. These trends are far greater and more persistent than necessary to determine statistical significance; they are clearly of environmental importance.

4. COMPARISON WITH WELFARE PARAMETERS

Chloride results provided additional evidence of contamination at three wells: 106, 114 and 116. Contamination, however, was also evident at these wells based solely on responses of indicator parameters.

5. COMPARISON WITH SITE-WIDE ESTIMATORS AS NP FLAGS

As discussed previously, Figures 15 through 21 include summary statistics for the Sauk County landfill site. The figures show that the site-wide estimates of median and IQR for chloride, COD, and pH agree closely with their category I counterparts. In contrast, the site median for hardness overestimates the category I median value by about 25%; but the IQR estimate is quite good. Similarly, for conductivity, the site-wide estimates for center and scale overestimates the corresponding category I values by about 20 %. An indicator parameter flag based on these site-wide estimates for hardness and conductivity would result in a generally more lenient standard than a flag based on a subset of clean wells. The site-wide estimates for alkalinity have the greatest discrepancy: the site median value overestimates the category I median value by about 63%;

the site IQR underestimates the category I IQR by about 50%.

Despite these uneven results, an NP value of 4 was used to see how effectively such a value might flag potentially contaminated wells for review. The results are shown in Table 5.

6. COMPARISON WITH VOC RESULTS AND SIMPLE NP FLAG

Table 5 compares the results of: (1) the well categories; (2) sampling for VOC's; and (3) using a simple flag based on an NP value of 4.

Based on these results, we see that, for the Sauk site, evidence of contamination provided by the analysis of inorganic constituents and COD agrees quite well with the limited VOC sampling results. The exception is for well 115, which is categorized as a II. Well 115 was so categorized primarily because the trends evident at the well were of fairly low magnitude and levelled off in about mid-1984 (recall that downward vertical gradients were estimated at well 115 and 116). The inorganic data clearly shows contamination at well 116, which is located near to well 115, but at a greater depth. Finally, we see that the simple NP flag compares reasonably well with the more labor-intensive, well-by-well categorization.

TABLE 5

SUMMARY RESULTS FOR SAUK CO. LANDFILL

CATEGORY/ well #	VOC RESULTS	NP FLAG
CAT. I		
104	--	COD--probable sample contamination; chloride--too early to check for trend
105	--	COD--probable sample contamination
107	--	COD--likely sample contamination
110	no detect	none
118	--	none
	0 Det.; 1/1	
CAT. II		
115	17 Det.; 2/2	alkalinity, hardness, conductivity--all show possible trend
117	--	none
119	--	chloride--too early to check for trend
	17 Det.; 2/2	
CAT. III		
111	--	alkalinity, hardness, conductivity--all show strong trend
CAT. IV		
106	--	chloride, COD, alkalinity, conductivity, sulfate--all too early to check for trend; initial concentrations are 5-19 NP units above site median
114	31 Det., 2/2	chloride, COD, alkalinity, hardness, conductivity--too early to check for trend; all show highly elevated initial concentrations (6-36 NP units); well first sampled in '82 vs. waste disposal in '73
116	23 Det., 2/2	alkalinity, hardness, conductivity--all show possible trend, but too early to be certain
	54 Det.; 4/4	

Notes:

1) "--" means no VOC samples taken at this well.

2) Well 114 has 7 individual VOC results over 100 ppm.

3) Well 25d has 17 Det., 1/1. The corrsponding DNR well number is not available.

B. CASE STUDY: MARATHON COUNTY LANDFILL SITE #2892

1. SITE CHARACTERISTICS

DESIGN: Clay lined with leachate collection. Landfill is 11 acres in size.

WASTES ACCEPTED: Municipal solid waste and industrial waste.

HYDROLOGIC SETTING: All wells are screened within the Horicon formation (composed of sandy, gravelly till), which is considered to be relatively homogeneous within the site vicinity. (Thus, the boxplots for this site are arranged by flow field only).

FLOW FIELD INFORMATION: Groundwater generally flows from the northwest to the southeast near the site. The water table elevations measured in 1983 suggest a horizontal gradient of about 8.5×10^{-3} ft./ft. This gradient steepens at the immediate southeast, downgradient corner of the landfill. (See Figure 33 for a site map).

SAMPLING DATES: The landfill began operation in 1980. Groundwater quality sampling began in May, 1980 for most of the wells reviewed. The constituents reviewed are: chloride, COD, pH, alkalinity, hardness, conductivity and sulfate.

2. ANALYSIS

TOTAL # OF WELLS REVIEWED: 6

NUMBER OF WELLS PER CATEGORY: All wells at the Marathon landfill are in Category I, indicating no evidence of contamination.

The boxplots for this site are shown in Figures 34-40; time-series plots are shown in Figures 36-41. A quick review of the

Figure 33
Plan View of Marathon
County Landfill

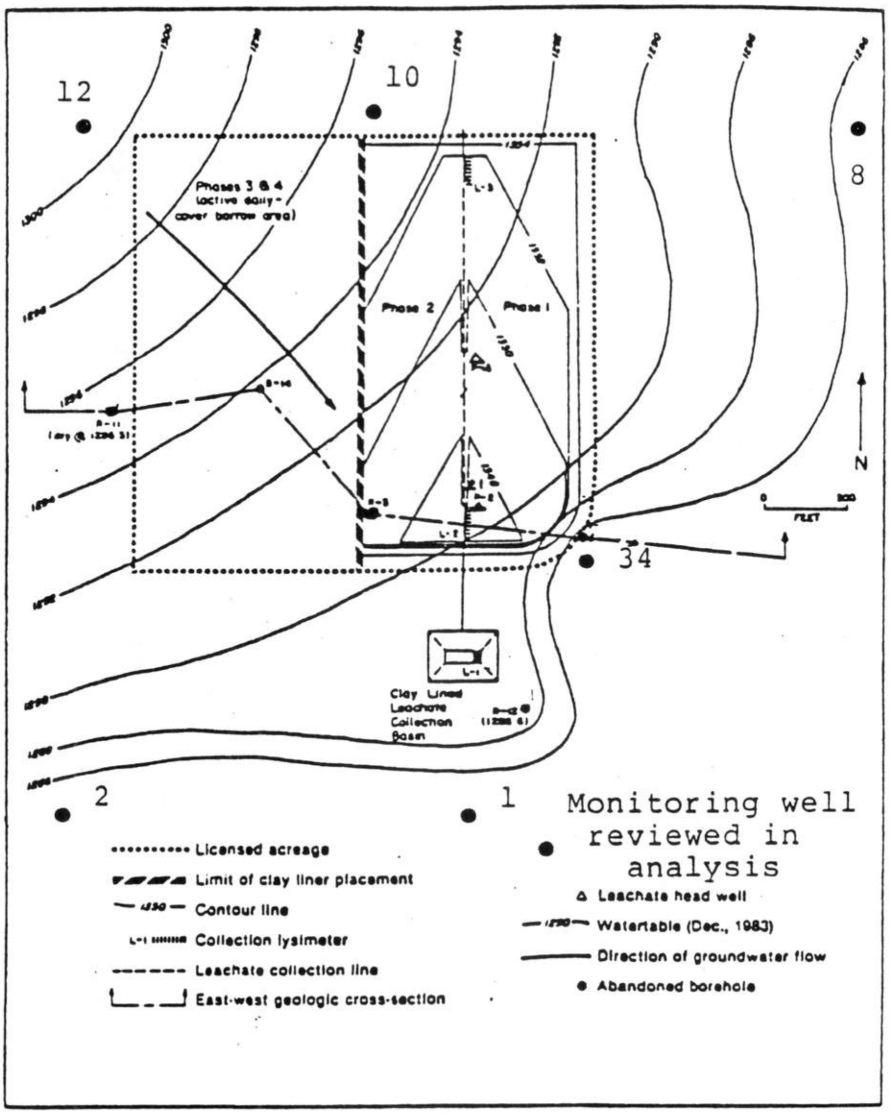


FIGURE 34

MARATHON COUNTY LANDFILL

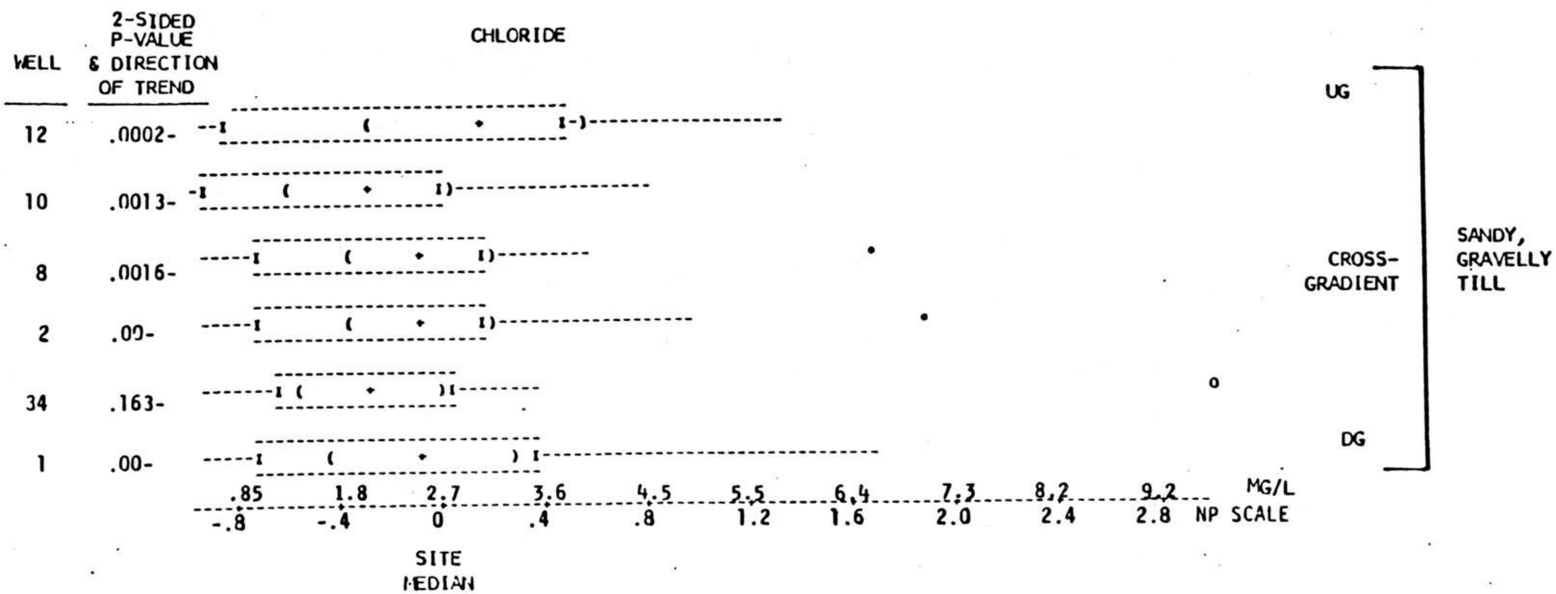


FIGURE 35

MARATHON COUNTY LANDFILL

SULFATE

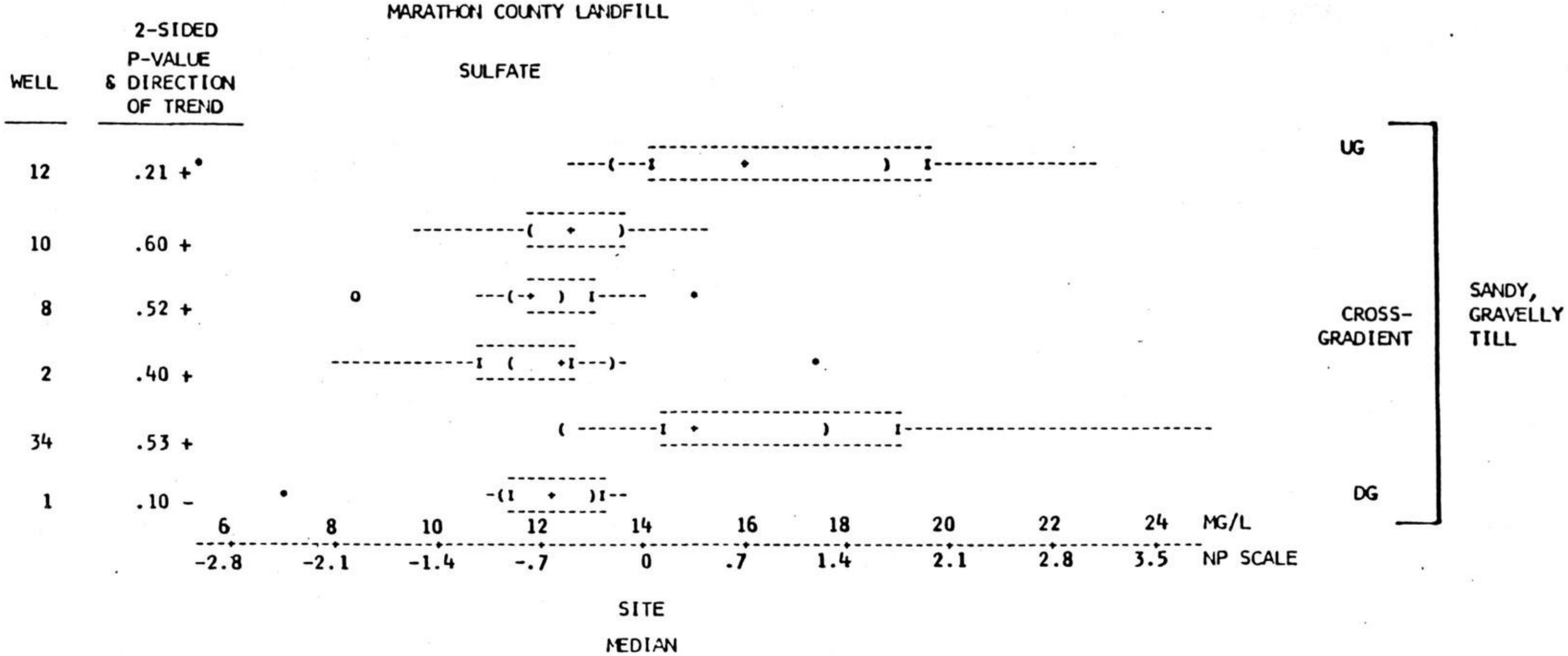


FIGURE 36

MARATHON COUNTY LANDFILL

COD

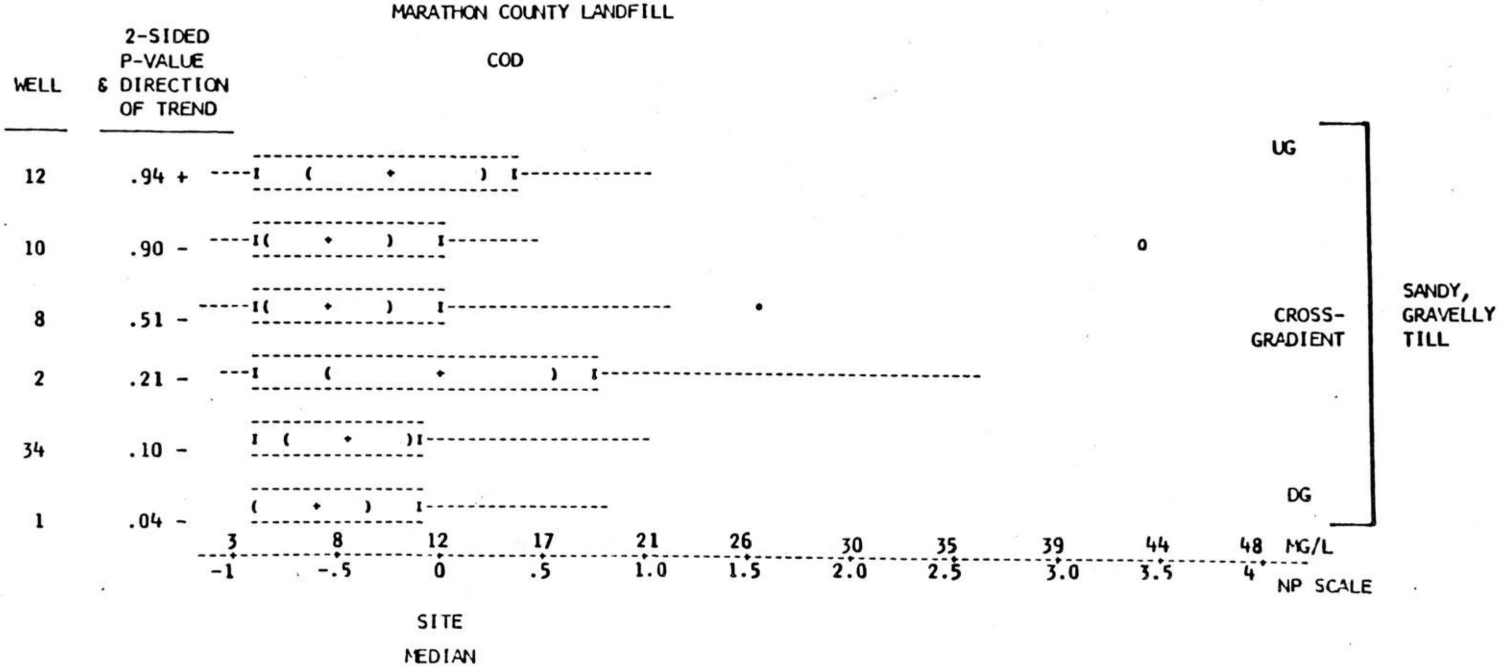


FIGURE 37

MARATHON COUNTY LANDFILL

pH

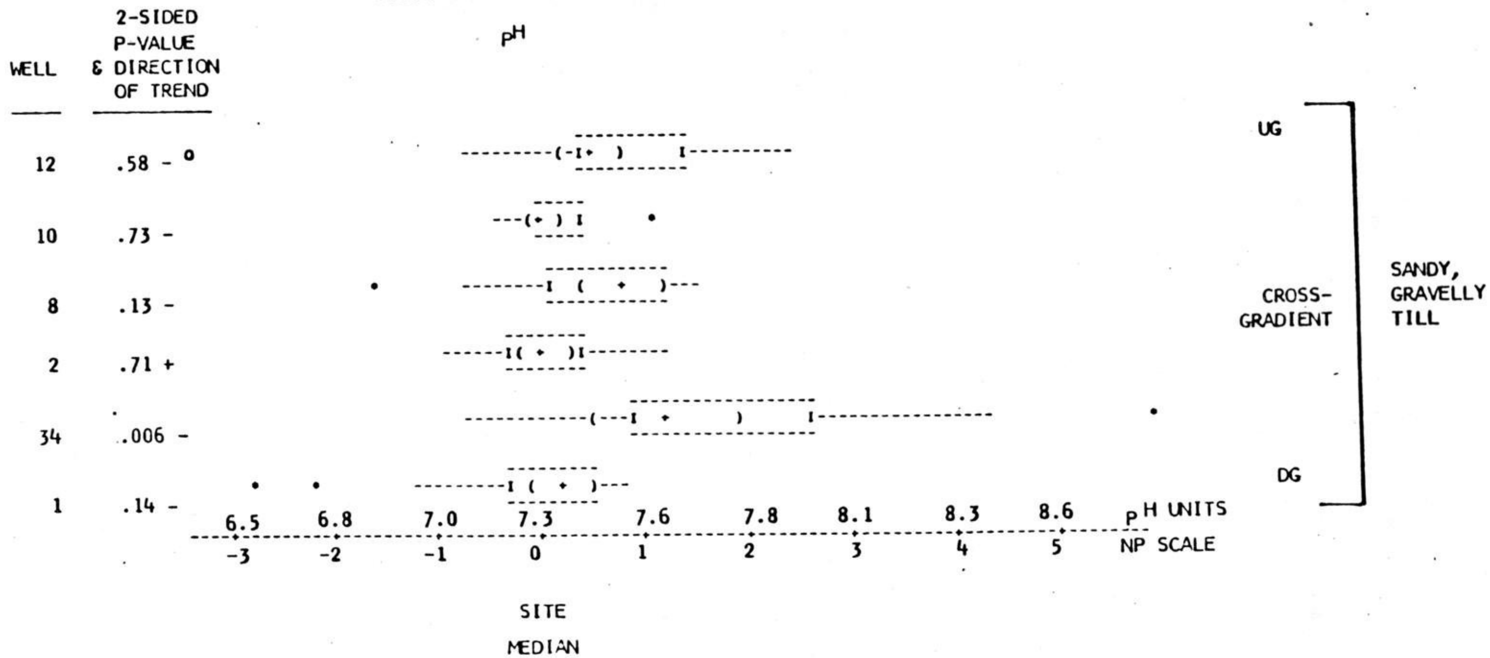


FIGURE 38

MARATHON COUNTY LANDFILL

ALKALINITY

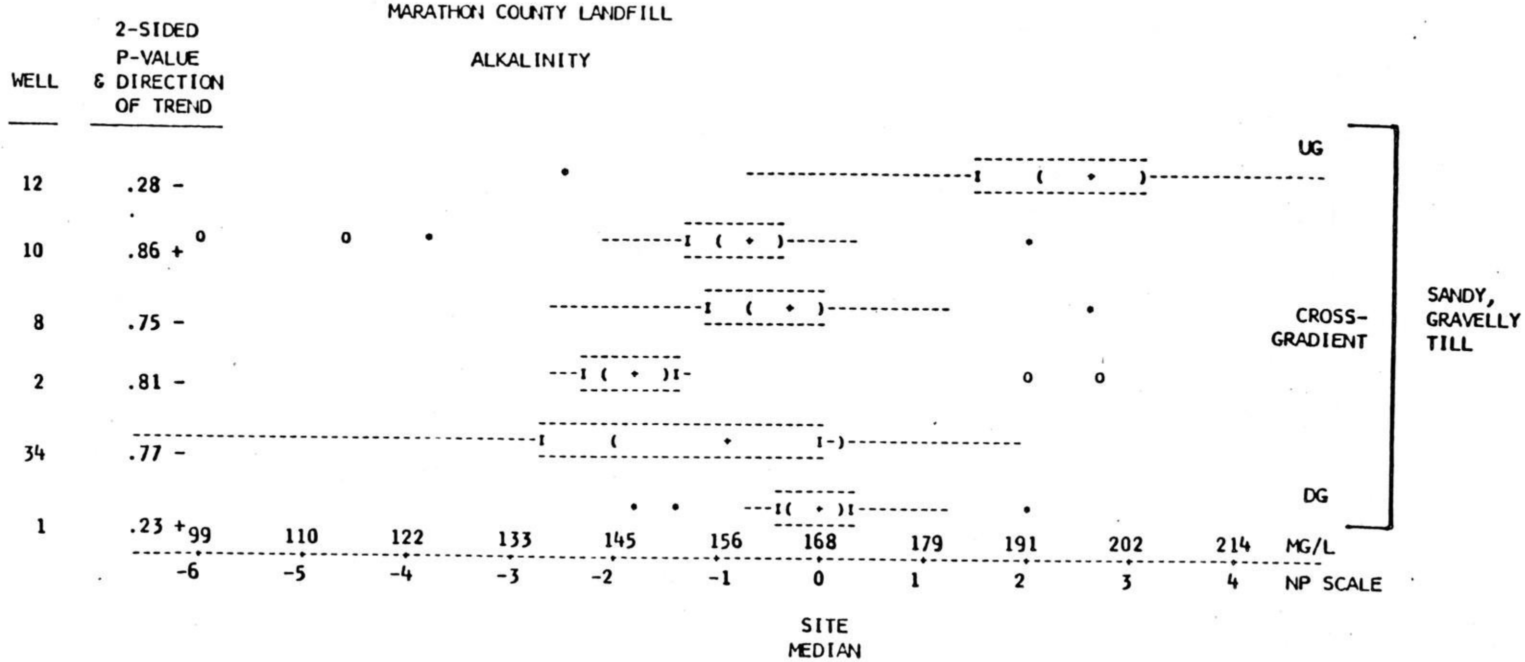


FIGURE 39

MARATHON COUNTY LANDFILL

HARDNESS

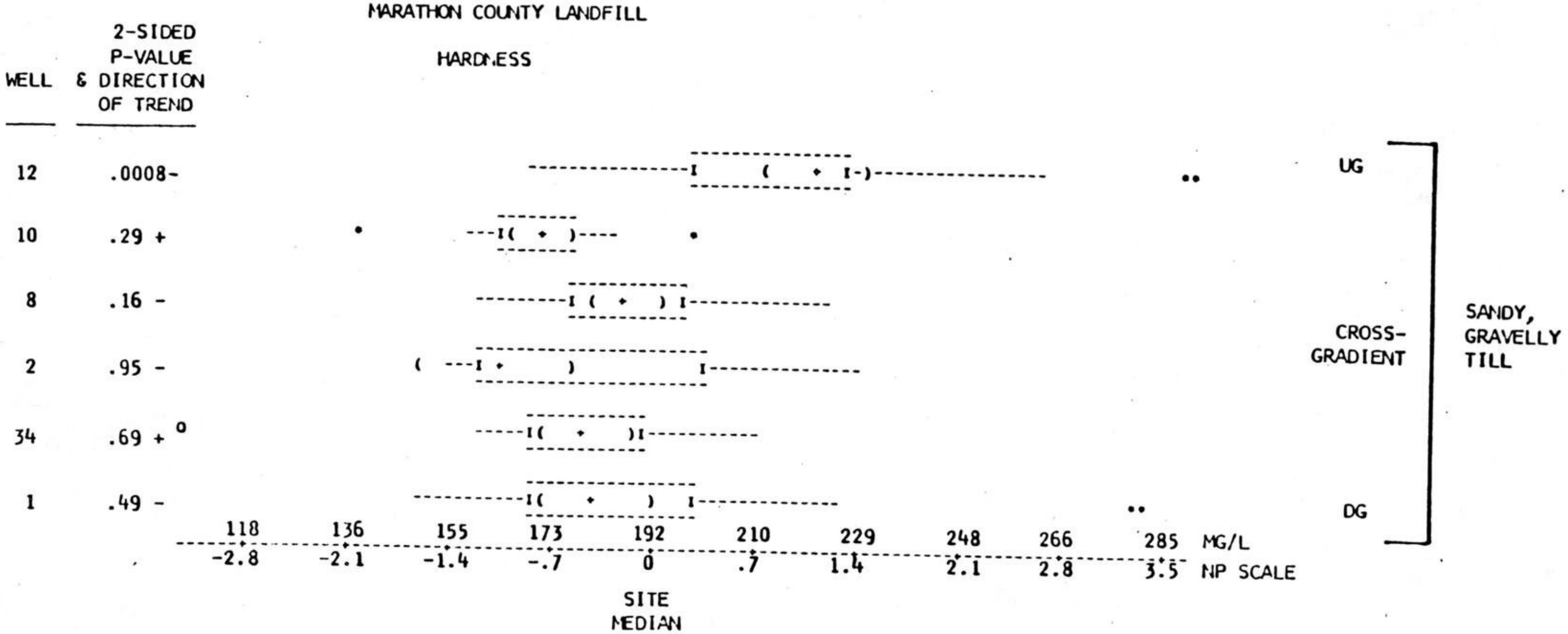


FIGURE 40

MARATHON COUNTY LANDFILL

CONDUCTIVITY

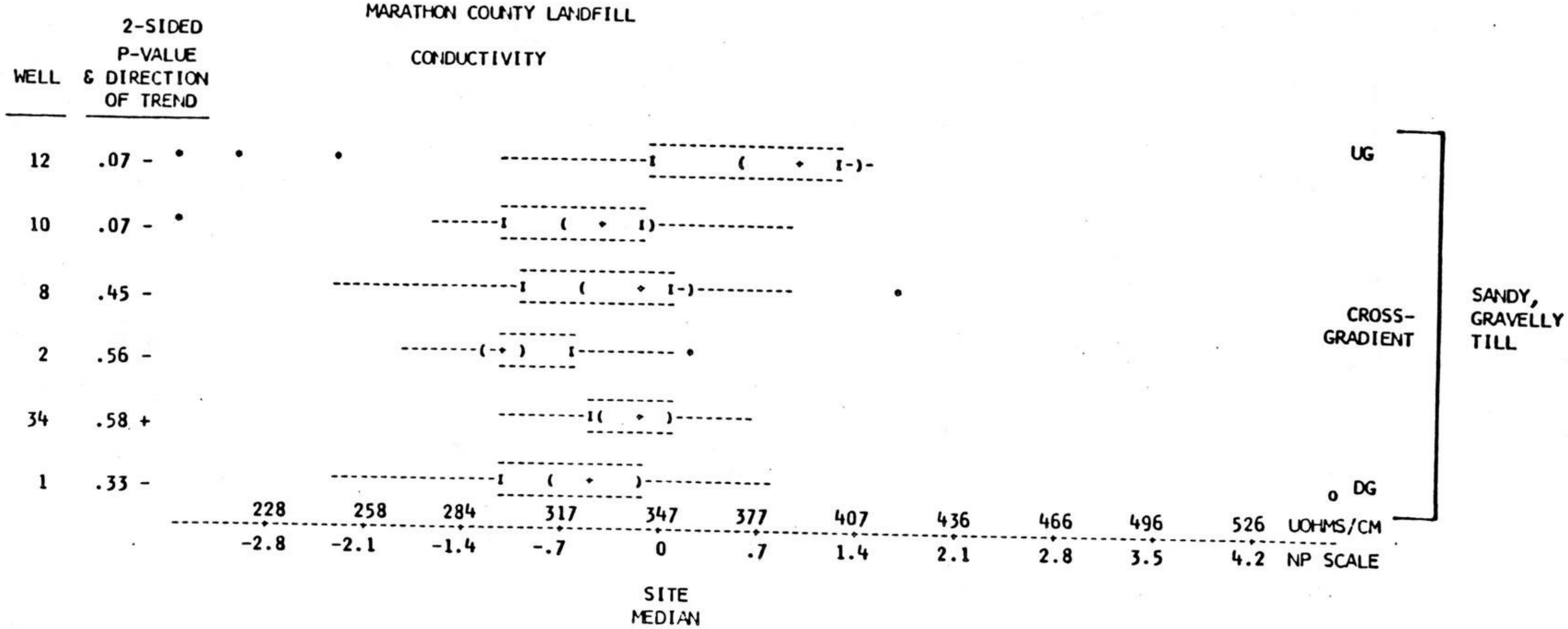


FIGURE 41

MARATHON COUNTY LANDFILL

WELL 12

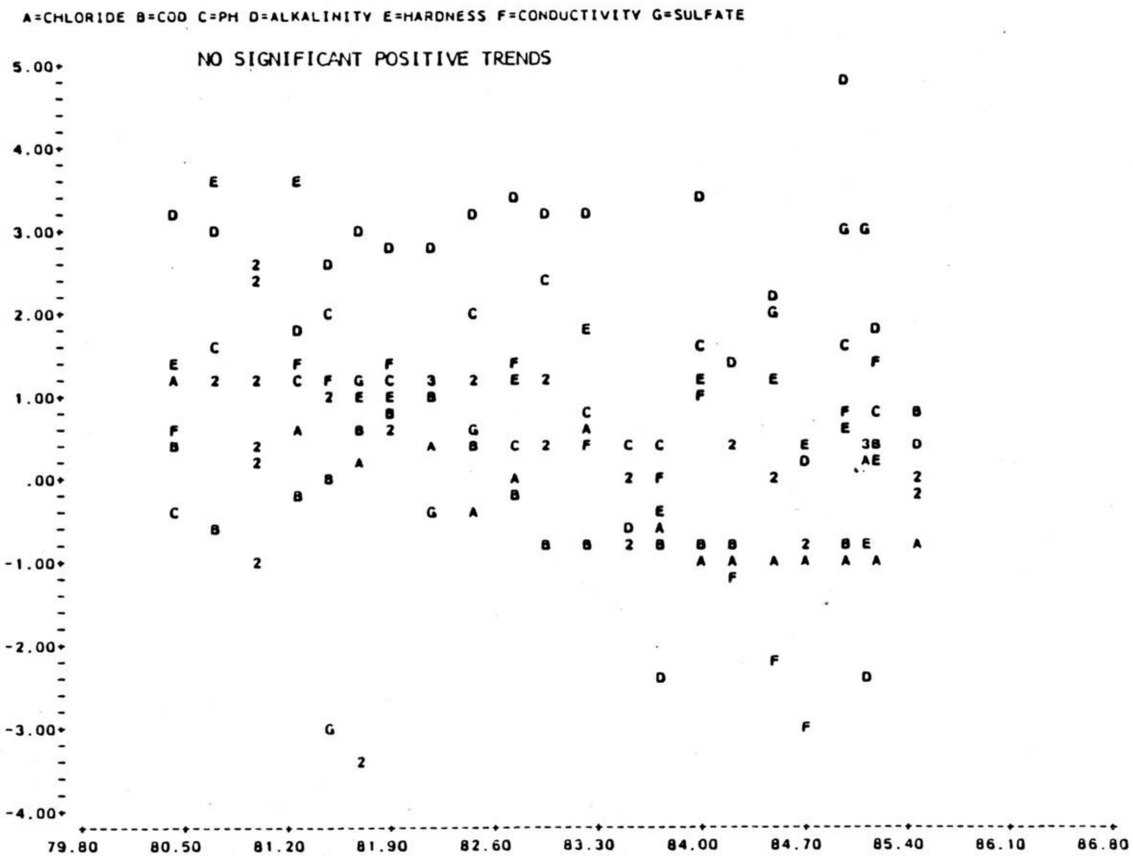


FIGURE 42

MARATHON COUNTY LANDFILL

WELL 10

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY G=SULFATE

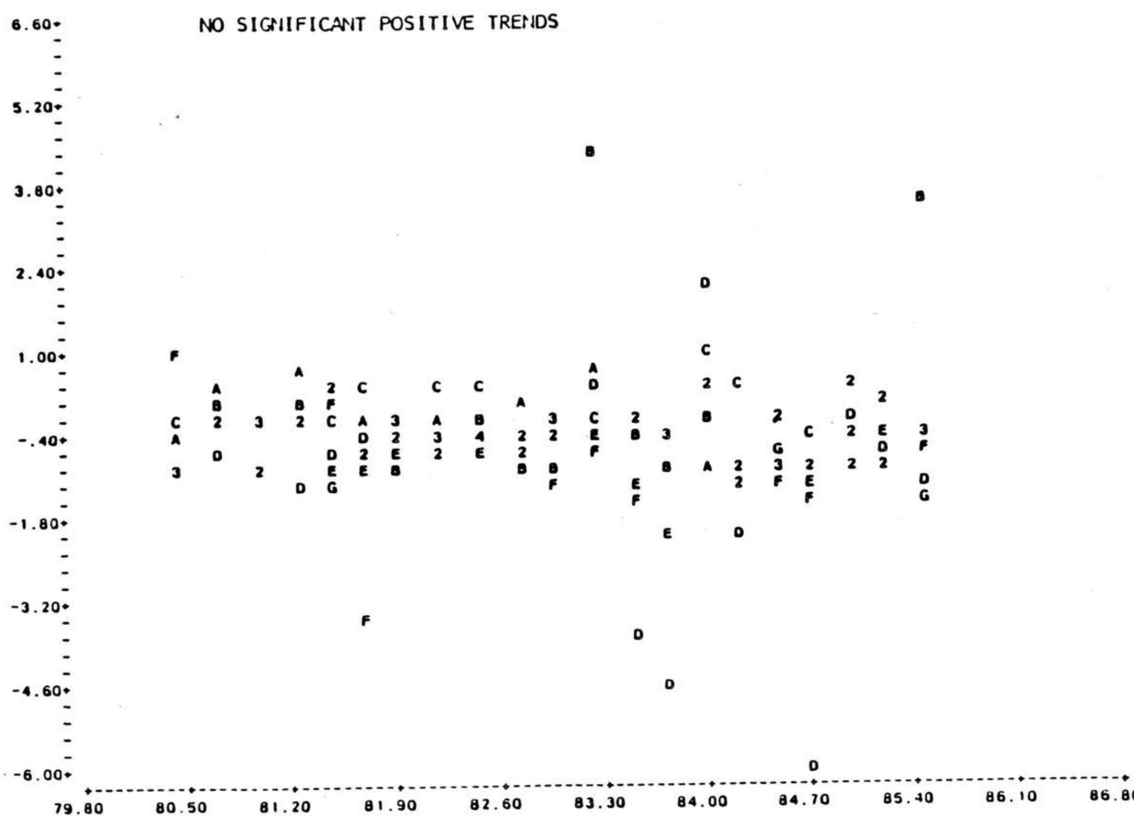


FIGURE 43

MARATHON COUNTY LANDFILL

WELL 8

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY G=SULFATE

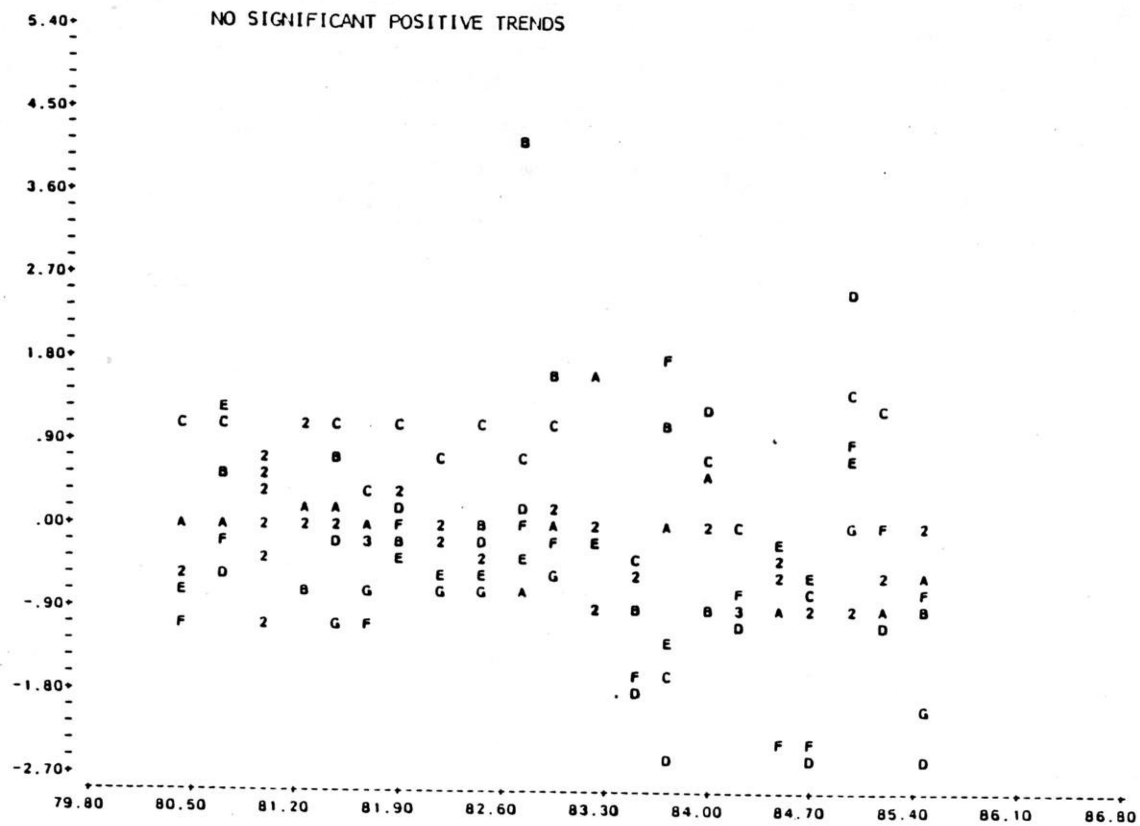


FIGURE 44

MARATHON COUNTY LANDFILL

WELL 2

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY G=SULFATE

NO SIGNIFICANT POSITIVE TRENDS

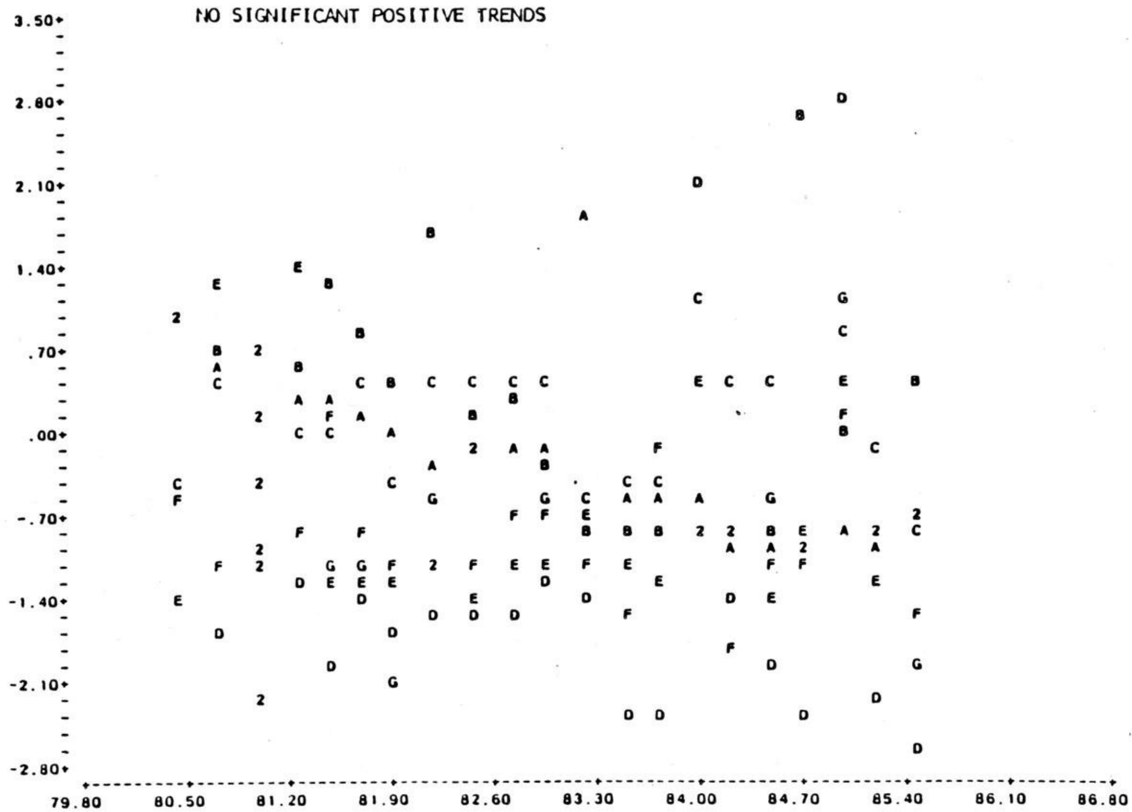


FIGURE 45

MARATHON COUNTY LANDFILL

WELL 34

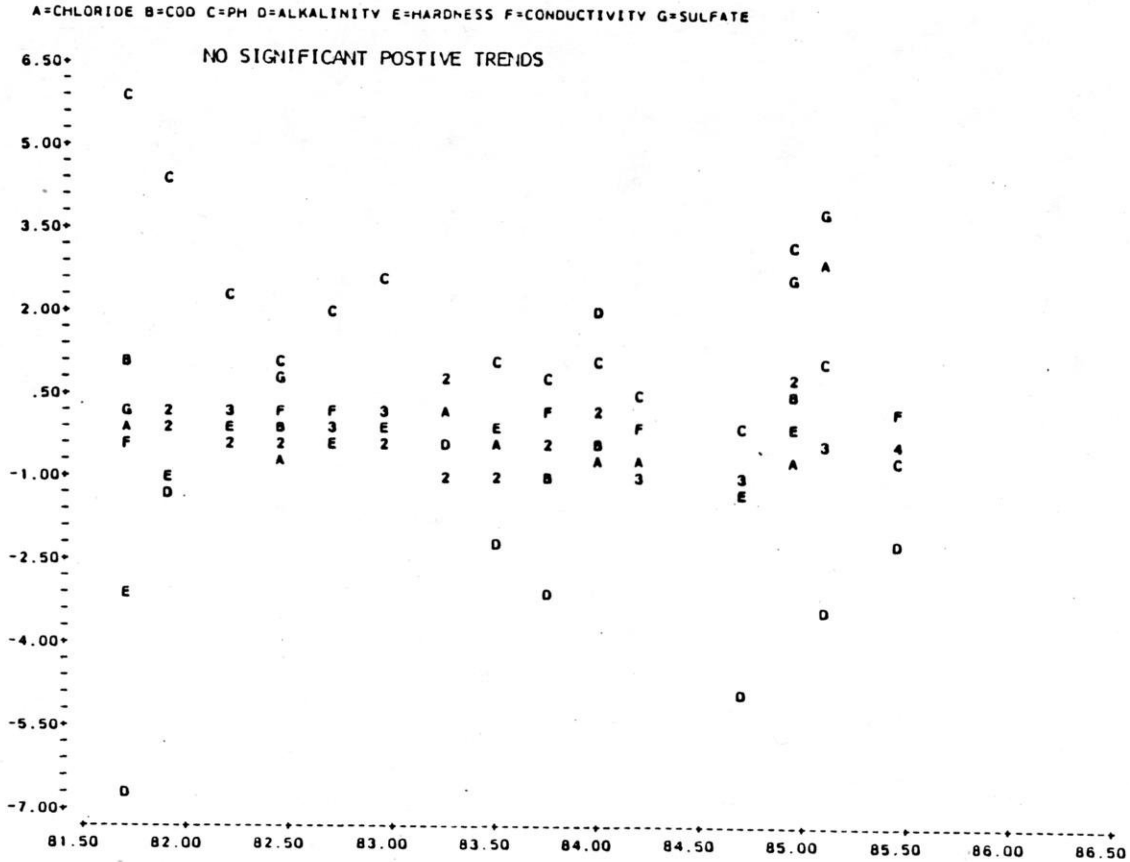
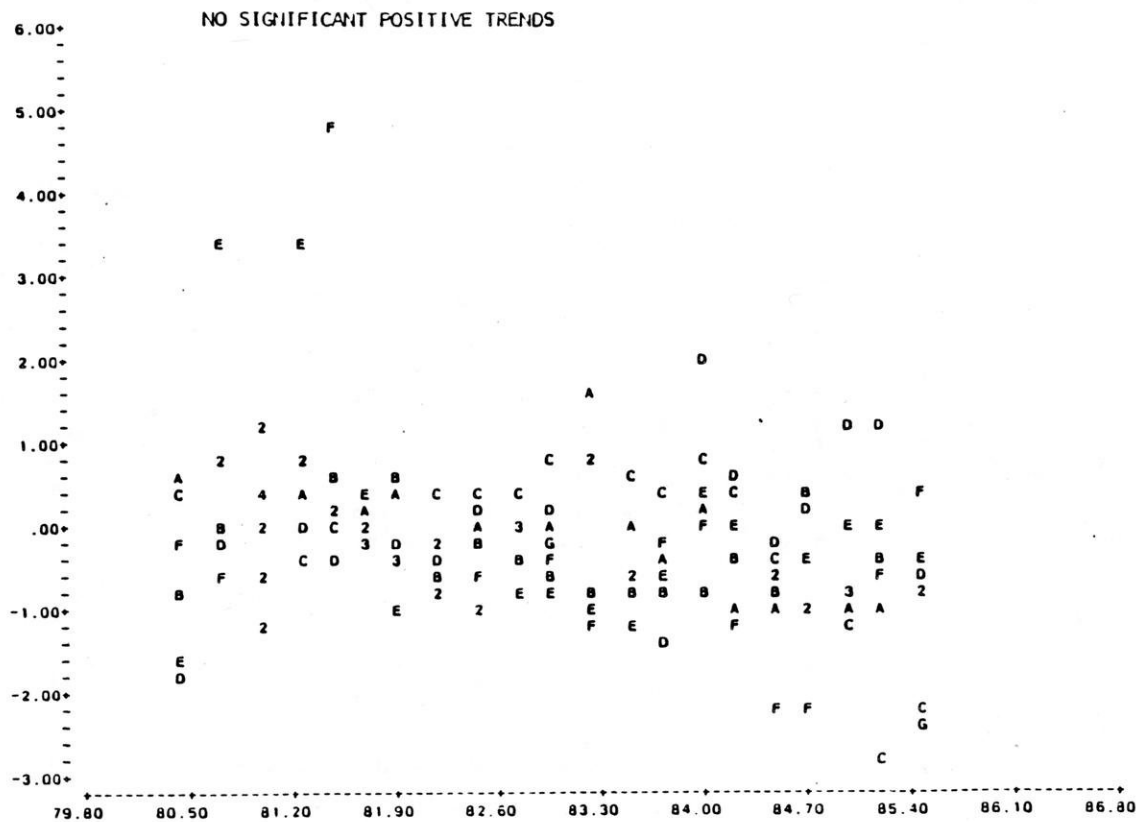


FIGURE 46

MARATHON COUNTY LANDFILL

WELL 1

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY G=SULFATE



boxplots confirms that constituent behavior at all wells is quite stable, regardless of well location. A visual test of medians shows that only a few wells have statistically discernible differences in their median values for alkalinity, hardness, conductivity and sulfate. None of these differences, however, are important according to the criteria for an environmentally important median displacement.

Similarly, there are no significant positive trends for any constituent. If there were, the spatial overview provided by the boxplots shows that none would be environmentally important. Well 34 shows a significant decline in pH. The boxplot suggests that this may be due to initially high pH conditions, which is confirmed by the time-series plot. The boxplot also shows that even with the decrease in pH, the median value at the well remains higher than for the other wells.

Just as with the Sauk site, we can draw these conclusions without specific stratigraphic or location information.

3. COMPARISON WITH WELFARE PARAMETERS

Chloride and sulfate were reviewed at this site. Neither provided information regarding contamination beyond that provided by the indicator parameters. Concentrations of both constituents are well below their legally defined PALs. Unmistakable trends, or signals of degrading water quality, in either of these constituents could easily be detected long before the statutory PAL would be reached.

4. COMPARISON WITH SITE-WIDE NP FLAG

Since all wells are in Category I, the site-wide estimates are identical to the Category I estimates.

5. COMPARISON WITH VOC RESULTS AND SIMPLE NP FLAGS

Table 6 summarizes the results of the VOC analysis and NP flags. As shown in the table, sampling for VOCs was conducted twice at only two wells included in this analysis. There were no detections of VOCs at these wells. VOCs were detected in the landfill leachate in concentrations between 2 - 3200 ppb.

The limited VOC data agree well with the inorganic and COD data. The data, however, are too scant for meaningful comparison. Finally, we see that the site-wide NP screening value flagged only six isolated observations. As Table 6 shows, three of these were likely outliers, two might be related to well completion, and the sixth was not part of an increasing trend.

TABLE 6

SUMMARY RESULTS FOR MARATHON CO. LANDFILL

CATEGORY/ well#	VOC RESULTS	NP FLAG
CAT.I		
1	--	conductivity--too early to check for trend
2	--	none
8	--	COD; possible sample contamination
10	--	COD; possible sample contamination
12	no detect	alkalinity--no trend
34	no detect	pH--initially high concentration (6 NP above site median)

Notes:
"--" means no VOC sample taken at the well.

C. CASE STUDY: DANE CO. LANDFILL--VERONA SITE #2680

1. SITE CHARACTERISTICS:

DESIGN: Natural attenuation design; partial leachate collection system installed during Phase II expansion. Leachate head well is #127.

Site is 49 acres; approved capacity is about 2 million cubic yards of waste.

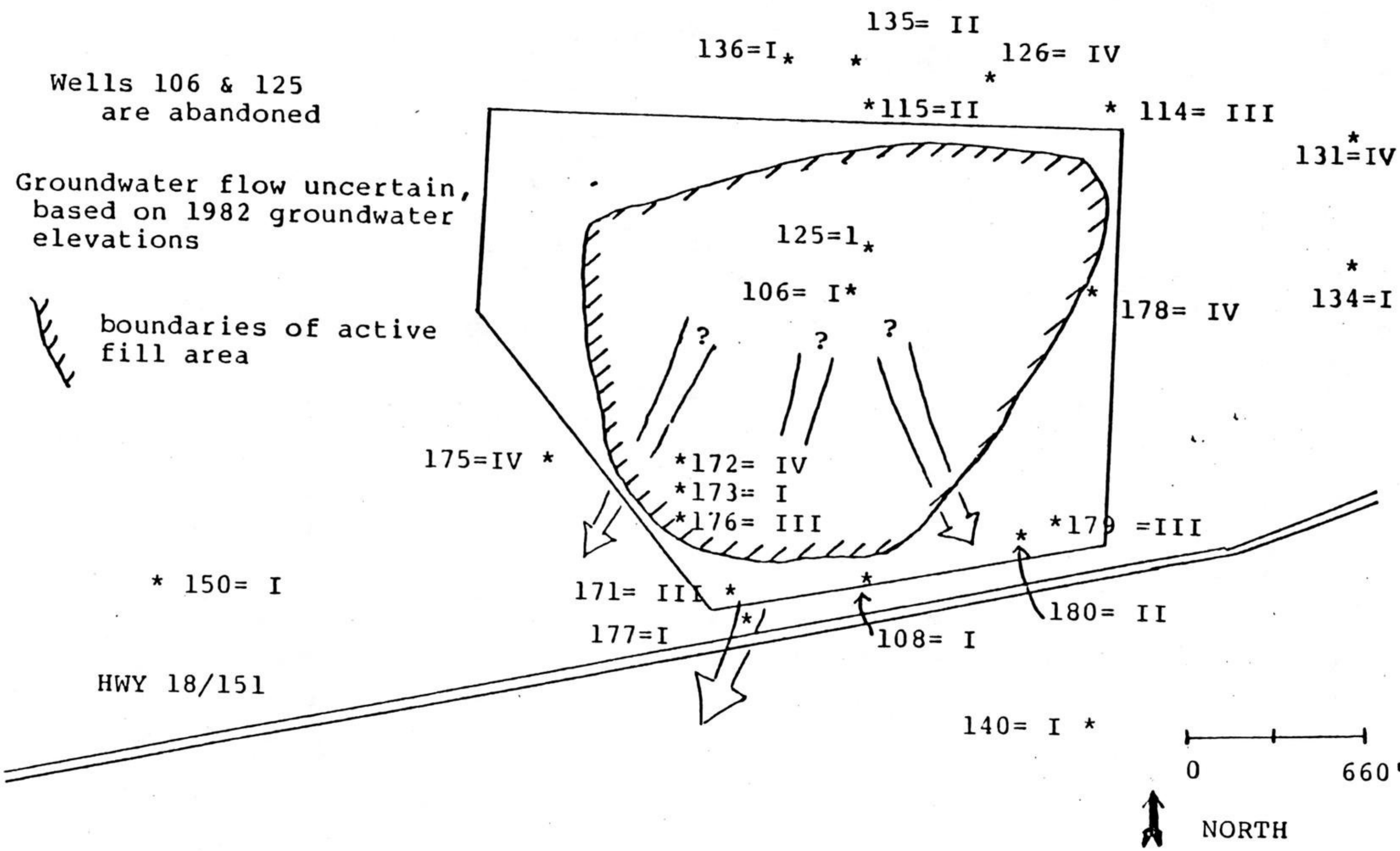
WASTES ACCEPTED: Municipal solid waste.

SAMPLING INFORMATION: Five constituents are reviewed: chloride, COD, pH, hardness, and conductivity. Sampling began in March, 1977. The landfill was licensed and began operations in July 1977. Water quality sampling began in 1977 at the first wells. Sampling at additional wells (#171-181) began between 1980 and 1984.

FLOW FIELD: Figure 47 shows a map of the site. Plans submitted in 1975 suggested that groundwater flowed to the southwest and had a "small" horizontal gradient. Additional groundwater elevation measurements taken in 1982 showed that flow may be to southwest, south, or southeast. In December, 1983, the horizontal gradient across northern part of site was estimated at 0.001 feet/foot; the horizontal gradient across southern part of site was estimated at 0.022 feet/foot.

DNR staff conclude that there are insufficient wells to determine if groundwater mound exists near Phase I area of site and that additional

FIGURE 47
DANE COUNTY - VERONA LANDFILL SITE MAP



wells are needed to verify groundwater flow along northern and northeastern edges of site. An EIS prepared for site projected maximum leachate quantities of 10,000 gallons per day. No estimates are readily available regarding leachate quantities removed by the system.

2. ANALYSIS

TOTAL NUMBER OF WELLS REVIEWED: 22, including some private wells.

NUMBER OF WELLS PER CATEGORY:

Category	#Wells
I	10
II	5
III	2
IV	5

STRATIGRAPHY: The available stratigraphic information enables grouping the wells into 5 formations. Well depths vary from 22-110 feet.

The following table compares the results of well categorization for the different strata. Only two divisions-- clean and contaminated-- are used in order to maintain generality.

Formation type:	I, II	III, IV
Silty sand, till (upppermost)	9	3
Sand & gravel	1	
Dolomite bedrock (Pr.du Chien)	1	2
Cambrian sandstone (underlies dolomite bedrock).	1	
Limestone:		
a. bedrock	1	1
b. weathered	2	
c. fractured		1

As shown in the table, there is no clear pattern of contamination related to strata characteristics; there are clean and contaminated wells in each strata. In particular, the three wells screened in the silty sand (wells 171, 172 and 175) and identified as likely contaminated are all located downgradient of the site. In contrast to the Sauk landfill, however, there are wells located apparently up and cross-gradient of the site for which the data indicate contamination. Two of these (wells 114 and 126) are screened in dolomite bedrock; two others (wells 179 and 178) are screened in limestone.

3. DISCUSSION OF SELECTED WELLS:

Wells which show evidence of contamination despite an apparent upgradient location or which have highly elevated concentrations despite lack of significant trend results require special consideration. Such anomalous wells are discussed below, with reference to the attached graphical and trend analyses. The boxplots for the site are shown in Figures 49 - 53 (in order to fit the boxplots for all wells on a single page, a modified boxplot format was used. This format omits the upper and lower horizontal lines of the "box," but contains the same information as the previously described plots. If a confidence interval "bracket" falls in the same location as "I" indicating the edge of box, the "I" is overridden and only the confidence interval bracket is plotted). Selected time-series plots are shown in Figures 54-60; the remaining time-series plots are in Appendix D.

A. Upgradient wells:

Well 114 (MW-14): This well is located about 180 feet upgradient of northeast corner of site. It is screened in dolomite bedrock at a depth of 47 feet and it is a water table well.

Well 126 (MW-26): This well is also screened in dolomite bedrock at a depth of 42 feet at the northeastern, upgradient edge of site. It is also a water table well.

In assessing groundwater quality at these wells, we can use information from all of the other wells monitored at the site. In addition, we can make comparisons with groundwater quality at well

FIGURE 49

DANE COUNTY - VERONA LANDFILL

CHLORIDE

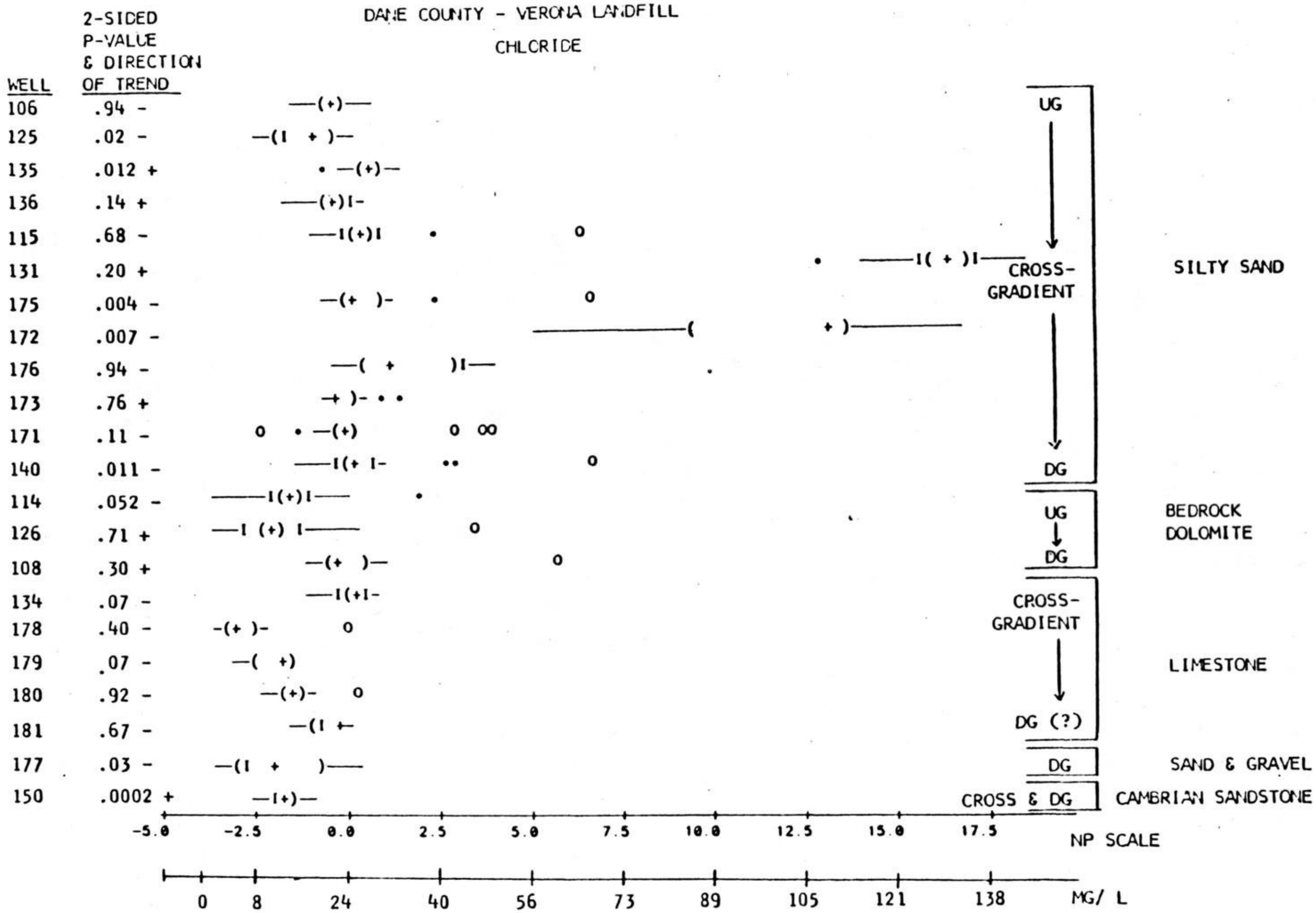


FIGURE 50

DANE COUNTY - VERONA LANDFILL

COD

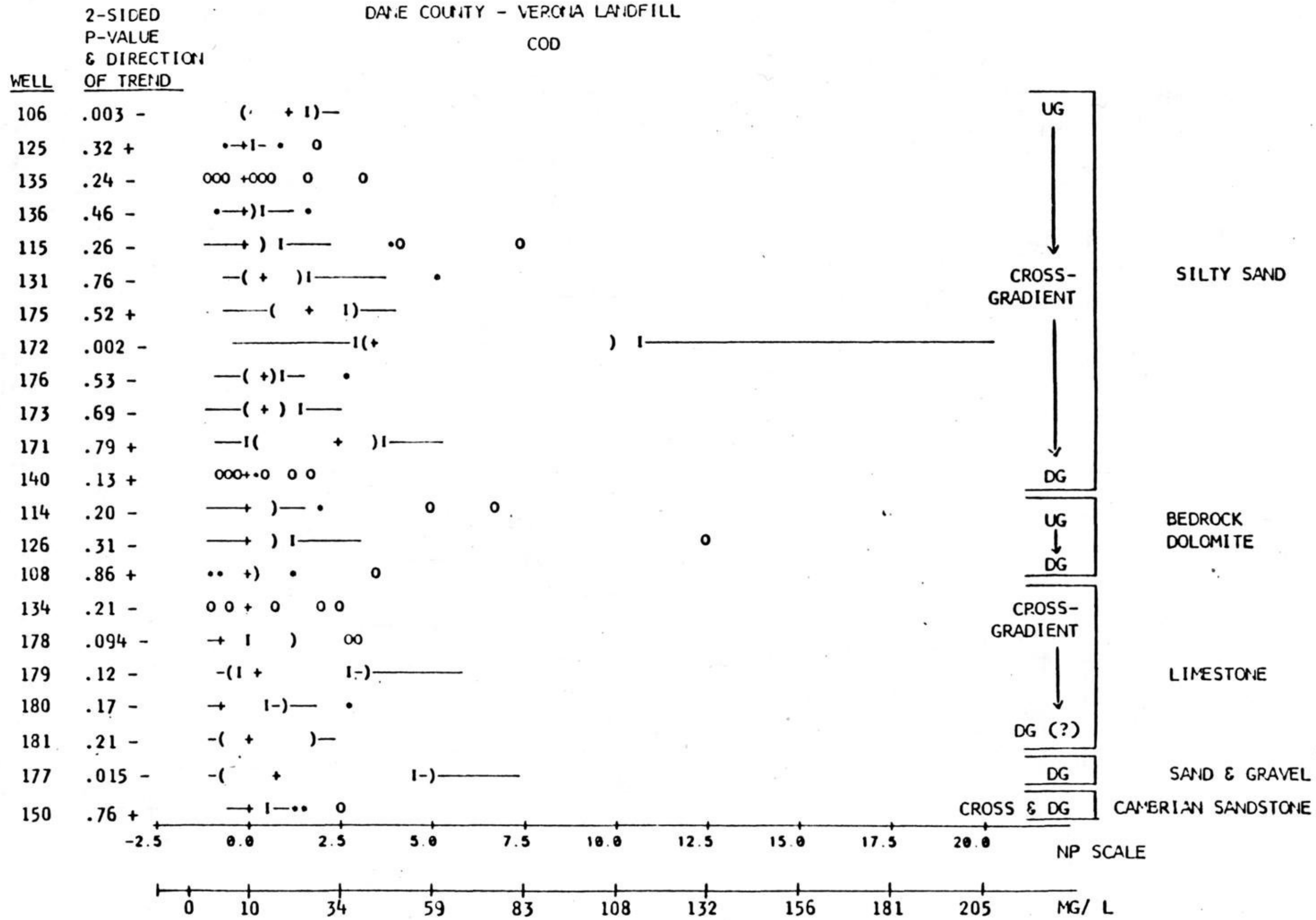


FIGURE 51

DANE COUNTY - VERONA LANDFILL

pH

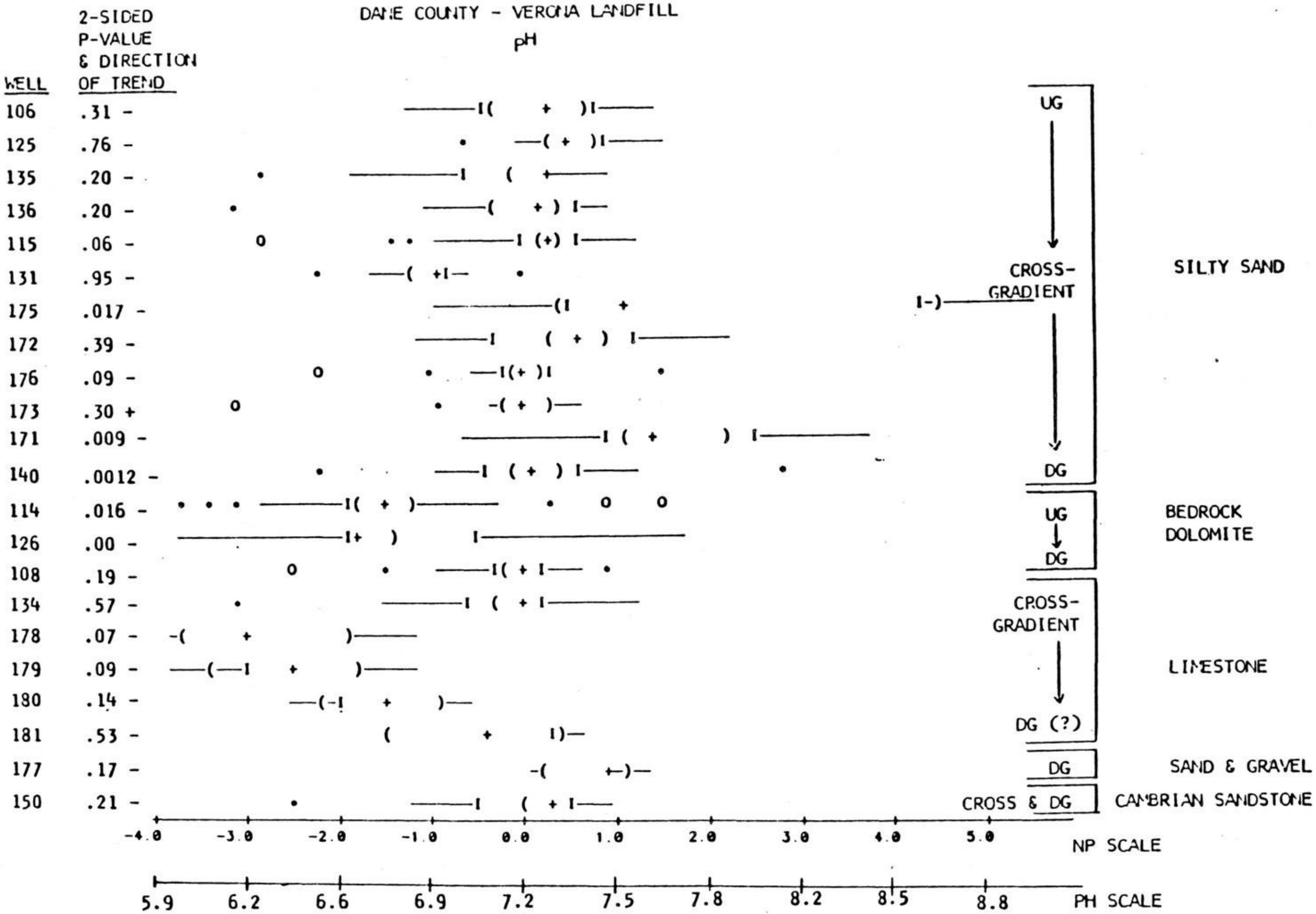


FIGURE 52

DANE COUNTY - VERONA LANDFILL

HARDNESS

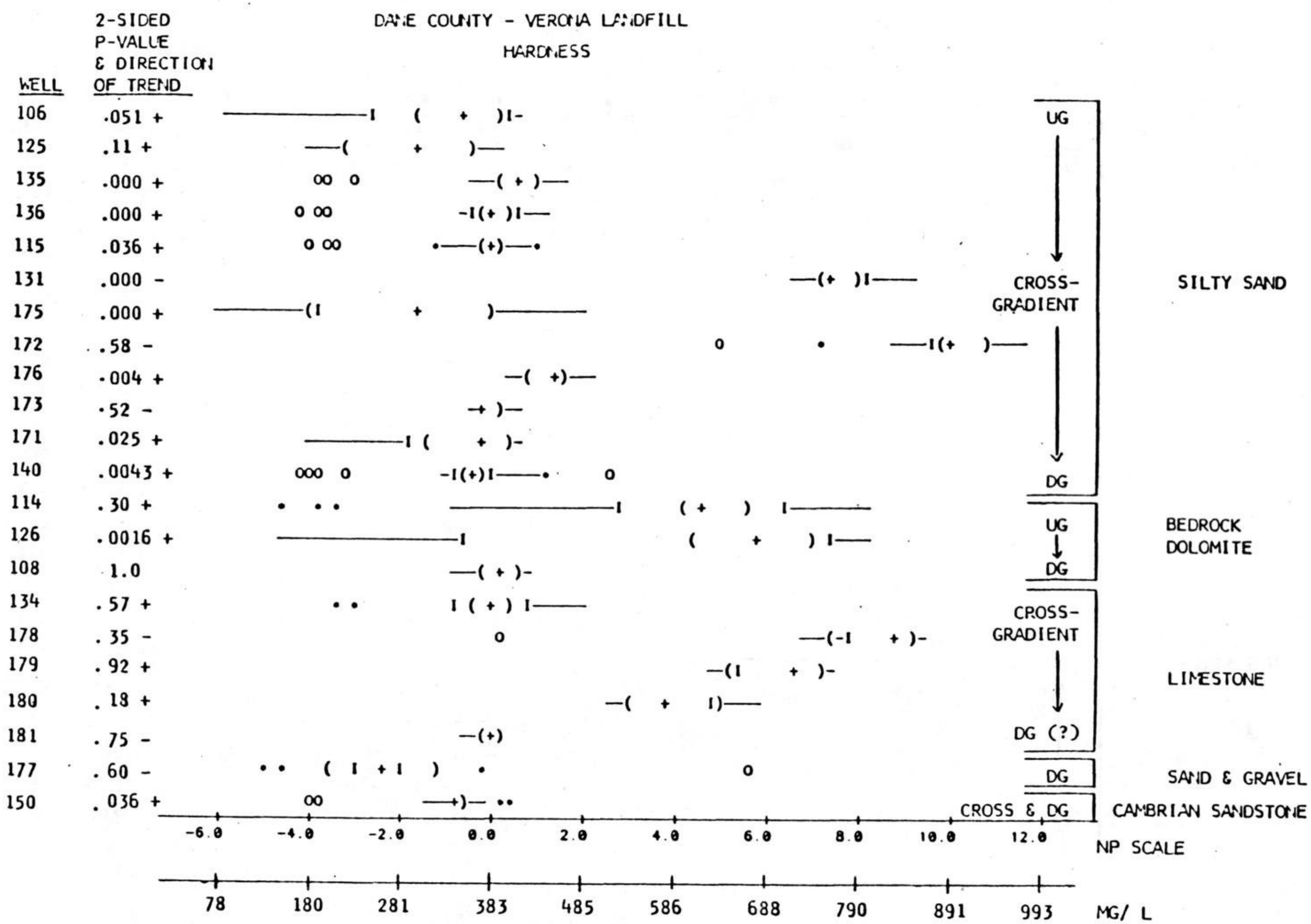


FIGURE 53

DANE COUNTY - VERONA LANDFILL

CONDUCTIVITY

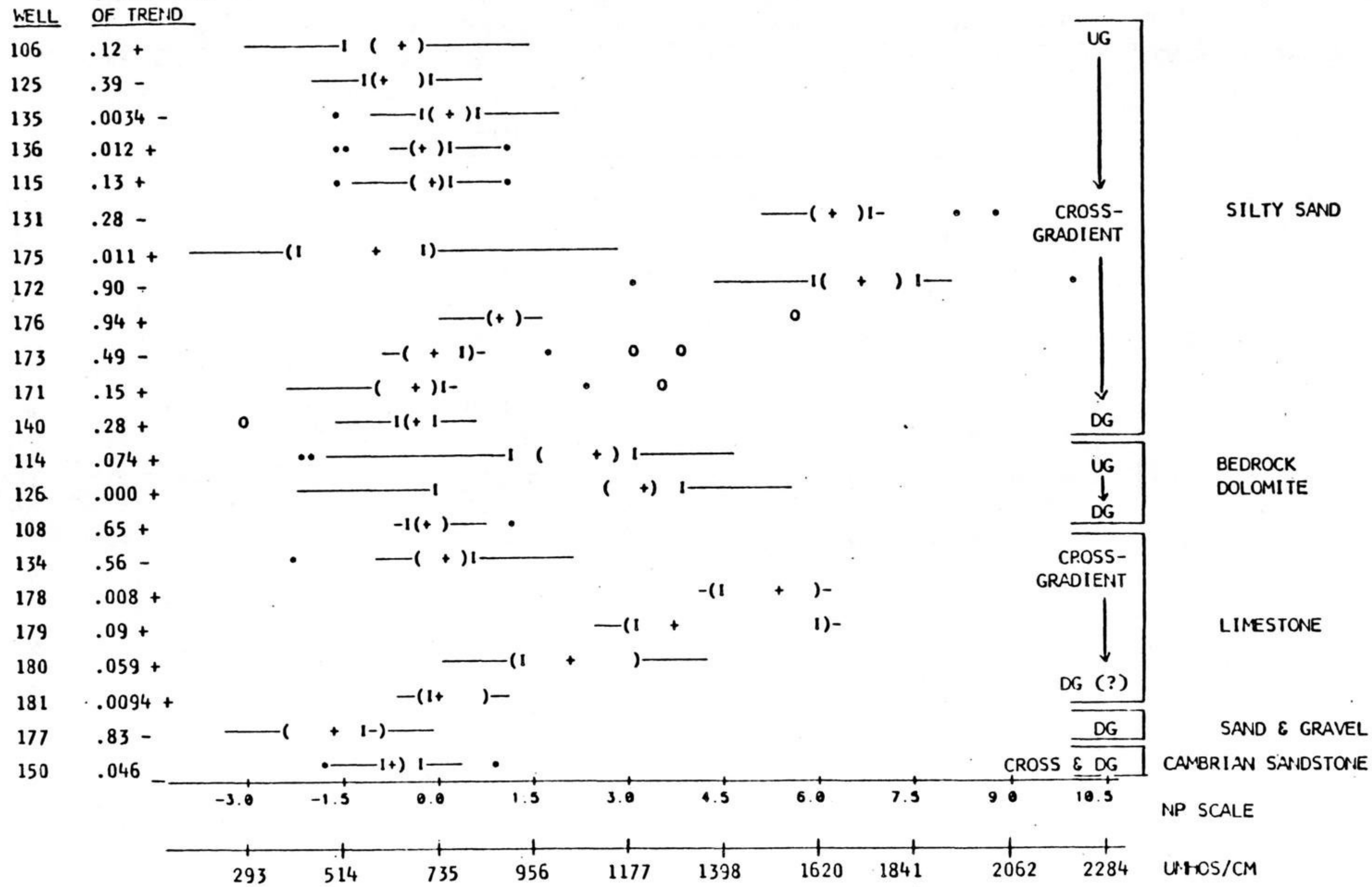


FIGURE 54
DANE COUNTY - VERONA LANDFILL
WELL 108 CATEGORY I BEDROCK DOLOMITE
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

NO SIGNIFICANT POSTIVE TRENDS

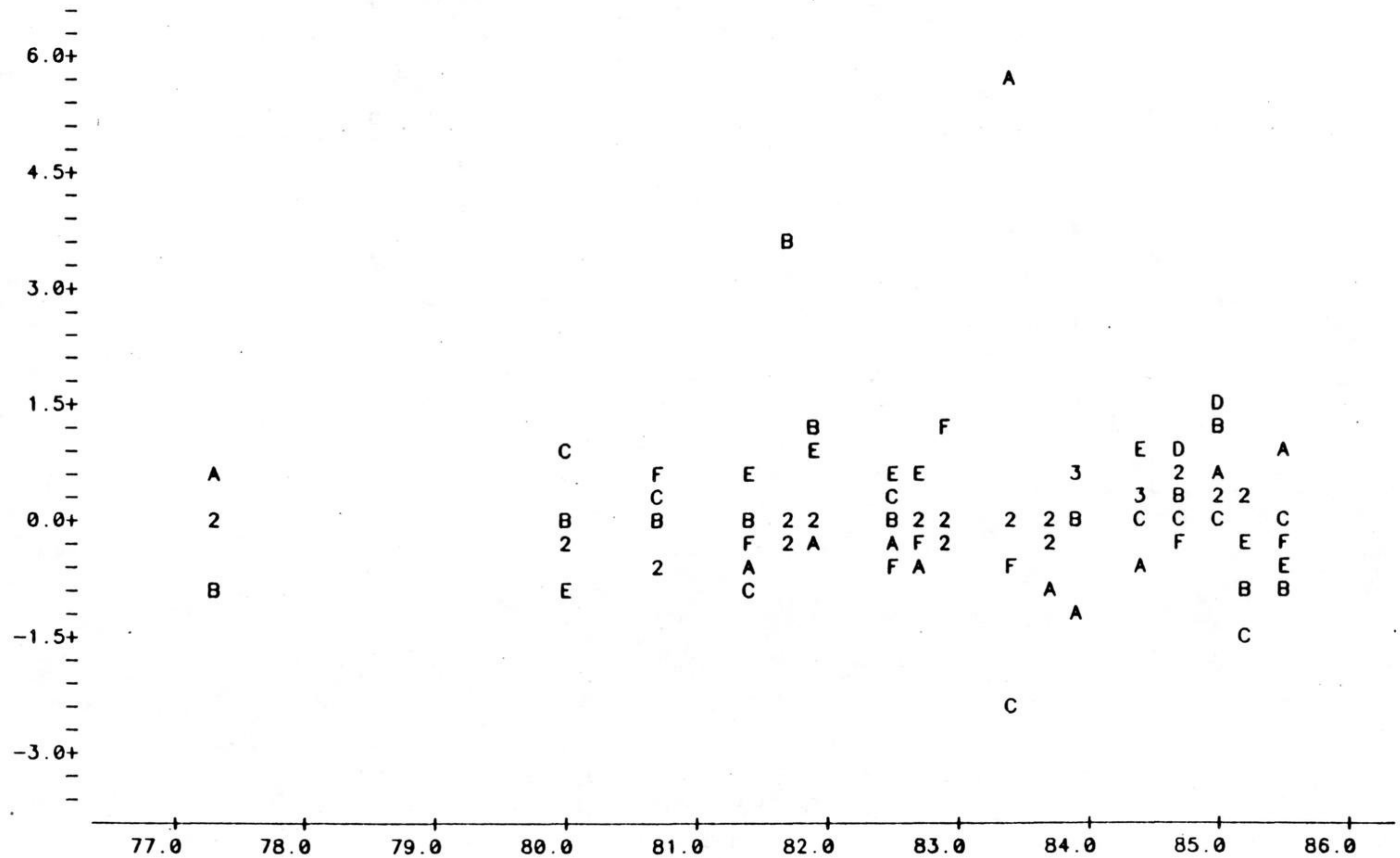


FIGURE 55
DANE COUNTY - VERONA LANDFILL
WELL 114 CATEGORY III BEDROCK DOLOMITE
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

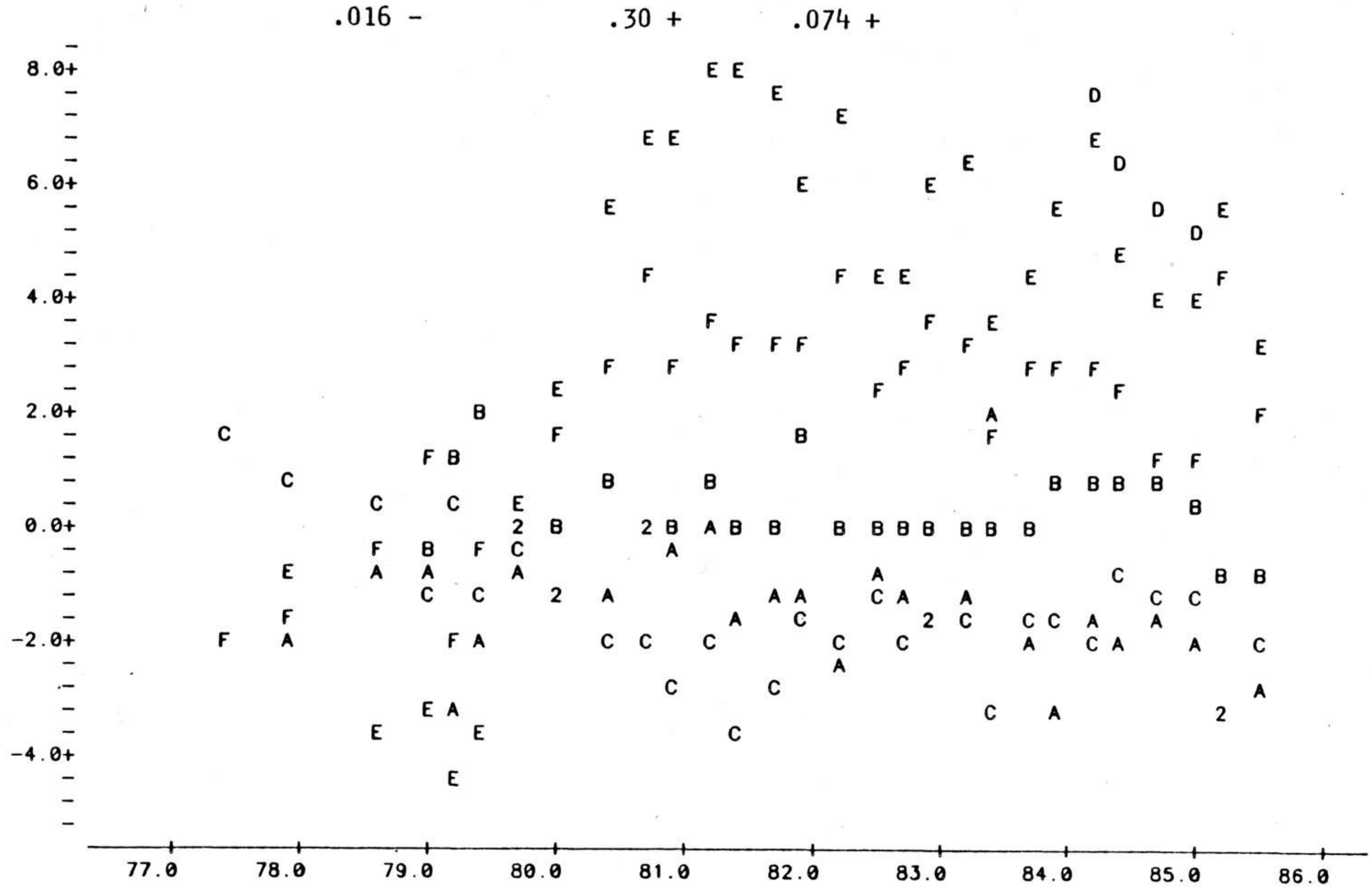


FIGURE 56
DANE COUNTY - VERONA LANDFILL
WELL 126 CATEGORY IV BEDROCK DOLOMITE

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

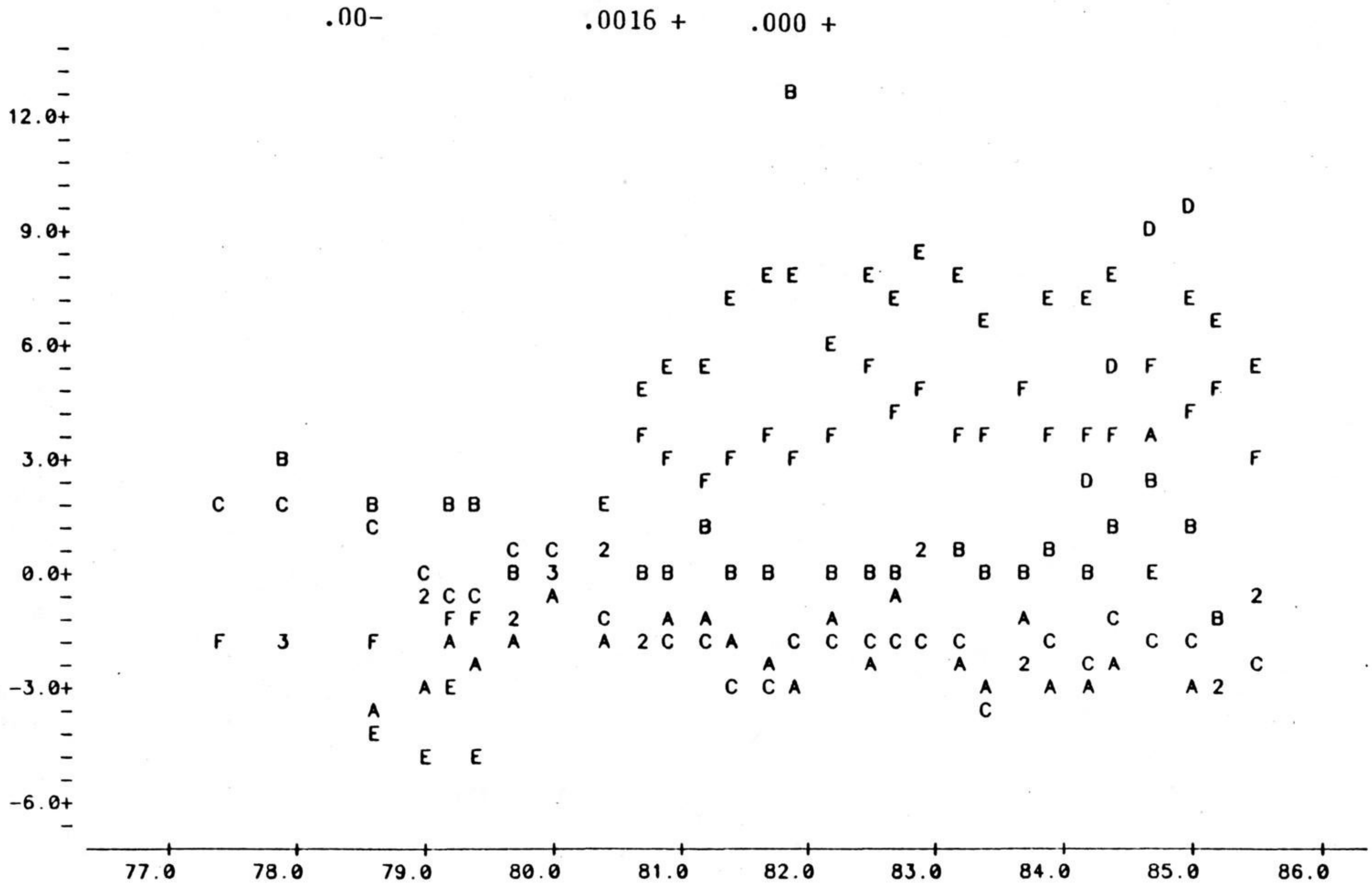


FIGURE 57

DANE COUNTY - VERONA LANDFILL

WELL 178 CATEGORY IV LIMESTONE

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

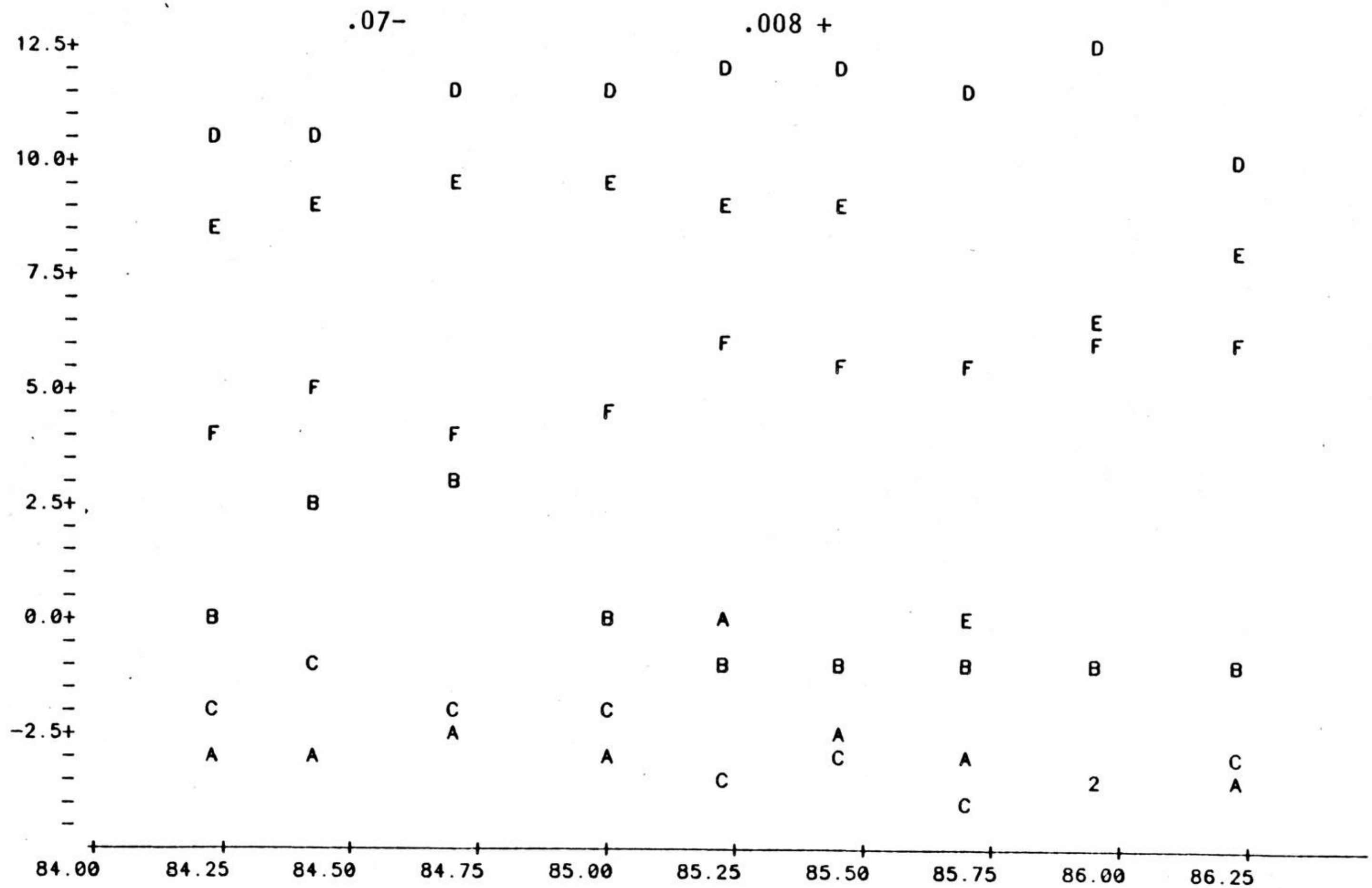


FIGURE 5.8

DANE COUNTY - VERONA LANDFILL

WELL 131 CATEGORY IV SILTY SAND

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

NO SIGNIGICANT POSITIVE TRENDS

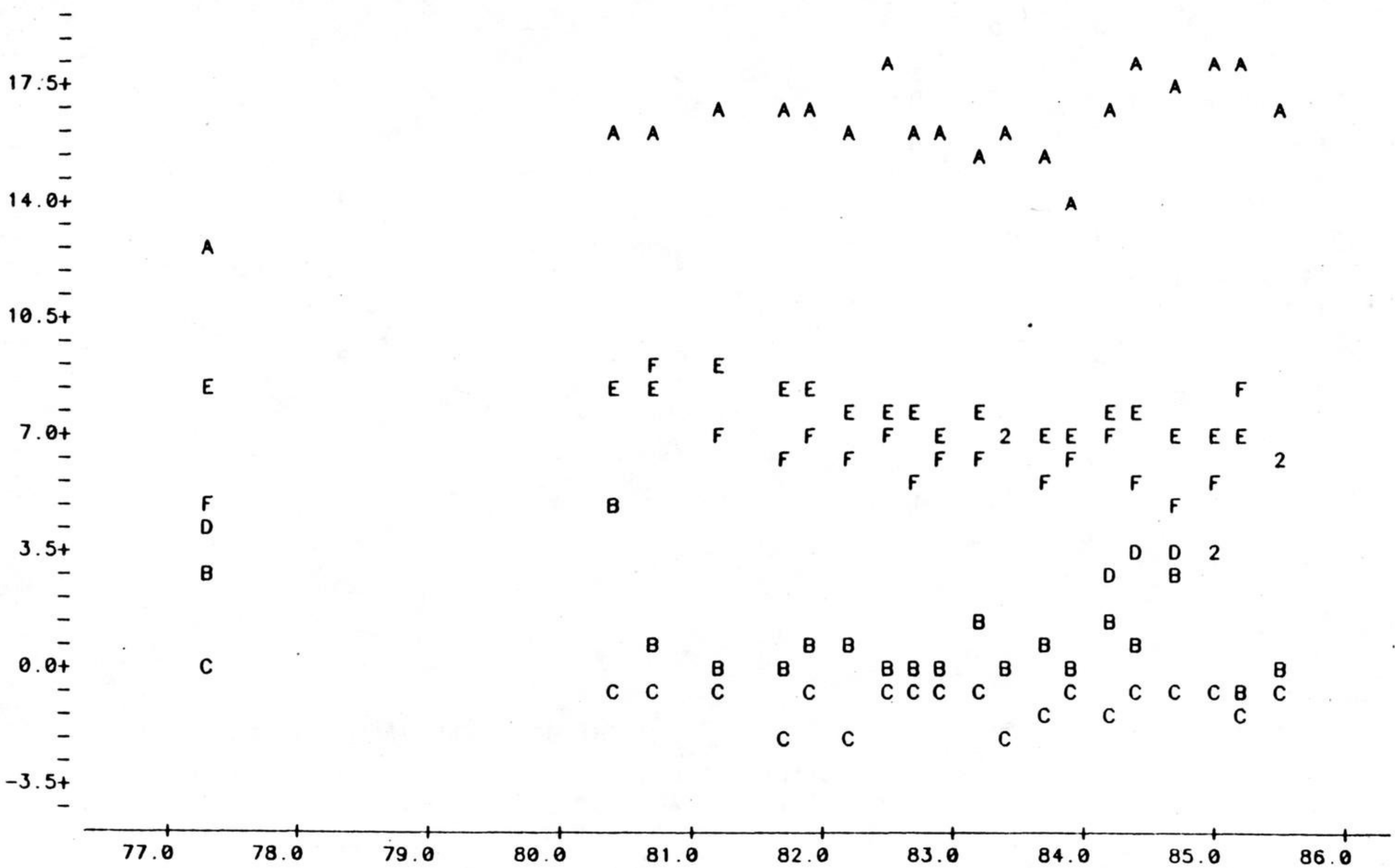


FIGURE 59
DANE COUNTY - VERONA LANDFILL
WELL 172 CATEGORY IV SILTY SAND
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

NO SIGNIFICANT POSITIVE TRENDS

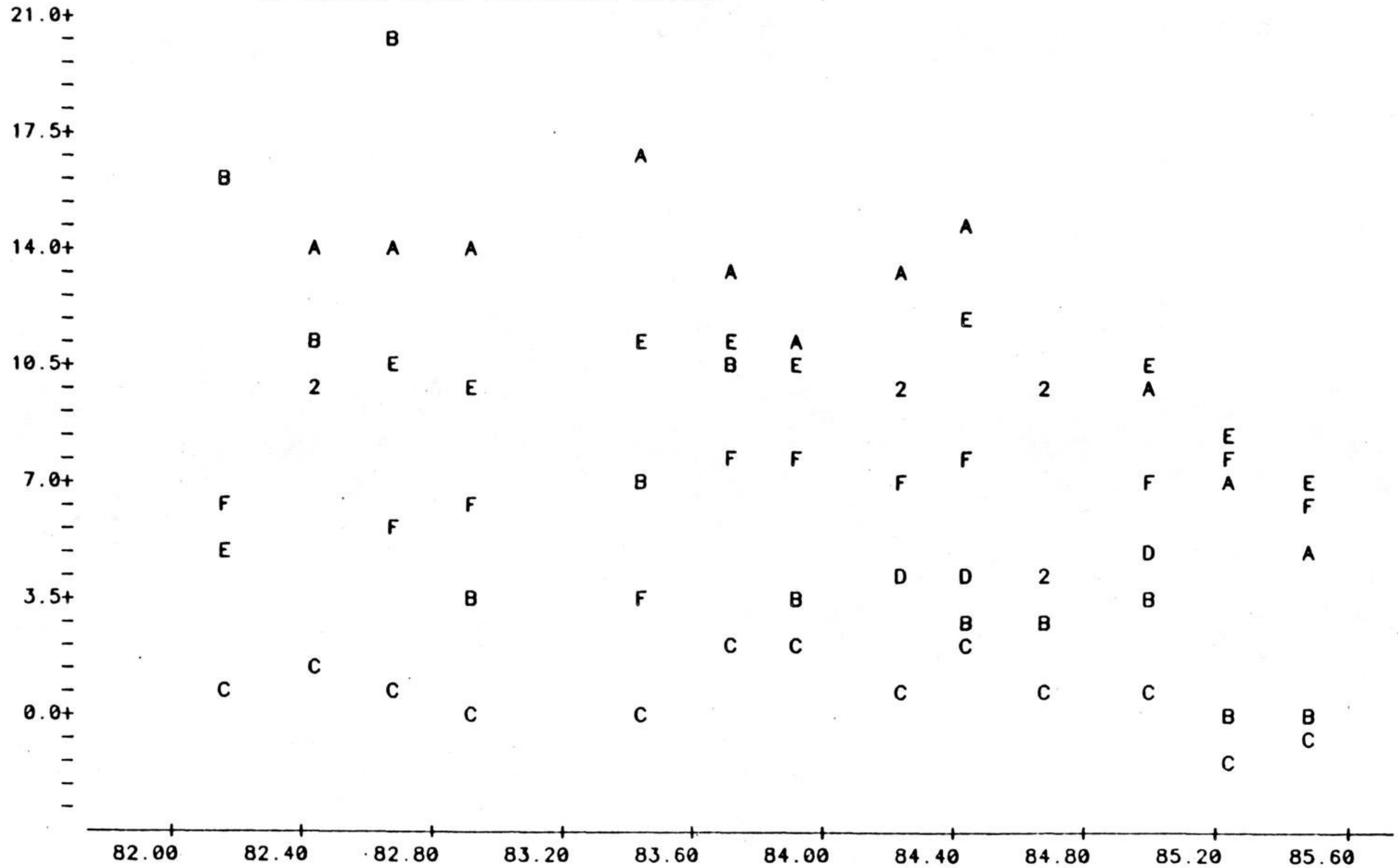
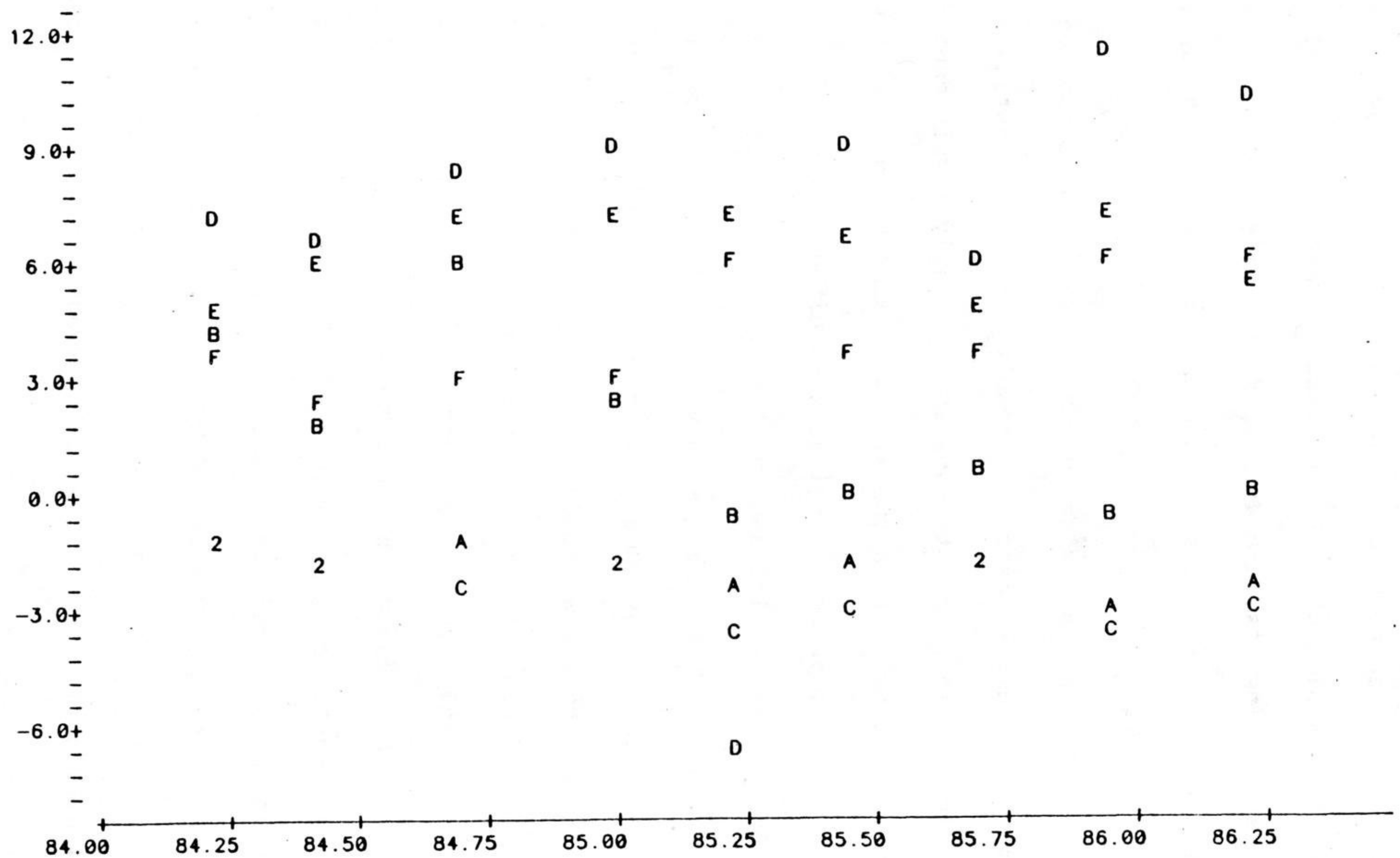


FIGURE 60
DANE COUNTY - VERONA LANDFILL
WELL 179 CATEGORY III LIMESTONE
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

.09- .09 +



108, which is also screened in bedrock dolomite due downgradient of site. Well 108 is in category II, because it shows little evidence of being affected by landfill leachate. It thus provides information about possible inherent groundwater characteristics within this strata; using its data as a specific "reference distribution" helps to reduce possible errors in inference due to inherent spatial variability.

Well 114: The following summarizes the information contained in the boxplots and the trend test results for each constituent at well 114..

chloride: The boxplots show statistically significant displacement of the median below both the site-wide median and that of well 108. The trend result shows a marginally signifcant downward trend (p-value of 0.052). The boxplot and the time-series plot show that this is a low magnitude trend, relative to variations in chloride at the rest of the site; thus, we conclude it might be related to well stabilization.

COD: the boxplots for COD are less informative, since many of the reported values for COD for the whole site are reported as less than 10 mg/l (and several are reported as less than 2 or 5 mg/l, apparently reflecting a change in the laboratory test method). This results in many ties for ranking, which distorts the boxplot and decreases the power of the trend test. The time-series plot shows that early COD values were likely affected by start-up contamination of samples; COD values stabilize later. The p-value for the trend test is .20 (not significant) and is supported by visual inspection of the time series plot.

pH: The trend result shows a significant downward trend (p-value .0157). The boxplots show a significant displacement below both the site median of 7.2 and the median value for well 108 (also pH 7.2). The boxplots and time-series plot show that the most extreme values are about ph 6. The data suggest that the downward trend has resulted in this shift of median, thus this affect appears to be related to the landfill rather than to strata differences.

Hardness: The boxplots show that the median value is highly displaced (by about 3 NP units) above the site median and from well 108. The trend result is not signifcant (p-value .30 +), but does indicate a general pattern of greater concentrations at later time periods (as shown by the "+" following the p-value). The time-series plot reveals that the trend result is an artifact of the test statistic which was computed for the entire sampling record: from 1978 to 1981 there is a easily recognizable high magnitude trend for hardness. Even if we ignore the unusually low initial concentrations, we see a trend that spans about 8.5 NP units (equivalent to about 6.3 standard deviations), beginning at about -1.0 and peaking at about 7.5 NP units. From 1981 to 1986, hardness concentrations show a gradual decline; which accounts for the non-significant trend result.

Conductivity: The boxplots show that the median value is moderately displaced (about 2 NP units) from the site median and from well 108. The trend result shows a marginally significant increasing trend (p-value .074). The time series plot confirms this significant trend from mid-1977 to 1981; the overall magnitude of this trend is quite large, about 6 NP scale units. Again, there is a gradual decline in conductivity concentrations, which results in the marginally significant trend result, as calculated for the entire sampling record.

<u>Conclusions regarding well 114</u>: Categorize as a III, indicating that the data show it is likely contaminated. This designation--rather than a IV--incorporates the slight ambiguity in the constituent responses, while still indicating that the well requires further investigation. The time-series plot summarizes the strength of the evidence: had we conducted trend tests in 1981, we would have seen three significant trends; moreover, the parallel nature of the trends (hardness and conductivity increasing while pH decreases) suggests that what we see in the data are the effects of landfill leachate, not inherent strata characteristics.

<u>Well 126</u>: The pattern of constituent responses at well 126 are very similar to those for 114. To summarize briefly, well 126 shows unambiguously significant trends in 3 constituents: pH, conductivity and hardness. The trend magnitudes are large, reflecting a change in concentration of about 4 to 10 NP scale units. Finally, as shown the time-series plot, the steepest portions of the trend slopes coincide at between mid-1979 and mid-1980. Because the evidence is strong, we

categorize the well as a IV, indicating the well is contaminated.

<u>Well 178</u>: This well is located about 75 feet from the northeast edge of the landfill, and is screened in limestone at a depth of 54 feet. Well 178 is one of the later installed wells for which sampling first began in early 1984, almost 7 years after the landfill began operation. Since there are only 9 observations in the record, the boxplots and trend tests should be interpreted with caution. In addition to the general site-wide information, we use well 180 as a specific reference well, since it is also screened at a similar depth in limestone, and is categorized as a II.

<u>Chloride</u>: The well median is displaced below the site median, and the median value for well 180. The concentrations are quite stable, fluctuating only between about 5-10 mg/L. The trend result is not significant.

<u>COD</u>: The well median is centered on the site median. The trend result shows a marginally significant decreasing trend.

<u>pH</u>: Although the well median is significantly less than the majority of the individual well median values, it cannot be statistically distinquished from the median of well 180. pH shows a marginally significant decreasing trend (p-value 0.07) and decreases to almost ph 6.0 by the end of the sampling record.

Hardness: The well median is highly displaced above the site median by over 8 NP units; and moderately displaced from well 180 (by about 3 NP units). The data are fairly stable at these high concentrations, varying only from about 700 - 880 mg/L.

Conductivity: the well median is also highly displaced above the site median by about 4 NP units and statistically significant from well 180 by about 1 NP unit or 148 mg/L. The trend result shows a significant (p-value 0.008) increase in conductivity, and while the total magnitude of the trend is fairly low, it begins at an unusually high concentration relative to wells for which earlier data exist.

Conclusions regarding well 178: The well is categorized as a IV (i.e., contaminated) on the basis of the high median shifts for hardness and conductivity, the significant trend for conductivity and the extremely low pH values relative to the rest of the site. This designation is warranted given the location of the well and its short sampling record. For example, alkalinity and hardness concentrations are unusually high relative to the other wells. If earlier data were available, it is possible that alkalinity and hardness would have shown positive trends. Instead, the available data suggests that the well is showing a pattern of stable contamination.

B. Discussion of wells without significant trends:

There are 4 wells which were categorized as III or IV, without the benefit of the additional strong inference imparted by significant trends for several constituents. In all cases, the categorization was based primarily on the high displacement of the individual well medians from the site median estimates.

One such well, #131, is remarkably different from the other wells. This is readily apparent in both the boxplots and its time-series plot, which shows continually high values for hardness and chloride concentrations which exceed even the statutory PAL limit of 125 mg/L. Discussion with DNR staff revealed that the well is a privately owned farm well. DNR staff attribute the well's contamination to farming practices rather than to the landfill operation.

The three remaining wells are 172, 178, and 179. As is shown in their time-series plots, water quality sampling did not begin at these wells until 5 to 7 years after the landfill began operating. In addition, these wells are all in locations vulnerable to contamination along the near downgradient southern edge or along the eastern and northeastern border of the site. Under these conditions (as discussed above for well 178), the presence of highly significant shifts of median concentrations for several constituents at a single well is nonetheless plausible evidence of landfill effects, even though it is not accompanied by trend in the data.

4. COMPARISON WITH WELFARE PARAMETERS

Chloride is the only constituent of welfare concern reviewed for this site. Two wells showed extreme chloride concentrations. At well 131, chloride concentrations exceeded the statutory PAL of 125 mg/L. As noted above, DNR attributes contamination at this well to farming activiies. Well 172 is located at the direct downgradient edge of the site and is screened in silty sand. By mid-1982, the time series plot shows that chloride concentrations were unusually high, 15 NP units above the site median value, equivalent to 122 mg/L.

At the other wells, chloride provided little additional information indicating contamination. Two wells (135 and 150), showed significant increasing trends, but these were of low magnitude (as were the associated trends in indicator parameters), and thus, were deemed from a policy standpoint as "environmentally unimportant."

It should be noted, however, that chloride is characeristically a very stable parameter, even low magnitude trends can be statistically discerned. Because of this, chloride can signal a change in concentrations more quickly than a highly variable constituent. In four wells at this site, low magnitude downward trends were detected. These decreasing trends may be related to the strata characteristics or perhaps to stabilization after well completion.

5. COMPARISON WITH VOC DETECTIONS

Table 7 summarizes the results of sampling for VOCs at Verona. In contrast to the Sauk site, the VOC resuts for the Verona landfill

TABLE 7

SUMMARY RESULTS FOR VERONA

CATEGORY/ well#	VOC RESULTS	NP FLAG
CAT. I		
106	--	none
108	9 Det.; 1/1	chloride--probable sample contamination
125	--	none
134	0 Det.; 0/1	none
136	0 Det.; 0/1	none
140	4 Det.; 2/5	none
150	--	none
173	10 Det.; 1/1 *	COD--too early to check for trend
177	8 Det.; 3/3 *	COD--too early to check for trend
181	4 Det.; 1/1	none
	35 Det.; 8/13	
CAT. II		
115	3 Det.; 1/1	COD--too early to check; chloride--possible sample contamination
135	0 Det.; 0/1	none
160	--	conductivity--too early to check for trend; hardness--possible trend or sample contamination
169	--	none
180	21 Det.; 2/2*	alkalinity, hardness--too early to check for trend
	24 Det.; 3/4	
CAT. III		
114	10 Det.; 1/1	alkalinity--data gap, high initial concentration (5-7 NP above site median); hardness--possible trend; conductivity--probable trend
171	10 Det.; 2/2	COD--no trend; chloride--no trend
176	16 Det.; 2/3	chloride--too early to check for trend; conductivity--too early to check for trend
179	33 Det.; 3/3 *	COD, hardness, conductivity--high initial concentrations (4-7 NP above site median)
	69 Det.; 8/8	

TABLE 7
SUMMARY RESULTS FOR VERONA (cont.)

CAT. IV well #	VOC RESULTS	NP FLAG
126	11 Det.; 1/3 *	alkalinity--data gap; high initial concentration (5-9 NP units above site median); hardness--probable trend; conductivity--probable trend
131	--	hardness and conductivity--too early to check, high initial concentrations (5-8 NP above site median)
172	0 Det.; 0/1	chloride, COD, alkalinity, hardness, conductivity--too early to check for trend; all highly elevated initial concentrations (NP 4-16 units) above site median
175	8 Det.; 3/3 *	pH--too early to check for trend; COD--no trend
178	16 Det.; 1/1 *	alkalinity, hardness, conductivity--high initial concentrations (4.5-11 NP above site median)
	35 Det.; 5/8	

Note:
"--" means VOCs were not sampled at the well.
Wells with individual VOC concentrations in excess of 100 ppb are shown by a *.

are far more variable and do not agree well with the inorganic and COD data. As shown in the table, there are numerous, repeated high concentations of VOCs in every well category.

6. COMPARISON WITH SITE-WIDE ESTIMATORS AS NP FLAGS

As noted previously, Figures 15-21 present summary statistics for constituents in category I and for the site in aggregate. As shown in the figures, the non-parametric median and IQR estimated directly from the entire site-wide data closely approximate the median and IQR for each constituent for wells in category I, i.e., uncontaminated wells.

As a result, for this site, using an NP value of 4 as a flag would have efficiently screened the entire site for likely contamination by inorganics and nonvolatile organics. Again, if the site-wide estimates actually approximate uncontaminated conditions, this "flag concentration" is equivalent to the median value plus approximately 3 F-psuedosigma. It is thus close in magnitude to the amount of contamination which DNR considers allowable before flagging a preventive action limit for indicator parameters. Finally, checking for trend at the point at which the concentrations exceeded the NP of 4 would have "caught" several trends which were not detected by testing for trend over the entire length of record.

V. CONCLUSIONS

As a result of applying the data evaluation procedures developed in this study to the selected landfill sites, several issues concerning groundwater quality analysis have been clarified. First, it appears that conventional statistical frameworks which impose an artificial upgradient/downgradient well classification as the basis for analysis are not only unnecessary, but can lead to incorrect inferences. Second, the analysis shows that common statistical assumptions of normality and independence are violated by several groundwater quality constituents, while seasonality in constituent concentrations is less important. The iterative process required to develop the data evaluation procedure, however, demonstrated that the more visible aspects of the data such as unequal sample size, data gaps, and erratic values are equally important factors limiting its analysis. While more sophisticated statistical techniques are available to account for serial dependence or for fitting probability models, they are not generally applicable to typical landfill data sets. Similarly, more sophisticated techniques cannot supplant policy decisions regarding amounts of various constituents considered acceptable in groundwater or appropriate signals of change in groundwater quality. A key contribution of this work is that the methods are tailored to the quality of the data and to the generally scant availabilty of supporting hydrologic information. The methods are insensitive to distributional assumptions and to erratic values. Further, the conclusion that contamination is present is generally based on highly significant results, which are little affected by small errors in estimating Type I error probabilities. Finally, it appears that very simple techniques of exploratory data analysis can be used to roughly approximate water quality characteristics under conditions of no contamination. These techniques show promise for devising a method for routine testing for compliance, based on principles similar to those used for quality control techniques. These issues are summarized below.

Establishing background conditions:

For most landfill sites, there are in fact more wells located down or cross-gradient from the facility than are located upgradient. These wells may or may not be affected by the landfill facility. That is, even if a plume of contamination is present, it is possible that none or only one or a few monitoring wells actually intercept it. Thus, for purposes of establishing the presence or absence of contamination it is more useful to classify wells as "presently affected" or "presently unaffected" than as upgradient or downgradient. This study has demonstrated a method to make such classifications. In this approach, each site provides its own internal reference distribution; wells judged as contaminated are those that differ substantially from the majority of other wells at the site. A common result is that there are more "background wells" at most sites than would be identified based on location alone.

With this approach, the only wells which are regarded as anomalous are those located beyond the estimated hydraulic influence of the site and which show signs of contamination; i.e., significant changes in central location, trends or highly elevated constituent concentrations. From a practical standpoint, it is important to note that groundwater quality data are generally updated more frequently than are groundwater elevation maps. It is also known that some types of landfills are likely to change the local groundwater regime, eg., by causing mounding of the water table due to recharge. The presence of anomalous wells suggests that the hydraulic influence of the site should be updated and verified before concluding that the anomalous behavior of a given well represents random variation.

It should be noted that the effectiveness of this evaluation procedure is limited if the majority of wells at a landfill site are sampling groundwater from areas which have already been contaminated by leachate (or other possible sources). The Janesville landfill represents such a situation; waste disposal at this site predates water quality monitoring by as much as seventeen years. Yet, even at this site the procedure proved useful for identifying contaminated wells regardless of the questionable availability of genuine baseline water quality information. In such situations, statistical approaches which do not rely on local groundwater quality conditions may be preferable. The present procedure, however, provides an important first step toward estimating background water quality conditions.

Comparing Responses of Constituents:

The focus of this research was to assess the presence of contamination based on the responses of indicator parameters historically used by DNR. An additional objective of the research was to assess, insofar as possible, how indicator parameter responses compare with responses for constituents of health and welfare concern. This assessment is inconclusive, due primarily to the scant data for sulfate and for VOC's. Several general observations are possible however.

Where the data permitted comparison, it was found that chloride and sulfate generally did not materially add to the evidence of contamination shown by indicator parameters. That is, trends or high elevations for these constituents either did or did not corroborate with the indicator parameters, but rarely did the response of chloride or sulfate provide the primary evidence of contamination. Rather, chloride and sulfate concentrations tended to be low and stable, perhaps showing low magnitude trends, or else they were highly elevated as at the Ft. Howard, Janesville and Oakcreek landfills. At all three of these sites, waste disposal predated water quality monitoring by several years. Chloride and sulfate are known for their mobility in groundwater and the first "pulse" of these constituents within a leachate plume could have occurred before monitoring began. If so, the concept of "first detection" implicit in the use of chloride and sulfate as tracers or of other indicator parameters no longer applies. For old landfill sites it may be preferable to use the historical data to determine which wells are contaminated and then to redirect the groundwater monitoring objective toward establishing whether a public health or welfare problem exists.

As shown in Table 8, there is not good agreement between the determination of contamination based on COD and inorganic data and contamination based on VOC data. These results are preliminary, however, and should be considered within the broad context of groundwater quality monitoring at landfills. First, the apparent lack of agreement should not be construed as showing that indicator parameters do not provide any indication of a landfill's effect on groundwater quality. The analysis has shown that monitoring for inorganic and COD can provide strong evidence that a landfill has grossly degraded groundwater quality at individual wells. Rather, the issue raised by the comparison is whether indicator parameters can provide information on the transport of synthetic organic contaminants for which there are health-based standards and for which there are theoretical and empirical grounds for expecting different transport mechanisms (Mackay, et al, 1985, Reinhard, et al, 1984).

Second, the VOC results were analyzed only by frequency, not by magnitude. Most of the quantitative VOC results were in the range of 1 - 100 ug/L; values which are above the legal standards but which are, from the perspective of analytical chemistry, quite low. More sophisticated analysis of the VOC data may show statistical differences for VOC responses among well categories.

More comparisons between the responses of inorganic constituents

TABLE 8

Summary of VOC Results for Seven Landfill Sites

CATEGORY I: (uncontaminated wells)

Landfill	VOC Results	"Average" detection for category
Eau Claire	0 Det.; 2/2	
Ft. Howard	2 Det.; 2/10	
Lacro	58 Det.; 13/13	
Marathon	0 Det.; 2/2	
Sauk	0 Det.; 1/1	
Veron	35 Det.; 8/13	95/41 = 2.3 Det./sampling event
	95 Det.; 28/41	

CATEGORY II: (probably uncontaminated)

Landfill	VOC Results	
Ft. Howard	6 Det.; 4/8	
Lacrosse	95 Det.; 13/13	
Sauk	17 Det.; 2/2	
Verona	24 Det.; 3/4	142/27 = 5.3 Det./sampling event
	142 Det.; 22/27	

CATEGORY III: (probably contaminated)

Landfill	VOC Results	
Eau Claire	0 Det.; 1/1	
Ft. Howard	2 Det.; 1/1	
Verona	69 Det.; 8/8	71/10 = 7.1 Det./sampling event
	71 Det.; 10/10	

CATEGORY IV: (contaminated)

Landfill	VOC Results	
Ft. Howard	20 Det.; 6/8	
Sauk	54 Det.; 4/4	
Verona	35 Det.; 5/8	109/20 = 5.5 Det./sampling event
	109 Det.; 15/20	

Note: The well categories regarding the likely presence or absence of contamination are based on the analysis of COD and inorganic data. The VOC analyses are used here as an independent measure of whether an individual well shows evidence of contamination.

and COD and VOC responses are required before firm conclusions are possible. The present analysis suggests a method for refining the allocation of VOC sampling based on analysis of historical data. That is, VOC sampling could be emphasized at Category I and Category IV wells. The goal of such selective sampling is similar to the concept of factorial design of experiments, which is to maximize the usefulness of limited experimental trials by comparing high (strong evidence of contamination by inorganics and COD) and low (little evidence of such contamination) combinations to the predicted hypotheses regarding the presence of VOCs. Such comparisons should permit more conclusive findings regarding the relationship, if any, between gross contamination of groundwater by major constituents of leachate and by associated synthetic organic compounds.

Implications for future work: Establishing Quality Control Procedures

In 1924, W.A. Shewhart introduced the use of statistical quality control (QC) charts for improving manufacturing processes. Quality control charts use time-series graphs to detect when a manufacturing process begins operating incorrectly--or goes "out of control"--such that the product no longer meets the desired specifications. For application to groundwater quality monitoring, the constituent concentration levels at individual wells are the product. Product specifications can be externally derived, as with standards for constituents of health and welfare concern, or can be internally derived, as is the case for indicator parameters.

For quality control techniques to be validly applied, the data-generating process must be in a state of statistical control (Burr, 1976). Statistical control, also referred to as stationarity, means that the statistical parameters of the process are constant over time. Determining whether statistical control exists must be based on historical data. If this state of control exists, then the central location of the data and the variations about it are used to define the acceptable "average" levels and acceptable variation. Thus, control charts are essentially a visual form of a hypothesis test which is repeated with each new observation.

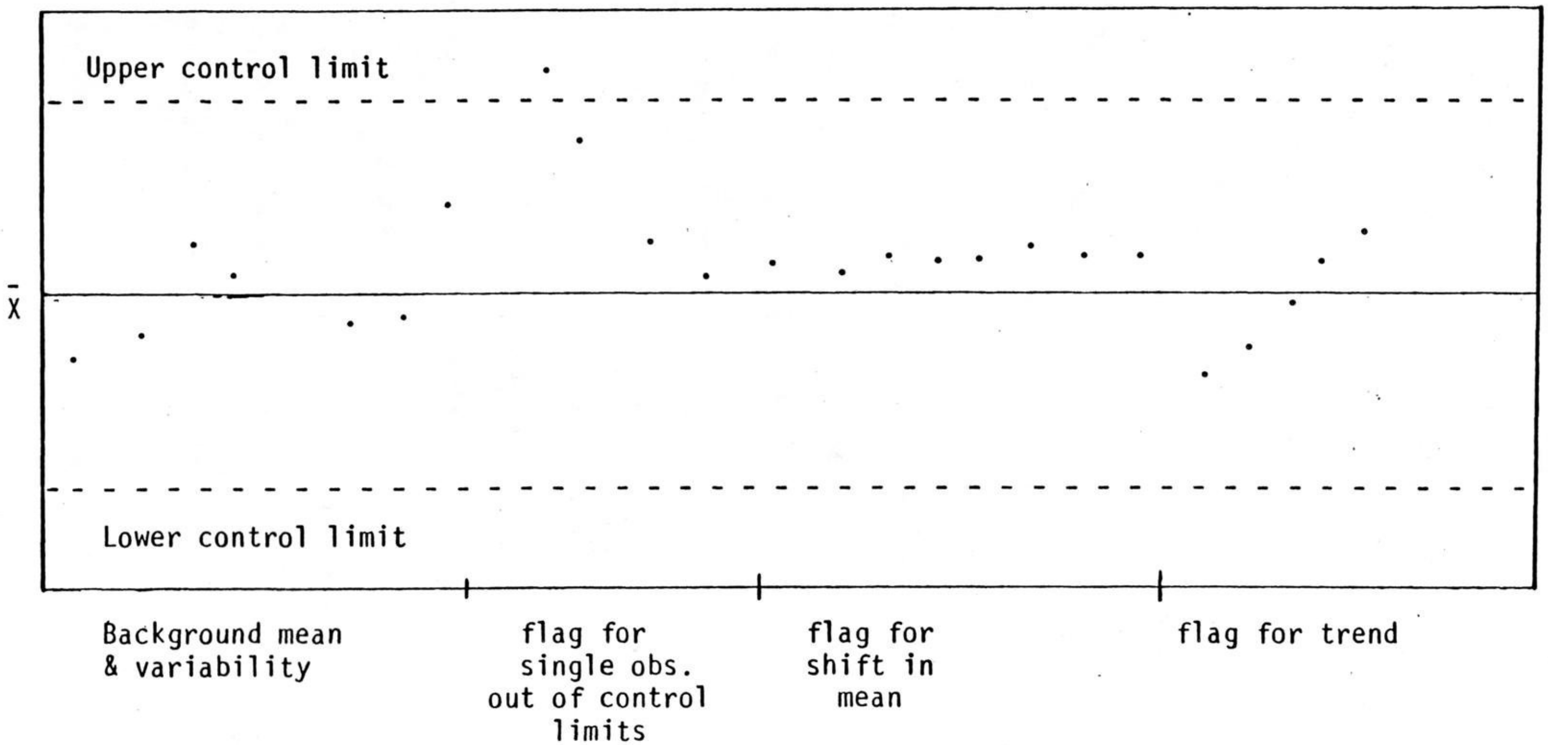
Quality control charts are widely used in industry and the literature on industrial applications is large. Several modifications have been made to the basic method, eg., establishing central location and variability for different populations. Their use has been suggested for interpreting monitoring data generally (Berthouex and Hunter, 1981) as well as specifically for groundwater monitoring data (Doctor, et al, May 1985). In September 1986, EPA proposed the use of QC charts for regulatory monitoring. The use of QC methods in groundwater quality monitoring, however, is generally far more constrained by limitations in the available data than is the case for industrial applications. There are also other important differences in application, in that when groundwater quality goes "out of control" there are far fewer options for bringing the contaminating process back into control. This fact argues for designing a procedure which quickly detects trends in constituent concentrations.

For groundwater quality applications, each well identified as "in control" would serve as its own future control. This approach eliminates many sources of error due to spatial and temporal variability and is analagous to the BSW's well-specific PAL approach. The major difficulty with the BSW's application of the concept is that they had little visual or quantitative information to assist them in making the initial determination of stationarity. (EPA's proposed test is similarly silent on this question).

There are several ways to interpret quality control charts, eg., simple visual tests or more sophisticated statistical criteria (Lucas, 1982). Several simple criteria for visual interpretation are suggested by Montgomery and by Berthouex and Hunter. Two examples of such visual tests proposed by EPA are illustrated in Figure 61. These visual checks are to detect: (1) a slight, but consistent increase in

Figure 61

Visual Evaluation of Water Quality Data Plotted on Control Charts for Individual Wells



contamination and (2) an increasing trend in constituent concentrations.

It should be apparent that a control chart for individual observations is similar to the time-series plots of constituents used in our recommended procedure. The difference is that the statistics plotted on a control chart are scaled by the mean and variability of the data at a single well. In contrast, the statistics plotted on the multiple time-series plots are scaled by estimates of central location and variability for the site. The practical importance of this difference may not be so great, however, given the lack of true baseline data for many wells and DNR's policy that statistical significance is not the predominant criteria for assessing responses in indicator parameters. A very simple, approximate control chart technique based on the NP values and time-series plots might enable DNR to quickly and routinely monitor groundwater quality at landfill sites. The case studies presented preliminary results of such an approach, using an NP value of 4 as a flag for indicator parameters. Empirical tests of other NP values, in conjunction with decision rules for in-house evaluation of these flags, would enable DNR to (1) more quickly issue PALs for indicator parameters; (2) take action only on flags that clearly signal serious contamination, thus reducing false positives; and (3) automatically revise PALS as needed, based on additional data.

References

Benjamin, J.R. and C.A. Cornell. 1970. Probability, Statistics and Decision for Civil Engineers. McGraw-Hill, Inc., U.S.A. (pp.48-49).

Berthouex, P.M. and W.G. Hunter. February 1981. Simple Statistics for Interpreting Environmental Data. J. Water Pollution Control Fedn., Vol. 53, no. 2., pp. 167-175 (p.1).

Berthouex, P.M. and W.G. Hunter. December 1983. How to Construct Reference Distributions to Evaluate Treatment Plant Effluent Quality. J. Water Pollution Control Fedn., Vol. 55, no. 12, pp.1417-1424.

Bhattacharyya, G.K. 1984. Tests of Randomness against Trend or Serial Correlation. Handbook of Statistics, Vol.4, pp. 89-111. P.R. Krishaiah and P.K. Sen, eds., Elsevier Science Publ.

Box, G.E.P., W.G. Hunter and J.S. Hunter. 1978. Statistics for Experimenters. John Wiley and Sons, Inc. (p. 81).

Bradley, J.V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Inc., Englewood Cliffs, New Jersey. (p. 31).

Burr, I.W. 1976. Statistical Quality Control Methods. Marcel Dekker, Inc. New York.

Dance, J.T. and E.J. Reardon. 1983. Migration of Contaminants in Groundwater at a Landfill. A Case Study: (5) Cation Migration in the Dispersion Test. J. Hydrology, Vol. 63, pp. 109-130.

Dietz, E.J. and T.J. Killeen. March 1981. A Nonparametric Multivariate Test for Monotone Trend with Pharmaceutical Applications. J. American Statistical Assoc., Vol. 76, no. 373, pp. 169-174. (app. p. 172).

Doctor, P.G., R.O. Gilbert and R.R. Kinnison. February 1985. Draft-Statistical Comparisons of Groundwater Monitoring Data. Milestone 2: Ranges of Variation in Ground-water at Hazardous Waste Sites. Interagency Agreement No. DW89931109-01-0. Pacific NW Laboratory, Richland, WA.

Doctor, P.G., R.O. Gilbert and R.R. Kinnison. May 1985. Draft-Statistical Comparisons of Groundwater Monitoring Data. Milestone 3: Ground-water Monitroing Plans and Statistical Procedures to Detect Leaking at Hazardous Waste Site Facilities. Interagency Agreement No. DW89931109-01-0. Pacific NW Laboratory, Richland, WA.

Dunker, P.N. and G.E. Beanlands. 1986. Environmental Significance: Exploration of the Concept. Environmental Management. Vol. 10, no. 1, pp. 1-10.

Electric Power Research Institute. July 1985. Final Report- Groundwater Data Analyses at Utility Waste Disposal Sites. EA-4165, Research Project 2283-2. Prepared by Tetra-Tech, Inc.

Ferrell, E.B. March 1953. Control Charts Using Midranges and Medians. Industrial Quality Control. pp. 30-34.

Groenveld, L. and R.D. Duval. June 1985. Statistical Procedures and Considerations for Environmental Management. Prepared for the Florida Dept. of Environmental Regulation, Tallahassee, FL.

Hirsch, R.M., J.R. Slack and R.A. Smith. Feb. 1982. Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research, Vol. 18, no. 1. pp. 107-121. (p. 109).

Hirsch, R.M. and J.R. Slack, 1984. A Nonparametric Trend Test for Seasonal Data with Serial Dependence. Water Resources Research, Vol. 20, no. 6, pp. 727-732. (p. 729).

Hoaglin, D.C., F. Mosteller and J.W. Tukey. 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons, Inc., NY.

Kammerer, Jr., P.A. Feb. 1984. An Overview of Ground-water Quality Data in Wisconsin. U.S.G.S. Water Resources Invest. Report, 83-4239. (p. 4).

Lettenmaier, D.P. October 1976. Detection of Trends in Water Quality Data from Records with Dependent Observations. Water Resources Research, Vol. 12, no. 5, pp. 1037-1046.

Loftis, J.C. and R.C. Ward. 1980. Water Quality Monitoring: Some Practical Sampling Frequency Considerations. Environmental Management, Vol. 4, no. 6, pp. 521-526.

Lucas, J.M. April 1982. Combined Shewhart-Cusum Quality Control Schemes. J. of Quality Technology, Vol. 14, no. 2., pp. 51-59.

Mackay, et al. 1985 Environmental Science and Technology, Vol. 19, No. 5.

McFarlane, D.S., J.A. Cherry, R.W. Gillham and E.A. Sudicky. 1983. Migration of Contaminants in Groundwater at a Landfill. A Case Study: (1) Groundwater Flow and Plume Delineation. J. Hydrology, Vol. 63, pp. 1-29. (p. 15-18).

Montgomery, D.C. 1985. Introduction to Statistical Quality Control. John Wiley & Sons. New York.

Plumb Jr., R.H. and C.K. Fitzsimmons. June 1984. Performance Evaluation of RCRA Indicator Parameters. Proceedings from the First Canadian/American Conference on Hydrogeology. Practical Applications of Ground Water Geochemistry. Banff, Alberta, Canada. Published by the National Water Well Assoc., Worthington, Ohio.

Plumb Jr., R.H. and A.M. Pitchford. November 1985. Volatile Organic Scans: Implications for Ground Water Monitoring. Presented at the National Water Well Assoc./American Petroleum Inst. Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground-Water. Houston, Texas.

Reinhard, M. et al. 1984. Environmental Science and Technology, Vol. 18, No. 12.

Salas, J.D., V.M. Yevjevich, and W.L. Lane. 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publications, Chelsea, Michigan.

Shapiro, S.S, M.B. Wilk and H.J. Chen. 1968. A Comparative Study of Various Tests for Normality. Journal Amer. Statistical Assoc., pp. 1343-1372.

Silver, C.A. 1986. Statistical Approaches to Groundwater Monitoring. J. Hazardous Materials, Vol. 13, pp. 207-216.

Smith, R.A., R.M. Hirsch and J.R. Slack. 1982. A Study of Trends in Total Phosphorus Measurements at NASQUAN Stations. U.S.G.S. Water Supply Paper No. 2190, pp. 1-34. (p.1).

Unwin, J.P. and R.A. Miner. June 1985. Groundwater Quality Data Analysis. Technical Bulletin No. 42. Natl. Council of the Paper Industry for Air & Stream Improvement, Inc. Central-Lake States Regional Center, W. Michigan, Univ., Kalamazoo, Michigan.

Velleman, P.F. and D.C. Hoaglin. 1981. Applications, Basics and Computing of Exploratory Data Analysis. Boston: Duxbury Press.

Appendix A: Overview of Hypothesis Testing

A statistical test of hypothesis is a method for testing whether an assertion about one or more random variables is false, based on information contained in a random sample. For interpretation of groundwater data, the water quality observations are samples of the random variable. The assertion to be tested is generally some formulation of the statement that groundwater quality is not contaminated.

As an overview, the general steps for testing a hypothesis are listed below. Each is discussed later in greater detail.

1. State the hypotheses to be tested (i.e., the null and alternative hypotheses) and the background assumptions implicit in the test.
2. Select the type I error probability for the test; thus establishing the critical values for the test.
3. Calculate a test statistic, having a known probability distribution, from the data.
4. Interpret the test statistic with respect to its probability distribution and the critical value corresponding to the preselected type I error.
5. If the value of the test statistic is within the type I error limits, do not reject the null hypothesis.
6. If the value of the test statistic is outside the type I error limits, reject the null hypothesis in favor of the alternative

hypothesis.

(Smith, et al, 1982).

Discussion

The null hypothesis is the assertion to be disproven. The alternative hypothesis is its mutually exclusive counterpart. Mathematically, this means that the areas defined on the sample space by the null and alternative hypothesis cannot intersect. Similarly, the union of the two hypotheses must represent the total sample space; i.e., their union must account for any possible realization of the random variable. (Benjamin and Cornell, 1967).

For any test of hypothesis, there are two possible errors which can be made. The type I error is the probability of rejecting the null hypothesis when it is true (a "false positive"). The type II error is the probability of mistakenly failing to reject the null hypothesis when it is not true (a "false negative"). There is an inverse relationship between these two errors, as long as the sample size remains constant. That is, by choosing to decrease the probability of a type I error, the probability of a type II error necessarily increases.

The significance level for a test of hypothesis is the probability of a type I error which the investigator decides is desirable or allowable, based on consideration of the consequences of the test result. Conventionally selected significance levels are often .01,

.05, or .10.

A concept related to these two types of errors is the power of the test. The power of a test is the probability of rejecting the null hypothesis when in fact it is false and should be rejected. Unlike the fixed probabilities for type I and II errors, power is a function of the sample size, magnitude of the effect being tested, variance of the sample data and construction of the alternative hypothesis. Other things being equal, a more powerful test is more likely to detect that an effect, such as degraded groundwater quality, has occurred.

The assumptions upon which the test are based must be established. If a distribution is assumed, the test is a "parametric" test. If a distribution is not required as an assumption, the test is "nonparametric" (Conover). Relative to nonparametric tests, a parametric test is usually more powerful when all assumptions are met. Nonparametric tests, however, are generally more powerful over a broader range of sample conditions (Bradley). It is important to note that the underlying assumptions in any test of hypotheses are not themselves tested as part of the hypothesis test. Moreover, in some cases, these assumptions cannot be adequately verified from the sample data. It is thus desirable for the test results to be little affected by the underlying assumptions. Such tests are known as robust tests.

The test statistic serves as the criterion for the test. This statistic is a function of the sample data and cannot involve any unknown parameters. The distribution of the test statistic under the null hypothesis must be known; i.e., since the null hypothesis is the statement we wish to disprove, we use its distribution as the probabilistic "frame of reference" for the test. For parametric tests, the distribution of the test statistic is determined by the parameters of the assumed population distribution. For nonparametric tests, the distribution of the test statistic can be specified from the magnitude or ranks of the sample data and the test assumptions. Alternatively, it can be approximated by parametric models (Bradley, 1968).

The test statistic is calculated from the sample data and compared with the limits, or critical values, associated with the significance level of the test and the distribution of the test statistic under the null hypothesis. The test statistic is interpreted as noted in points 5 and 6 above and a conclusion drawn regarding the whether the data support rejecting the null hypothesis in favor of the alternative hypothesis, or not rejecting the null hypothesis.

APPENDIX B

TABLE 1
SUMMARY RESULTS FOR EAUCLAIRE

CATEGORY/ well#	VOC RESULTS	NP FLAG RESULTS
CAT. I		
3	no det.	none
5	--	none
6	--	none
7	no det.	none
8	--	none
17	--	COD; probably sample contamination
CAT. II	--	
19		chloride, COD; probably sample contamination
CAT. III		
18	no det.	chloride--possible trend; conductivity--probable trend; alkalinity--cannot interpret due to gap in data; hardness--probable trend
CAT. IV		
no wells in this category		

Note: "--" means that VOCs were not sampled at the well. VOCs were detected only in the leachate from this site.

TABLE 2
SUMMARY RESULTS FOR FORT HOWARD

CATEGORY/ well#	VOC RESULTS	NP FLAG
CAT. I		
4	—	chloride, COD—too early to check for trend
7	—	none
10	0 Det.; 0/2	none
12	0 Det.; 0/2	none
15	2 Det.; 2/4	sulfate—highly elevated initial concentration (8-10 NP units above median), but below legal PAL
17	—	none
18	—	hardness—no trend; chloride—no trend; sulfate—no trend
19	—	none
21	0 Det.; 0/2	chloride, hardness, conductivity, sulfate—all too early to check for trend; all show high initial concentrations (4-9 NP above site median), still below legal PAL
22	—	none
33	—	sulfate—too early to check for trend
	2 Det.; 2/10	
CAT II		
1	—	COD—too early to check for trend; sulfate—too early to check for trend but shows highly elevated concentrations (9-24 NP units above site median), still below legal PAL
9	—	COD—probable sample contamination; hardness—probable sample contamination
13	0 Det.; 0/2	sulfate—too early to check for trend; high initial concentration, but still below legal PAL
14	5 Det.; 3/4	sulfate—too early to check for trend, high initial concentration but still below legal PAL; conductivity—probable trend
19	1 Det.; 1/2	none
	6 Det.; 4/8	

TABLE 2: FT. HOWARD (cont.)

CAT III		
5	2 Det.; 1/1	chloride, COD, conductivity--too early to check for trend; high initial concentrations (5-21 NP units above site median)
6	--	chloride, COD, conductivity--too early to check for trend; high initial concentrations (5-7 NP units above site median)
	2 Det.; 1/1	
CAT IV		
8	2 Det.; 2/2	chloride, alkalinity, hardness, conductivity--too early to check for trend; high initial concentrations (5-21 NP units above site median); COD--probable trend
11	0 Det.; 0/2	chloride, alkalinity, hardness, conductivity--too early to check for trend; COD--probably no trend
16	--	sulfate and chloride--too early to check for trend; hardness, conductivity--possible trends
20	18 Det.; 4/4	chloride, hardness, conductivity, sulfate--too early to check for trend, high initial concentration (4-9 NP units above site median)
23	--	chloride--possible trend; conductivity--strong trend; hardness--too early to check for trend
	20 Det.; 6/8	

Note:
"--" means not sampled for VOCs. Wells 14 and 20 show greatest VOC concentrations. DNR staff thinks well 14 has been outside the hydraulic influence of site.

TABLE 3
SUMMARY RESULTS FOR JANESVILLE

(no sampling for VOCs)

CATEGORY/well #	NP FLAG
CAT. I	
101	COD—possible trend
108	none
109	none
112	none
113	none
114	none
115	chloride—likely sample contamination
121	none
125	none
127	COD—too early in record to check for trend; chloride—probable sample contamination
128	none
CAT. II	
107	none
122	chloride—too early in record to check for trend; COD—probable sample contamination
123	chloride—possible trend, but occurs early in record
CAT. III	
126	COD—too early in record to check for trend; chloride—no trend
CAT. IV 106	chloride, COD, conductivity, sulfate—all show highly elevated initial concentrations (6 - 11 NP units above site median); well first sampled in 1981, many years after waste disposal in unlined area.

TABLE 4
SUMMARY RESULTS FOR LACROSSE

CATEGORY/ well #	VOC RESULTS	NP FLAG
CAT.I		
4	6 Det.; 2/2	none
9	--	none
10	--	none
11	10 Det.; 2/2	COD, likely sample contamination
12	25 Det.; 4/4	COD, possible sample contamination
13	--	chloride--likely sample contamination
14	0 Det.; 1/1	none
15	17 Det.; 4/4	none
	58 Det.; 13/13	
CAT. II		
6	10 Det.; 3/3	none
7	1 Det.; 1/1	chloride--no trend
8	0 Det.; 1/1	chloride--probably no trend; conductivity-- likely sample contamination
16	84 Det.; 8/8	none
	95 Det.; 13/13	

CAT. III
no wells in this category

CAT. IV
no wells in this category

Notes:

Most VOC detections are in the 1-10 ug/L range.
Well 16 not only showed numerous detections, but also many up to 100 mg/L, and dichloromethane detected 3 times at 2230-3660 ug/L.

TABLE 5
SUMMARY RESULTS FOR OAKCREEK

(no sampling for VOCs)

CATEGORY/well#	NP FLAG
CAT. I	
202	none
203	chloride; probably sample contamination
205	none
206	hardness, cond. & sulfate: all show highly elevated initial concentrations.
207	COD; probably sample contamination
208	none
209	COD; probably sample contamination
215	none
CAT. II	
210	pH; probably sample contamination
211	cond., sulfate; cannot interpret due to gap in data
CAT. III	
201	chloride--possible trend; cond. and sulfate--too early in record
228	hardness-- too early in record (note sampling began in '82 vs. waste disposal in '75)
229	pH, hardness, and cond. all show highly elevated initial concentrations (note sampling began in '82 vs. waste disposal in '75)
CAT. IV	
204	hardness, cond. and sulfate--too early in record to check; shows highly elevated initial concentrations
212	chloride--shows strong trend; cond.--shows strong trend
213	pH, COD--occurs too early in record, later chloride and hardness flags reveal earlier strong trends
214	chloride, hardness and sulfate--strong trend
227	COD--too early in record to check; chloride and pH show highly elevated initial concentrations

TABLE 6
SUMMARY RESULTS FOR WAUSAU

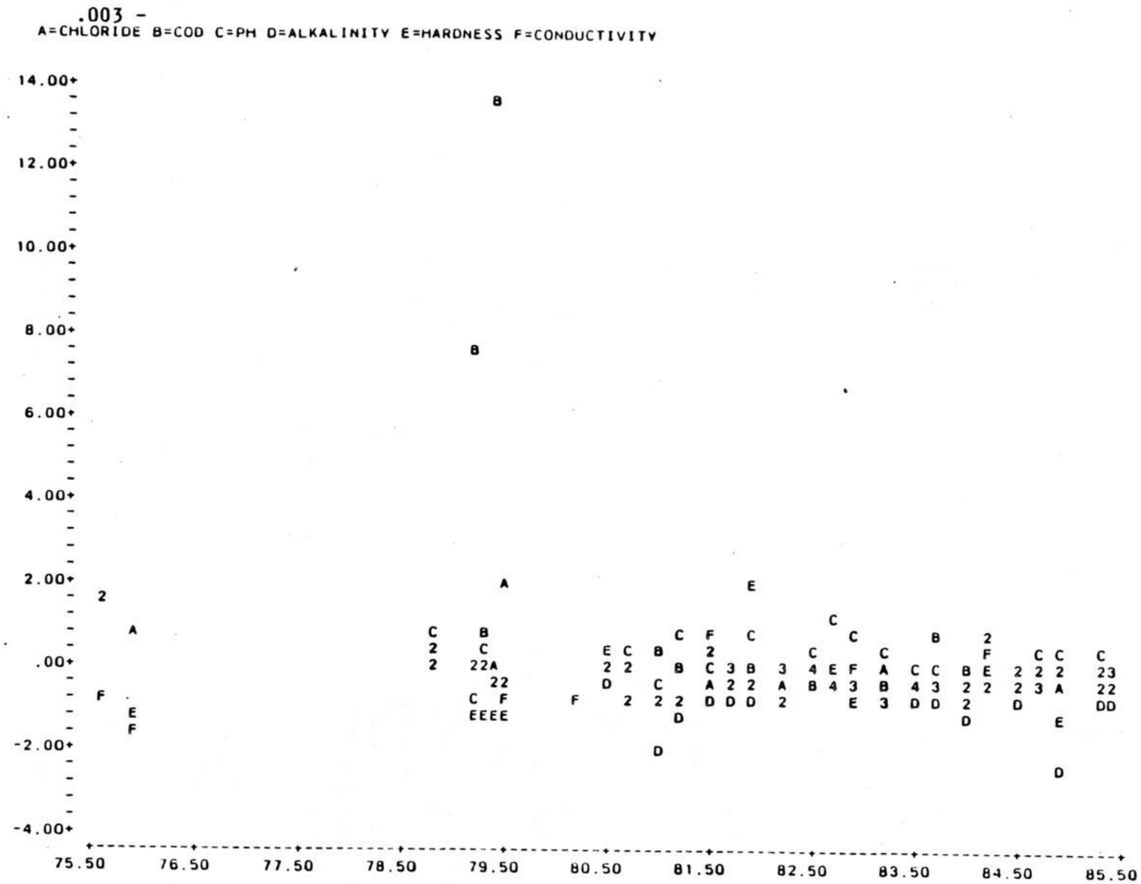
CATEGORY/ well #	VOC RESULTS[1]	NP FLAG RESULTS
CAT. I		
2	—	pH—possible sample contamination
3	—	none
4	—	none
CAT. II		
no wells in this category		
CAT. III		
no wells in this category		
CAT. IV		
1	—	chloride, COD, alkalinity, hardness and conductivity show highly elevated initial concentrations (i.e., 30-70 IQRs above site median[2]
5	—	as above, except initial concentrations are 20-140 IQRs above site median

Notes:
1) "—" indicates that VOCs were not sampled. VOCs were sampled at only two wells, for which neither had sufficient inorganic and COD data for review. Neither well was located on available site maps. The coordinates for one well, however, indicates that it is located near to well 3, which is in category I. No VOCs were detected at this well. The second well sampled for VOCs is apparently located at great distance from the landfill. VOCs were also not detected at this well.

2) Wells 1 and 5 are located directly downgradient of an unlined waste disposal cell which was filled to capacity during 1981. Water quality sampling at these wells began in early 1982. A review of the full time-series plots for these wells shows strong indication of a highly contaminated slug of leachate passing these sampling points.

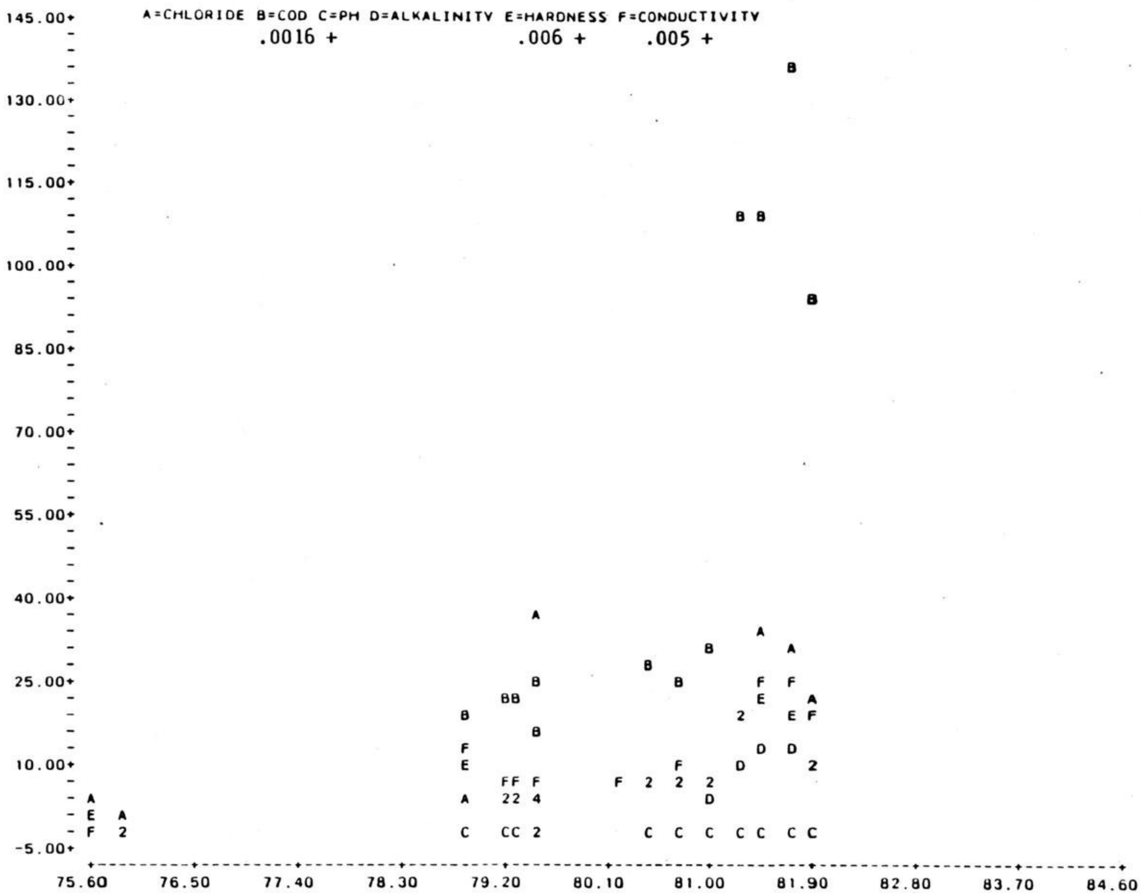
APPENDIX C: FIGURE 1

SAUK COUNTY LANDFILL
WELL 105 CATEGORY I SAND



APPENDIX C: FIGURE 2

SAUK COUNTY LANDFILL
WELL 106 CATEGORY IV SAND



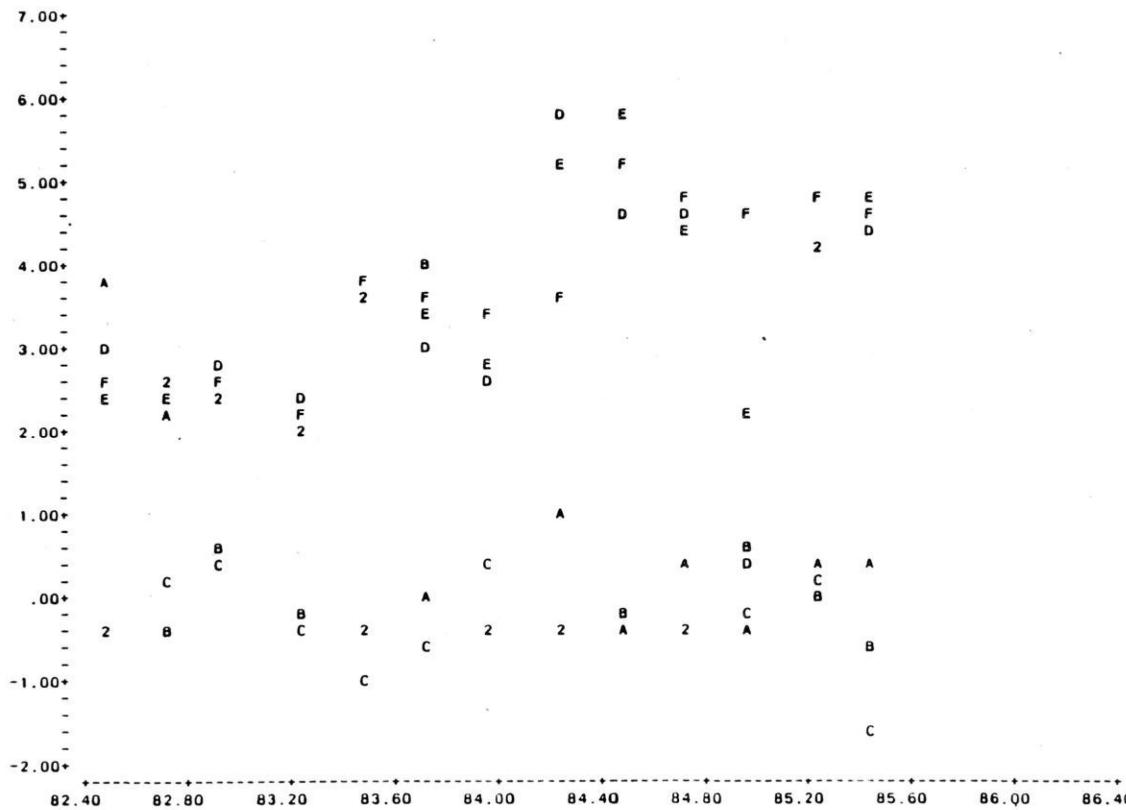
APPENDIX C: FIGURE 3

SAUK COUNTY LANDFILL

WELL 115 CATEGORY II SAND

.10 + .006 +

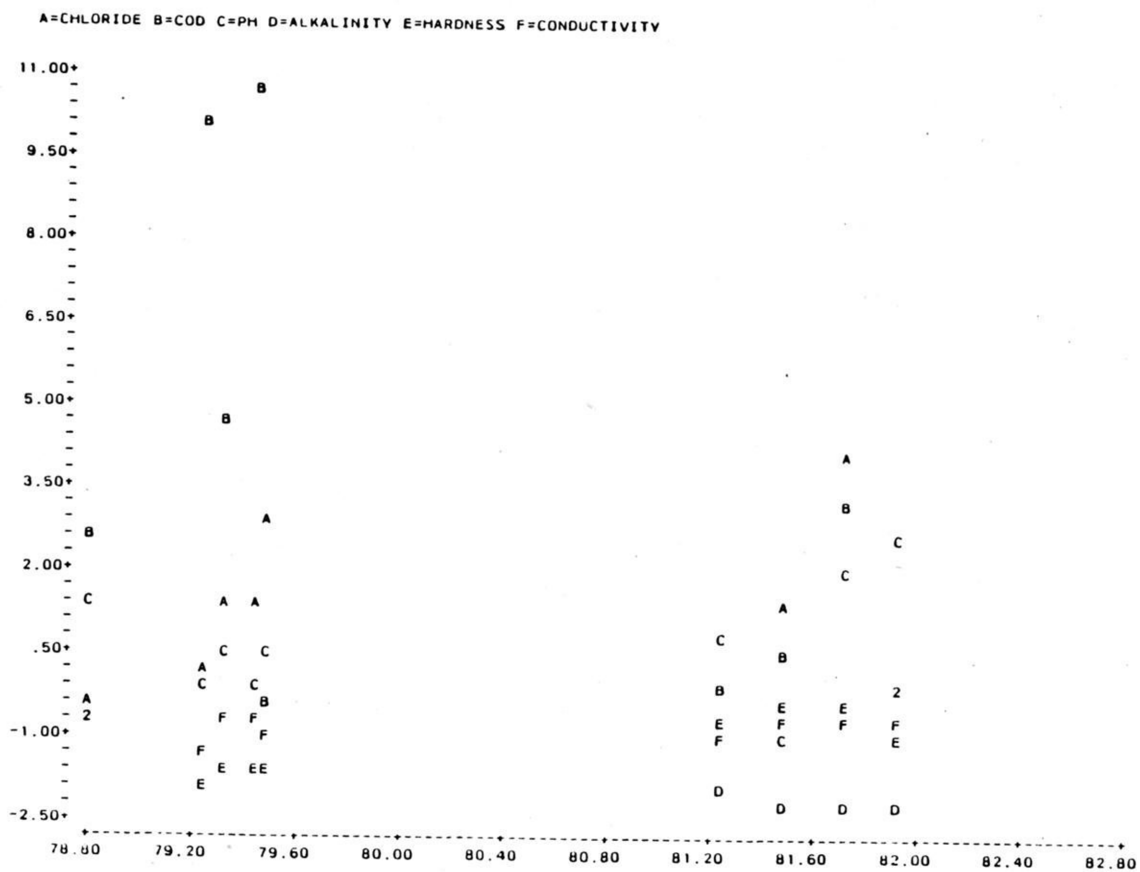
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY



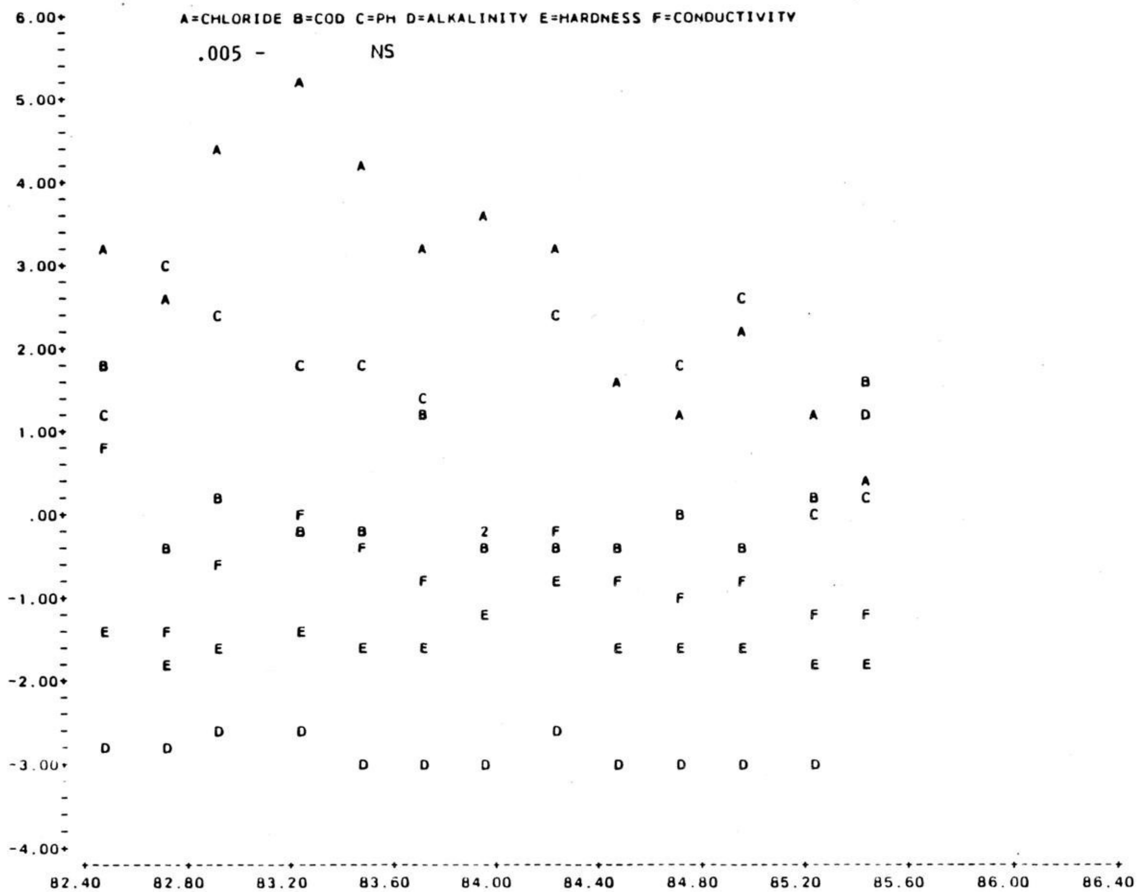
APPENDIX C: FIGURE 4

SAUK COUNTY LANDFILL

WELL 107 CATEGORY I SAND

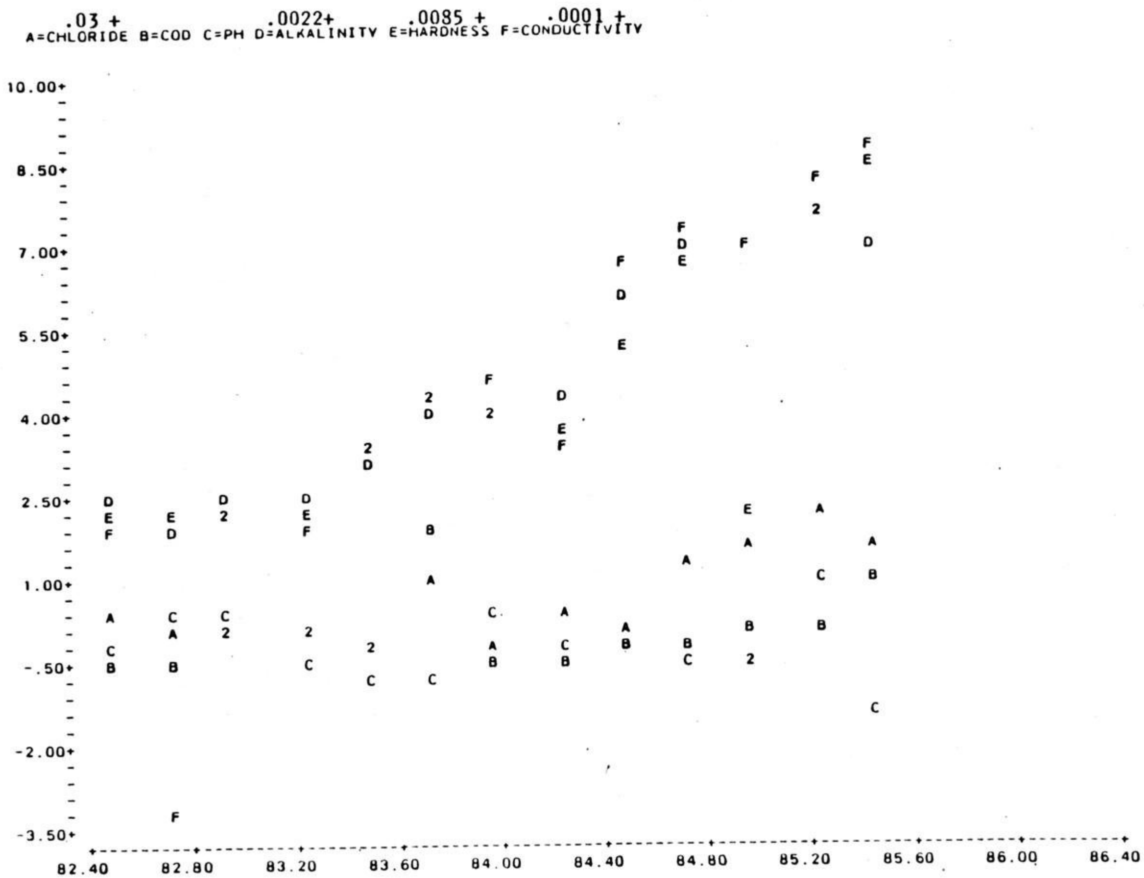


APPENDIX C: FIGURE 5
SAUK COUNTY LANDFILL
WELL 119 CATEGORY II SAND W/ GRAVEL



APPENDIX C: FIGURE 6
SAUK COUNTY LANDFILL

WELL 116 CATEGORY IV SANDSTONE

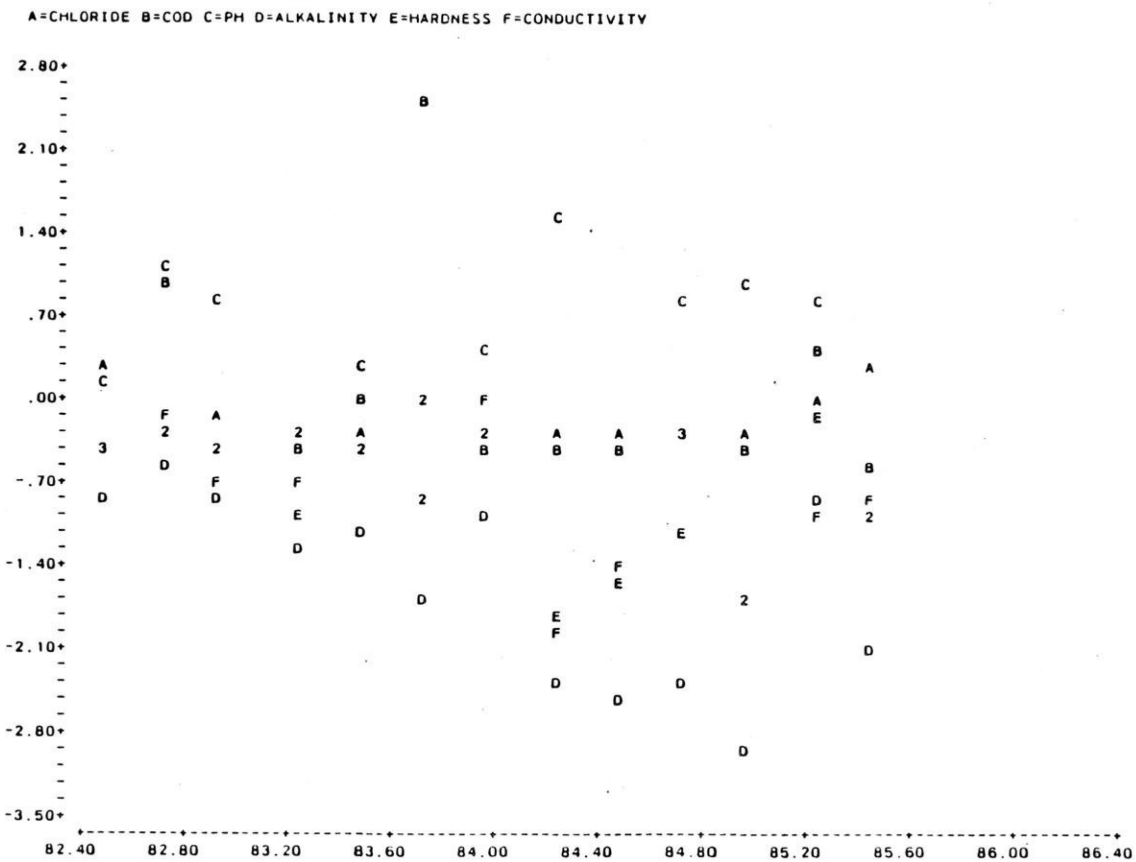


APPENDIX C: FIGURE 7

SAUK COUNTY LANDFILL

WELL 118 CATEGORY I FRACTURED SANDSTONE

.024 -

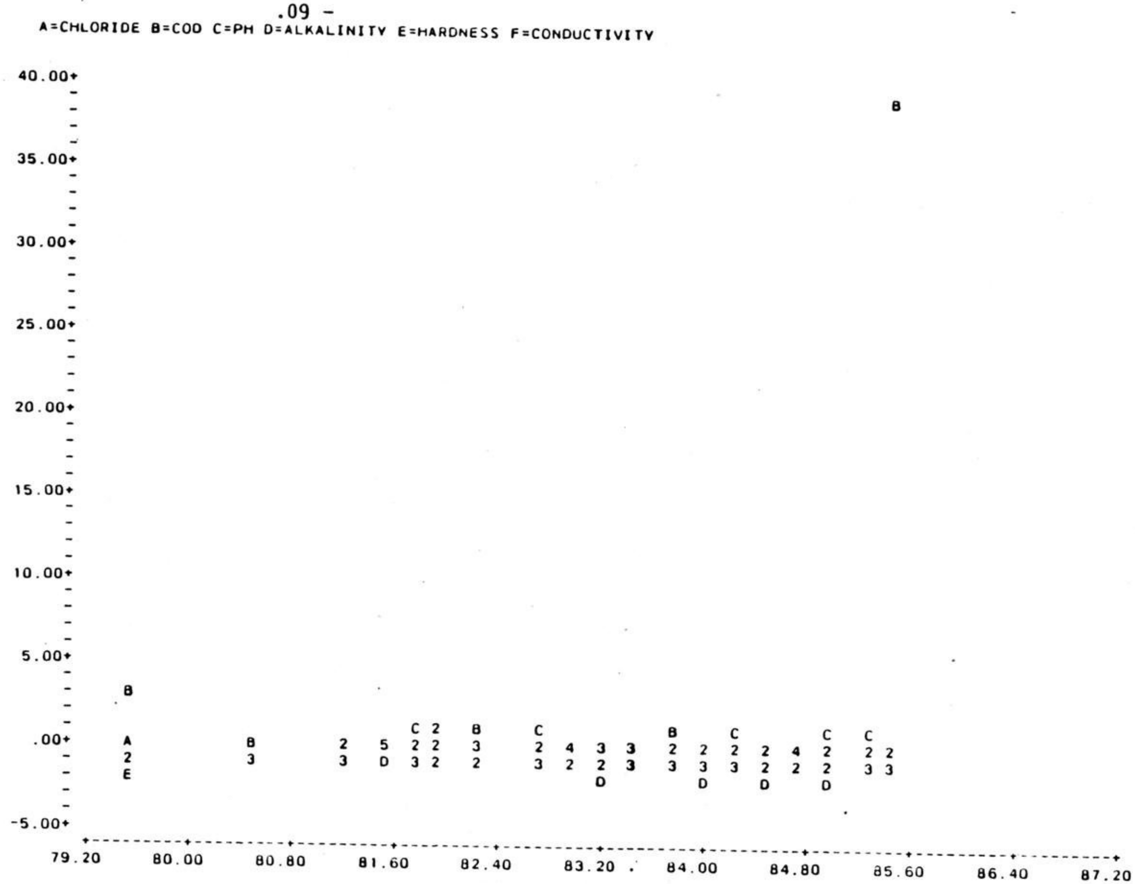


APPENDIX C: FIGURE 8

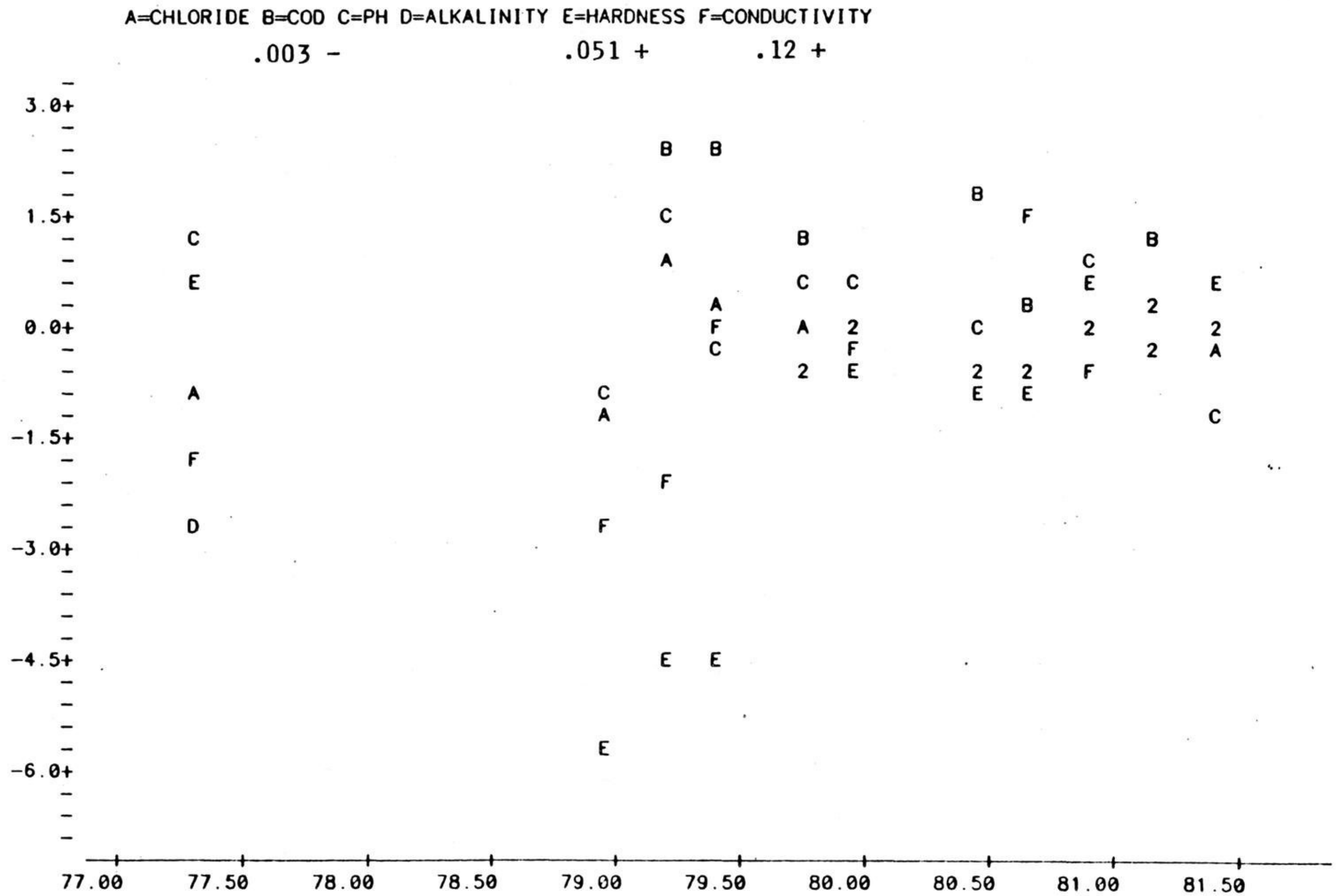
SAUK COUNTY LANDFILL

WELL 110 CATEGORY I BEDROCK SANDSTONE

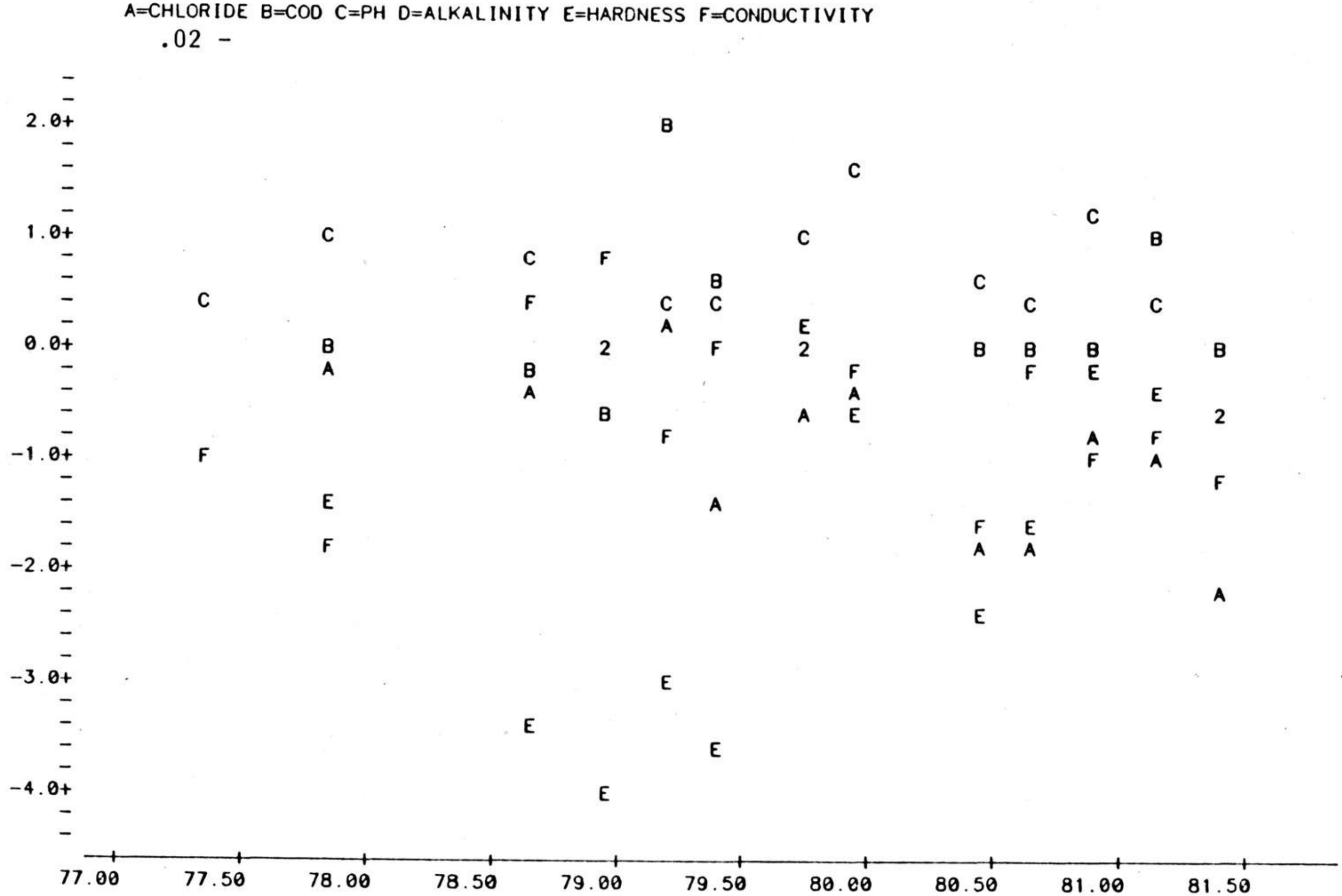
.09 -



APPENDIX D: FIGURE 1
DANE COUNTY - VERONA LANDFILL
WELL 106 CATEGORY I SILTY SAND



APPENDIX D: FIGURE 2
DANE COUNTY - VERONA LANDFILL
WELL 125 CATEGORY I SILTY SAND



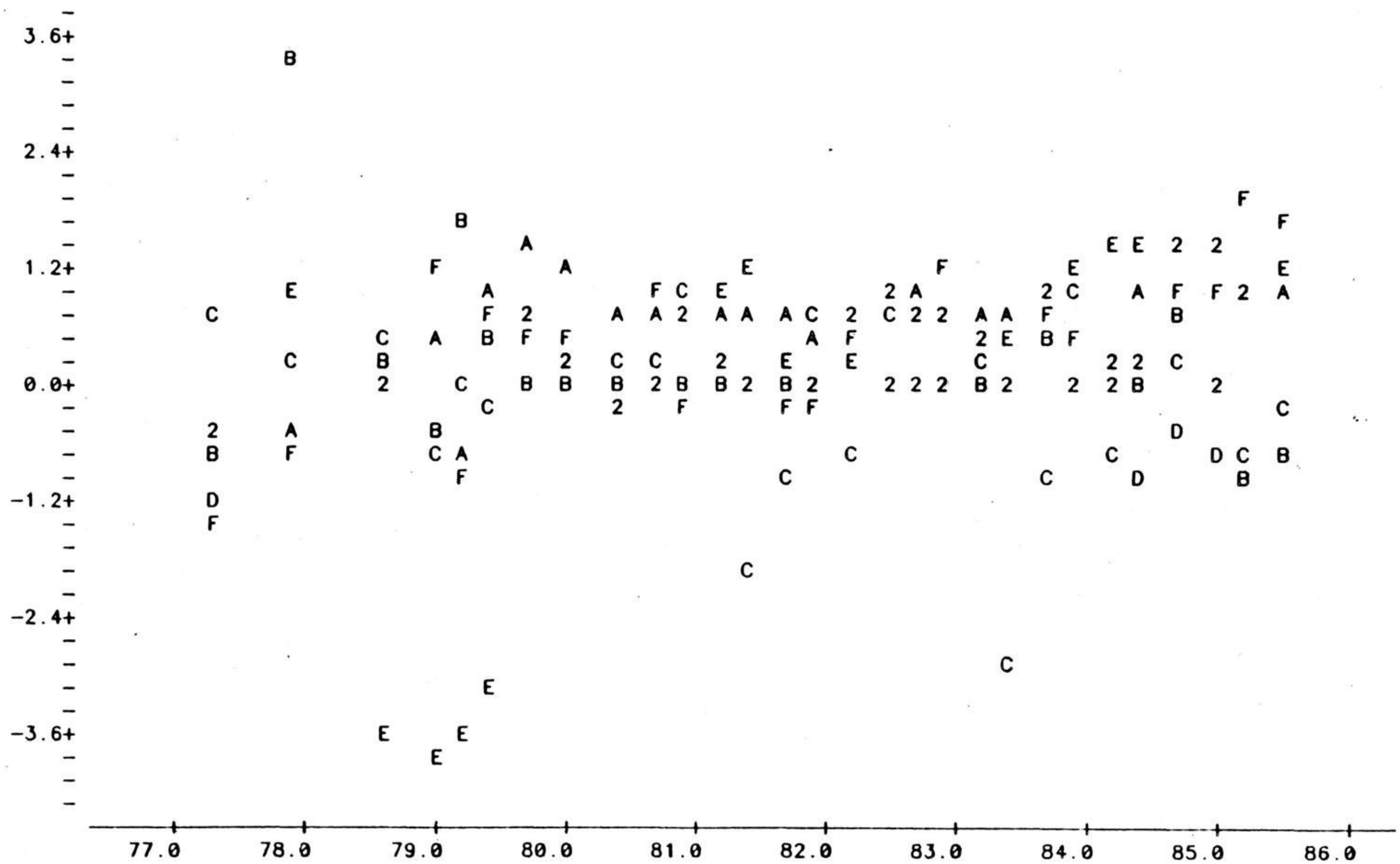
APPENDIX D: FIGURE 3

DANE COUNTY - VERONA LANDFILL

WELL 135 CATEGORY II SILTY SAND

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

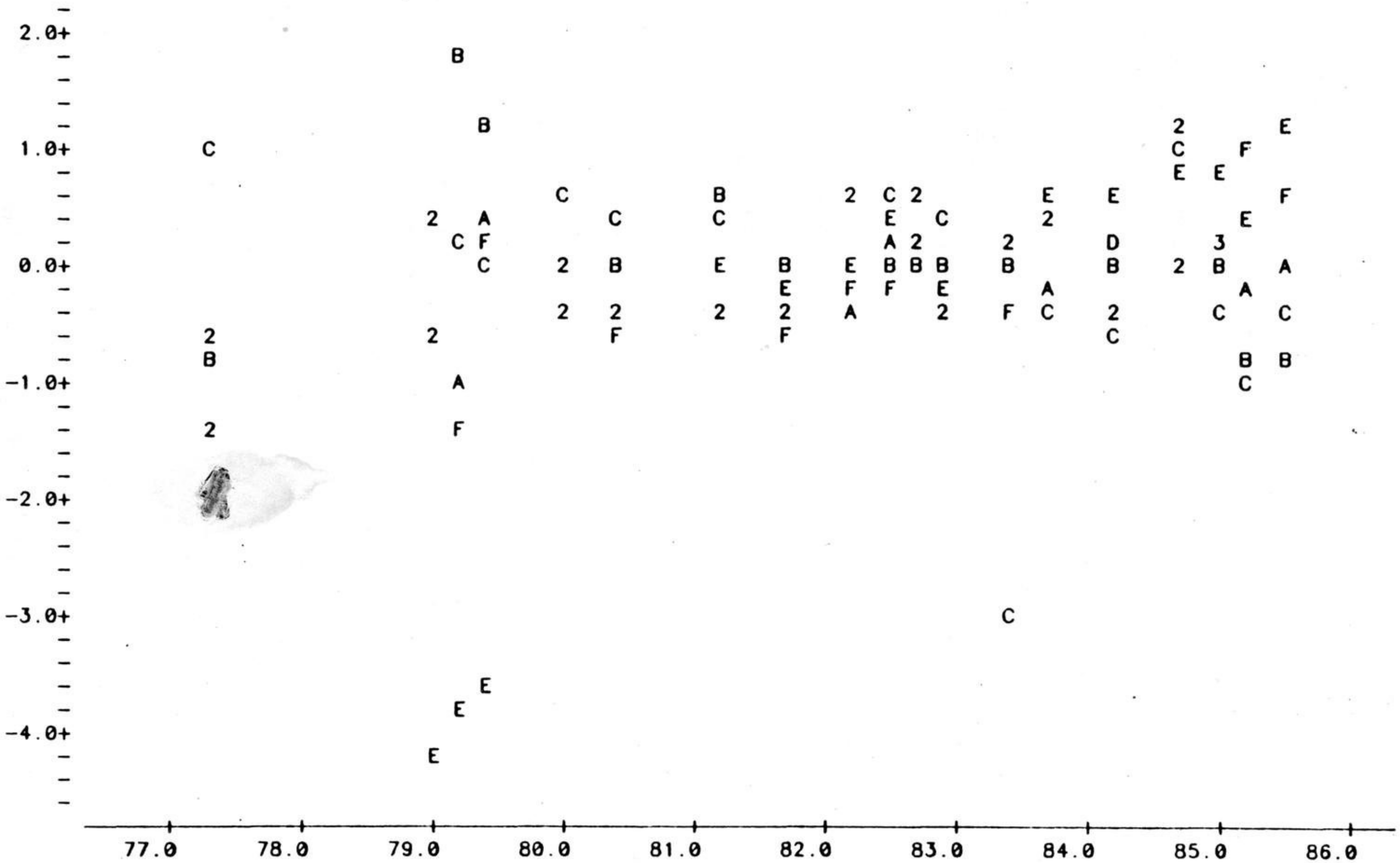
.012 + .000 + .0034 +



APPENDIX D: FIGURE 4
DANE COUNTY - VERONA LANDFILL
WELL 136 CATEGORY I SILTY SAND

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

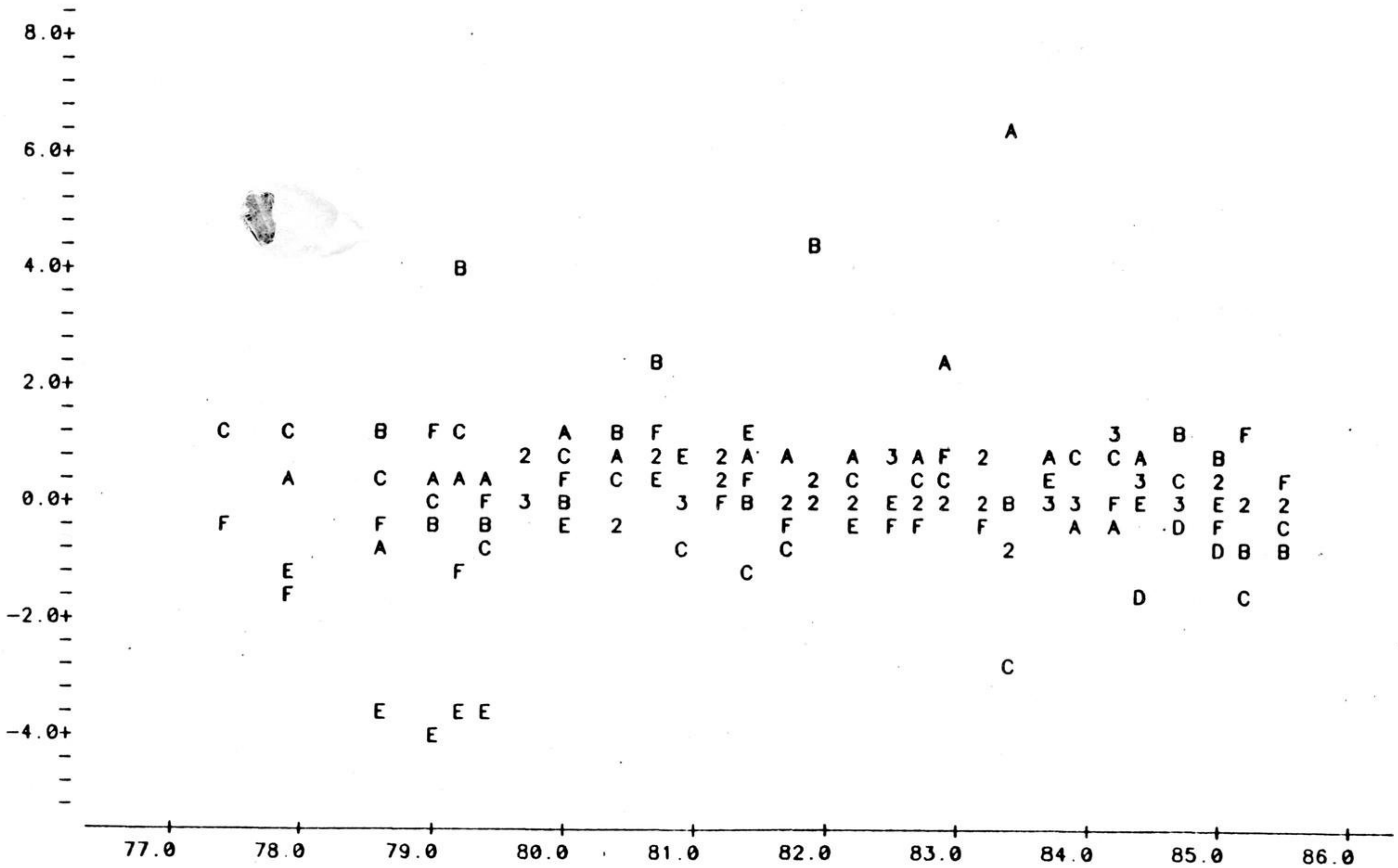
.000 + .012 +



APPENDIX D: FIGURE 5
DANE COUNTY - VERONA LANDFILL
WELL 115 CATEGORY II SILTY SAND

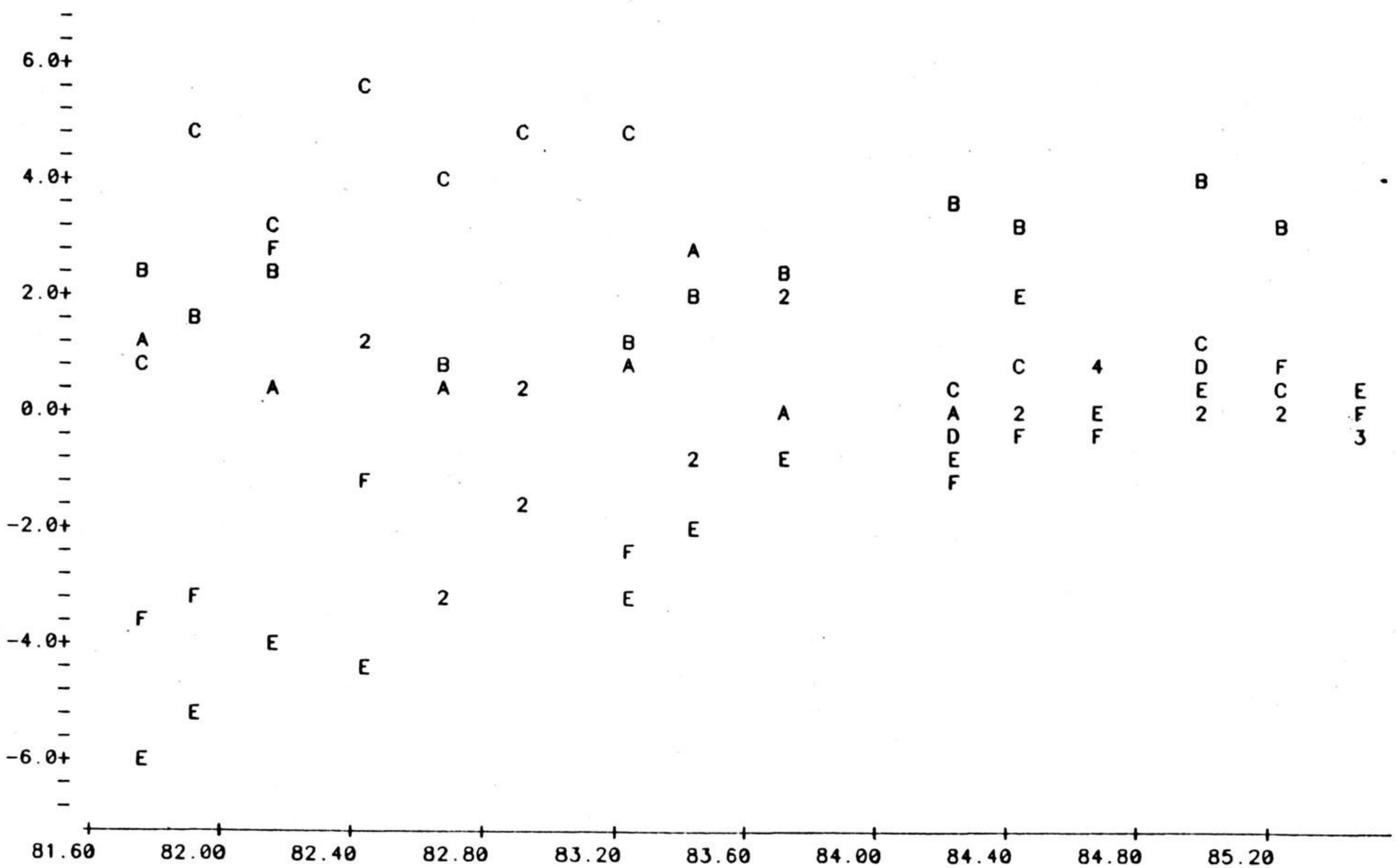
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

.06 - .036 + .13 +



APPENDIX D: FIGURE 6
DANE COUNTY - VERONA LANDFILL
WELL 175 CATEGORY IV SILTY SAND

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY
.004 - .017 - .000 + .011 +



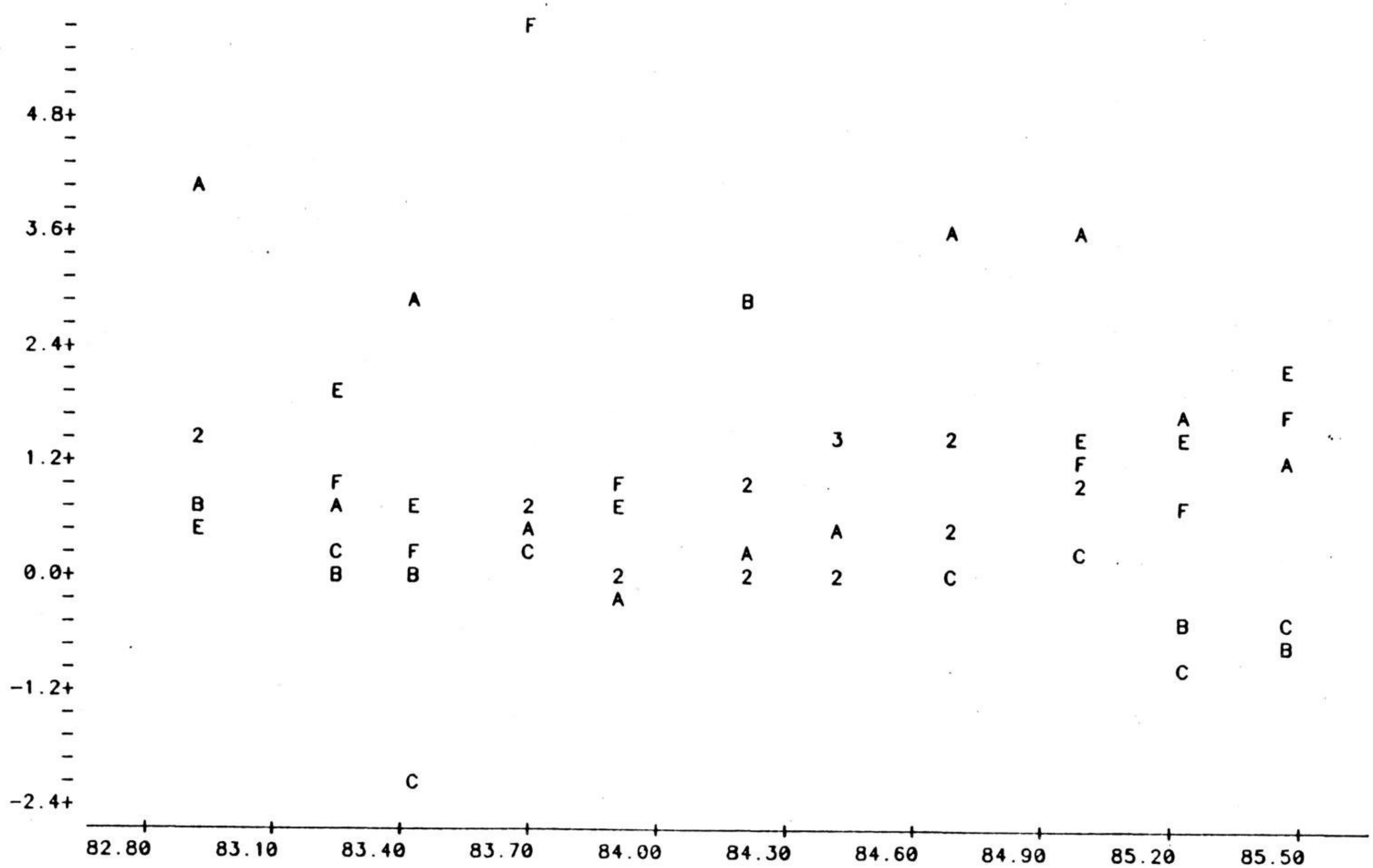
APPENDIX D: FIGURE 7

DANE COUNTY - VERONA LANDFILL

WELL 176 CATEGORY II SILTY SAND

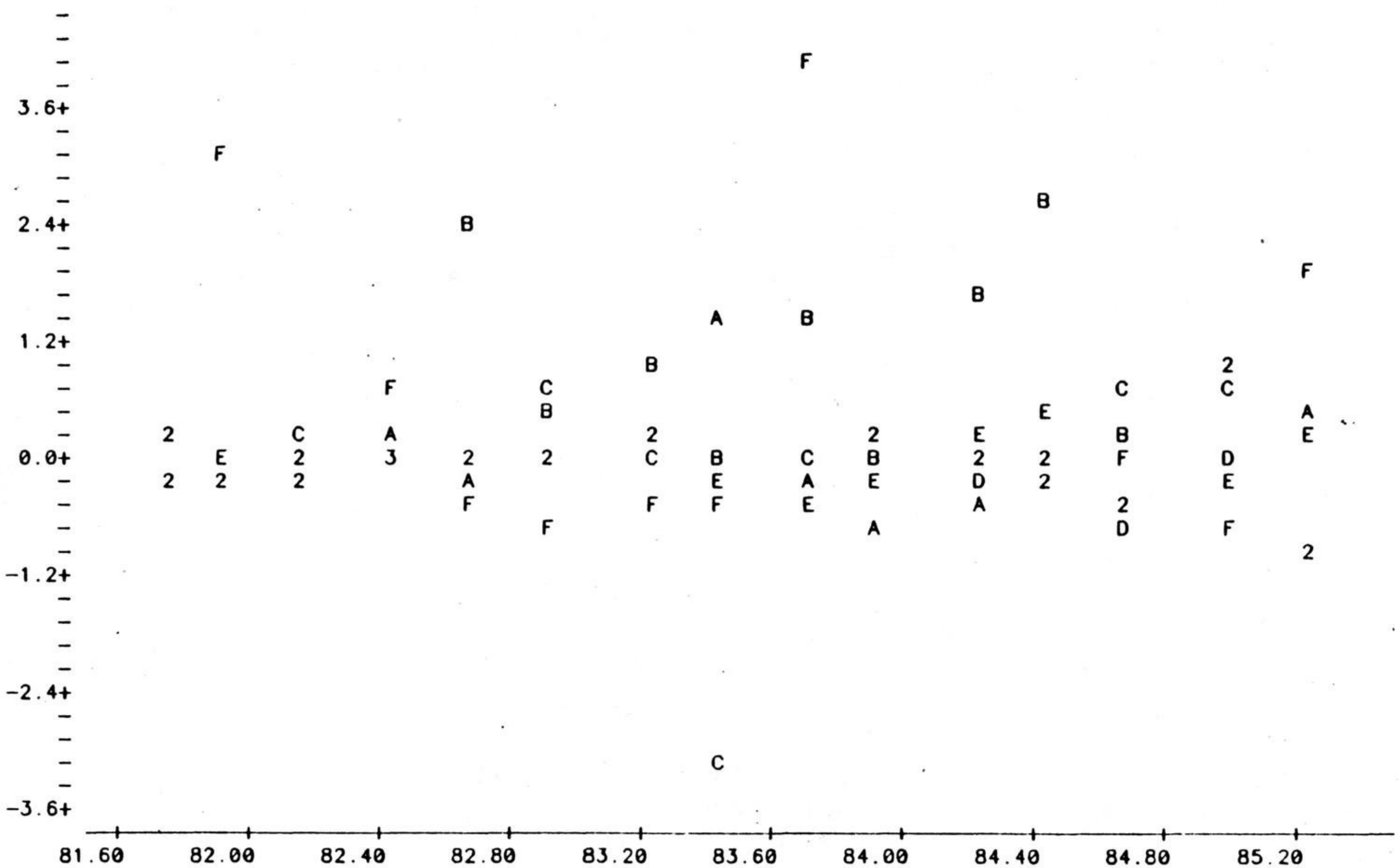
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

.09 - .004 +



APPENDIX D: FIGURE 8
DANE COUNTY - VERONA LANDFILL
WELL 173 CATEGORY I SILTY SAND

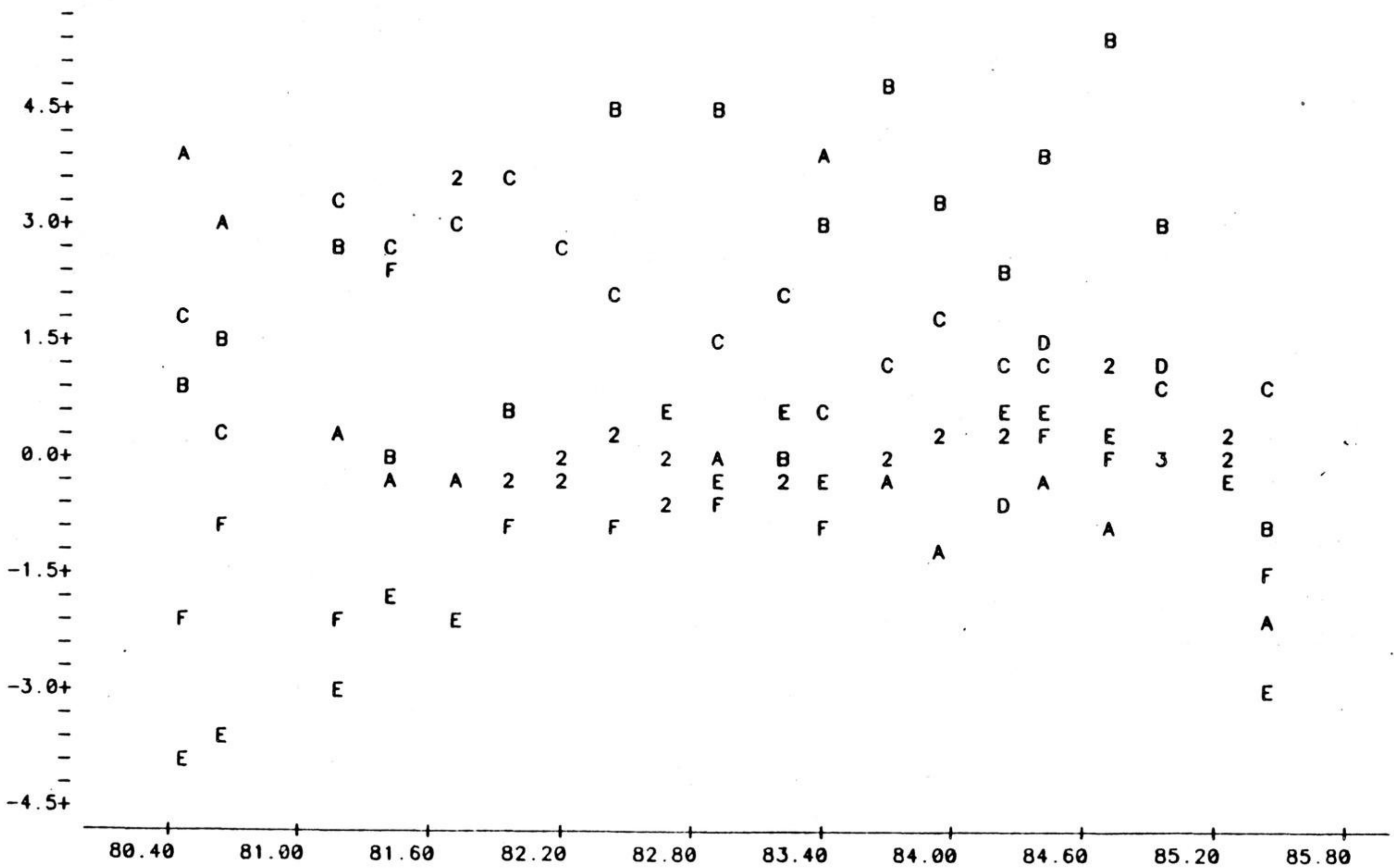
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY
NO SIGNIFICANT POSITIVE TRENDS



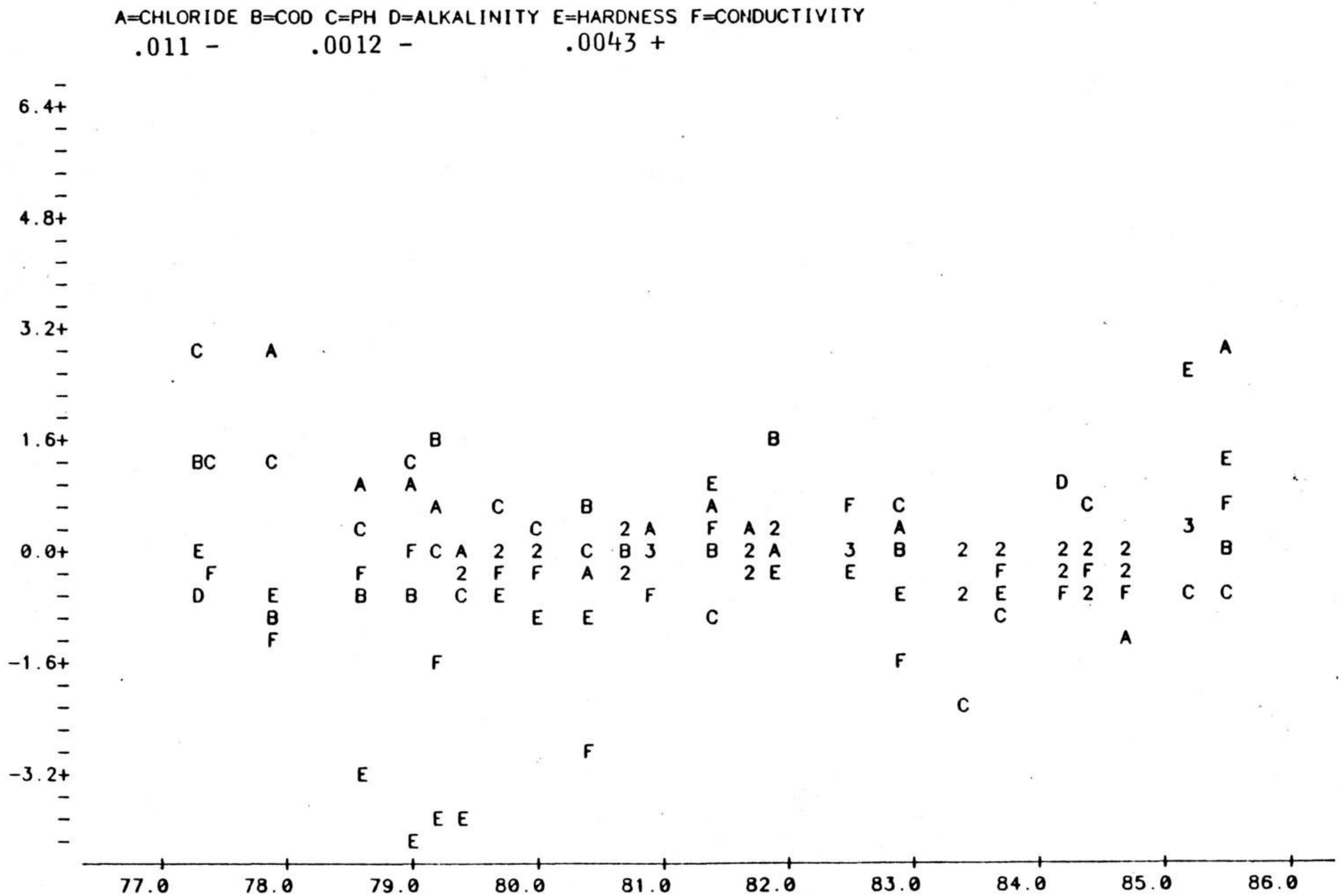
APPENDIX D: FIGURE 9
DANE COUNTY - VERONA LANDFILL
WELL 171 CATEGORY II SILTY SAND

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

.009 - .025 +



APPENDIX D: FIGURE 10
DANE COUNTY - VERONA LANDFILL
WELL 140 CATEGORY I SILTY SAND

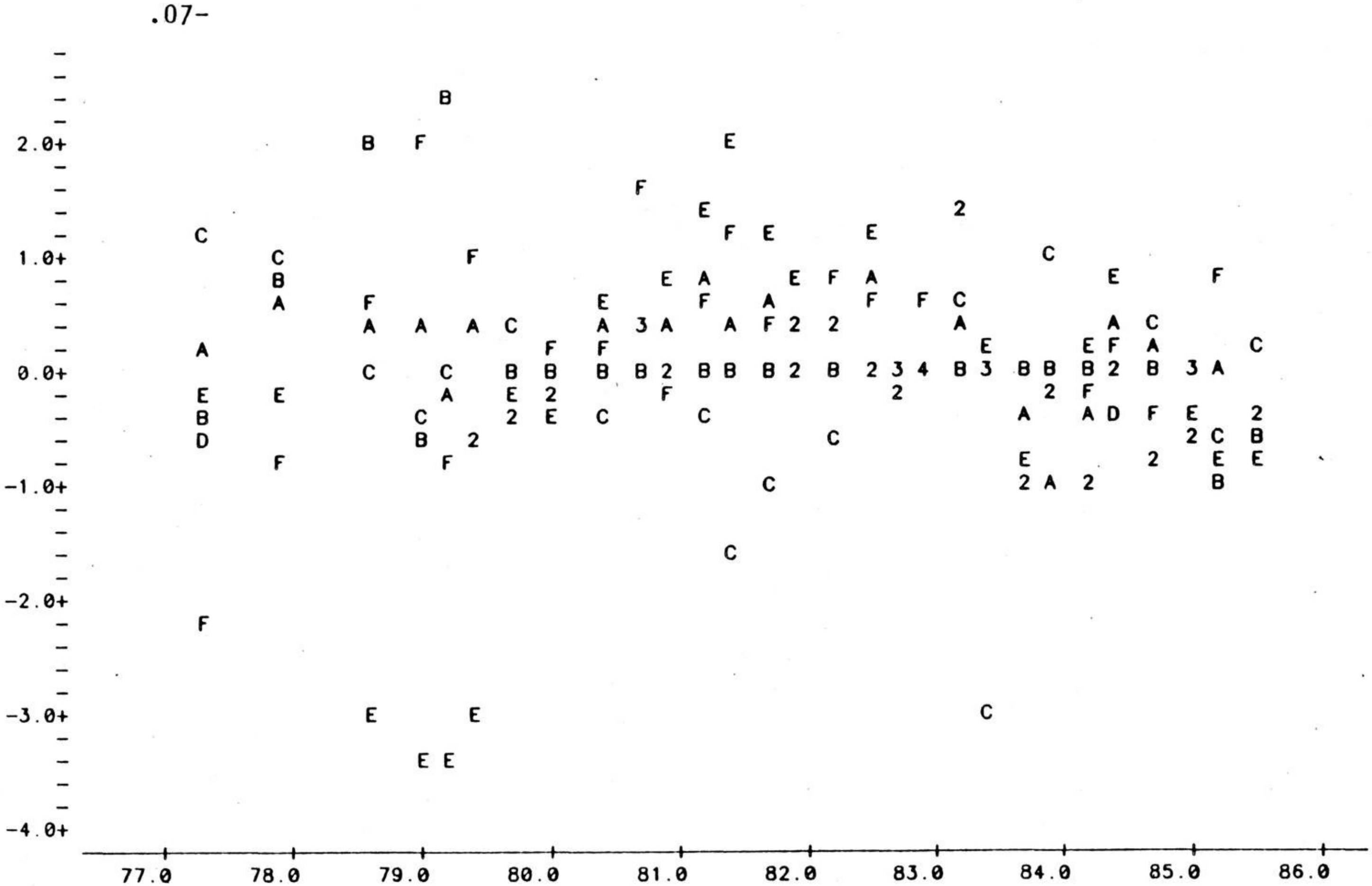


APPENDIX D: FIGURE 11

DANE COUNTY - VERONA LANDFILL

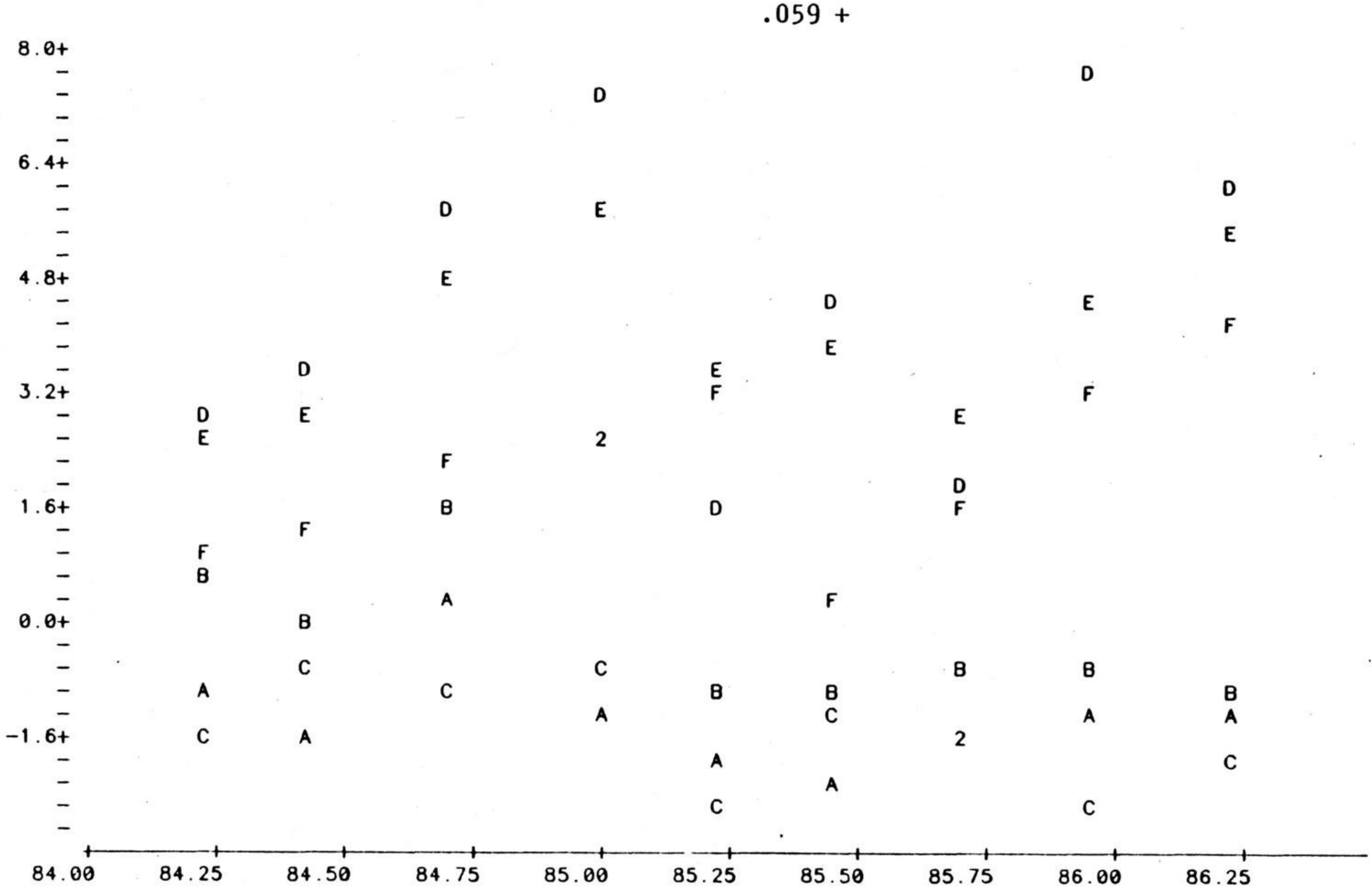
WELL 134 CATEGORY I LIMESTONE

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY



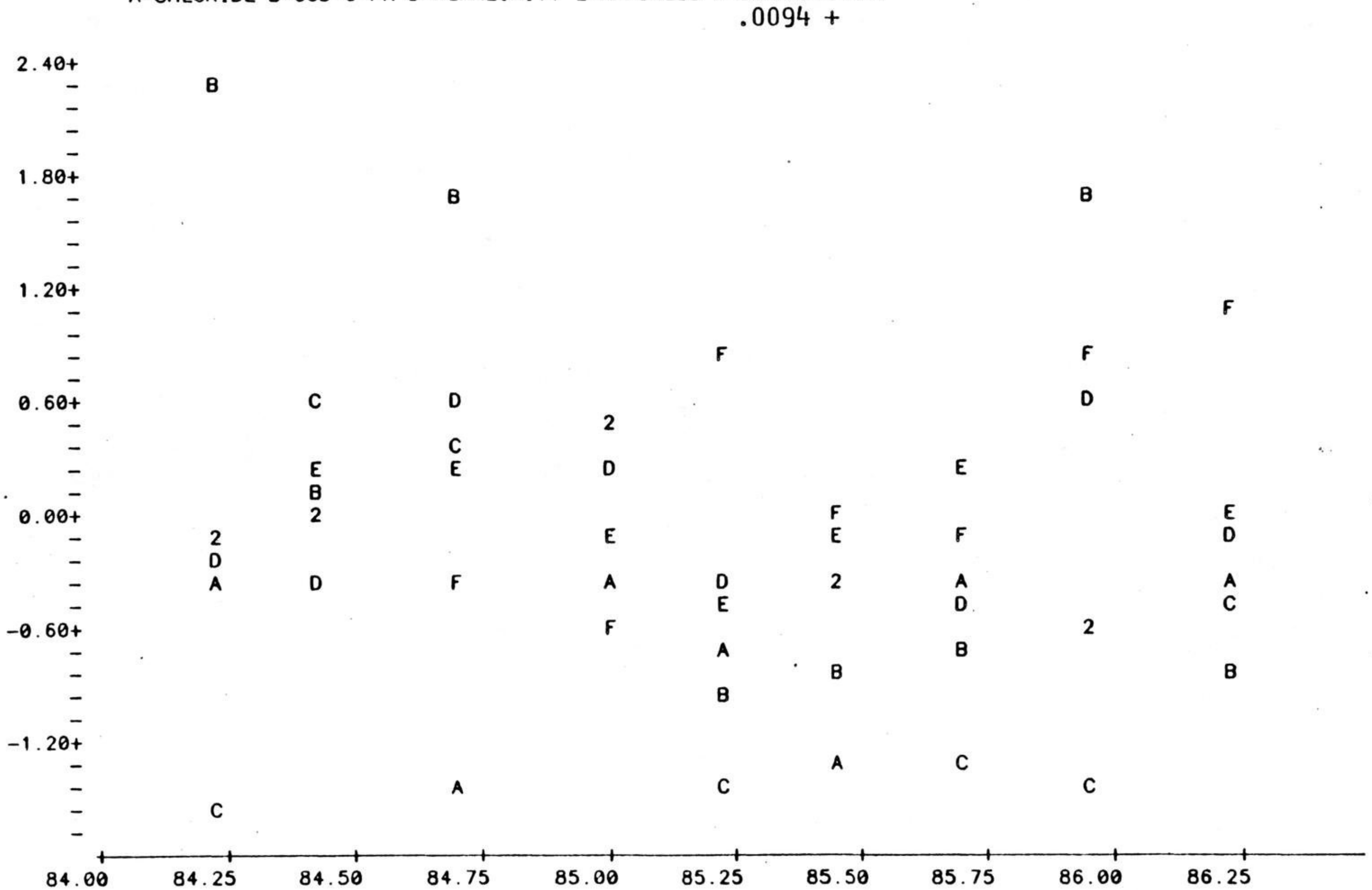
APPENDIX D: FIGURE 12
DANE COUNTY - VERONA LANDFILL
WELL 180 CATEGORY II LIMESTONE

A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY

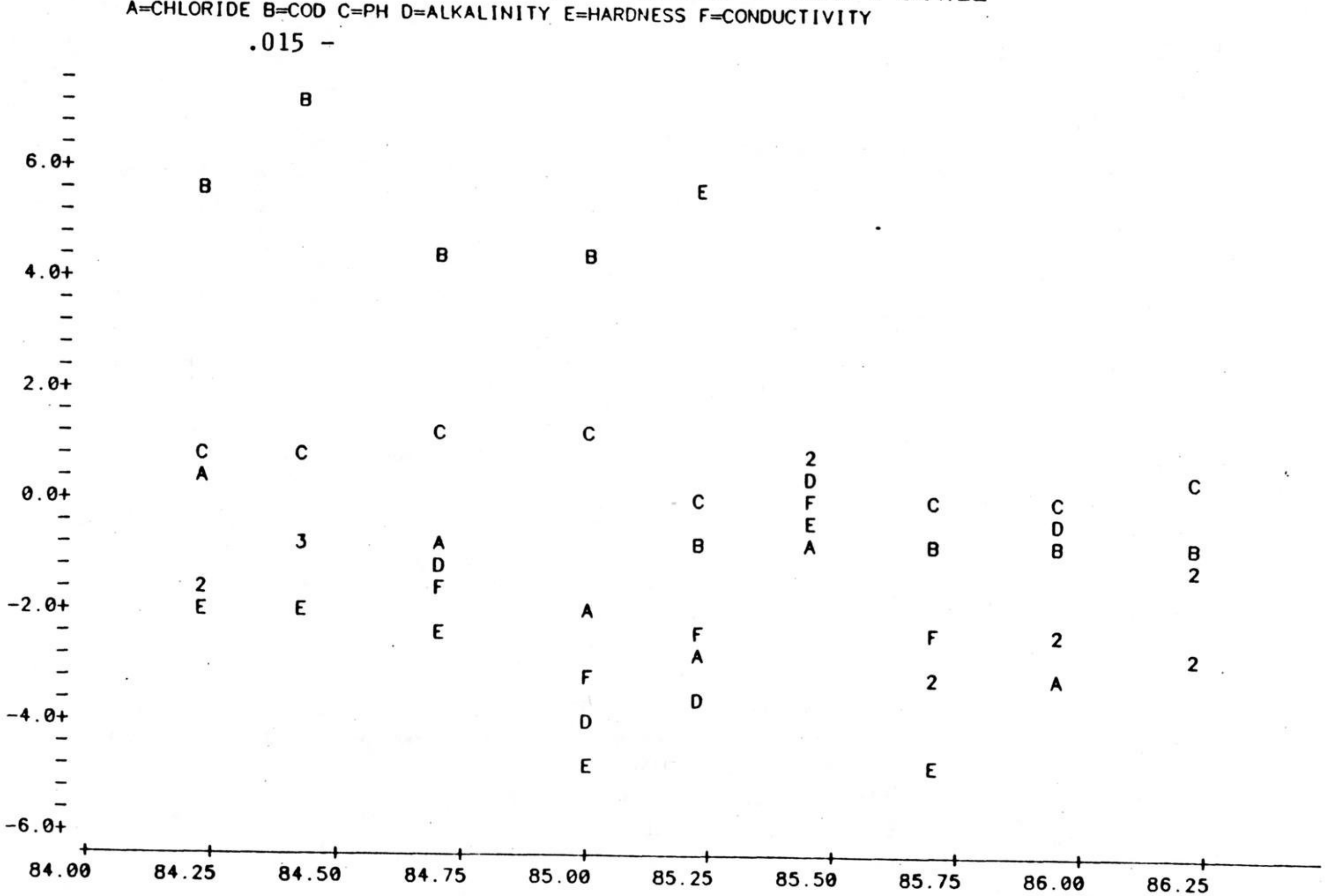


APPENDIX D: FIGURE 13
DANE COUNTY - VERONA LANDFILL
WELL 181 CATEGORY I LIMESTONE (?)

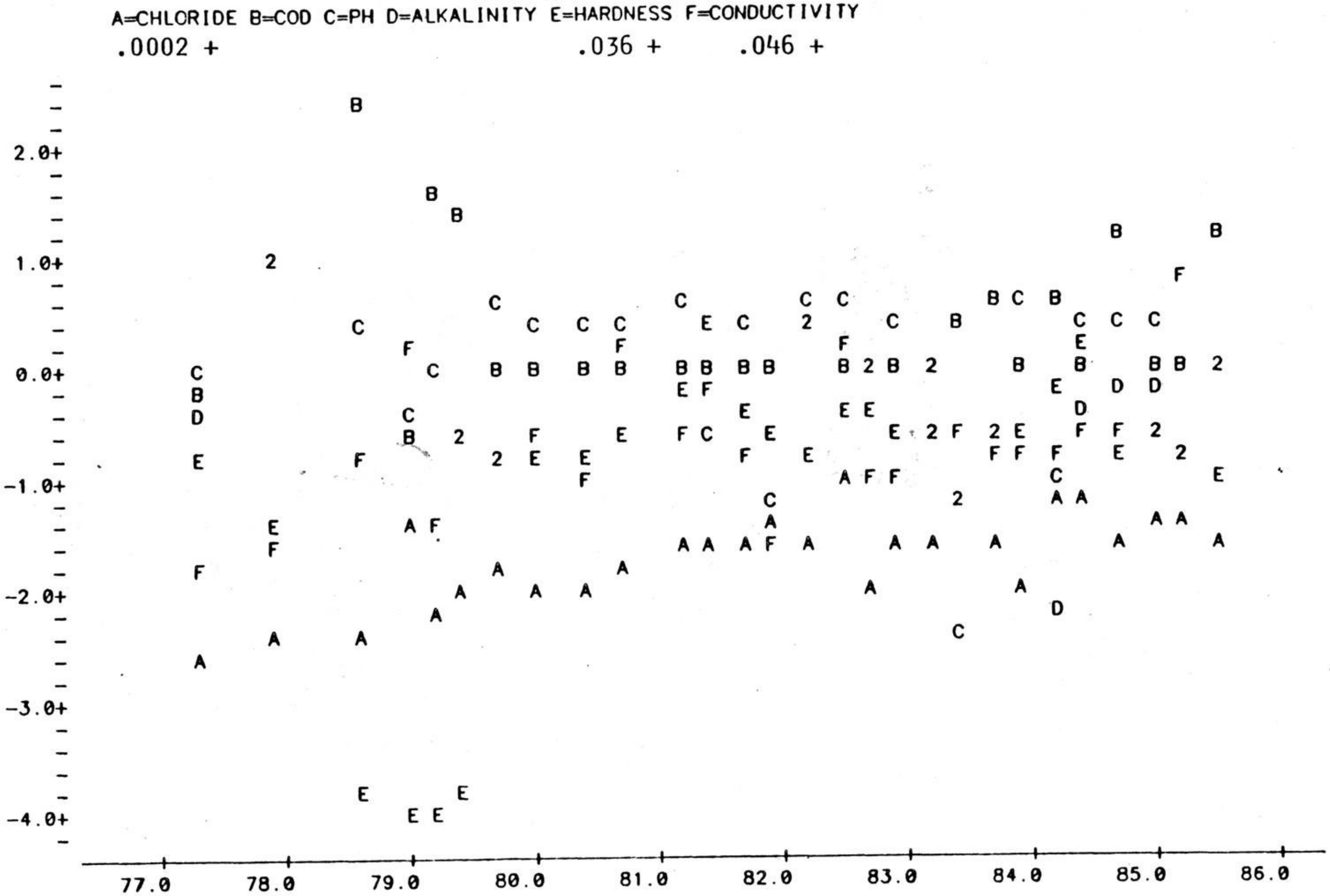
A=CHLORIDE B=COD C=PH D=ALKALINITY E=HARDNESS F=CONDUCTIVITY



APPENDIX D: FIGURE 14
DANE COUNTY - VERONA LANDFILL
WELL 177 CATEGORY I SAND & GRAVEL



APPENDIX D: FIGURE 15
DANE COUNTY - VERONA LANDFILL
WELL 150 CATEGORY I CAMBRIAN SANDSTONE



050847- Graphical and Statistical Methods to Assess the Effect of Landfills on Groundwater Quality