APPENDIX Q
Trend Analysis

Q-1. Introduction.

Q-1.1. This Appendix presents tools for detecting and estimating trends in environmental data. Trends may be spatial or temporal and can take various forms, including steady increases or decreases or a steep increase or decrease at a point in time or space. Detecting and estimating temporal or spatial trends are important for many environmental studies or monitoring programs. In cases where temporal or spatial patterns are strong, simple procedures such as time plots or linear regression over time can reveal trends. In more complex situations, sophisticated statistical models and procedures may be needed. The detection of trends may be complicated by the overlaying of long- and short-term trends, cyclical effects such as seasonal or weekly systematic variations, autocorrelations, or impulses or jumps from interventions or procedural changes. Trend is just one of several aspects of time series, the study of data with respect to time. Time series consists of trends, seasonal variation or seasonality, cyclical variation or repetitive trends, and irregular activity (Kvanli et al., 1996).

Q-1.2. The following subparagraphs present methods for detecting seasonal or temporal repetitive trends, correcting for seasonality, and testing procedures for trends using regression techniques and more robust trend estimation procedures. The investigations of trends in this Appendix are limited to one-dimensional domains, trends in a constituent concentration over time. This Appendix does not address spatial trends (with two- and three-dimensional domains) and trends over space and time (with three- and four-dimensional domains), which may involve sophisticated geostatistical techniques such as kriging (Appendix R). Gilbert (1987) and Gibbons (1994) provide additional resources for trend analysis.

Q-2. Identifying Seasonality and Other Repetitive Trends. Seasonality is one factor that accounts for changes in concentrations over time. Environmental monitoring data are likely to exhibit seasonality. According to Kvanli et al. (1996), seasonality is a predictable, periodic increase or decrease that occurs within a time period or cycle, such as 1 year. The key to identifying such trends is the repetition of the same pattern for each cycle. Identifying seasonality or other repetitive trends (i.e., persistent cyclic variations) is necessary before long-term increasing or decreasing temporal trends can be evaluated in environmental data. To identify these, a project team should visually inspect plots of data across time for seasonal or repetitive trends. Project teams should justify all seasonal trends identified visually with respect to site history, geology, chemistry, and professional judgment.

Q-2.1. *Introduction*.

Q-2.1.1. Generally, seasonality is not the primary focus of evaluating monitoring data for temporal trends. As such, data should be adjusted to remove the seasonal effects so that other temporal trends may be studied. For instance, if groundwater concentrations are diluted every spring by high recharge, true changes in groundwater may be masked by this effect. Likewise, if low water flow in fall leads to higher concentrations in groundwater that do not represent more leaching from a source area, then these effects should be accounted for in data evaluation. Seasonal effects may be removed by adjusting the sample data or using statistical methods unaffected by such relations. Adjustments to the sample data are described in this Paragraph. The subsequent Paragraph provides details about statistical tests that account for data with seasonal variability.

Q-2.1.2. There are various methods to de-seasonalize data. If the seasonal pattern is regular, it may be modeled with a sine or cosine function. Moving averages can be used, or differences (of order 12 for monthly data, for example) can be used. However, time series models may include rather complicated methods for de-seasonalizing the data. A simpler method is presented in EPA 530-SW-89-026 for applications to any seasonal cycle. For environmental data, seasonal cycles typically occur annually, monthly, or quarterly. Directions for the EPA method are presented in Paragraph Q-2.2, followed by an example in Paragraph Q-2.3. Although EPA's method assigns seasonality as a monthly cycle, this method can be applied with other seasonal or repetitive cycles by replacing "monthly" with the appropriate cycle.

Q-2.2. *Directions for Correcting Seasonality in Data*. To correct seasonality with time series data, directions are provided for monthly data that demonstrate a yearly cycle.

Q-2.2.1. Assume n years of monthly data are available.

Q-2.2.2. Let x_{ij} denote the unadjusted observation for the i^{th} month and the j^{th} year.

Q-2.2.3. Compute the average concentration for month i over the n-year period:

$$\bar{x}_i = \frac{(x_{i1} + ... + x_{in})}{n}.$$

This average represents the average of all observations taken in different years, but during the same month.

Q-2.2.4. Calculate the grand mean, $\bar{x}$, of all 12 n observations:

$$\bar{x} = \sum_{i=1}^{12} \frac{\bar{x}_i}{12}.$$

Q-2.2.5. Compute the adjusted concentrations,

$$y_{ij} = x_{ij} - \bar{x}_i + \bar{x} .$$

Q-2.2.6. The difference $x_{ij} - \bar{x}_i$ removes the average effect of month i from the monthly data. The grand mean ($\bar{x}$) must be added (on the right hand side of the equation) so that the mean of the adjusted y_{ij} values, $\bar{y}$, is equal to the grand mean ($\bar{x}$) of the unadjusted values.

Q-2.3. *Correcting Seasonality with Time Series Data (Based on Monthly Data with a Yearly Cycle)*. Consider evaluating seasonality for the monthly average temperature (in degrees Fahrenheit) in Austin, Texas, from 1995 through 1998 (Table Q-1). A time plot of the data is presented in Figure Q-1.

Table Q-1.
Monthly Average Temperature (°F) in Austin, Texas, from 1995 through 1998

Month-Year	Temperature	Month-Year	Temperature	Month-Year	Temperature	Month-Year	Temperature
Jan-95	50.03	Jan-96	47.10	Jan-97	46.00	Jan-98	53.06
Feb-95	53.00	Feb-96	53.38	Feb-97	50.15	Feb-98	52.21
Mar-95	57.00	Mar-96	52.84	Mar-97	60.68	Mar-98	55.90
Apr-95	62.23	Apr-96	62.77	Apr-97	59.57	Apr-98	62.70
May-95	71.94	May-96	73.67	May-97	67.87	May-98	73.68
Jun-95	74.23	Jun-96	77.13	Jun-97	74.97	Jun-98	79.60
Jul-95	79.26	Jul-96	81.06	Jul-97	78.45	Jul-98	82.10
Aug-95	78.45	Aug-96	77.42	Aug-97	77.94	Aug-98	80.19
Sep-95	74.07	Sep-96	72.93	Sep-97	75.03	Sep-98	78.73
Oct-95	66.06	Oct-96	66.13	Oct-97	65.84	Oct-98	68.10
Nov-95	55.77	Nov-96	56.55	Nov-97	53.83	Nov-98	60.37
Dec-95	51.37	Dec-96	51.93	Dec-97	47.50	Dec-98	49.81

Q-2.3.1. The plot indicates that seasonality plays a role in this data. There are $n = 4$ years of monthly data. The average temperature for each month and the grand average for all months are presented below:

Month	**Average Temperature**
January	49.05
February	52.19
March	56.61
April	61.82
May	71.79
June	76.48
July	80.22
August	78.50
September	75.19
October	66.53
November	56.63
December	50.15
Grand Average	64.60

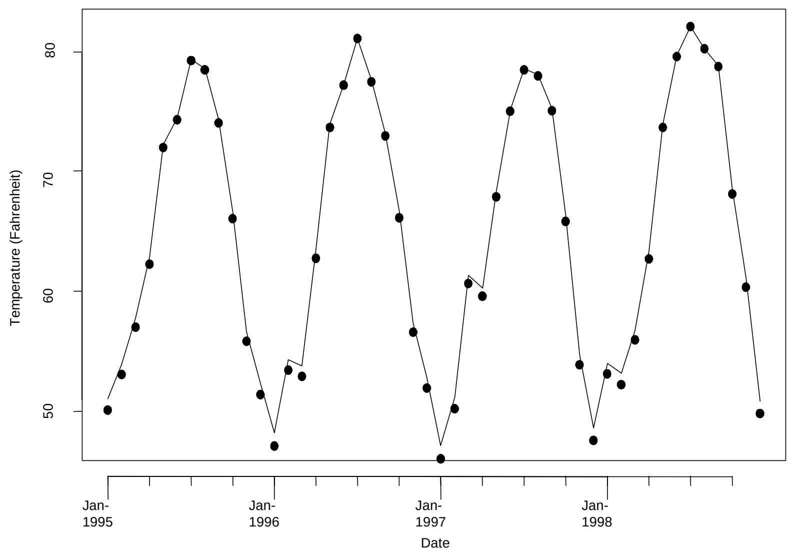

Figure Q-1. Monthly average temperature (°F) in Austin, Texas, from 1995 through 1998.

Q-2.3.2. The average January temperature is simply the average of all the January temperatures, no matter the year:

$$\bar{x}_{January} = \frac{50.03 + 47.10 + 46.00 + 53.06}{4} = 49.05\,.$$

Q-2.3.3. The other monthly averages are estimated in the same fashion. The grand average is simply the average of all of the monthly averages:

$$\bar{x} = \frac{49.05 + 52.19 + 56.61 + 61.82 + 71.79 + 76.48 + 80.22 + 78.50 + 75.19 + 66.53 + 56.63 + 50.15}{12} = 64.60$$

Q-2.3.4. The adjusted averages are presented in Table Q-2. The adjusted Jan-1995 temperature, for example, was estimated by the following: adjusted temperature = 50.03 – 49.05 + 64.60 = 65.58. Figure Q-2 is a plot of the adjusted temperatures. The vertical scale of the plot is the same as the plot of the adjusted data to emphasize that the seasonal variation has been smoothed out.

Table Q-2.
Adjusted Monthly Average Temperature (°F) in Austin, Texas, from 1995 through 1998

Month-Year	Temperature	Monthly average temperature	Grand average temperature	Adjusted temperature	Month-Year	Temperature	Monthly average temperature	Grand average temperature	Adjusted temperature
Jan-95	50.03	49.05	64.60	65.58	Jan-97	46.00	49.05	64.60	61.55
Feb-95	53.00	52.19	64.60	65.41	Feb-97	50.15	52.19	64.60	62.56
Mar-95	57.00	56.60	64.60	64.99	Mar-97	60.68	56.60	64.60	68.67
Apr-95	62.23	61.82	64.60	65.01	Apr-97	59.57	61.82	64.60	62.35
May-95	71.94	71.79	64.60	64.74	May-97	67.87	71.79	64.60	60.68
Jun-95	74.23	76.48	64.60	62.35	Jun-97	74.97	76.48	64.60	63.08
Jul-95	79.26	80.22	64.60	63.64	Jul-97	78.45	80.22	64.60	62.83
Aug-95	78.45	78.50	64.60	64.55	Aug-97	77.94	78.50	64.60	64.03
Sep-95	74.07	75.19	64.60	63.47	Sep-97	75.03	75.19	64.60	64.44
Oct-95	66.06	66.53	64.60	64.13	Oct-97	65.84	66.53	64.60	63.90
Nov-95	55.77	56.63	64.60	63.73	Nov-97	53.83	56.63	64.60	61.80
Dec-95	51.37	50.15	64.60	65.81	Dec-97	47.50	50.15	64.60	61.95
Jan-96	47.10	49.05	64.60	62.64	Jan-98	53.06	49.05	64.60	68.61
Feb-96	53.38	52.19	64.60	65.79	Feb-98	52.21	52.19	64.60	64.62
Mar-96	52.84	56.60	64.60	60.83	Mar-98	55.90	56.60	64.60	63.89
Apr-96	62.77	61.82	64.60	65.55	Apr-98	62.70	61.82	64.60	65.48
May-96	73.67	71.79	64.60	66.47	May-98	73.68	71.79	64.60	66.49
Jun-96	77.13	76.48	64.60	65.25	Jun-98	79.60	76.48	64.60	67.71
Jul-96	81.06	80.22	64.60	65.44	Jul-98	82.10	80.22	64.60	66.47
Aug-96	77.42	78.50	64.60	63.52	Aug-98	80.19	78.50	64.60	66.29
Sep-96	72.93	75.19	64.60	62.34	Sep-98	78.73	75.19	64.60	68.14
Oct-96	66.13	66.53	64.60	64.20	Oct-98	68.10	66.53	64.60	66.16
Nov-96	56.55	56.63	64.60	64.52	Nov-98	60.37	56.63	64.60	68.33
Dec-96	51.93	50.15	64.60	66.37	Dec-98	49.81	50.15	64.60	64.25

Q-2.4. *Summary*. Corrections for seasonality should be used with great caution because they represent extrapolation into the future. There should be good scientific explanation and good empirical evidence for the seasonality before corrections are made. For instance, larger than average rainfalls for two or three Augusts in a row does not justify the belief that there will never be a drought in August, and this idea extends directly to any monitoring system. In addition, the quality (bias, robustness, and variance) of the estimates of the proper corrections must be considered even in cases in which corrections are called for. If seasonality is suspected, adjusting for seasonality may not be necessary to evaluate long-term trends when appropriate statistical methods are utilized. Such methods will be discussed in the following Paragraph.

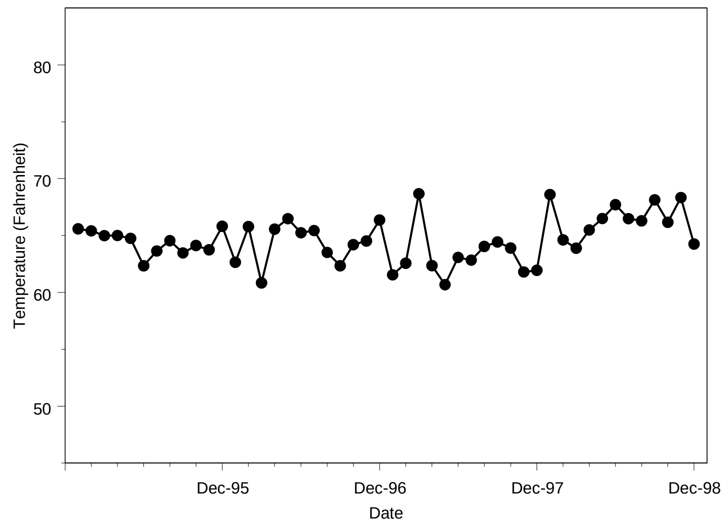

Figure Q-2. Adjusted monthly average temperature (°F) in Austin, Texas, from 1995 through 1998.

Q-3. Methods for Trend Assessment.

Q-3.1. *Introduction*. As a first step in evaluating trends, graphical representations are recommended to identify possible trends. A plot of the data versus time is recommended for temporal data, as it may reveal long-term trends and show other major types of trends, such as cycles or impulses.

Q-3.1.1. A posting plot is recommended for spatial data to reveal spatial trends such as areas of high concentration or areas that were inaccessible. (See Appendix J for further discussion of posting plots.) Gilbert (1987) recommends smoothing time series to identify cycles and long-term trends that may be obscured by natural variation in the data. Gilbert also mentions using control charts as an effective graphical tool of trends. Control charts are presented at the end of this section.

Q-3.1.2. Most of the statistical tools presented below are applicable to environmental data; the focus is on monotonic, long-term trends (i.e., trends that are exclusively increasing or decreasing, but not both), as well as other sources of systematic variation, such as seasonality.

Q-3.1.3. There are numerous tests for trends. Trend tests, like other statistical tests, can be divided in terms of distributional assumptions. Parametric trend tests, which assume data follow a normal distribution, involve regression-based methods for estimating trends and determining if a significant trend exists. Nonparametric trend tests, which do not make assumptions about the underlying data distributions, are based on the Mann-Kendall trend test.

Q-3.1.4. Independence is crucial for parametric and nonparametric tests. The departure from independence (if data are correlated) can result in incorrect conclusions (Gibbons, 1994). To minimize the possibility that samples are *not* independent, Gibbons recommends a sampling frequency of no more than one sample per quarter. In practice, sampling frequency may be based on knowledge of site conditions such as groundwater flow rates.

Q-3.1.5. Regression-based methods usually are not recommended for environmental studies as a general tool for estimating and detecting trends, although they may be useful as a quick and easy-to-use screening tool for identifying strong linear trends. Regression analyses can be misleading if seasonal cycles are present, the data are not normally distributed, or the data are serially correlated (Gilbert, 1987). In such cases, Gilbert suggests that the non-parametric seasonal Kendall test is preferable to regression methods. Non-parametric trend tests are more appropriate when data do not conform to a particular distribution and when there are data below the detection limit. For groundwater monitoring, Gibbons (1994) states that non-parametric analyses are the most reasonable estimators of trend.

Q-3.2. *Regression-Based Methods*. Classic procedures for assessing linear trends use regression. Linear regression is a common procedure in which calculations are performed on a data set containing pairs of observations (x_i, y_i). For temporal trends, the x_i values represent time and the y_i values represent the observations, such as contaminant concentrations. "If plots of data versus time suggest a simple linear increase or decrease over time, a linear regression of the variable against time may be fit to the data. A *t*-test may be used to test that the true slope is not different from zero (Gilbert, 1987)."

Q-3.2.1. Regression procedures are easy to apply but entail several limitations and assumptions. For example, simple linear regression (the most commonly used method) is designed to detect linear relationships between two variables; other types of regression models generally are necessary to detect non-linear relationships, such as cyclical or non-monotonic trends. Regression is also very sensitive to extreme values (outliers) and presents difficulties in handling data below the detection limit, which are commonly encountered in environmental studies.

Q-3.2.2. A regression model is of the form:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where:

Y = response/dependent variable
X = independent/explanatory variable (e.g., time)
β_0 = "true" intercept
β_1 = "true" slope
ε = random error.

Q-3.2.3. If not for the random error, ε, all of the points (x_i, y_i) would lie precisely on the line $Y = \beta_0 + \beta_1 X$. The regression model assumes that the error is a normally distributed random variable (ε) with a mean of zero and constant variance (i.e., the variance does not depend on X). In practice, β_0 and β_1 are unknown quantities and a set of n measured values (x_i, y_i) is used to estimate a regression line of the form:

$$y_i = b_0 + b_1 x_i + e_i$$

where b_0 is an estimate of β_0, b_1 is an estimate of β_1, and e_i estimates ε_i. The slope and intercept can be estimated as follows:

$$b_1 = r\left(\frac{s_y}{s_x}\right)$$

$$b_0 = \overline{y} - b_1\overline{x} \quad .$$

Q-3.2.4. The estimated "residuals" (e_i) are calculated from the equation:

$$e_i = y_i - (b_0 + b_1 x_i).$$

Q-3.2.5. Tests for normality (for example, normal probability plots as discussed in Appendix J) are required to verify the normality of the set of results $\{e_i\}$. A plot of e_i versus x_i is required to verify that the variance of the residuals is constant (i.e., not dependent upon X). Figure Q-3 shows two commonly seen residual patterns. In Figure Q-3a, the residuals show no pattern, so the assumption of constant variance is met. In Figure Q-3b, the variance of the residuals increases as the independent variable (X) increases so the assumption of constant variance is not met. Statistical software is often used to verify the normality of the residuals and constant variance because it is burdensome to do so manually. Moreover, the analyst must ensure that time plots of the data do not possess any cyclical patterns, outlier tests show no extreme data values, and data validation reports indicate that nearly all of the measurements are above detection limits.

Q-3.2.6. Because of these limitations, regression is *not* recommended as a general tool for estimating and detecting trends, although it may be useful as a screening tool for identifying strong linear trends. For situations in which regression methods can be applied appropriately, a solid body of literature on hypothesis testing is available that uses the concepts of statistical linear models as a basis for inferring the existence of temporal trends.

Q-3.2.7. For simple linear regression, the statistical test of whether the slope is significantly different from zero is equivalent to testing if the correlation coefficient is significantly different from zero; that is, if $r = 0$, the slope $b_1 = 0$ (for more details on the correlation coefficient test see Appendix O). Directions are provided in Paragraph O-2.2, followed by an example in Paragraph O-2.3.

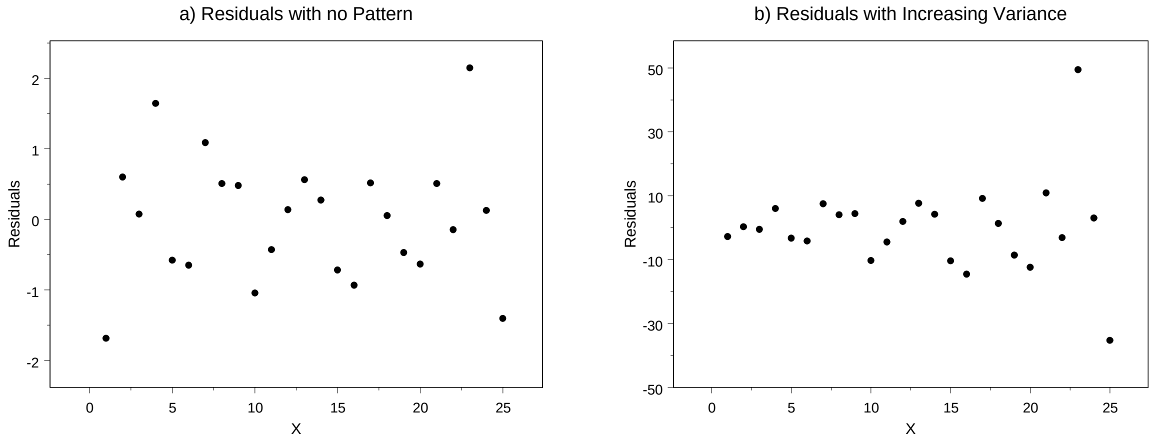

Figure Q-3. Residuals versus the independent variable.

Q-3.2.8. This test assumes a linear relation between X and Y with independent, normally distributed errors and constant variance across all X and Y values. Censored values (below the detection limit) and outliers may invalidate the tests.

Q-3.2.9. If a linear trend is present, based on visual inspection or results from testing for trends, the true slope (change per unit time) may be estimated. An estimate of the magnitude of trend can be obtained by performing a regression of the data versus time (or some function of the data versus some function of time) and using the slope of the regression line that best fits the data as a measure of strength in the trend.

Q-3.3. *Non-parametric Methods*.

Q-3.3.1. *Introduction*. Kendall's tau (Appendix O) can be used to evaluate trends. An alternative method is presented here to use for a single set of observations, $x_1, x_2, \ldots, x_n$, which have been ordered by time of measurement. The test statistic S is calculated by:

$$S = S^+ - S^-$$

where S^+ is the number of pairs (x_i, x_j) with $i < j$ and $x_i < x_j$. Likewise, S^- is the number of pairs (x_i, x_j) with $i < j$ and $x_i > x_j$.

Q-3.3.1.1. It can be shown that there are a total of $n(n-1)/2$ possible pairwise comparisons for a set of n pairs (x_i, x_j). The sample statistic Kendal1's tau, τ , is:

$$\tau = \frac{S}{n(n-1)/2}$$

Note that differences of zero are not included in the test statistic (and should be avoided, if possible, by recording data to sufficient accuracy). However, an adjustment for ties may be made (i.e., when many ties occur), for a series of measurements $x_1, x_2, \ldots, x_n$ performed sequentially in time, by calculating Kendall's tau-b, τ_b:

$$\tau_b = \frac{S}{\sqrt{\left(\frac{n(n-1)}{2} - n'_X\right)\left(\frac{n(n-1)}{2}\right)}}\ .$$

The quantity n'_X is the number of tied pairs (x_i, x_j), where $j > i$, for $i = 1, 2, \ldots n$. The tie adjustment increases the magnitude of Kendall's tau and is useful for evaluating trends (or correlation) when measurements are censored.

Q-3.3.1.2. The Mann-Kendall test does not assume any particular data distribution and accommodates censored values. Non-detected results should be assigned a value smaller than the lowest measured value when the detection limit is small. Otherwise, when calculating *S*, pairs of results such as (3, <10), (<3, <10), and (<3, <3) should be considered to be ties and assigned a value of zero. For example, for the set of n = 4 sequential measurements {30, <10, < 20, <25}, the number of tied pairs $n'_X = 3$ for the calculation of τ_b: (<10, <20), (<10, <25), and (<20, <25). As the test only depends upon signs of differences between data points (or the ranks), information about magnitude of these differences is not used. As such, the test possesses less power than its parametric counterpart, Pearson's *r* (i.e., a larger number of data points are required to identify a correlation using Kendall's tau). However, Mann-Kendall is advantageous because assumptions about the underlying data distribution are not required, and it is more robust (i.e., insensitive) than a parametric test to outliers and censored values. Kendall's tau is also invariant with respect to monotonic transformations of the variable *X*. For example, the value of τ calculated for *X* will be identical to that calculated for *Ln*(*X*).

Q-3.3.1.3. Conducting the Mann-Kendall test for small sample sizes is appropriate for data with fewer than 40 samples (Gilbert, 1987); the EPA suggests using this method with data sets having fewer than 10 samples. If the number of samples becomes too large, the calculations become cumbersome by hand. Directions for the Mann-Kendall trend test for a small sample size (less than 10 samples) are presented in Paragraph Q-3.3.2, followed by an example in Paragraph Q-3.3.3.

Q-3.3.1.4. The Mann-Kendall test is essentially a significance test under the hypothesis γ = 0 (refer to Appendix O). A trend exists if the sample statistic τ is significantly different from zero at some specified level of confidence. If there is an underlying upward trend, the differences will tend to be positive (S will be a large value), so a sufficiently large positive value of the sample statistic τ (e.g., a value near 1) suggests an upward trend. Conversely, if the differences tend to be negative (S will be a large negative value), a sufficiently large negative value of τ (e.g., a value near –1) suggests a downward trend. If the statistic τ is nearly zero (i.e., not significantly different from zero), there is no evidence of a trend. The slope of the time-ordered data plotted versus time is zero. The significance test for $\gamma = 0$ is a nonparametric test for zero slope (Gilbert, 1987). For a two-sided test the null and alternative hypotheses are:

$H_0 : \gamma = 0$: No upward or downward trend.

$H_A : \gamma \neq 0$: An upward or downward trend.

For a one-sided test

$H_0 : \gamma \leq 0$ (or $\gamma \geq 0$): No upward (or no downward trend).

$H_A : \gamma > 0$ (or $\gamma < 0$): An upward trend (or a downward trend).

Q-3.3.1.5. In practice, it is not convenient to calculate a value of τ for the data set and to compare this to a critical value of τ for the desired level of significance, τ_p (so that, for example, if $\tau > \tau_p$, there is an increasing trend at the p100% level of confidence). The calculations for the Mann-Kendall test are done differently for large versus small data sets. For small data sets (Paragraph Q-3.3.2), the value of S for the data set (rather than τ) is calculated and compared to a critical value of S taken from a statistical table. For large data sets, the standard normal distribution is used to determine the statistical significance of τ (Paragraph Q-3.3.4).

Q-3.3.1.6. Note that irregularly spaced measurement periods are permitted with the Mann-Kendall test (Gibbons, 1994). The test can also be modified to deal with multiple observations per time period and generalized to deal with multiple sampling locations and seasonality (Gilbert, 1987). The Mann-Kendall test for the situation in which one observation per time period is taken from one sampling location (e.g., groundwater monitoring well) is presented in Paragraph Q-3.3.2.

Q-3.3.1.7. For large sample sizes, the normal approximation to the Mann-Kendall test is used. If there are more than 10 samples, as long as there are not many tied data values, Gilbert (1987) suggests this normal approximation is quite accurate. Directions for this approximation are provided in Paragraph Q-3.3.2.4, followed by an example in Paragraph Q-3.3.2.5. Tied observations (when two or more measurements are equal) degrade the statistical power and should be avoided, if possible, by recording the data to sufficient accuracy. If the sample size is 10 or more, a normal approximation to the Mann-Kendall procedure may be used.

Q-3.3.2. *Directions for the Mann-Kendall Trend Test for a Small Sample Size*. List the data in the order collected over time: $x_1, x_2, \ldots, x_n$ where x_i is the datum at time t_i.

Q-3.3.2.1. Assign a proxy value to values reported as below the detection limit (DL). Note that this proxy value should be less than any measured value. Construct a Data Matrix similar to the top half of the Table Q-3.

Q-3.3.2.2. Determine the sign for each possible difference and compute the Mann-Kendall statistic, S, which is the number of positive signs minus the number of negative signs in the triangular table: $S = S^+$ (i.e., total number of + signs) – S^- (i.e., total number of – signs).

Q-3.3.2.3. Use Table B-10 of Appendix B to determine the probability (p) using the sample size (n) and the absolute value of the statistic S if $n \leq 10$.

Q-3.3.2.3.1. For testing H_0, no trend against H_A: upward trend, reject H_0 if $S > 0$ and $p < \alpha$.

Q-3.3.2.3.2. For testing H_0, no trend against H_A: downward trend, reject H_0 if $S < 0$ and $p < \alpha$.

Q-3.3.2.4. Table Q-3 presents the resulting matrix of differences when applying the steps above.

Table Q-3.
Basic Mann-Kendall Trend Test with a Single Measurement at Each Time Point

Time x_i	t_2 x_2	t_3 x_3	t_4 x_4	. . .	t_{n-1} x_{n-1}	t_n x_n	**No. of Differences > 0**	**No. of Differences < 0**
x_1	$x_2 - x_1$	$x_3 - x_1$	$X_4 - x_1$	. . .	$x_{n-1} - x_1$	$x_n - x_1$		
x_2		$x_3 - x_2$	$X_4 - x_2$	. . .	$x_{n-1} - x_2$	$x_n - x_2$		
					.	.		
					.	.		
.					.	.		
x_{n-2}					$x_{n-1} - x_{n-2}$	$x_n - x_{n-2}$		
x_{n-1}						$x_n - x_{n-1}$		
Total							(S^+)	(S^-)

Q-3.3.2.5. The number of positive and negative differences are recorded for each row (two right most columns) and the values (within the two right most columns) are summed to obtain S^+ and S^-. Differences equal to zero are ignored.

Q-3.3.3. *Example of a Mann-Kendall Trend Test for Small Sample Sizes (n < 10).* Evaluate the linear trend of benzene taken from quarterly groundwater samples at well MW01 in Site A from 2000–2001.

Q-3.3.3.1. Benzene has been detected during all of these sampling events, so no proxy concentrations were derived. At the 90% level of confidence ($\alpha = 0.10$), test:

H_0: No trend; H_A: Downward trend.

Q-3.3.3.2. Figure Q-4 is a plot of the concentrations over time. It does appear that a downward trend is present. This test, though, will identify if a statistically significant trend is present (Table Q-4).

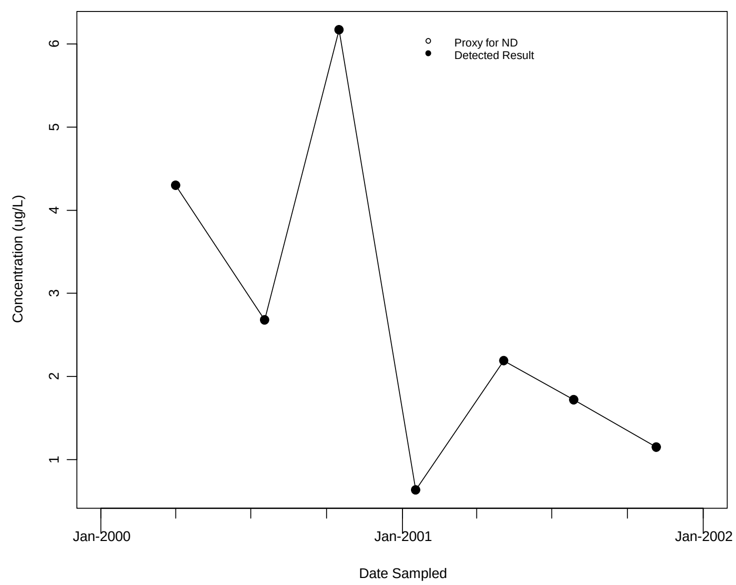

Figure Q-4. Trend for benzene in groundwater (small sample size).

Q-3.3.3.3. The Mann-Kendall test statistic, $S = 5 - 16 = -11$.

Q-3.3.3.4. Using Table B-10 of Appendix B, the p value for $n = 7$ and $|S| = 11$ is $p = 0.068$.

Q-3.3.3.5. As $S < 0$ and $p < \alpha = 0.10$, we reject H_0 and conclude there is significant evidence of a downward trend.

Q-3.3.4. *Directions for a Normal Approximation to the Mann-Kendall Test Procedure.* List the data in the order collected over time. Assign a proxy value to values reported as below the DL. Note that this proxy value should be lower than any measured value. Construct a Data Matrix similar to the top half of the data table below (Table Q-5).

Table Q-4.
"Upper Triangular" Data for Basic Mann-Kendall Trend Test with a Single Measurement at Each Time Point—Data Table

Time	7/00	10/00	1/01	5/01	7/01	11/01	**No. of Differences > 0**	**No. of Differences < 0**
x_i	2.68	6.17	0.64	2.19	1.72	1.15		
$x_1 = 4.3$	–1.62	1.87	–3.66	–2.11	–2.58	–3.15	1	5
$x_2 = 2.68$		3.49	–2.04	–0.49	–0.96	–1.53	1	4
$x_3 = 6.17$			–5.53	–3.98	–4.45	–5.02	0	4
$x_4 = 0.64$				1.55	1.08	0.51	3	0
$x_5 = 2.19$					–0.47	–1.04	0	2
$x_6 = 1.72$						–0.57	0	1
Total							***5***	***16***

Table Q-5.
Data for Example Q-3.3.5

Jun-98	**Apr-98**	**Jul-98**	**Oct-98**	**Apr-99**	**Jul-99**	**Oct-99**	**Apr-00**	**Jul-00**	**Oct-00**	**Jan-01**	**May-01**	**Jul-01**	**Nov-01**	**(Time: earliest to latest)**	
														Benzene concentrations	
x_i	3.79	3.42	5.47	0.81	1.78	7.56	4.3	2.68	6.17	0.64	2.19	1.78	1.15	#of + Diff.	#of – Diff.
12.2	–8.41	–8.78	–6.73	–11.4	–10.4	–4.6	–7.9	–9.52	–6.0	–11.6	–10.0	–10.4	–11.1	0	13
3.79		–0.37	1.68	–2.98	–2.01	3.77	0.51	–1.11	2.38	–3.15	–1.6	–2.01	–2.64	4	8
3.42			2.05	–2.61	–1.64	4.14	0.88	–0.74	2.75	–2.78	–1.23	–1.64	–2.27	4	7
5.47				–4.66	–3.69	2.09	–1.17	–2.79	0.7	–4.83	–3.28	–3.69	–4.32	2	8
0.81					0.97	6.75	3.49	1.87	5.36	–0.17	1.38	0.97	0.34	8	1
1.78						5.78	2.52	0.90	4.39	–1.14	0.41	0.00	–0.63	5	2
7.56							–3.26	–4.88	–1.39	–6.92	–5.37	–5.78	–6.41	0	7
4.3								–1.62	1.87	–3.66	–2.11	–2.52	–3.15	1	5
2.68									3.49	–2.04	–0.49	–0.90	–1.53	1	4
6.17										–5.53	–3.98	–4.39	–5.02	0	4
0.64											1.55	1.14	0.51	3	0
2.19												–0.41	–1.04	0	2
1.78													–0.63	0	1
1.15												*Total*		***28***	***62***

Q-3.3.4.1. Compute the sign of all possible differences as shown in the bottom portion of Table Q-5.

Q-3.3.4.2. Compute the Mann-Kendall statistic, *S*, as shown in Paragraph Q-3.3.2. *S* is the number of positive signs minus the number of negative signs in the triangular table: $S = S^{+} - S^{-}$.

Q-3.3.4.3. If there are no ties, calculate the variance of *S*:

$$V(S) = \frac{n(n-1)(2n+5)}{18}.$$

Q-3.3.4.4. If ties occur, let g represent the number of tied groups and w_j represent the number of data points in the j^{th} tied group. For ties, the variance of S is:

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} w_j(w_j-1)(2w_j+5) \right].$$

Q-3.3.4.5. Calculate the following statistic:

$$z = \begin{cases} \dfrac{S-1}{\sqrt{V(S)}}, & S > 0 \\ 0, & S = 0 \\ \dfrac{S+1}{\sqrt{V(S)}}, & S < 0 \end{cases}.$$

Q-3.3.4.6. Note that tied values do not affect the calculation of S but affect only $V(S)$ and the calculation of z using the large sample approximation.

Q-3.3.4.7. Use Table B-15 of Appendix B to find the critical value $Z_{1-\alpha}$ (if testing H_0: No trend against H_A: Upward trend) or the critical value $-Z_{1-\alpha}$ (if testing H_0, no trend against H_A: downward trend) such that (1 – α)100% of the normal distribution lies to the left of $Z_{1-\alpha}$.

Q-3.3.4.7.1. For testing H_0, no trend against H_A: upward trend, reject H_0 if $z > Z_{1-\alpha}$.

Q-3.3.4.7.2. For testing H_0, no trend against H_A: downward trend, reject H_0 if $z < -Z_{1-\alpha}$.

Q-3.3.5. *Example of The Mann-Kendall Procedure Using Normal Approximation for Larger Samples*. Consider evaluating whether or not there is a significant trend for benzene using a set of samples taken from quarterly groundwater samples at well MW01 in Site A from 1998–2001. Benzene has been detected during all of these sampling events, so no proxy concentrations were derived.

Q-3.3.5.1. Test H_0, no trend against H_A: downward trend based on a 90% level of confidence ($\alpha = 0.10$).

Q-3.3.5.2. Figure Q-5 is a plot of the concentrations over time. It does appear that a downward trend is present (Table Q-5).

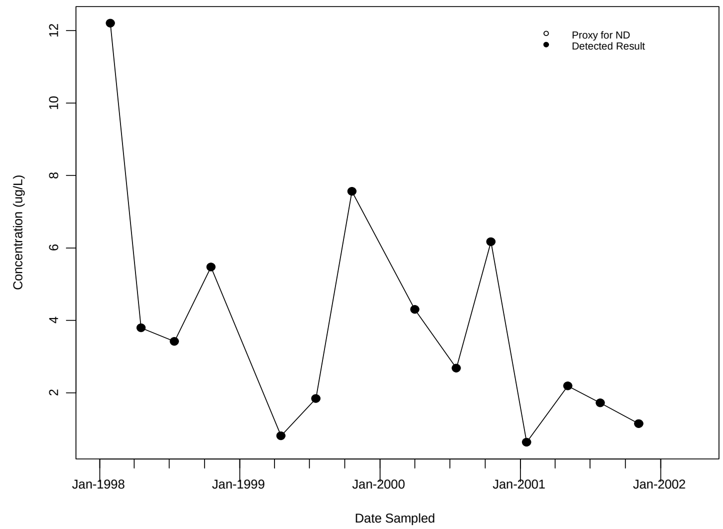

Figure Q-5. Trend for benzene in groundwater (large sample size).

Q-3.3.5.3. The Mann-Kendall statistic, $S = 28 - 62 = -34$.

Q-3.3.5.4. Since there are two observations with a value of 1.78, there are $g = 1$ tied groups and $w_1 = 2$. The calculated variance of S is

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{j=1}^{g} w_j \left(w_j - 1\right)\left(2w_j + 5\right)}{18}$$

$$= \frac{14(13)(33) - 2(1)(9)}{18} = 332.7 \ .$$

Q-3.3.5.5. Because $S < 0$, the approximate z test statistic is

$$z = \frac{S+1}{\sqrt{V(S)}} = \frac{-34+1}{\sqrt{332.7}} = -1.809.$$

Q-3.3.5.6. Using Table B-15 of Appendix B, find the critical value $-Z_{0.90} = -1.28$.

$-1.809 < -1.28$, so we can reject H_0.

Q-3.3.5.7. That means there is significant evidence of a downward trend.

Q-3.3.6. *Multiple Observations.* Often, more than one sample is collected for each time period. There are two ways to deal with such multiple observations. One method is to compute a summary statistic, such as the median, for each time period and to apply one of the Mann-Kendall trend tests to the summary statistic. The summary statistic would be used instead of the individual data points in the triangular table. The steps given for the Mann-Kendall for small sample sizes or larger samples could then be applied to the summary statistics.

Q-3.3.6.1. An alternative approach is to consider all of the multiple observations within a given time period as being essentially equal (tied) values within that period. The S statistic is computed as before, with n being the total of all observations. The variance of the S statistic is changed to:

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} w_j(w_j-1)(2w_j+5) - \sum_{k=1}^{h} u_k(u_k-1)(2u_k+5)\right] + \frac{\sum_{j=1}^{g} w_j(w_j-1)(w_j-2)\sum_{k=1}^{h} u_k(u_k-1)(u_k-2)}{9n(n-1)(n-2)} + \frac{\sum_{j=1}^{g} w_j(w_j-1)\sum_{k=1}^{h} u_k(u_k-1)}{2n(n-1)}$$

where g represents the number of tied groups (i.e., number of groups that have tied observations), w_j represents the number of data points in the tied j^{th} group, h is the number of time periods that contain multiple data, and u_k is the sample size in the k^{th} time period where $k = 1, 2, \ldots, h$. For example, let four X measurements be made for the first time period, three for the second, two for the third, and one for each of the subsequent time periods. The value of h will be 3 for the three time periods with multiple measurements, and the value of u_k will be 4, 3, and 2 for k = 1, 2, and 3 respectively. The values of g and w_j will depend on actual X measurements. For the special case of ties and multiple measurements for a time period, the reader is referred to Gilbert (1987).

Q-3.3.6.2. The preceding variance formula assumes that the data are not correlated. If correlation within single time periods is suspected, it is preferable to use a summary statistic (i.e., the median) for each period and then apply either the Mann-Kendall for small sample sizes or larger samples to the summary statistics.

Q-3.3.6.3. The preceding methods involve a single sampling location (station). However, environmental data often consist of sets of data collected at several sampling locations (e.g., groundwater monitoring wells). For example, data are often systematically collected at several fixed sites on a lake or river, or within a region or basin. The data collection plan (or experimental design) must be systematic in the sense that approximately the same sampling times should be used at all locations. In this situation, it may be desirable to simultaneously evaluate all of the sampling locations for the presence of a common characteristic or "regional trend." However, there must be consistency in behavioral characteristics across sites over time for a single summary statement to be valid across all sampling locations. A useful plot to assess the consistency requirement is a single time plot of the measurements from all stations in which a different symbol is used to represent each station. Paragraph Q-3.3.7 illustrates such data sets.

Q-3.3.6.4. If the stations exhibit approximately steady trends in the same direction (upward or downward), with comparable slopes, a single summary statement across stations is valid, implying that two relevant sets of hypotheses should be investigated.

Q-3.3.6.4.1. *Comparability of Stations.*

H_0: The trends at all K stations are homogeneous.

H_A: At least two stations exhibit different dynamics.

Q-3.3.6.4.2. *Testing for Overall Monotonic Trend.*

H_0^* : Contaminant levels do not change over time.

H_A^* : There is an increasing or decreasing trend consistently exhibited across all stations.

Q-3.3.6.5. Therefore, the analyst must first test for homogeneity of stations and then, if homogeneity is confirmed, test for an overall monotonic trend.

Q-3.3.6.6. Ideally, the stations should have equal numbers. However, the numbers of observations at the stations can differ slightly because of isolated missing values, but the overall time periods spanned must be similar. The EPA recommends that an equal number of observations (a balanced design) be required for fewer than three time periods. For four or more time periods, up to one missing value per sampling location may be tolerated.

Q-3.3.6.7. When only one measurement is taken for each time period for each station, a generalization of the Mann-Kendall statistic can be used to test the above hypotheses. Directions for this condition are presented in Paragraph Q-3.3.8, followed by an example in Paragraph Q-3.3.9.

Q-3.3.6.8. Gilbert (1987) states: "The validity of these chi-squared tests depends on each of the z_k values having a standard normal distribution. [T]his implies that the number of data (over time) for each station should exceed 10. Also, the validity of the tests requires that the z_k values be independent, meaning data from different stations must be uncorrelated."

Q-3.3.6.9. If multiple measurements are taken at some time and station, the previous approaches are still applicable. However, the variance of the statistic S_k must be calculated using the equation for calculating $V(S)$ based on multiple observations within a given time period. Note that S_k is computed for each station, so n, w_j, g, h, and u_k are all station-specific.

Q-3.3.7. *Illustration of Data Taken from Multiple Stations and Multiple Times*. Let i = 1, 2,..., n represent time, let k = 1, 2,..., K represent sampling locations or stations, and $x_{i,k}$ represent the measurement at time i for location k. These data can be summarized in matrix form, as shown below:

		Station			
		1	2	…	K
Time	1	$x_{1,1}$	$x_{2,1}$	…	$x_{K,1}$
	2	$x_{1,2}$	$x_{2,2}$	…	$x_{K,2}$
	.	.	.	…	.
	.	.	.	…	.
	.	.	.	…	.
	n	$x_{1,n}$	$x_{2,n}$	…	$x_{K,n}$
		S_1	S_2	…	S_K
		$V(S_1)$	$V(S_2)$	…	$V(S_K)$
		z_1	z_2	…	z_K

where

$S_{k,}$ = Mann-Kendall statistic for station k

$V(S_k)$ = variance for S statistic for station k

z_k = $S_k \big/ \sqrt{V(S_k)}$.

Q-3.3.8. *Directions for the Mann-Kendall Statistic Used to Test a Monotonic Trend*. Let i = 1, 2,..., n represent time, k = 1, 2,..., K represent sampling locations or stations, and $x_{i,k}$ represent the measurement at time i for location k. Let α represent the significance level for testing homogeneity and α^* represent the significance level for testing an overall trend.

Q-3.3.8.1. Calculate the Mann-Kendall statistic S_k and its variance $V(S_k)$ for each of the K stations using the methods for larger sample sizes.

Q-3.3.8.2 For each of the K stations, calculate

$$z_k = S_k \big/ \sqrt{V(S_k)} \ .$$

Q-3.3.8.3 Calculate the average

$$\bar{z} = \sum_{k=1}^{K} z_k / K .$$

Q-3.3.8.4. Calculate the homogeneity chi-square statistic

$$\chi_h^2 = \sum_{k=1}^{K} z_k^2 - (K \bar{z}^2) .$$

Q-3.3.8.5. Using a chi-squared table, find the critical value, $\chi^2_{1-\alpha,\nu}$ the (1 – α)100th percentile of the chi-squared distribution with $\nu = K - 1$ degrees of freedom.

Q-3.3.8.5.1. If $\chi_h^2 > \chi^2_{1-\alpha,\nu}$, the stations are not homogeneous (have different dynamics at different stations) at the significance level α. Therefore, individual α*-level Mann-Kendall tests should be conducted at each station using the methods presented previously. That is, test each of the *K* wells individually as described in Paragraphs Q-3.3.3 or Q-3.3.5.

Q-3.3.8.5.2. If $\chi_h^2 \le \chi^2_{1-\alpha,\nu}$, there are comparable dynamics across stations at significance level α. Using a chi-squared table, find the critical value for the chi-squared distribution with 1 degree of freedom at the $\alpha *$ significance level, $\chi^2_{1-\alpha^*,1}$.

Q-3.3.8.6. If $K \bar{z}^2 > \chi^2_{1-\alpha^*,1}$, then reject H_0^* and conclude that there is a significant (upward or downward) monotonic trend across all stations at significance level $\alpha *$. The signs of the S_k indicate whether increasing or decreasing trends are present.

Q-3.3.8.7. If $K \bar{z}^2 \le \chi^2_{1-\alpha^*,1}$, there is not significant evidence at the $\alpha *$ level of a monotonic trend across all stations; that is, the stations appear approximately stable over time.

Q-3.3.9. *Example of Comparability of Stations and an Overall Monotonic Trend*. The following wells at Site A are to be evaluated to determine if the benzene concentrations show decreasing trends consistently across these wells based on a 95% level of confidence. Data for benzene at these wells are shown in the Table Q-6. The flag "ND" is applied to sample for which benzene was not detected. For non-detected concentrations, proxy values are presented in the table and are set to the sample's detection limit.

Q-3.3.9.1. For this example, $K = 3$.

Q-3.3.9.2. The average of the *z* values is

$\bar{z} = (-1.916 - 1.040 + 2.135)/3 = -0.2737$.

Q-3.3.9.3. The homogeneity chi-square statistic is

$$\chi_h^2 = \sum_{k=1}^{K} z_k^2 - (K\,\bar{z}^2) = \left[(-1.916)^2 + (-1.040)^2 + (2.135)^2\right] - 3(-0.2737)^2 = 9.086 .$$

Q-3.3.9.4. The critical value is $\chi^2_{0.95,2} = 5.991$, with $\nu = K - 1 = 2$ degrees of freedom and 95% level of confidence (from Table B-2 of Appendix B).

Q-3.3.9.5. Because $\chi_h^2 \geq \chi^2_{0.95,2}$, the stations are not homogeneous based on a 95% level of confidence, and each should be tested using the technique presented in Paragraph Q-3.3.5 as $n >$ 10.

Table Q-6.
Benzene Data for Example Q-3.3.9

Time	Well (Site A)					
	MW01		**MW03**		**MW05**	
1	12.2		0.062	ND	2.17	
2	3.79		1.78		2.75	
3	3.42		0.04	ND	6.91	
4	5.47		2.31		8.64	
5	0.81		7.24		11.0	
6	1.84		1.85		14.1	
7	7.56		0.31		3.45	
8	4.30		2.00		36.7	
9	2.68		0.14		20.2	
10	6.17		0.23		8.34	
11	0.64		0.065	ND	17.0	
12	2.19		0.76		21.8	
13	1.72		0.22		2.01	
14	1.15		0.05	ND	29.1	
S_k	–35		–19		39	
$V(S_k)$	333.7		333.7		333.7	
z_k	–1.916		–1.040		2.135	

Q-3.3.10. *Multiple Observations over Extended Time Periods*. Temporal data are often collected over extended time periods. Within the time variable, data may exhibit periodic cycles, patterns in the data that repeat over time. For example, temperature and humidity may change with the season or month and affect environmental measurements. For this discussion, the term “season” represents one time point in the periodic cycle, such as a month within a year or an hour within a day. There are two approaches for testing for trends—the seasonal Kendall test and Sen’s test for trends—if seasonal cycles are anticipated. The seasonal Kendall test may be used for large sample sizes, and Sen's test for trends may be used for small sample sizes. In either case, the data are analyzed separately by season, and the results are compared among seasons. Both of these estimation techniques are described below. If different seasons manifest similar slopes (rates of change) but different intercepts, the Mann-Kendall technique for multiple sampling locations with multiple observations is applicable, replacing station by season. For example, Figure Q-6 shows a time plot of a series that appears to be decreasing although it is somewhat masked by a seasonal cycle that repeats every four time periods. The data could be analyzed by the Mann-Kendall technique presented in Paragraph Q-3.3.8 if they are broken out by season (e.g., data points 1, 5, 9, 13, and 17 would constitute one season series).

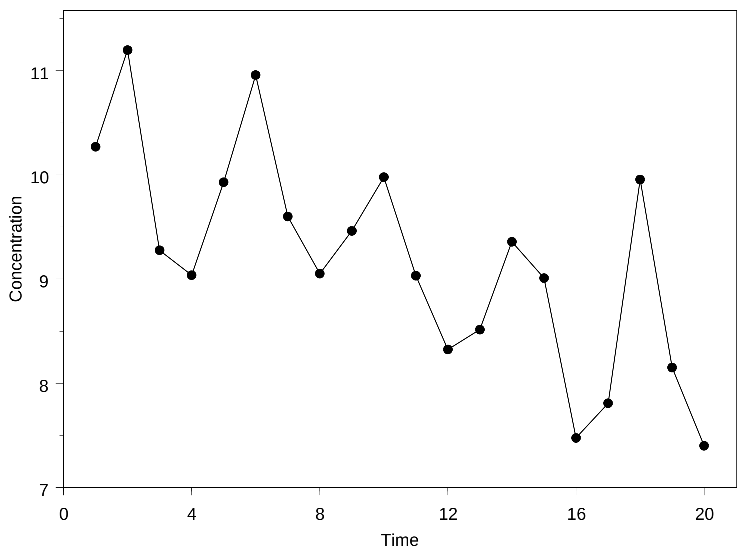

Figure Q-6. Time plot of seasonal series with decreasing trend.

Q-3.3.10.1. For data with seasonality, the seasonal Kendall test, an extension of the Mann-Kendall test, involves calculating the Mann-Kendall test statistic, S, and its variance separately for each "season" (e.g., month of the year, day of the week). The sum of the S's and the sum of their variances are then used to form an overall test statistic that is assumed to be approximately normally distributed for larger size samples.

Q-3.3.10.2. For data at a single site, collected at multiple seasons within multiple years, the techniques for multiple sampling locations with multiple observations can be used to test for homogeneity of time trends across seasons. The methodology follows the explanation below of Sen's slope estimator exactly, except "station" is replaced by "season" and the inferences refer to seasons.

Q-3.3.10.3. If a linear trend is observed when some variable of interest is plotted against time, based on a visual inspection or the results of a statistical test for a trend, the magnitude of the slope of the line is a measure of the “strength” of the trend and the sign of the slope provides the direction of the trend. The true slope (change per unit time) may be estimated using a parametric or non-parametric method. Linear regression analysis is a parametric method for estimating a slope. Sen’s slope estimator is a non-parametric method for estimating the slope of a line.

Q-3.3.10.4. This approach involves computing slopes for all pairs of ordinal time points and using the median of these slopes as an estimate of the overall slope. As such, it is insensitive to outliers and can handle a moderate number of values below the detection limit and missing values.

Q-3.3.10.5. Directions are presented in Paragraph Q-3.3.11, followed by an example in Paragraph Q-3.3.12.

Q-3.3.11. *Directions for a Sen’s Slope Estimator*. Assume that there are n time points (or n periods of time), and let x_i denote the data value for the i^{th} time point. If there are no missing data, there will be $N' = n(n-1)/2$ possible pairs of time points (i, j), in which $i > j$ (i.e., x_i was taken at a time after the measurement x_j).

Q-3.3.11.1. For non-detected results, the detection limit may be used as the data value (Gibbons, 1994) or one-half the detection limit may be used as the data value (Gilbert, 1987). Note that this proxy value should be lower than any measured value.

Q-3.3.11.2. Define the slope for each pair, called a pairwise slope, as

$$b_{ij} = \frac{(x_i - x_j)}{(i - j)}.$$

Q-3.3.11.3. Sen’s slope estimator is the median of the $n(n-1)/2$ pairwise slopes.

Q-3.3.12. *Example of a Sen’s Slope Estimator.* The Sen’s slope estimate is calculated to evaluate the linear trend for benzene in Paragraph Q-3.3.3 (seven groundwater samples collected quarterly from 2000–2001 from well MW01 at Site A). Because benzene was detected for all the sampling events, proxy concentrations were not derived.

Q-3.3.12.1. There are 7(6)/2 = 21 possible pairs of time points (i, j) in which $i > j$. The slope for each pair will be estimated and displayed in a data matrix similar to the one presented in Paragraph Q-3.3.3, except each cell in the matrix represents the pairwise slope

$$b_{ij} = \frac{(x_i - x_j)}{(i - j)}.$$

Q-3.3.12.2. If there is no underlying trend, then a given x_i is just as likely to be above another x_j as it is to be below. If there is no underlying trend, there would be an approximately equal number of positive and negative slopes and Sen's slope would be near zero.

Q-3.3.12.3. If the data exhibit cyclic trends, the Sen's slope estimator can be modified to account for the cycles. For example, if data are available for each month for a number of years and the length of a cycle is one year, 12 separate sets of slopes would be determined (one for each month of the year using all of the data for that particular month); similarly, if daily observations exhibit weekly cycles, seven sets of slopes would be determined, one for each day of the week. In these estimates, the above pairwise slope is calculated for each time period and the median of all of the slopes is an estimator of the slope for a long-term trend. This is known as the seasonal Kendall slope estimator, which is rarely calculated by hand owing to the number of calculations required.

Table Q-6.
Pairwise Slopes Data Table

Original Time Measure	t_1=4/00 x_1=4.3	t_2=7/00 x_2=2.68	t_3=10/00 x_3=6.17	t_4=1/001 x_4=0.64	t_5=5/01 x_5=2.19	t_6=7/01 x_6=1.72	t_7=11/01 x_7=1.15
x_1=4.3		–1.62	0.935	–1.22	–0.528	–0.516	–0.525
x_2=2.68			3.49	–1.02	–0.163	–0.24	–0.306
x_3=6.17				–5.53	–1.99	–1.483	–1.255
x_4=0.64					1.55	0.54	0.17
x_5=2.19						–0.47	–0.52
x_6=1.72							–0.57
x_7=1.15							

Ordered pairwise slopes (smallest to largest):	–5.53	–1.99	–1.62	–1.483	–1.255	–1.22	–1.02
	–0.57	–0.528	–0.525	–0.52	–0.516	–0.47	–0.306
	–0.24	–0.163	0.17	0.54	0.935	1.55	3.49

Q-3.3.12.4. The median of these 21 pairwise slopes is –0.52, the 11th ordered result when the results are sorted from smallest to largest.

Q-3.3.13. *Testing a Trend Using Confidence Limits for Sen's Slope Estimator.* Gilbert (1987) presents a simple method, based on the normal distribution, to estimate the $(1 - \alpha)100\%$ confidence interval about the true slope. This "large sample" estimate is appropriate for data sets with at least 10 samples. Directions for estimating such confidence intervals are presented below. Aside from estimating the confidence limits for the slope associated with a trend that has been previously identified (e.g., using Mann-Kendall's test), this approach can be used to determine if a trend is presented. If the confidence interval for the slope contains zero, there is no evidence of an underlying trend. However, if the confidence interval does not contain zero, there is evidence to suggest a trend. Directions are presented in Paragraph Q-3.3.14, followed by an example in Paragraph Q-3.3.15.

Q-3.3.14. *Directions for Creating Confidence Limits for Sen's Slope Estimator*. Compute $N' = n(n-1)/2$ if there is just one result in each time period, and $N' =$ the number of possible data pair combinations among the time periods (and results from the time period cannot be considered data pairs) if there is more than one result in each time period.

Q-3.3.14.1. Based on the desired two-sided confidence level (1 – α)100%, find $Z_{1-\alpha/2}$.

Q-3.3.14.2. Compute the variance of S as

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} w_j(w_j-1)(2w_j+5) \right]$$

when one observation per time period is available (g represents the number of tied groups and w_j represent the number of data points in the j^{th} group) or

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} w_j(w_j-1)(2w_j+5) - \sum_{k=1}^{h} u_k(u_k-1)(2u_k+5) \right] + \frac{\sum_{j=1}^{g} w_j(w_j-1)(w_j-2)\sum_{k=1}^{h} u_k(u_k-1)(u_k-2)}{9n(n-1)(n-2)} + \frac{\sum_{j=1}^{g} w_j(w_j-1)\sum_{k=1}^{h} u_k(u_k-1)}{2n(n-1)}$$

when multiple observations per time period are available (g represents the number of tied groups, w_j represents the number of data points in the j^{th} group, h is the number of time periods containing multiple data, and u_k is the sample size in the k^{th} time period).

Q-3.3.14.3. Compute $C_\alpha = Z_{1-\alpha/2}\sqrt{V(S)}$.

Q-3.3.14.4. Compute $M_1 = (N' - C_\alpha)/2$ and $M_2 = (N' + C_\alpha)/2$.

Q-3.3.14.5 The lower and upper limits of the confidence interval are the M_1^{th} largest and $(M_2 + 1)^{\text{th}}$ largest of the N' ordered slope estimates (from lowest to highest), respectively. If M_1 and $M_2 + 1$ are not whole numbers, use linear interpolation (Gilbert, 1987).

Q-3.3.15. *Example of Confidence Limits for Sen's Slope Estimator*. Consider estimating a two-sided 95% confidence interval for Sen's slope estimated in Paragraph Q-3.3.12, where:

$n = 7,\ S = -0.52$ and $N' = n(n-1)/2 = 21$.

Q-3.3.15.1. For $\alpha = 0.05,\ Z_{1-\alpha/2} = Z_{0.975} = 1.96$.

Q-3.3.15.2. The following are calculated:

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} w_j(w_j-1)(2w_j+5)\right] = \frac{1}{18}\left[7(6)(19) - 0\right] = 44.33$$

$$C_\alpha = Z_{1-\alpha/2}\sqrt{V(S)} = 1.96\sqrt{44.33} = 13.05$$

$$M_1 = (N' - C_\alpha)/2 = (21 - 13.05)/2 = 3.975$$

$$M_2 = (N' + C_\alpha)/2 = (21 + 13.05)/2 = 17.025.$$

Q-3.3.15.3. From the list of ordered results in Paragraph Q-3.3.15 , the interpolated value between the 3rd and 4th ordered result is –1.486 and the interpolated value between the 18th (17 + 1) and 19th ordered result is 0.550. Therefore, the confidence interval for the slope is (–1.486, 0.550). As this interval contains zero, there is insufficient evidence of an underlying trend (even though the slope of –0.52 suggests a negative trend).

Q-4. Control Charts.

Q-4.1. *Introduction*. Control charts are a quality control procedure that can be applied to environmental monitoring data, such as data from air or groundwater monitoring systems. Control charts provide a visual means of monitoring constituent concentrations at a given well or location over time, identifying slight or sudden fluctuations over time and detecting deviations from a "state of control." A process is in-control if the observed variation is attributable to small, uncontrollable changes. A process is out-of-control if a relatively large variation is introduced that can be traced to an assignable cause (Kvanli et al., 1996).

Q-4.1.1. Control charts are most frequently used in groundwater monitoring detection programs for intra-well comparisons, in which data are collected for a single well over some period of time. Control charts are useful for areas with no previous contamination because detecting contamination may require a significant change. This is particularly applicable to monitoring down-gradient of waste cells or landfills, because it can highlight whether there has been a release to groundwater. If contamination was historically present, it will take a significant increase in concentrations relative to historical values to show a detection (Gibbons, 1994). Control charts, however, are not constructed for making precise probability statements; they are constructed as a guide for determining when investigative action is needed (Gilbert, 1987). Furthermore, contamination may be present intermittently or may increase in a step function. The absence of an increasing trend does not necessarily support that a release has not occurred.

Q-4.1.2. Control charts are designed for a given constituent and well in which concentrations are plotted against time with horizontal lines called "control limits." Control limits are based on meaningful and sufficient historical data with no outliers and trends over time. As new data become available, those concentrations are also plotted. The EPA recommends, and current RCRA regulations specify, developing control limits with data consisting of at least eight independent samples over a 1-year period. As with most statistical applications, more historical data are desirable but, in practical terms, are rarely available.

Q-4.1.3. The assumptions underlying control charts are that when the process is in-control, data are independent and normally distributed with a fixed mean and constant variance. Independence is crucial. Control charts are not robust with respect to the departure from independence (i.e., when data are correlated). To minimize the possibility that samples are dependent, Gibbons (1994) recommends a sampling frequency of no more than one sample per quarter. To identify serial correlation, a sample's serial correlation coefficient can be calculated. (Details are provided in Appendix O.) A correlogram may be plotted to determine if serial correlation is large enough to create problems. (Details are provided in Appendix J.) A quick method for determining if serial correlation is large is to compare the autocorrelation coefficients to

$$\pm 2 / \sqrt{n}$$

where n is the number of time periods when data were collected. Autocorrelation coefficients that exceed either of these values require further investigation.

Q-4.1.4. The assumption of normality is not nearly as crucial, but the data's distribution should still be investigated. To achieve normality, data transformations (such as natural-log transformations or square-root transformations) should be applied to sample data, as appropriate. Gilbert (1987) suggests that as long as data are normally distributed and the correlation associated with the data is not too large, control chart methods work well. Gilbert goes on to say that although environmental data are typically non-normal, control charts are still useful for indicating where concentrations are not likely to be from the same distribution as in the past.

Q-4.1.5. Seasonality, a component of the data's variability, should also be considered before control charts are developed. Seasonality can be addressed by removing seasonal effects from the data, if sufficient data are available for at least two seasons of the same type. Removing seasonality was previously discussed in Paragraph Q-2. Gilbert (1987) recommends two other methods to circumvent seasonality issues. If data are available for a number of complete cycles, separate control charts for each season can be prepared. If the data do not span a long duration and the magnitude of the cycles is relatively small, a moving-average control chart may be constructed.

Q-4.1.6. In terms of proxy concentrations appropriate for control charts, Gibbons (1994) suggests that if at least 25% of samples are detections, a proxy concentration based on just the sample-specific method detection limit is adequate for control charts.

Q-4.1.7. Several types of control charts are discussed in this section: Shewart control charts, CUSUM control charts, and Shewart-CUSUM control charts. The advantage to Shewart control charts is that they are immediately sensitive to large changes. The advantage to CUSUM control charts is that they are sensitive to small and gradual changes. Shewart-CUSUM control charts are a combination of the other two. As such, their benefit is that they can detect both sudden and gradual changes in concentrations.

Q-4.2. *Shewart Control Charts*.

Q-4.2.1. *Introduction*. Shewart control charts, which are the oldest and simplest charts (Gibbons, 1994), are sensitive to sudden changes and focus on the current monitoring value. Current data (not historical data) are first plotted against time. Control limits are subsequently placed on the same plot as horizontal lines. The control limits are calculated using historical data from a period of time when the system under study was stable. New data that fall outside of the control limits indicate that current conditions have changed from the historical ones used to establish the control limits. Although lower control limits are used in other fields, only the upper control limit is typically established for environmental data, as the objective is to identify dramatically increasing concentrations. An upper control limit can be developed from historical data using the equation $\mu + Z\sigma$, where μ is the population mean, σ is the population standard deviation, and Z is an upper percentage point of the normal distribution. For this case, Z is typically equal to 3, which corresponds to a confidence level of $1-\alpha = 0.9987$ for a single new comparison.

Q-4.2.1.1. However in most cases, long-run historical data are unavailable and a sample estimate of the mean ($\bar{x}$) and standard deviation (s) must be used. In this case, the equation for the upper control limit is $\bar{x} + Zs$. When using the sample estimates to calculate an upper control limit with as few as eight historical samples, however, the control limit only provides an overall 95% confidence for five new comparisons and the overall confidence decreases as the number of future observations increases (Gibbons, 1994). As such, EPA 530-SW-89-026 recommends setting control limits to $\bar{x} + 4.5s$ for routine groundwater monitoring situations. "Overall confidence levels for this control limit are 95% with n = 8 and 35 future comparisons; however, verification resampling further reduces false positive rates to acceptable levels for most monitoring programs" (Gibbons, 1994), avoiding the problem of multiple comparisons discussed in Appendices M and N. It should be noted that 4.5 is a generic value recommended by the EPA to be protective in most monitoring situations. Gibbons, 1994 warns "[t]he reader should note that unlike prediction limits which provide a fixed confidence level (e.g. 95%) for a given number of future comparisons, control charts do not provide explicit confidence levels, and they do not adjust for the number of future comparisons." See Appendix K for information on developing prediction limits to cover a specific number of future observations and tolerance limits to cover an indefinite number of future observations.

Q-4.2.1.2. If more than eight historical samples are available, it is reasonable to use only the most recent eight. Once a control limit is developed, the current monitoring value is compared to the control limit. If the value exceeds the control limit, the groundwater system should be investigated for causes associated with the increase in concentration. Directions for preparing a Shewart control chart are given in Paragraph Q-4.2.2, followed by an example in Paragraph Q-4.2.3.

Q-4.2.2. *Directions for Preparing a Shewart Control Chart.*

Q-4.2.2.1. Verify the following assumptions:

Q-4.2.2.1.1. For each sampling location (e.g., a well for groundwater monitoring), data are available from at least eight independent samples from previous sampling events to estimate the mean and standard deviation.

Q-4.2.2.1.2. Determine if data are correlated.

Q-4.2.2.1.3. Identify if data or transformed data are normally distributed.

Q-4.2.2.1.4. Check if seasonality is affecting data, and, if so, remove the seasonality.

Q-4.2.2.2. At a given location or well, take independent samples over n historical sampling events ($n \geq 8$).

Q-4.2.2.3. Calculate the mean ($\bar{x}$) and standard deviation (s) of the n samples.

Q-4.2.2.4. Calculate an upper control limit by the equation $\bar{x} + Zs$, where Z is set to 4.5 for routine groundwater monitoring programs. Note that setting $Z = 4.5$ ensures a 95% overall confidence level when $n = 8$ and 35 future comparisons are made to this upper control limit (Gibbons, 1994).

Q-4.2.2.5. Plot the current concentrations with respect to time and superimpose the upper control limit.

Q-4.2.2.6. Identify if the system is in-control or out-of-control by identifying if concentrations are below the upper control limit or above the upper control limit, respectively.

Q-4.2.2.7. Investigate any situation in which a concentration is above the upper control limit.

Q-4.2.3. *Example of a Shewart Control Chart*. Benzene is measured from quarterly groundwater samples at well MW01 in Site A from 1998–2000 to develop a control chart to compare to the 2001 sampling results (Table Q-7).

Q-4.2.3.1. Verifying assumptions are as follows.

Q-4.2.3.1.1. $n = 10$. Samples were taken with at least a 3-month interval; therefore, the samples should be independent.

Q-4.2.3.1.2. This set of data is the same as that used to calculate the serial correlation for the example in Paragraph O-2.6.2. From that example, the following summary statistics were estimated: $\bar{x} = 4.824$ and $s_x = 3.284$, and the serial correlation coefficient = –0.2527. The correlogram for these data is shown in Figure Q-7.

Table Q-7a.
Historical Data for Upper Control Limit in Example Q-4.2.3

Time	Jan-98	Apr-98	Jul-98	Oct-98	Apr-99	Jul-99	Oct-99	Apr-00	Jul-00	Oct-00
Time Period	1	2	3	4	5	6	7	8	9	10
Conc. (µg/L)	12.2	3.79	3.42	5.47	0.81	1.84	7.56	4.32	0.68	6.17

Table Q-7b.
Current Data to Use to Compare to Control Limit in Example Q-4.2.3

Time	Jan-01	May-01	Jul-01	Nov-01
Conc. (µg/L)	0.64	2.19	1.72	1.15

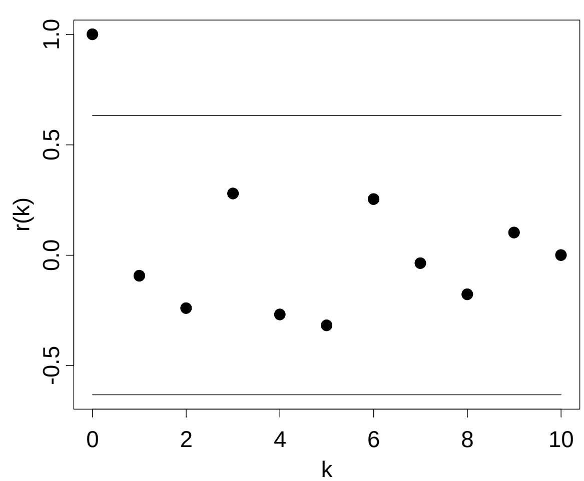

Figure Q-7. The correlogram.

Q-4.2.3.2. Serial correlation does not appear to be a problem, even though the default at $k = 0$ (where k is the autocorrelation coefficient) is greater than the $\pm 2/\sqrt{n}$ bounds (± 0.632).

Q-4.2.3.2.1. To test the assumption of normality, the Shapiro-Wilk test was performed with the data based on a 95% level of confidence. Results of this test provide evidence to suggest that the data follow a normal distribution because the p value is 0.3363, which is greater than the significance level of $\alpha = 0.05$ (there is not enough evidence to reject the null hypothesis of normality).

Q-4.2.3.2.2. There are not enough results to adequately identify seasonal trends and no obvious trend is visible in the previous time plot. For this example, we will assume that the data are not affected by seasonality.

Q-4.2.3.3. There are not enough results to adequately identify seasonal trends and no obvious trend is visible in the previous time plot. For this example, we will assume that the data are not affected by seasonality. Calculate the control limit as follows.

Q-4.2.3.4. The upper control limit $= \bar{x} + Z\,s = 4.824 + (4.5 \times 3.284) = 19.06$. None of the samples taken in 2001 exceeds this upper control limit, as shown in Figure Q-8.

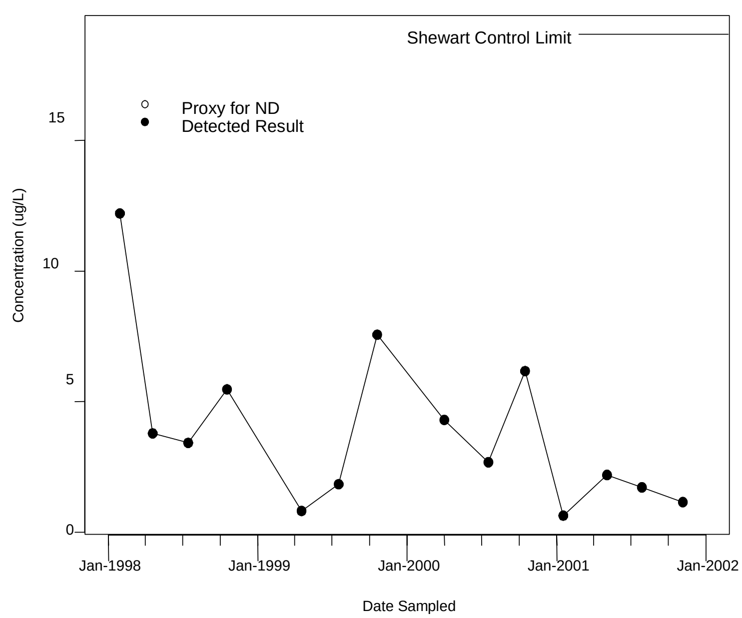

Figure Q-8. Historical data (1998–2000) and 2001 data with Shewart Control limit for benzene (SW8260B) in groundwater at Site A, MW-01.

Q-4.3. *CUSUM Control Charts*. CUSUM control charts are more sensitive than Shewart control charts to small and gradual changes. They incorporate current and historical information by calculating a cumulative sum, *S*, for the i^{th} sample. Directions for preparing a CUSUM control chart are provided in Paragraph Q-4.3.1, followed by an example in Paragraph Q-4.3.2. See Gibbons (1994) for more information.

Q-4.3.1. *Directions for a CUSUM Control Chart.* Verify that the assumptions required for CUSUM charts are met.

Q-4.3.1.1. *Assumptions.*

Q-4.3.1.1.1. At least eight independent samples (from previous sampling events) were collected for each sampling location (groundwater monitoring well) to estimate the mean, $\bar{x}$, and sample standard deviation, *s*.

Q-4.3.1.1.2. The data cannot be correlated; determine if the data are correlated.

Q-4.3.1.1.3. The data must be normal; determine whether the data or transformed data are normally distributed.

Q-4.3.1.1.4. Determine whether seasonality is affecting the data; if so, remove the seasonality.

Q-4.3.1.2. Calculate the mean and standard deviation for the historical data results.

Q-4.3.1.3. Choose an appropriate value for *k* (one-half the size of a difference worth detecting). The EPA recommends setting *k* = 1, which means that a difference of two units of standard deviation is meaningful.

Q-4.3.1.4. At a given location or well, determine the cumulative sum for each independent sample. Define

$$S_i = \begin{cases} 0, & i = 0 \\ \max[0,\ z_i - k + S_{i-1}], & i = 1, 2, \ldots, n \end{cases}$$

where $z_i = \dfrac{x_i - \bar{x}}{s}$. The function max[*a*, *b*] means to use the value *a* or *b*, whichever is higher.

Q-4.3.1.5. Choose the appropriate control limit, *h*. EPA recommends setting *h* = 5. The value of 5 is based on simulations and recommendations contained in Lucas (1982), Hockman and Lucas (1987), and EPA 600/4-88-/040. Essentially, *h* is the upper control limit. (One way to determine whether S_i exceeds five is to plot *S* versus *i* for the data.)

Q-4.3.1.6. Identify if the system is in-control or out-of-control by identifying whether each S_i is less than *h* (in-control), or greater than *h* (out-of-control).

Q-4.3.1.7. Investigate any situation in which a concentration is out-of-control. Ideally, additional samples would determine if the out-of-control condition is real and persistent.

Q-4.3.1.8. EPA 530-SW-89-026 recommends detecting a difference of two standard deviations, or $k = 1$. CUSUM control charts are developed by plotting each S_i against the iteration *i*. Each S_i is compared to an appropriate control limit, *h*. EPA guidance recommends $h = 5$. If any S_i value exceeds *h*, the groundwater system should be investigated for causes associated with the increase in concentration.

Q-4.3.2. *Preparing a CUSUM Control Chart*. Consider evaluating the same data used in the example for developing Shewart Control Charts. Benzene concentrations taken from quarterly groundwater samples at well MW01 in Site A from 1998–2000 will be used as a basis for comparison to the 2001 sampling results.

Q-4.3.2.1. The assumptions for developing CUSUM control charts are the same as developing Shewart control charts. As explained in Paragraph Q-4.2.3, all of these assumptions have been met.

Q-4.3.2.2. Set $k = 1$, $S_0 = 0$, and $h = 5$.

Q-4.3.2.3. For each of the current results in 2001, S_i is calculated as

$$S_i = \max[0,\ z_i - k + S_{i-1}]$$

where

$$z_i = \frac{x_i - \mu}{\sigma}$$

μ is estimated by $\bar{x} = 4.824$, and σ is estimated by $s_x = 3.284$. (Specify what data are being used to calculate the mean and standard deviation.) Each S_i value is then compared to $h = 5$; cases in which $S_i \geq h$ are defined as samples out-of-control. (**Note:** Both the mean and standard deviation come from 10 historical samples.)

Q-4.3.2.4. Results are presented in Table Q-8 and show that none of the current results are out-of-control.

Q-4.3.2.5. As an example of these calculations, consider the July 2001 concentration, where $i = 3$:

$$z_3 = \frac{1.72 - 4.824}{3.284} = -0.945$$

$$S_3 = \max[0, (-0.945 - 1 + 0)] = \max[0, -1.945] = 0 .$$

Q-4.3.2.6. Because $S_3 = 0 < h = 5$, the sample is in-control. In this example, there are no out-of-control events because $S_i < 5$ for all i.

Q-4.4. *Combined Shewart-CUSUM Control Charts*. Combined Shewart-CUSUM control charts can be used to detect sudden and gradual changes in concentrations. These control charts combine the benefits of the Shewart and CUSUM charts, as illustrated in Paragraph Q-4.4.1.

Table Q-8.
Current Data

Time	Jan-01	May-01	Jul-01	Nov-01
Concentration (µg/L)	0.64	2.19	1.72	1.15
i	1	2	3	4
z_i	–1.274	–0.802	–0.945	–1.119
S_i	0	0	0	0
Out-of-control?	No	No	No	No

Q-4.4.1. Consider evaluating the same data used in the example for developing the Shewart and CUSUM control charts. Benzene concentrations taken from quarterly groundwater samples at well MW01 in Site A from 1998–2000 will be used to develop a control chart to compare to the 2001 sampling results.

Q-4.4.2. The assumptions for developing Shewart-CUSUM control charts are the same as for developing Shewart and CUSUM control charts. As explained above, all of these assumptions have been met.

Q-4.4.3. Set $h = 5$, $k = 1$, and use the Shewart chart control limit $\text{SCL} = 4.5$ as recommended by the EPA.

Q-4.4.4. The standardized values for each of the current results are estimated, as shown in Table Q-9. The standardized values, z_i, are developed using the historical average and standard deviation of $\bar{x} = 4.824$ and $s = 3.284$.

Q-4.4.5. Then, each z_i value is compared to $\text{SCL} = 4.5$, and each S_i value is compared to $h = 5$. If $z_i > \text{SCL}$ or $S_i > h$, the result is out-of-control.

Q-4.4.6. Results are presented in Table Q-9 and indicate that none of the current results are out-of-control.

Q-4.4.7. As an example of these calculations, consider the July 2001 concentration, where $i = 3$:

$$z_3 = \frac{1.72 - 4.824}{3.284} = -0.945 \ .$$

Q-4.4.8. $S_3 = \max[0, (-0.945 - 1 + 0)] = \max[0, -1.945] = 0$.

Q-4.4.9. As $z_3 = -0.945 < SCL = 4.5$ and $S_3 = 0 < h = 5$, this sample is in-control.

Q-4.4.10. A plot of the standardized results (z_i) versus the time interval (i) can be designed to illustrate this information, as shown in Figure Q-9.

Table Q-9.
Current Data

Time	Jan-01	May-01	Jul-01	Nov-01	
Concentration (µg/L)	0.64	2.19	1.72	1.15	
i	1	2	3	4	
z_i	–1.274	–0.802	–0.945	–1.119	Compare z_i to SCL = 4.5.
S_i	0	0	0	0	Compare S_i to h = 5.
Out-of-control? (i.e., z_i > SCL = 4.5 or, S_i > h =5)?	No	No	No	No	

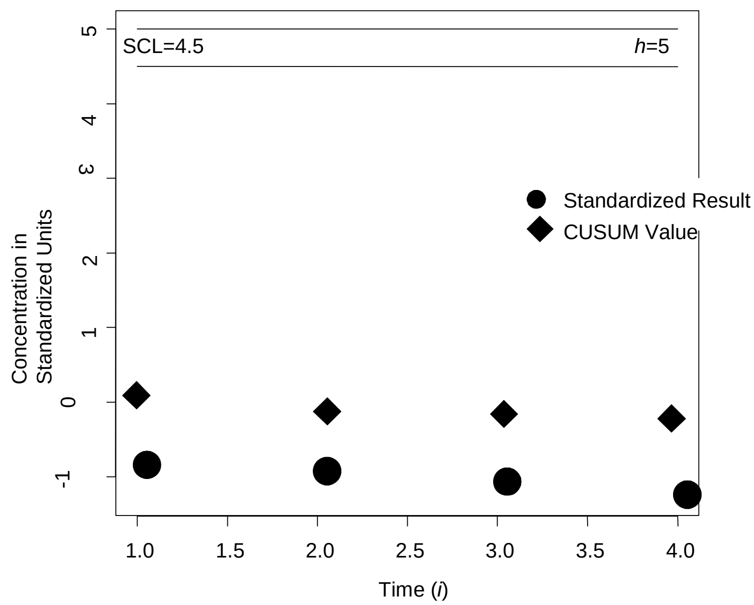

Figure Q-9. Combined Shewart-CUSUM Control Chart (mean = 4.824, standard deviation = 3.284, *k* = 1, *h* = 5, SCL = 4.5).