Assumptions and Procedures for Calculating Water Quality Status and Trends In Tidal Waters of the Chesapeake Bay and its Tributaries

A cumulative history

Prepared for the Tidal Monitoring and Analysis Workgroup
(Previously the Data Analysis Workgroup)
Chesapeake Bay Program
by
Elizabeth Ebersole, MD/DNR
Mike Lane, VA/Old Dominion University
Marcia Olson, NOAA Chesapeake Bay Office
Elgin Perry, Statistical Consultant
Bill Romano, MD/DNR

Updated June 2002

Introduction

The Chesapeake Bay Monitoring Program Analysis Methods were compiled at the direction of the Tidal Monitoring Analysis Work Group (TMAW, formerly the Data Analysis Work Group–DAWG) of the Monitoring Subcommittee. This document summarizes the data analysis methods used by the Monitoring Program investigators to determine status (current condition) and trends (overall increases or decreases over time). This document also describes the adjustments made over time and as necessitated by the individual peculiarities associated with analyzing water quality, living resource or benthic data.

Status

Status is a measure of current condition compared to some benchmark. For some water quality and living resource parameters, reference levels such as restoration target levels or goals have been established and the current condition of an area can be assessed with respect to that level. Status assessment determines if current levels "meet," "fail," or are “borderline” with respect to the target level. Because reference levels are not available for many parameters and because there is some interest in how areas compare to others of similar type, efforts to develop a relative measure of status have been ongoing. Relative status compares recent data for a specific parameter at a particular station or segment to all stations and segments of the same salinity regime in a benchmark dataset. Based on this comparison, the station or segment is given a ranking of "good", "fair" or "poor" for the parameter in question. For most measures of status in the TMAW analyses, using either reference or relative benchmarks, "recent" or "current" data are those collected during the most recent three years.

Reference status – The number of water quality parameters and living resources that have specific goals, target levels, or regulatory criteria is limited, but growing. Methods of assessing status with respect to these levels are different, depending both on the parameter and how the reference level is defined. For example, habitat requirements for submerged aquatic vegetation (SAV) have been determined, and acceptable levels of five parameters critical to SAV growth (light, DIN, DIP, chlorophyll, and suspended solids) established for the various salinity zones during the SAV growing season. The requirements apply only to surface waters. Initially, status was assessed by comparing the 3-year seasonal median values to the requirement value. For example, the requirement for suspended solids is met if seasonal median concentrations are at or below 15 mg/L. More recently, a more rigorous approach has been used in order to give statistical confidence to the assessment. The Wilcoxon Signed Rank Test uses the individual monthly values to determine if the location is significantly ($p<0.05$) above or below the requirement level or not significantly different (borderline).

Goals for dissolved oxygen (DO) have also been established. These apply to spring spawning and summer seasons and set target levels specific to above- and below-pycnocline waters. The goals and methods for assessing attainment are described in (Alden et al., 2000). However, new DO criteria and compliance measures currently being developed as part of the TMDL (Total Minimum Daily Load) process will no doubt supercede these habitat restoration goals.

Relative status - The first version of the relative status method was developed and implemented for the 1997 Re-evaluation effort (Alden and Perry, 1997). Briefly, data from all stations, basin-wide, for the period 1985-1996 were assembled and each datum categorized as tidal fresh, oligo-, meso- or polyhaline, depending on the salinity associated with the data record. For each station and (if relevant) depth layer, separately, the data were averaged by month within year. Then, the data within each salinity regime were pooled and the 5th and 95th percentiles calculated. The 5th and 95th percentiles were empirical endpoints representing the extremes of "good" and "bad" within each salinity regime as observed over the history of the Monitoring Program. This constituted the benchmark data set.

To assess current condition, each datum from the most recent three years was scored relative to these benchmarks; that is, each datum was categorized according to its associated salinity and scored as a percentage (between 1 and 100) of the distance between the benchmark endpoints for that salinity regime. A composite score for the station or segment was obtained by finding the median monthly score over the three-year period.

The method was simple to implement, it resulted in a score between 0 and 100 (which could then be equated to a categorical qualitative assessment), and it could be applied broadly to both water quality and living resource parameters. In retrospect, however, the method did not work exactly as anticipated. One of the underlying assumptions of this relative characterization is that the basin wide distribution of measurements for any particular parameter can be partitioned generally into thirds, each third equating to a status of "good", "fair" and "poor". As implemented, however, that method yielded unequal distributions, i.e., in some cases the method resulted in too many areas characterized as "good" when they were clearly unsatisfactory and vice versa.

For the 1998, 1999 and year 2000 status updates, a modified version of the method was used (Perry, 1999). The modifications are several:

- The benchmark period and benchmark data set are reduced from the entire period of record and include only the first six full years of data: data collected between January 1985 and December 1990. Additionally, both the benchmark and status data are log transformed prior to analysis. (Note, that for water quality parameters, the log and square root transformations are about equal in effecting a normal distribution of the data, and more effective than an inverse transformation or using untransformed data. This may not be the case for other parameters.)
- The benchmark and current status data (from the most recent three years), grouped as before by station or segment, depth layer and salinity zone, are partitioned using a beta cumulative distribution function and the status data set is scored using the logistic probability integral transform.
- The score is then adjusted based on sample size to account for interdependence of observations. The lack of independence in observations at a site tends to result in too many observations in the ends of the distribution, i.e., too many in the good and poor categories, too few in the midrange. The adjustment effects a more equal distribution of scores in the benchmark data set. In the status data set, however, the scores may have quite a different distribution, since the two data sets are independent of one another. For example, if improvements in a parameter were substantial and widespread, then a larger proportion of

recent data would be "fair" or "good" relative to the benchmark period and a smaller proportion would be "poor."

Trend

The implications of trend direction differ with each parameter. For instance, in Chesapeake Bay, an increasing trend in phosphorus concentration is considered degradation (and therefore bad) while an increasing trend in bottom water dissolved oxygen concentration is an improvement (good). Certain biological groups are considered nuisance species or otherwise undesirable. Increases in these are termed bad. Others are thought of as beneficial, and increases in these are termed good. The dynamics of the physical/chemical and biological parameters are complex and these simplistic terms of good and bad may not always hold true.

A generalized description of the trend detection methods used by investigators is given below. Modifications or variations on these methods are discussed in their respective chapters.

Linear trend - The Chesapeake Bay Program has conducted trend analyses since the early 1990s. The primary objective of these analyses has been to detect and quantify water quality responses to nutrient reduction actions, i.e., to measure progress toward Bay restoration goals. The year 1985 was the first full calendar year of the CBP Monitoring Program and the base year for evaluating progress. Fairly simplistic tests of linear trend, i.e., tests of whether a parameter has increased or decreased significantly since the base year, have been used to address this objective. The Seasonal Kendall test (see below) is used to identify trends; Sen's slope estimator (Sen, 1968) is used to estimate the magnitude of change over time when a significant trend is present. These analyses have been conducted annually since 1990 under the auspices of DAWG/TMAW.

Nonlinear trend - As the monitoring program timeline has lengthened, additional analyses have been added or explored. The utility of identifying non-linear trends was investigated in 1998-99 (Alden et al., 2000). While linear trends provide the "bottom line" of change over time, non-linear trends provide a picture of how a parameter changed within the period of evaluation and can serve as early warning signals for managers. These analyses identify whether the trend is primarily linear (generally unidirectional), U-shaped (decreasing early in the time series, increasing later in the time-series) or the reverse (inverse-U-shaped). They also can determine where the critical point, if any, is located (either within or outside the period of evaluation). Non-linear trend results for the 1985-1999 period are available for selected water quality parameters and regions.

Observed versus flow-adjusted trends - Because nutrient loading to the Bay and tributaries is correlated with freshwater inflow, and because most living organisms have specific salinity tolerances and salinity is highly dependent on freshwater inflow, most analyses take flow or salinity or both into account. To that end, trend analyses may be performed on observed data and on data with the effects of freshwater input removed ("flow-adjusted" data). The flow-adjustment technique (Alden et al., 1997) used in the 1996 through 1999 reporting cycles is described below. The methodology is currently under review and is likely to be superseded by another method.

Flow-adjustment (1996-99) - The flow-adjusting step finds the best predictive flow variable among several and removes the variance associated with this flow variable by subtracting the least squares prediction from the observed response. Flow-adjusted residuals are produced by conducting regression analyses employing four of the flow transformations developed for the USGS NASQAN program (Smith et al., 1982). In all cases, a 7-day moving average of flow is used as the independent variable in the flow regression. The four transforms of the flow variable are the logarithm of flow and three hyperbolic transforms given by the expressions

$$h3 = 1/(1+(10-1.5x\beta^* \times F)$$
$$h4 = 1/(1+(10-1.0x\beta^* \times F)$$
$$h5 = 1/(1+(10-0.5x\beta^* \times F),$$

where F is the freshwater discharge and ß***** is the integer value of the log-transformed grand mean flow. USGS fall line daily freshwater discharge estimates for the Maryland and Virginia tributaries for the full period of record are used to create the flow transformations. For tidal fresh stations, 3- and 7-day lags are evaluated. For oligohaline stations, 3-, 7- and 14-day lags are evaluated. For meso- and polyhaline stations, 7-, 14- and 21-day lags are evaluated. The flow variable that results in the best predictive model as determined by the highest R-square is chosen to flow-detrend the observed data. The flow-adjustment is most valid on an annual basis, thus only annual flow-adjusted trends are reported. This contrasts with annual and seasonal trends reported for observed data.

Trend analysis - To calculate trends, a combination of statistical tests is used. First, the Mann-Kendall statistic for trend over years is computed for each "season". In this context, each of the 12 months of the year is defined as a season. The Seasonal Kendall procedure is then used, which provides two statistics for each station (or segment, if station data are pooled):

- a combined test for trend at the station based on combining the trend information from the individual month Mann-Kendall statistics into a single statistic, and
- a test for homogeneity of trend among months, which is used to determine if it is reasonable to use the combined statistic.

For segment-wide analyses, the van Belle and Hughes (1984) procedure is also applied. It is an extension of the Seasonal Kendall test that is used when there is more than one station in a segment. It provides a test for trend based on combining the trend information from the individual stations in the segment. It also provides a test for homogeneity of trend among stations to determine if it is reasonable to use this combined statistic. To quantify the magnitude of trend, Sen's slope estimator and confidence interval are computed. These procedures are described in detail in Gilbert (1987).

The form of the input data depends on the parameter analyzed. Medians, geometric means, or log transformations of observed or flow-detrended data are computed for each month for each year for each station, segment, as appropriate. By combining the Mann-Kendall statistics over selected months, tests for trend in different seasons are also possible. (See individual sections for the purpose of seasonal analyses and season definitions.)

The trend statistics reported in the results tables include the p-value of the Seasonal Kendall score for significant trend, the median trend slope, and the p-value for homogeneity, a baseline value, and the magnitude of change from the baseline value. For these analyses, the magnitude of trend is expressed as the percent change since the beginning of the period of record. Where medians are used, the formula is

pct change = [(slope*nyrs)/base median)]*100**.**

For log-transformed data and geometric means, the formula is

$$\text{pct change} = \frac{[\exp(\text{base logmean} + \text{slope*nyrs}) - \exp(\text{base logmean})]*100}{\exp(\text{base logmean})},$$

where slope is the median trend slope per year, nyrs is the number of *intege*r years in the analysis, and the base statistic is computed from the first 24 months of the data record, in most cases starting with January 1985. The direction and magnitude of change are included in results tables only when a significant trend is found. An alpha level of 0.01 is used for all trend tests using observed data, and an alpha level of 0.05 is used for flow-adjusted data. Results for water quality parameters are subject to additional screening based on the proportion of observations below the laboratory analysis detection limit.

When the p-values for station or seasonal heterogeneity are significant, indicating non-uniform changes among stations or among months within season, then it rests with best professional judgement whether to report the results. The default position is to report the result as the best available assessment of the overall pattern.

TMAW Source Library

TMAW maintains an online library of reference materials at *www.chesapeakebay.net/data/wqual/workshop/workshop.htm*. These materials include, among other things: work plans; reports, "white papers" and other documentation prepared for the workgroup; computer programs for calculating such things as status and trends, "percent light at leaf (PLL)"; as well as diagnostic tools and other products developed for or used by the workgroup.

References

Alden, R.W. III, E.S. Perry, and M.F. Lane. 2000. A Comparison of Analytical Techniques for Determining Trends in Chesapeake Bay Water Quality Monitoring Program Data. AMRL Technical Report #3114. Applied Marine Research Laboratory, Norfolk, VA.

Alden, R.W. III and E. S. Perry. 1997. *Presenting Measurements of Status.* A "white paper" written for and presented to the Chesapeake Bay Program Data Analysis Workgroup. 15 pp.

Alden, R.W. III, L. Haas, M.F. Lane, and E.S. Perry. 1997. *Recommendations for a Unified Approach to Flow Corrections in the Maryland and Virginia Tributaries*.

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

Jordan, S., C. Stenger, M. Olson, R. Batiuk, and K. Mountford. 1992. Chesapeake Bay Dissolved Oxygen Goal for Restoration of Living Resource Habitats. Chesapeake Bay Program Office, Annapolis, MD. Reevaluation Report #7c. CBP/TRS 88/93.

Perry, E.S. 1999. Personal Communication.

Sen, P.K. 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Ass.* **63**:1379-1389.

Smith, R.A., Hirsch, R.M., and Slack, J.R. 1982. *A Study of Trends in Total Phosphorus Measurements at NASQAN Stations*. U.S. Geological Survey Water-Supply Paper 2190. 34 p.

van Belle, G. and J.P. Hughes. 1984. Nonparametric tests for trend in water quality. *Water Resources Res.* **20**(1):127-136.

Maryland and Virginia Mainstem and Tidal Tributaries

A Little History

Over the years of the Monitoring Program, analytical methods have changed or been modified. Some of the changes have been due to changes in parameters and laboratory techniques, others have been due to new statistical techniques and/or new thinking; still other changes have followed because of technological advances in data management and communications. In the wake of such change, comparability and consistency issues have been and will continue to be challenges to the workgroup.

Historically, responsibility for water quality status and trend analyses was divided among the primary Monitoring Program partners, albeit under the auspices and guidance of the analytical workgroup. Maryland state staff or grantees performed the analyses for Maryland tidal tributaries; Virginia commonwealth staff or grantees performed the analyses for Virginia tidal tributaries; and USEPA Bay Program staff performed the analyses for the mainstem Bay. Although analyses were performed by different entities, the same methods, conceptually, were used and modified as necessary to conform to the individual sampling programs. With the advent of the CIMS database and universal access to data through the web, it was thought that cost efficiencies and consistency could be gained by centralizing the analyses. In year 2000, preliminary data preparation was performed by the separate partners, but the status and trend analyses (covering data through 1999) for the Bay and tributaries were all done by Bay Program staff. Although benefits were derived from that exercise, the responsibility for review and interpretation of the results still resided, rightly, with the separate partners, and the back and forth of data sets and results proved cumbersome and time-consuming. In 2001 and 2002, Maryland staff performed the analyses for the Maryland tributaries and Bay Program staff performed the analyses for the mainstem and Virginia tributaries. Both groups used the same computer programs for all aspects of analysis and reporting.

Note: in the following sections, terms such as “1997 update” refer to analyses of data records that have been updated with data collected in the year named, e.g., 1997. The analyses were actually performed and the results reported in the following year.

Parameters

The core parameters for which status and trend analyses are conducted each year are listed below.

Four nutrient parameters:

- total nitrogen (TN);
- dissolved inorganic nitrogen (DIN);
- total phosphorus (TP); and
- dissolved inorganic phosphorus (DIP).

Eight additional parameters:

- total suspended solids (TSS);
- active chlorophyll a (CHLA), as a response indicator of nutrient enrichment and habitat quality;
- bottom dissolved oxygen (DO), as a response indicator of nutrient enrichment and habitat quality;
- Secchi depth (SECCHI), as a measure of water clarity;
- "percent light at the leaf (PLL)," a calculated estimate of light reaching submerged aquatic vegetation (SAV) at various depths. PLL is derived from the measurements of DIN, DIP and TSS. For this update, PLL at 0.5 m and at 1 m were analyzed.
- KD, a measure of light penetration;
- salinity; and
- water temperature.

Similar analyses for additional parameters may be available as well: e.g., particulate phosphorus (PP), nitrite/nitrate (NO23), ammonia (NH4), silicate (SI) and carbon compounds (e.g., PC); also nutrient ratios, such as TN:TP and DIN:DIP.

Flow-adjusted trend analyses have been conducted only on the four nutrient parameters TN, DIN, TP, and DIP, and on TSS, CHLA and bottom DO. The most recent flow-adjusted trends for water quality are for the 1985-1998 period. The flow-adjustment methodology is currently under review.

<u>Spatial and Temporal Scales</u>

Water samples for laboratory analysis of nutrients, chlorophyll and suspended solids are collected at surface and bottom and at 1 m above and 1 m below the pycnocline, if one exists. For status and trend analyses, where both surface and above-pycnocline samples are collected, measurements are averaged, resulting in one value for the surface-mixed layer. Likewise, where both bottom and below-pycnocline samples are collected, measurements are averaged, resulting in one value for the bottom-mixed layer. Trend analyses are done separately for surface-mixed and bottom-mixed layers. In the Virginia tributaries, chlorophyll is measured only at the surface and in some regions, the number of missing values for other parameters preclude analysis of bottom measurements.

[Note: in 1997 and 1998, status assessments for surface chlorophyll used only surface measurements, even when above-pycnocline measurements were available. Chlorophyll is measured only in surface waters in the Virginia tributaries, and this was intended to equalize data handling among segments. In subsequent years, it was decided to use all available data and treat as indicated above.]

Water temperature, salinity and dissolved oxygen are measured *in-situ* at 1- to 2-m intervals through the water column. In the case of dissolved oxygen, only bottom concentrations are analyzed for status and trends. For salinity and water temperature, only surface and bottom measurements are analyzed for trends; status is not evaluated for these two parameters.

Annual routine status and trend analyses are conducted using water quality data collected from the Chesapeake Bay mainstem and tidal tributaries from January 1985 (or from the beginning date if the program began later) through December of the most recent year. The core seasonal analyses include:

- the *annual* season or calendar year (months 1-12);
- the *SAV growing season* (months 4-10 in tidal fresh, oligohaline and mesohaline regions, and months 3-5 and 9-11 in polyhaline regions);
- *spring* (months 3-5 in polyhaline regions and 4-6 in other salinity zones); and
- s*ummer*, which is defined differently for different parameters. For dissolved oxygen, summer includes months 6-9; for chlorophyll a, summer includes months 7-9. For most parameters, analyses are done for all season definitions.

The flow-adjusted data are analyzed for trend only over the annual season (months 1-12). Flow-adjusted data are not assessed for status.

For a regional picture of status and trends, stations are aggregated for analysis into segments. Prior to 1997, by-segment analyses used the original CBP segmentation scheme. The segmentation was modified for the1997 Reevaluation Effort to reflect more closely the salinity conditions of the evaluation period, i.e., 1985 and subsequent years. Status and trend analyses for the 1997 update used this station aggregation. In 1998, the segmentation scheme was further modified slightly. This scheme is the basis for 1998 to present by-segment analyses. Documentation of the chronology and definition of the several schemes (Olson, 2000) is available on line in the TMAW Source Library.

<u>Status Calculations</u>

As described in the introduction (page 1)

<u>Trend Analysis and Flow-Adjusting Procedures</u>

As described in the introduction (page 1), with the following additional details.

Flow adjustment in the mainstem Bay – The mainstem Bay receives discharges from large and small tributaries up and down its length and it is difficult to remove the effects of flow for main Bay stations in the same way as done for the tributary stations. At the request of the data analysis workgroup, Ray Alden and colleagues developed a flow-adjustment for the main Bay in time for the 1997 trend update. Like the flow adjustment procedure for the tributaries, this method is also currently under review. The flow-adjusted analysis was not performed for the 2000 update.

The "adjustment" for the mainstem is, in fact, segment-specific regression models that include a flow factor as well as various pre-selected month, depth, salinity and/or water temperature factors, if they added significantly to the model fit. The input value for daily flow is the sum of the daily flows of the major tributaries discharging to the Bay at or north of the segment being analyzed. Similar to the procedure applied in the tributaries, the procedure finds the best predictive flow variable among several and removes the variance associated with flow and associated variables by subtracting the least squares prediction from the observed response. Copies of these programs are archived online in the TMAW Source Library.

Decision Rules for Reporting Trends With Observations Below Detection Limit - In the CBP water quality monitoring database, parameters whose levels are below the detection limit of the analytical method are assigned the value of the detection limit. Over the history of the Monitoring Program, many of the laboratory analytical techniques have changed or improved and lowered their limit of detection. An artifact of this advance is that the lower values of the BDL measurements later in the data record may be falsely detected as a downward trend. To avoid this, water quality values are censored to the highest detection limit of the analysis period as part of the data handling prior to analysis. Censoring is based on the detection-limit history of each station for the individual station analyses. For segment analyses, however, where stations within a single segment are monitored by different organizations and have different detection limits, the censoring level is the highest detection limit of the stations in the segment. After censoring, all censored data are set to one-half the detection limit value.

Data sets having large numbers of values below detection limit (BDLs) may create statistical problems for trend analyses. The Seasonal Kendall test for trend, and similar sign tests such as the Van Belle and Hughes test, adjust variance estimates upward for ties in magnitude (e.g., Gilbert, 1990). Since BDL values in the raw data set produce such ties, trend analyses of data sets with high percentages of BDLs will be based upon greater variances than those without BDLs, all else being equal. Thus, the power of the trend analyses for the data sets with BDLs will be reduced compared to those without detection limit censoring.

There is an additional wrinkle to flow-adjusted data. When a data set with BDL values is flow-adjusted by the procedures previously described, many ties in magnitude disappear, since each datum is adjusted based upon the flow measurement from the day of collection. As a result, the trend analyses conducted after flow adjustment will, in all probability, have fewer ties in magnitude, lower variances and an artificial increase in power compared to the trend analysis based upon the observed data. This increase in power is an artifact of the flow adjusting process and is not based on changes in the magnitude of trends that are due to flow.

The DAWG guidelines for reporting Seasonal Kendall trend test results, with respect to BDLs, have changed over the years. For the 1985-1997, -1998 and -1999 updates, the following rules applied:

- If a significant trend result is obtained and more than 5% of the data are below the worst-case detection limit, then report the direction of trend, but not the magnitude (percent change).
- If more than 20% of the data is censored, then report neither the direction nor magnitude of trend.
- If results are significant only for flow adjusted data and more than 5% of which are BDL, confirm the results through the use of a Tobit analysis procedure (Tobin, 1958). Tobit analysis is a regression-based procedure that is designed to handle left censoring such as occurs with lower detection limits.

For the 1985-2000 trend updates, DAWG adopted different decision rules:

- If the percentage of BDL observations is 15 or less, report the Seasonal Kendall trend test p-value and direction as well as the Sen slope estimator of the magnitude of the trend (e.g., 35 %).
- If the percentage of BDL observations is greater than 15 and less than or equal to 35, report the Seasonal Kendall trend test p-value and direction, but do not report the Sen slope estimator of trend magnitude.
- If the percentage of BDL observations is greater than 35 and less than or equal to 50 and the Seasonal Kendall trend test p-value indicates a significant trend, report the Seasonal Kendall trend test p-value and direction, but do not report the Sen slope estimator of trend magnitude.
- If the percentage of BDL observations is greater than 35 and less than or equal to 50 and the Seasonal Kendall trend test p-value does not indicate a significant trend, report nothing, noting that there are too many observations below the detection limit to determine the presence or absence of trend.
- If the percentage of BDL observations is greater than 50, report nothing, noting that too many observations were below the detection limit to determine the presence or absence of trend.

Rationale - The rationale for these rules is based on findings demonstrated by simulation analysis for the Seasonal Kendall test and Sen slope estimator (Alden, Perry and Lane, 2000) and is briefly summarized here: 1) The false positive rate of the Seasonal Kendall test does not seem to be affected by the level of censoring of the data; 2) The power of the Seasonal Kendall test begins noticeably to decline when censoring exceeds 35 %; 3) The Sen slope estimator begins noticeably to exhibit bias when censoring exceeds 15%. At levels of censoring of 15% or less, both the Seasonal Kendall test results and the Sen slope estimator are reliable and should be reported. At levels of censoring greater than 15%, the Sen slope estimator should not be reported because it becomes biased. The Seasonal Kendall test retains a robust type I error rate and a flat power response up to 35% censoring and thus should be reported up to that level. If the Seasonal Kendall test produces a significant result when the level of censoring exceeds 35%, one may infer that this result is obtained in spite of the loss of power and therefore is a valid result and should be reported. If the Seasonal Kendall procedure fails to produce a significant result when censoring is in the 35% to 50% interval, this failure may have resulted from a loss of power and should be reported as a non-significant result, which carries the implication that the trend is below the level that we have power to detect with an uncensored data set. While the Seasonal Kendall procedure continues to exhibit the nominal type I error rate for levels of censoring that are greater than 50%, and thus significant results for these high levels of censoring might be judged reliable, the risk that the uncensored data are unduly influenced by a large scale stochastic event (e.g., drought, hurricane, etc.) becomes large and these results should not be reported.

Determining percent BDL – This aspect of the analytical methodology seemed self-evident at first and was not formally discussed or delineated in detail. The several reporting entities used the same censoring procedure for each datum (i.e., setting values lower than the highest theoretical detection limit to one-half the detection limit value), but otherwise each "did their own thing." For the 1998 and 1999 updates, the workgroup defined a more detailed procedure. It was modified somewhat for the 2000 update.

- Flag and censor each value below the highest detection limit over the trend period.

  For parameters that are *directly measured* during the whole time period, the detection limit is simply the highest measured detection limit used for that parameter over the time period. For example, the highest detection limit for orthophosphate (PO4) at stations in Maryland minor tributaries between 1985 and the present is 0.01 mg/L. This was the detection limit at the analytical laboratory from 1985 to May 31, 1986.

  For *calculated parameters*, i.e., parameters derived by addition or subtraction from directly measured parameters, the theoretical detection limit is the sum of the detection limits of the constituent parameters. The highest theoretical detection limit, then, is the highest of such sums over the trend period. For example, total nitrogen (TN) is obtained from TKNW+NO23 and/or from PN + TDN, depending on which constituents are measured. Both methods have been used over the history of most stations in the Monitoring Program. For example, at mainstem Bay stations, TN was obtained by the first method from the beginning until October 1987, and by the second thereafter. At a station in the lower Bay sampled, say, by VIMS, the highest detection limits for TKNW and NO23 at any given time in those years sum to 0.11 mg/L; the highest for PN+TDN is 0.10. Thus, the data are censored to one half of the higher of the two methods, to 0.11/2 mg/L.

- Censor monthly mean values. The Seasonal Kendall trend test is performed on monthly mean values for separate depth layers of various definition, e.g., surface, surface-mixed, lower-mixed, and bottom layers. In the first version of the procedure, the monthly value for each layer was considered BDL if the number of censored (flagged) measurements was >= 50% of the individual sample values. For example, two sampling events at a deep water station in July yields four total values in the mean for the surface-mixed layer: one sample each from surface and above-pycnocline layers, times two events. If two or more of those four values are BDL, then the monthly value was considered BDL in this version.

  This decision rule was modified in the 2000 update. Since the BDL issue is actually an issue of "ties" between months in the Seasonal Kendall test, the possibility of a tie due to censoring is eliminated if any one of the values in the monthly mean is *not* BDL. Thus, in the modified version, a monthly mean value is flagged as BDL only if all of the values in the mean are BDL.

- Compute BDL percentage based on station/segment-layer-season group**.** The trends are calculated either by segment or by station for a given layer for various defined seasonal groupings. The percent BDL for a parameter is the number of BDL monthly mean values divided by the total number of monthly mean values in the station/segment-layer-season group of values for that parameter. The results, with respect to BDLs, are reported or suppressed based on the decision rules given above.

Reporting Rules for Non-homogeneous Trend Results - The procedure by van Belle and Hughes (1984) is used to test for homogeneity of trend among months to see if the trend is consistent across months within seasons and across stations within segments. Homogeneity within seasons is tested and considered significant at $p <= 0.01$. If trends among months within seasons for a given station are not homogeneous, then the analyst reviews the data and uses professional judgment to determine if the overall annual trend is considered valid and can be reported with confidence. The default rule is to report the trend. If the analyst finds reason for no confidence, then "no trend" is reported.

Related Information

For details on water quality field sampling or laboratory analysis methods in Maryland**,** see http://www.dnr.state.md.us/bay/tribstrat/status_trends methods.html or contact Elizabeth L. Ebersole, Tidewater Ecosystem Assessment, Maryland Department of Natural Resources, 580 Taylor Avenue, Annapolis, MD 21401, bebersole@dnr.state.md.us. In Virginia, contact F. A. Hoffman, VA Dept. of Environmental Quality, P.O. Box 10009, Richmond, VA 23240, fahoffman@deq.state.va.us.

References

Alden, R.W., E.S. Perry, and M.F. Lane. 2000. *A Comparison of Analytical Techniques for Determining Trends in Chesapeake Bay Water Quality Monitoring Program Data.* AMRL Technical Report # 3114. Applied Marine Research Laboratory, Norfolk, VA.

Gilbert, R.O. 1990. *Statistical Methods for Environmental Pollution Monitoring.* Van Nostrand Reinhold: New York, NY.

Olson, M.M. 2000. *Analytical Segmentation Scheme for the 1997 Re-evaluation and Beyond*, prepared for the USEPA Chesapeake Bay Program Office by the Monitoring Subcommittee Data Analysis Workgroup.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*. **26**:24-36.