Quality Assurance Monitoring Plan

Stream Ambient Water Quality Monitoring

by
Dave Hallock
and
William Ehinger

Environmental Monitoring & Trends Section (EMTS)
Environmental Assessment Program
Washington State Department of Ecology
Olympia, Washington 98504-7710

Revision of 1995 Version
April 2003

Publication No. 03-03-200

This report is available on the Department of Ecology home page on the World Wide Web at http://www.ecy.wa.gov/biblio/0303200.html

Quality Assurance Project Plan

Stream Ambient Water Quality Monitoring

Revision of 1995 Version
April 2003

303 (d) listings addressed in this study:
NA

Approvals

Approved by:	June 12, 2003
Dave Hallock, Investigator, Environmental Monitoring and Trends Section	Date
Approved by:	June 10, 2003
Rob Plotnikoff, Unit Supervisor, Environmental Monitoring and Trends Section	Date
Approved by:	June 20, 2003
Ken Dzinbal, Manager, Environmental Monitoring and Trends Section	Date
Approved by:	April 15, 2003
Stuart Magoon, Director, Manchester Environmental Laboratory	Date
Approved by:	May 27, 2003
Cliff Kirchmer, Ecology Quality Assurance Officer	Date

Table of Contents

Table of Contents iii

Tables iv

Figures iv

Organization and Schedule 1
- Organization 1
- Schedule 1

Background and Problem Statement 2

Project Description 4
- Goals and Decision Statement 4
- Objectives 4
- Constraints 6

Quality Objectives 6
- Measurement Quality Objectives 6
- Data Quality Objectives 7
  - Precision 7
  - Bias 12
  - Reporting Limits 12

Sampling Design 13
- Representativeness 16
- Comparability 17

Field Procedures 17

Laboratory Procedures 18

Quality Control 19
- Laboratory QC 19
  - Check Standards 19
  - Analytical Duplicates 20
  - Matrix Spikes 20
  - Blanks 20
- Field QC 20
  - Replicates 20
  - Blanks 21
  - Instrumentation 21
  - Corrective Action 22

Data Management Procedures 22

Reports 22

Data Review, Verification, and Validation 23
- Missing Data 24

Data Quality Assessment 24

Trend Power Assessment 25
Characterizing Water Quality and Analyzing Trends 26
References 26

Tables

Table 1. Ambient monitoring personnel and areas of responsibility. 1
Table 2. Schedule for annual tasks. 2
Table 3. Measurement Quality Objectives 7
Table 4. The relationship between sample size (n) and minimum relative detectable trend (δ). 9
Table 5. Calculated maximum permissible error ($s_{error(mp)}$) values to detect a trend given $\beta = 0.1$, $\alpha = 0.1$, $\phi = 0.17$, and $n=60$. 11
Table 6. Results below reporting limits collected from September 1997 to September 2002 13
Table 7. Location of long-term ambient monitoring stations (as of WY 2003). Figure 2 is a map of the station locations. 14
Table 8. Water quality constituents monitored monthly as part of Ecology's ambient stream monitoring program. 16
Table 9. Parameters measured in the field 18
Table 10. Container type, water volume required, method of preservation and maximum permissible holding times for lab-analysed samples are listed below. 18
Table 11. Laboratory analytical methods and reporting limits. 19
Table 12. Historical ranges and 90th percentiles for stream monitoring data based on WY 1998 through 2002 (all stations). 20
Table 13. Field quality control samples required annually (lab QC is specified in MEL guidance documents). QC samples are divided equally among the four sampling regions. 21
Table 14. Example for estimating required trend magnitude ($\Delta\mu$) or sample size (n) to enable trend detection in a dataset (oxygen in mg/L at station 13A060, 1991/09/01-2002/09/01). 26

Figures

Figure 1. The relationship between sample size (n) and minimum relative detectable trend (δ). . 9
Figure 2. Location of long-term ambient monitoring stations (as of WY 2003). See Table 7 for a key to the stations. 14

Organization and Schedule

Organization

One person is assigned to each of the four Ecology regions. This person is a contact person for regional staff and responsible for sample collection and data analysis for the basins in that region (Table 1). Cross-regional duties have also been assigned to specific individuals.

Table 1. Ambient monitoring personnel and areas of responsibility.

Personnel	Region	Phone	Other duties
Rob Plotnikoff	Statewide	407-6687	Unit lead, substitute sampler.
Dave Hallock	Statewide	407-6681	Data management, misc. data analyses and reports, substitute sampler.
Chris Coffin	Central	509 454-4257	TMDL Effectiveness Monitoring.
Jim Ross	Eastern	509 329-3425	TMDL Effectiveness Monitoring, metals monitoring quality control assessment.
Bill Ward	Northwestern	407-6621	Continuous temperature monitoring, equipment procurement.
Chad Wiseman	Southwestern	407-6682	Bioassessment.

Schedule

About 82 stations are sampled monthly statewide. For sampling purposes, these stations are divided into four "runs" approximately corresponding to Ecology regions: Central, Eastern, Northwestern, and Southwestern. (For logistical reasons, some stations in one region may be sampled during the adjoining region's run.) Runs may occur consecutively, in the order listed in Table 1, during the first four weeks of each month. However, we may also schedule the eastern and western runs to go out concurrently--but only after consultation with the lab.

Preparations for sampling trips are made on Thursdays. Sampling takes place Monday through Wednesday and post-sampling activities, including field data entry, on the following day. At times, some runs may take four days to complete. In these cases, we must begin sampling on Sunday so that the lab will have sufficient time to complete their bacteriological analyses by the close of business Friday. Samples are delivered to the lab by the morning following sampling either by bus, by airfreight, or by delivery to the headquarters walk-in cooler.

Data are usually available from the lab within six weeks of sampling. Once a full month's data are available, they are entered into the database, a preliminary data validation is done, and then results, designated "preliminary," are reported to the web. Field data entry is checked quarterly and, once corrections are made, field and lab data for the quarter are uploaded to the Environmental Information Management (EIM) database.

Sampling is scheduled by water year (WY; October through September). A number of tasks are necessary on an annual basis to prepare for an upcoming WY and to close out a completed WY (Table 2).

Table 2. Schedule for annual tasks.

Date	Task
1 October	New WY begins; new "basin" stations (see "Sampling Design") will be sampled for the first time this month.
30 November	Last of previous WY's data should be available (except for flow) from the lab. Summary statistics are recalculated; annual reporting and quality control (QC) review can begin.
31 March	Ideally, flow data should be available by this date (though it can sometimes take as much as a year). Begin planning for coming WY basin stations (coordinate with regions, etc.).
30 June	Complete annual report and update "finalized" data on the internet. This may be delayed up to six months if flow data are unavailable.
Summer	Scope unfamiliar upcoming basin stations; enter new station information into database.
30 September	After completing field data entry for previous WY, update sampling schedules in database for upcoming WY.

Background and Problem Statement

There are a number of regulations in state and federal law that require ambient water quality monitoring. For example, Washington State requires water quality monitoring relative to forest practices (RCW 90.48.420, paragraph (2)), salmon recovery (RCW 70.85.210), and receiving waters (173-201A-170). Section 305(b) of the federal "Clean Water Act" (Title 33 U.S.Code Chapter 26) requires that states report on how well its waters support their designated uses and section 303(d) requires states to identify waters that do not meet water quality standards.

In addition, our ambient monitoring data are used to support a number of other activities. Among these activities are Total Maximum Daily Loads (TMDL); waste discharge permitting; water and watershed management by local governmental entities and others; and reporting of general water quality to the public.

Ecology and its predecessor agencies have conducted ambient water quality monitoring across the state since the 1950s. Procedures prior to WY 1978 were largely undocumented and monitoring activities were inconsistent. The objectives, specific methods, and quality of specific data collection activities prior to 1978 are unknown. Sampling ranged from daily to quarterly at fixed stations for various durations of time (weeks, months, years), and included a variety of constituents.

From WY 1978 through WY 1988, the Ambient Monitoring Section (AMS) of the Washington Department of Ecology collected samples at monthly intervals from numerous rivers and streams throughout Washington State. This monitoring was conducted more consistently with respect to schedule and the constituents and stations being sampled, though procedures were still mostly undocumented. In WY 1989, a quality control (QC) procedure much like that described in this document was implemented and annual documentation began with WY 1991 Annual Report (Hopkins, 1993).

Beginning in WY 1991, the station network was re-designed to increase the number of stations we could assess over the long-term. The new design included 33 “core” stations (monitored each year; now called "long-term" stations), 33 “rotating” stations (monitored every third year), and 12 “floating” stations (monitored for one year only).

In 1993, Ecology initiated a “basin approach” to water quality management (Wrye, 1993). This approach specified a five-year cycle of management activities. Consequently, beginning in WY 1995 (October 1994) we modified our network once again to the current design of 62 long-term stations, and about 20 “basin” stations (Hopkins, 1993):

Long-term stations were chosen for both trend analysis and to characterize water quality (see the “Project Description” section). Stations were selected

a) *near the mouth of major rivers*-to monitor most major systems in the state;

b) *where major rivers enter Washington State*-to monitor the quality of water before it is impacted by activities in Washington;

c) *downstream of major areas likely to impact water quality*-to detect trends in water quality that may be a result of the effects of urban centers or land use activities; and

d) *in the upper reaches of major rivers*-to determine expected water quality that may be due to natural (or at least less impacted) conditions.

Basin stations are chosen to characterize water quality and to address specific needs for monitoring data. Stations are selected to

a) support planned TMDL activities;

b) confirm suspected water quality problems;

c) partition sources of water quality degradation;

d) characterize waterbodies where we have not previously monitored; and

e) support the waste discharge permitting process.

Our monitoring is focused primarily on conventional constituents (e.g., sediment, nutrients, bacteria; see the “Sampling Design” section) and not toxics. We do, when funding allows, monitor metals concentrations at some stations; that monitoring is described elsewhere (Hopkins, 1995). As of October 1, 2002, our database (see "Data Management Procedures" section) contained nearly 600,000 results; more than 17,000 are added annually.

In summary, Ecology's stream monitoring program is intended to characterize the status and trends in ambient water quality statewide. The resulting data and information are used for a number of purposes.

Project Description

Goals and Decision Statement

The role of the ambient monitoring network is to provide timely water quality data and periodic data analysis reports to clients within the Department of Ecology and elsewhere, and to make these data and reports available to other potential users (other federal, state, and local governmental agencies, educational institutions, consulting firms, and individuals). Data collected through this monitoring program are used for a variety of purposes (see "Background and Problem Statement"), but in broad terms, uses may be summarized as the determination of status and trends in water quality in streams statewide.

Objectives

Specific objectives of the stream monitoring program are as follows:

1) *Determine whether water quality at sampling sites exceeds water quality standards*. This objective is intended to address the 303(d) section of the Clean Water Act. Results are compared to water quality standards according to listing rules maintained by Ecology's Water Quality Program (WQP). Listing rules are modified frequently and are, therefore, not included here. In some cases, individual results are compared to a criterion, in other cases, aggregation of data may be required. Monitoring may be expanded beyond the program discussed in this document to better address listing rules. For example, we recently implemented a "continuous" temperature monitoring program.

2) *Assess the status of water quality in Washington*. This objective is intended to address the 305(b) section of the Clean Water Act. A monitoring program with this objective might best be designed to sample a randomized (non-biased) subset of all possible stream segments (the EMAP approach; *e.g.*, McDonald *et al*., 2002). However, this approach is expensive and, because access to the randomly chosen sites is often difficult, year-round monitoring is impractical if not impossible. In practice, our monitoring design will evaluate *major* streams only. Poor water quality at a particular station will indicate an overall, cummulative problem in the watershed, but we will not necessarily be able to identify the extent of the problem. We are pursuing funding to add a randomized component to our monitoring.

3) *Provide analytical water quality information that describes present conditions and changes (trends)*. Long-term monitoring at fixed stations followed by periodic statistical analysis of the data and interpretive reports of the results are one of the mainstays of the ambient monitoring network. The data requirements for trend analysis are quite rigorous. Five or more years of monthly data--a long-term commitment of resources--is required (Lettenmaier, 1977). However, these data are extremely valuable because they provide the most efficient and sensitive means for the early detection of emerging water quality problems. The data

quality objectives are based primarily on the objective of early detection of deteriorating water quality conditions in Washington's less impacted rivers and streams. These requirements are also adequate for the detection of improving water quality conditions in degraded water bodies as well as for meeting the other objectives stated here.

4) *Provide timely and high-quality data for other users*. Specific uses of data collected through this program are as varied as the number of entities studying or managing water quality in Washington. Each use will have its own minimum data quality requirements, but our data quality will be appropriate for most uses. Other uses of our data include

    a) TMDL analyses: our data are used to refine and verify TMDL models.
    b) Supporting the waste discharge permitting system: permit writers require receiving water data.
    c) Development of water quality standards: our data are often the cornerstone for technical analysis leading to revisions of the state's water quality standards (WAC 173.201A).
    d) Cooperative projects with other governmental entities: for example, our data have been used to support various Conservation District projects.

We may monitor any stream in Washington State, with a few exceptions. Long-term stations are already sited on major streams statewide (see "Background and Problem Statement"). Basin stations should be amenable to monthly monitoring for a full WY. The streams should, for example, have permanent access and be perennial or nearly so. We generally avoid sample sites on federal or tribal land. Other than these and logistical considerations, stations may be sited virtually anywhere.

To address the above objectives, we measure several conventional water quality constituents. Four constituents can be readily compared to state standards: temperature, pH, dissolved oxygen, and fecal coliform bacteria. (The latter may change to *E. coli*.) We also measure constituents susceptible to change due to anthropogenic sources: nutrients (total phosphorus (TP), soluble reactive phosphorus (SRP), total nitrogen (TN), nitrate plus nitrite nitrogen (NO23), ammonia nitrogen (NH3)), total suspended sediment, conductivity, and turbidity.

A discussion of decision rules is appropriate (more or less) for the first three objectives.

- The question posed by the first objective is "are water quality standards violated at each monitored station?" The decision rules to answer this question change regularly. Currently, rules (termed "303(d) Listing Criteria") are under development by Ecology's WQP. Final answers to this question are, ultimately, determined by the WQP after a public review process. Answers of "yes" lead to TMDL and Waste Load Allocation analyses.
- The second objective poses the question "what is the quality of Washington's streams?" Procedures for addressing this question are also fluid. Most recently, Ecology used a statistical approach to extrapolate conditions at our monitored stations statewide. See Butkus (2002) for a more thorough discussion of specific procedures. No management actions are predicated on the outcome of this objective.

- The third objective asks about current conditions and trends in water quality. Assessments of current conditions are site-specific and include various measures of central tendency and dispersion, non-parametric statistics such as cumulative frequency plots, and an aggregation technique called the "Water Quality Index" (Hallock, 2002b). Trend assessment is most commonly performed using the non-parametric seasonal Kendall test for trend (see Hirsh *et al.*, 1982) with a confidence level specified at 90 or 95%. Usually, findings are provided to the public (via the stream monitoring program web pages) and used in Ecology's "State of the Environment" reports. More detailed analyses are sent to Ecology regional offices and local entities such as Conservation Districts.

Constraints

Sampling has rarely been cancelled or rescheduled because of poor weather. To do so regularly could impart a bias to the final data. We do occasionally reschedule runs for personal reasons (though we have enough backup samplers that this can usually be avoided), or miss samples because of temporary road or bridge closures. We also sometimes fail to measure a constituent at a few stations when equipment fails in the field (we have backup equipment available at central locations to minimize this problem). Unlike weather, however, these occurrences, presumably, are random relative to water quality and will not affect long-term data analyses.

Quality Objectives

Measurement Quality Objectives

The Environmental Protection Agency (EPA) defines Measurement Quality Objectives (MQOs) as "'acceptance criteria' for the quality attributes measured by project data quality indicators. [They are] quantitative measures of performance…" (Environmental Protection Agency, 2002). In practice, these are often the precision, bias, and accuracy guidelines against which laboratory (and some field) QC results are compared. Precision may be assessed by the analysis of laboratory duplicates or check standard replicates, and bias by comparing the mean of blank and check standard results to known values. The acceptable levels listed in Table 3 are to be applied to batch-level data and may be assessed by only a few QC samples. Failing to meet these criteria would trigger corrective action (see that section).

Table 3. Measurement Quality Objectives.

Analyte	Accuracy (deviation or % deviation from true value)	Precision (% relative standard deviation)	Bias (% deviation from true value)	Lower Reporting Limit
Field Constituents				
Conductivity	± 5 µs/cm at 100 µs/cm	NA	NA	NA
Oxygen	± 0.2 mg/L	NA	NA	NA
pH	± 0.10 std. units	NA	NA	NA
Temperature	± 0.2 ºC	NA	NA	NA
Lab Constituents				
Ammonia-N	20%	7 %RSD	5%	0.01 mg L^{-1}
Fecal coliform	NA	28 %RSD	NA	1 colony 100 mL^{-1}
Nitrate+Nitrite-N	20%	7 %RSD	5%	0.01 mg L^{-1}
Soluble Reactive Phosphorus	20%	7 %RSD	5%	0.003 mg L^{-1}
Suspended Solids	20%	7 %RSD	5%	1 mg L^{-1}
Total Nitrogen	20%	7 %RSD	5%	0.025 mg L^{-1}
Total Phosphorus	20%	7 %RSD	5%	0.01 mg L^{-1}
Turbidity	20%	7 %RSD	5%	0.5 NTU

Data Quality Objectives

EPA defines DQOs as "qualitative and quantitative statements that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors…." (Environmental Protection Agency, 2002). DQOs may be used to evaluate whether the data are adequate to address the project's objectives. Among our objectives, the ability to detect changes in water quality (trends) is the cornerstone of our sampling design. A historical perspective, which only long-term records can provide, is necessary in order to make informed decisions regarding TMDL development, water quality assessments, or the effects of regulatory actions on water quality. The data quality objectives, below, were developed to address statistical requirements for trend analysis. They will also be adequate to address our other objectives.

Precision

Linear trend analysis is a form of hypothesis testing of the model (Lettenmaier, 1977)

$$y_t = \mu + \Delta\mu * t/t_l + \varepsilon_t \qquad 1)$$

where

y_t = the value of the monitored water quality variable at time, t
μ = the mean at the beginning of the time period
$\Delta\mu$ = the change in the mean over the time period,
t_l = the length of the time period,

t = the time elapsed since the beginning of the time period,
ε_t = a stochastic error term.

The hypothesis to be tested is:

H_0 (null hypothesis): $\Delta\mu = 0$ (no change in the mean value), and
H_a (alternate hypothesis): $\Delta\mu \neq 0$ (a change has occurred).

The size of trend ($\Delta\mu$) that can be detected depends on the degree of confidence one desires in one's conclusion, the number of independent samples collected, and the variability in the data.

Power, confidence level, and sample size are related so that both α (the probability of detecting a change when one has not occurred, *i.e.*, falsely rejecting the null hypothesis) and β (the probability of not detecting a change when one has occurred, *i.e.*, falsely failing to reject the null hypothesis) decrease with increasing sample size. Also, when one chooses a smaller α (*i.e.*, one assumes a stricter criterion before rejecting H_0), β increases (assuming sample size stays the same). For the purposes of this power analysis we have chosen $\alpha = 0.10$ (10% chance of wrongfully detecting a trend, *i.e.,* one which does not exist) and $\beta = 0.10$ (10% chance of not detecting a trend when one is present).

Given values for α, β, and sample size (n), one can calculate the magnitude of the trend that can be detected relative to the standard deviation of the data (Lettenmaier, 1977). (Note that n in this discussion refers to *independent* samples, in our case collected monthly. One cannot increase n simply by collecting more frequent samples if successive samples are correlated.) Figure 1 and Table 4 show the relationship between the minimum relative detectable trend (δ; Equation 2) and sample size for a two-tailed trend test with both α and $\beta = 0.10$ (see Smith *et al*., 1989). From Figure 1, n=180 (*i.e.*, 15 years of monthly samples) would appear about optimum. More samples than this will not reduce δ much; with fewer samples, δ increases rapidly. Ideally, however, trends should be detected as early as possible so that remedial action can be taken. Also, too long a period will hide short-term trends. We would like to be able to detect trends after 10 years (n=120). For a sample size of 120 (ten years of monthly data, assuming that no significant auto-correlation exists), δ is 0.93. In other words, when the ratio of trend magnitude to standard deviation of the detrended, deseasonalized data is at or above 0.93, there is a high probability (90%) that it will be detected. This analysis applies to normally distributed data. (There are different ways to deseasonalize and detrend data. We use WQHYDRO software (Aroner, 2002) to deseasonalized data by subtracting an estimate of the seasonal median and detrend by subtracting the seasonal Sen slope estimate. There are also other variance-reduction techniques, such as flow adjusting, that we will not discuss here.)

$$\delta = \Delta\mu / s_{obs} \qquad 2)$$

where s_{obs} is the total standard deviation of the deseasonalized, detrended data.

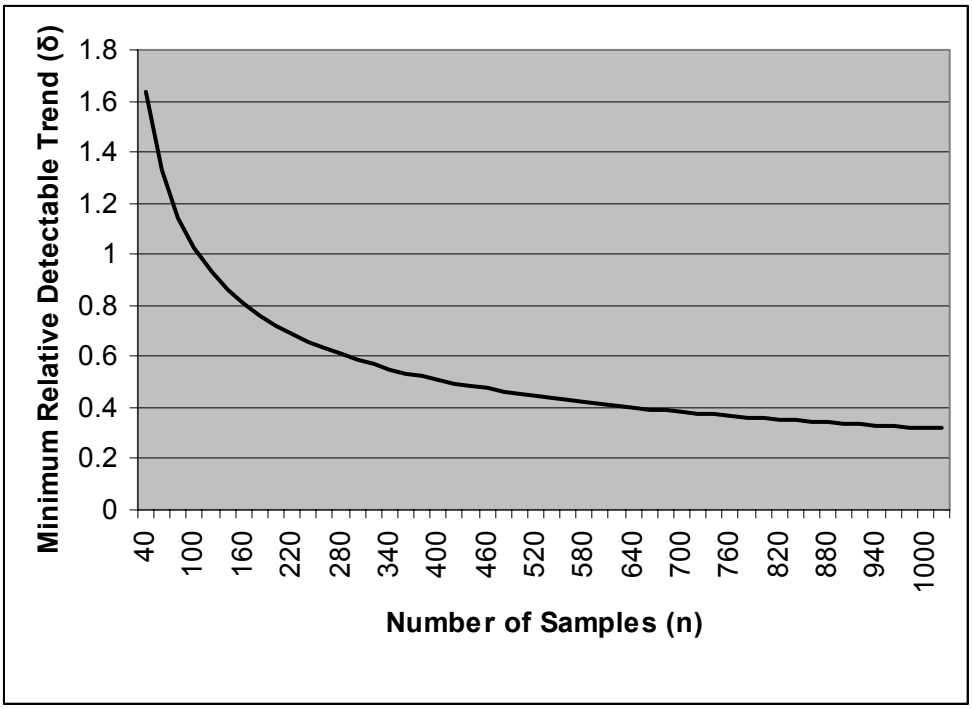

Figure 1. The relationship between sample size (*n*) and minimum relative detectable trend (δ).

Table 4. The relationship between sample size (*n*) and minimum relative detectable trend (δ).

Sample Size (*n*)	Years	Minimum Relative Detectable Trend (δ)
60	5	1.33
120	10	0.93
180	15	0.76
240	20	0.66

We must now specify the absolute magnitude of the trend we wish to detect. Because the ability to detect trends is related to the variance of the data, which for many constituents increases with increasing concentration, we have identified different concentration ranges for the constituents we monitor (Table 5). This is also consistent with a desire to detect trends earlier in high-quality (low-concentration) systems where the ecological impacts of a given $\Delta\mu$ are greater and earlier mitigation is more cost-efficient. For most constituents, we have set the desired trend magnitude ($\Delta\mu$) to 20 percent of the upper bound for each range. (This is over the ten-year period evaluated, not the annual change.)

We may now express error due to field and laboratory procedures in terms of its effect on our ability to detect trends. If we accept that error will reduce our ability to detect trends by 10 percent, the proportion of the total variance (s^2_{obs}) in the detrended, deseasonalized data due to error, ϕ, will be 0.17 (see Ehinger, 1995, or Smith *et al.*, 1989 for the derivation). That is,

$$\phi = s^2_{error} / s^2_{obs} \quad 3)$$

where s^2_{error} is variance due to error.

Combining Equations 2 and 3, the *maximum permissible* standard deviation due to error will be

$$s_{error(mp)} = \Delta\mu * (\sqrt{\phi})/ \delta \quad 4)$$
$$(= \Delta\mu * 0.44 \text{ for } \phi = 0.17 \text{ and } \delta = 0.93)$$

We have collected sufficient quality control data over the five years prior to this writing to evaluate the actual error attained ($s_{error(att)}$). Our error goals ($s_{error(mp)}$) and the actual errors obtained for different constituents and concentration ranges are shown in Table 5. While $s_{error(att)} > s_{error(mp)}$ indicates that we did not meet our *a priori* error goal, it does not necessarily indicate that trends cannot be identified at the specified $\Delta\mu$. (Nor does meeting our error goal guarantee that we can detect trends for any particular data set.) The critical parameter is the *total* observed variance: s_{obs} determines $\Delta\mu$ for a given δ (Equation 2). Even if $s^2_{error(att)}$ is a higher proportion of s^2_{obs} than the 0.17 we specified (ϕ) when developing $s_{error(mp)}$, s_{obs} may still be sufficiently low to allow trend detection.

See the "Data Quality Assessment/Trend Power Assessment" section for cautions on applying the above procedures.

Table 5. Calculated maximum permissible error ($s_{error(mp)}$) values to detect a trend given $\beta = 0.1$, $\alpha = 0.1$, $\phi = 0.17$, and n=120. Actual error ($s_{error\ (att)}$) from data collected during the last five years is also shown. Actual errors not meeting our *a priori* objectives (*i.e.*, $s_{error\ (att)} > s_{error\ (mp)}$) are shown in bold.

Variable (units)	Desired $\Delta\mu$ [a]	Conc. Range (μ)	$s_{error\ (mp)}$ [b]	Empirical	
				$s_{error\ (att)}$ [c]	No. [d]
Electrical conductivity (μ S/cm)	10	< 50	4.4	0.99	39
	20	>50-100	8.8	1.6	67
	30	>100-150	13.2	3.7	43
	60	>150	26.4	5.6	51
Fecal coliform bacteria (colonies /100 mL)	200	<1-1000	88	12	665
	400	>1000	176	**178**	5
NH_3-N (μ g N /L)	4	<20	1.76	**2.5**	165
	20	>20-100	8.8	3.1	29
	40	>100	17.6	1.5	4
Nitrogen, total (μ g N/L)	20	<100	8.8	8.2	40
	40	>100-200	17.6	10.3	42
	100	>200-500	44	15.0	50
	200	>500	88	70.1	67
NO_3NO_2-N (μ g N /L)	20	<100	8.8	2.5	76
	40	>100-200	17.6	10.4	30
	100	>200-500	44	3.5	37
	200	>500	88	28.6	56
Oxygen, dissolved (mg O_2 /L)	1.6	<8	0.70	0.11	4
	2.0	> 8-10	0.88	0.10	40
	2.4	> 10-12	1.06	0.10	107
	4.8	>12	2.11	0.12	51
pH	1.5	N/A	0.66	0.13	0.13
Phosphorus, soluble reactive (μ g P L^{-1})	10	<50	4.4	0.65	176
	20	>50-100	8.8	**11.4**	18
	40	>100	17.6	**20.7**	5
Phosphorus, total (μ g P/L)	10	<50	4.4	4.7	140
	20	>50-100	8.8	5.9	37
	40	>100	17.6	15.0	21

Variable (units)	Desired $\Delta\mu$ [a]	Conc. Range (μ)	$s_{error\ (mp)}$ [b]	Empirical	
Solids, suspended (mg /L)	2	<10	0.88	0.49	303
	4	>10-20	1.76	1.2	95
	10	>20-50	4.4	2.5	99
	20	>50	8.8	8.6	60
Temperature (ºC)	6	N/A	2.64	0.13	191
Turbidity (NTU)	2	<10	0.88	0.17	525
	4	>10-20	1.76	0.45	71
	10	>20-50	4.4	0.88	64
	20	>50	8.8	6.5	33

[a] $\Delta\mu$ has been set to 20 percent of the upper end of the concentration range or 40 percent for the upper-most range. ($\Delta\mu$ is the change over the entire sample period, *i.e.*, 10 years.)

[b] $s_{error(mp)} = \Delta\mu * 0.44$ (Equation 4).

[c] Attainable error calculated as the root-mean-square (RMS) error from field splits. For sediment and fecal coliform bacteria where there is no field processing of samples, lab splits were used. For temperature, pH, and conductivity, where field splits are impractical, sequential samples were used (for these constituents, some of the variability is due to instream processes and not sampling or analytical error). Because results below reporting limits are censored by the laboratory, $s_{error(att)}$ for the lowest concentration ranges, particularly for nutrients, may be biased low.

[d] Number of pairs in the RMS calculation.

Bias

A consistently biased data set will not affect nonparametric trend analysis. However, if a bias is corrected (or imparted) at some mid-point in the sampling period, then the statistical analysis will be compromised. Overlapping new and old procedures for several months prior to abandoning the old method will assess bias due to changes in analytical or sampling procedures. Because this is an ongoing, long-term project, we assume that any batch-specific bias in the chemical analyses will be corrected so that long-term bias will not occur within a single method. Sampling bias will be minimized by strictly adhering to the protocols discussed and referenced herein. Bias due to the time (of day) of sample collection is discussed in the "Sampling Design/Representativeness" section.

Reporting Limits

A certain proportion of results below reporting limits is expected and will not impair our ability to analyze the data. However, a large percentage of results below reporting limits for certain constituents at certain stations will impair our analyses (Table 6).

Statistical analyses of data sets with a large percentage of results below the reporting limit can be problematic. The empirical results in Table 5, for example, are biased low for the lowest

concentration ranges because the calculated variance between any two results below the reporting level is always 0. Also, especially for constituents that are log-normally distributed, changing detection limits can impart an artificial trend in a data set with a large number of near-detection limit concentrations.

At times, we may select a method with a lower reporting limit for some stations and analyses, but due to logistical and financial constraints, we are usually forced to accept limitations on the data. For example, we are unlikely to be able detect trends in ammonia at most stations, though even results below reporting limits can be useful in characterizing water quality.

Table 6. Results below reporting limits collected from September 1997 to September 2002.

Analyte	Percent of samples below reporting limit for worst-case station	Worst-case station	Percent of all samples below reporting limit	Reporting Limit
Turbidity	42	Finch Creek at Hoodsport	2.6	0.5 NTU
Suspended Solids	58	Finch Creek at Hoodsport	6.5	1 mg L^{-1}
Total Phosphorus	45	Various	13.1	0.01 mg L^{-1}
Soluble Reactive Phosphorus	100	Various	42.4	0.01 to 0.003 mg L^{-1}
---WY 2002 only	100	Various	17.3	0.003 mg L^{-1}
Nitrate+Nitrite-N	71	Pend Oreille @ Metaline Falls	9.8	0.01 mg L^{-1}
Ammonia-N	100	Various	61.7	0.01 mg L^{-1}
Total Nitrogen	8.3	Various	0.5	0.025 mg L^{-1}
Fecal coliform	79	Columbia River at Grand Coulee	9.0	1 colony 100 mL^{-1}

Sampling Design

Stations and station selection criteria are discussed in the "Background and Problem Statement" section. Long-term monitoring stations are shown in Figure 2 and Table 7. The approximately twenty basin stations change with each new WY; locations of basin stations for any particular WY may be found in our annual reports (*e.g.*, Hallock, 2002; http://www.ecy.wa.gov/biblio/0203047.html).

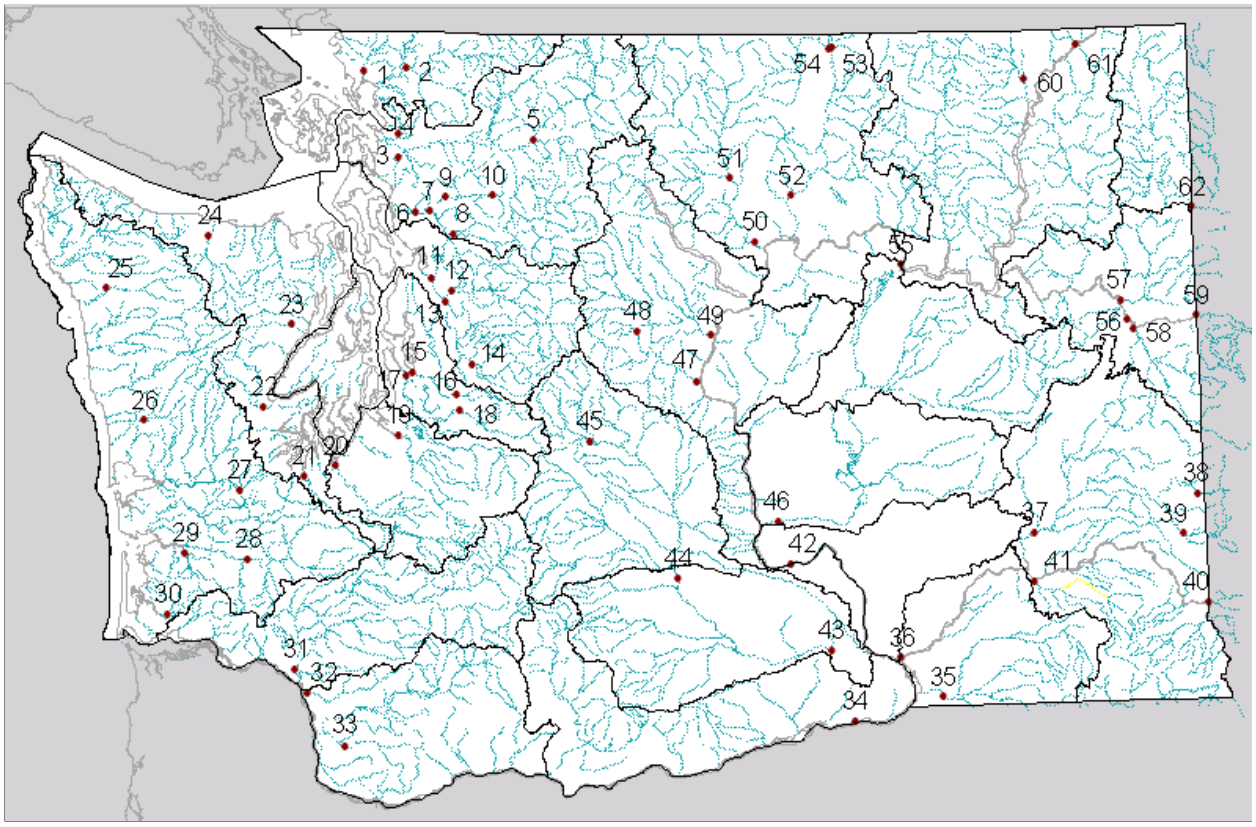

Figure 2. Location of long-term ambient monitoring stations (as of WY 2003). See Table 7 for a key to the stations. The intra-state boundaries define Water Resource Inventory Areas.

Table 7. Location of long-term ambient monitoring stations (as of WY 2003). Figure 2 is a map of the station locations.

ID	STATION	STANAME	ID	STATION	STANAME
1	01A050	Nooksack R. @ Brennan	32	27B070	Kalama R nr Kalama
2	01A120	Nooksack R @ No Cedarville	33	27D090	EF Lewis R nr Dollar Corner
3	03A060	Skagit R nr Mount Vernon	34	31A070	Columbia R @ Umatilla
4	03B050	Samish R nr Burlington	35	32A070	Walla Walla R nr Touchet
5	04A100	Skagit R @ Marblemount	36	33A050	Snake R nr Pasco
6	05A070	Stillaguamish R nr Silvana	37	34A070	Palouse R @ Hooper
7	05A090	SF Stillaguamish @ Arlington	38	34A170	Palouse R @ Palouse
8	05A110	SF Stillaguamish nr Granite Falls	39	34B110	SF Palouse R @ Pullman
9	05B070	NF Stillaguamish @ Cicero	40	35A150	Snake R @ Interstate Br
10	05B110	NF Stillaguamish nr Darrington	41	35B060	Tucannon R @ Powers
11	07A090	Snohomish R @ Snohomish	42	36A070	Columbia R nr Vernita
12	07C070	Skykomish R @ Monroe	43	37A090	Yakima R @ Kiona
13	07D050	Snoqualmie R nr Monroe	44	37A205	Yakima R @ Nob Hill
14	07D130	Snoqualmie R @ Snoqualmie	45	39A090	Yakima R nr Cle Elum
15	08C070	Cedar R @ Logan St/Renton	46	41A070	Crab Cr nr Beverly
16	08C110	Cedar R nr Landsburg	47	45A070	Wenatchee R @ Wenatchee

ID	STATION	STANAME	ID	STATION	STANAME
17	09A080	Green R @ Tukwila	48	45A110	Wenatchee R nr Leavenworth
18	09A190	Green R @ Kanaskat	49	46A070	Entiat R nr Entiat
19	10A070	Puyallup R @ Meridian St	50	48A070	Methow R nr Pateros
20	11A070	Nisqually R @ Nisqually	51	48A140	Methow R @ Twisp
21	13A060	Deschutes R @ E St Bridge	52	49A070	Okanogan R @ Malott
22	16A070	Skokomish R nr Potlatch	53	49A190	Okanogan R @ Oroville
23	16C090	Duckabush R nr Brinnon	54	49B070	Similkameen R @ Oroville
24	18B070	Elwha R nr Port Angeles	55	53A070	Columbia R @ Grand Coulee
25	20B070	Hoh R @ DNR Campground	56	54A120	Spokane R @ Riverside State Pk
26	22A070	Humptulips R nr Humptulips	57	55B070	Little Spokane R nr Mouth
27	23A070	Chehalis R @ Porter	58	56A070	Hangman Cr @ Mouth
28	23A160	Chehalis R @ Dryad	59	57A150	Spokane R @ Stateline Br
29	24B090	Willapa R nr Willapa	60	60A070	Kettle R nr Barstow
30	24F070	Naselle R nr Naselle	61	61A070	Columbia R @ Northport
31	26B070	Cowlitz R @ Kelso	62	62A150	Pend Oreille R @ Newport

Near-surface grab samples are collected at all stations once each month. This sampling frequency was chosen in order to optimize the probability of statistically detecting trends while minimizing both auto-correlation between consecutive samples and the cost of collection (Lettenmaier, 1977). Sample collection generally occurs at a set time each month during a particular WY (*e.g.,* station x is sampled on the morning of the second day of the central region run which is done during the first full week of the month). Logistics sometimes require that the schedule change, particularly between WYs. In these cases, we may statistically account for the effects of time of sampling on results for temperature, dissolved oxygen concentration, and pH. One way of doing this is to monitor diurnal changes with *in situ* instruments. (Quality Assurance Project Plan (QAPP) and other protocol documents for this "continuous" monitoring are not included here.)

We generally sample 12 water quality constituents at each station (Table 8). We also record barometric pressure, which is used to determine percent oxygen saturation. At most long-term stations and some basin stations, flow is measured using one of several techniques. (Flow-monitoring is a separate program and will not be discussed further here; see http://www.ecy.wa.gov/programs/eap/flow/shu_main.html). At a few stations, when funding allows, we may also monitor metals concentrations. This monitoring also has a separate QAPP (Hopkins, 1995). We may at times conduct more frequent or more dense (more stations in a basin) monitoring in order to address specific needs (e.g., upstream/downstream turbidity measurements). In these cases, methods and QC requirements will be similar to those specified here.

Table 8. Water quality constituents monitored monthly as part of Ecology's ambient stream monitoring program.

Standard constituents monitored at all stations:		
electrical conductivity	suspended solids, total	phosphorus, total
oxygen, dissolved	turbidity	ammonia, total
pH	fecal coliform bacteria	nitrate + nitrite, total
temperature	phosphorus, soluble reactive	nitrogen, total

Representativeness

As discussed in objective 2), stations are not selected to represent any larger population, such as a watershed or the state as a whole. We do, however, have long-term stations in most of the state's 62 Water Resource Inventory Areas (WRIA) (Figure 2). These stations are usually located near the bottom end of the WRIA and thus are expected to represent the impact of cummulative effects in the watershed. Long-term stations may not necessarily be representative of "typical" water quality in the watershed except, perhaps, on a flow-weighted basis.

Water collection consists of a single, near surface water sample taken from bridges or, in a few cases from the riverbank, where the river appears to be well mixed vertically and horizontally. Although vertical heterogeneity of sediment-associated chemical species does occur, especially in large rivers, homogeneity is assumed for our objectives of characterization and trend analysis. This assumption should not be made for some constituents if our data are to be used for loading analyses.

Likewise, temporally, we assume 12 monthly samples are representative for our purposes. However, those using our data for annual loading estimates should evaluate this assumption, particularly for constituents that tend to be "patchy," like sediment, or highly associated with "first flush" effects, like total phosphorus.

The time of day when samples are collected is determined by the logistics of sampling all stations and delivering the samples to the lab for timely analysis. We attempt to sample a particular station at the same time each trip during a given WY, but the time of sample collection often changes between WYs. Results for constituents with large diurnal variations (temperature, pH, and dissolved oxygen) may need to be adjusted statistically to a common time prior to trend analysis, for example. Also, a single monthly measurement for these constituents will not necessarily capture a maximum or minimum. In WY 2001, at 42 stations where we deployed continuous temperature instruments, on average, our grab sample underestimated the summer maximum by 3.7°C and underestimated the maximum seven-day average of daily maximums by 2.9°C.

Comparability

All measurement and analytical procedures are documented so that the data will be comparable with samples collected and analyzed in a like manner.

Field Procedures

An overview of field procedures is presented below. Ward (2001; http://www.ecy.wa.gov/biblio/0103036.html) includes considerably more detail, particularly with respect to instrument calibration procedures.

Water samples for dissolved oxygen, pH, specific conductivity, and turbidity determinations are collected by lowering a stainless steel bucket (APHA, 1998) to about 30 cm below the water surface. The sampler is dropped quickly through the surface layer to minimize the collection of floating or micro-layer contaminants. Water for total suspended solids is collected directly in a sample bottle attached to the bucket. Water for nutrient analyses is collected in an acid washed bottle attached to the bucket. Water for fecal coliform bacteria determinations is collected by lowering an autoclaved bottle inserted into a bottle holder designed to orient the mouth of the bottle to the flow. The dissolved oxygen sample is collected in a 300 mL bottle held inside the bucket. Temperature is measured directly in the stream using a long-line thermistor.

Samples are returned to the van for processing within 5 to 10 minutes after collection. The sediment and bacteria samples are labeled and placed on ice. Dissolved oxygen is fixed by adding $MnSO_4$ and sodium azide to the bottle. The bottle is stoppered, capped with a water seal and stored in the dark. Dissolved oxygen samples are titrated (modified Winkler titration; APHA, 1998) upon returning to the office, from 12 to 96 hours after collection.

Aliquots are poured from the stainless bucket into cups for pH and conductivity measurements (Table 9), and into a Nalgene bottle that is sent to the lab for turbidity measurement. While the meters are equilibrating, the nutrient bottle is agitated and an aliquot poured into an acid washed Nalgene bottle for total nutrient determinations at the lab. The rest of the water in the nutrient bottle is filtered in the field through a 0.45 µm membrane filter into a brown Nalgene bottle for dissolved nutrient determinations at the lab. All samples requiring laboratory analyses are placed in the containers provided by the lab and labeled with the date, sample site, sample identification number (previously assigned by the lab for each sample), sampler's initials, and the chemical analyses requested. Preservatives, if required, are typically added to the bottle by the lab prior to sampling. Samples are then placed on ice and delivered to the lab according to procedures pre-arranged with the lab. Shipment of samples, preservatives, and sample holding times conform to Table 1 in Manchester's User's Manual (Ecology, 2002) (Table 10).

Field measurements and comments are recorded on a form prepared prior to the sampling trip (Ward 2001). Stream height measurements are also recorded.

Sampling equipment is rinsed thoroughly with de-ionized water after processing samples. The nutrient sampler is acid-rinsed. Detailed pre- and post-sampling cleaning and meter operation and calibration procedures are described in Ward (2001).

Table 9. Parameters measured in the field

Variable	Method	Resolution
Temperature	Thermistor	0.1°C
pH	Glass electrode	0.1 unit
Dissolved oxygen	Titration	0.1 mg L^{-1}
Electrical conductivity	Electrode	1 µS/cm

Table 10. Container type, water volume required, method of preservation and maximum permissible holding times for lab-analysed samples are listed below.

Determinand	Container Type	Sample Volume (mL)	Preservation	Holding Time
Turbidity	poly	500	cool to <4°C	48 hrs
Suspended Solids	poly	1000	cool to <4°C	7 days
Total Phosphorus	poly	125	Adjust to pH<2 w/ H_2SO_4 and cool to <4°C	28 days
Soluble Reactive Phosphorus	brown poly	125	filter in field and cool to <4°C	48 hrs
Nitrate+Nitrite-N	poly	125	adjust to pH <2 w/ H_2SO_4 and cool to <4°C	28 days
Ammonia-N	poly	125	adjust to pH<2 w/ $H_2SO_4$and cool to 4°C	28 days
Total Nitrogen	poly	125	adjust to pH<2 w/ $H_2SO_4$and cool to <4°C	28 days
Fecal coliform	Autoclaved glass or poly	250	cool < 4°C	24 hrs

Laboratory Procedures

Manchester Environmental Laboratory (MEL) conducts our laboratory analyses and laboratory procedures following Standard Operating Procedures and other guidance documents. Analytical methods and lower reporting limits are listed in Table 11.

Table 11. Laboratory analytical methods and reporting limits.

Analyte	Sample Matrix	Number of Samples[a]	Method	Reference[b]	Lower Reporting Limit
Ammonia-N	Total	984	Automated phenate	SM4500NH3H	0.01 mg L^{-1}
Fecal coliform	NA	984	Membrane filter	SM9222D	1 colonies 100 mL^{-1}
Nitrate+Nitrite-N	Total	984	Automated cadmium reduction	SM4500NO3I	0.01 mg L^{-1}
Soluble Reactive Phosphorus	Dissolved	984	Automated Ascorbic acid	SM4500PG	0.003 mg L^{-1}
Suspended Solids	Total	984	Gravimetric	EPA160.2	1 mg L^{-1}
Total Nitrogen	Total	984	Persulfate digestion, cadmium reduction	SM4500NB	0.025 mg L^{-1}
Total Phosphorus	Total	984	Persulfate digestion, ascorbic acid	SM4500PI	0.01 mg L-1
Turbidity	Total	984	Nephelometric	SM2130	0.5 NTU

[a] Approximate annual total based on 12 samples per station, 82 stations per year. Does not include quality control samples.

[b] SM=Standard Methods (APHA, 1998); EPA=Environmental Protection Agency (EPA, 1983)

Quality Control

Laboratory QC

MEL operates their own standard QC program, documented in Ecology (2002), Standard Operating Procedures for individual analyses, and their Quality Assurance Manual (Ecology, 2001). MEL's QC program includes the analysis of reference materials, check standards, duplicates, matrix spikes, and blanks.

Check Standards

Precision is addressed by the analysis of check standards (water with a known concentration of analyte) equal to about 10% of the total number of analyses. The mean value for a statistically significant number of check standard results may be used to judge whether there is any bias due to calibration. If the 95% confidence limit on the mean value does not include the true or reference value then bias due to calibration may be present.

Generally, calibration standards are set by MEL as needed to bracket the concentration in particular samples. The check standards should equitably span the range of the expected results, ideally approximately 0.2 and 0.9 of the upper value for the range of calibration. The historical ranges of our data are shown in Table 12.

Table 12. Historical ranges and 90th percentiles for stream monitoring data based on WY 1998 through 2002 (all stations).

Analyte	Expected Range of Results	Approximate 90th percentile
Ammonia-N (mg/L)	<0.01 to 1.97	0.033
Fecal Coliform Bacteria (colonies/100mL)	<1 to 17,000	120
Nitrate+Nitrite-N (mg/L)	<0.01 to 17.1	0.995
Soluble Reactive P (mg/L)	<0.003 to 2.14	0.045
Suspended Solids (mg/L)	<1 to 1970	41
Total Nitrogen (mg/L)	<0.025 to 16.5	1.2
Total phosphorus (mg/L)	<0.01 to 2.44	0.104
Turbidity (NTU)	<1 to 1,900	22

Analytical Duplicates
Laboratory sample splits are analyzed on one of each pair of field-split samples. Using the same sample that was split in the field allows us to better partition sources of error between lab and field. Frequently, MEL will split additional samples as well.

Matrix Spikes
Matrix interference leading to bias is assessed by analyzing river water that has been spiked with a known quantity of the analyte. The quantity of analyte added should not produce a final concentration that is excessively high when compared to the historic range of data (Table 12). Spike amounts should approximately double the concentration in the sample prior to spiking.

Blanks
MEL's QC program includes analyzing blank samples according to their internal protocols.

Field QC

Water is collected according to standard operating procedures that are updated as necessary and reviewed annually with field personnel (Ward, 2001). Stations designated for QC sample collection are selected randomly, one each month in each of the four sampling regions for a total of 48 QC stations. Eight of these (two per region) are designated as blanks and the remaining 40 are split (for most constituents) for a total of 88 QC samples. This is 9 percent of the approximately 984 samples collected statewide annually.

Replicates
Short-term, temporal variability is assessed by collecting two samples sequentially, 15-20 minutes apart at 10 of the 12 monthly QC stations per region per year (the other two months are designated for blank samples). Results from the first sample are stored as the standard results. The second set of results is labeled as a QC sample, though the station is not identified to the lab. The difference between these results is used to calculate the expected variance that is due to

short-term in-stream factors, field collection and processing, and laboratory analyses. (The lab also splits this sample; see "Laboratory QC", above.)

For constituents receiving field processing (Table 13, footnote), the duplicate sample is split into two separate sub-samples for field measurements and processing, and submission to the lab. One set is given the "duplicate" label and the other is labeled "split" for data management purposes. These field-splits are used to calculate the variance that is due to field collection and processing, and laboratory analyses. The difference between split sample variance and the original sample variance is due to short-term in-stream processes.

Blanks

Sample contamination is assessed by the submission of eight field blanks at random intervals throughout each year. These are "transport blanks" for constituents where there is no field processing of the sample (total suspended solids), and "rinsate blanks" for other constituents. Fecal coliform bacteria blanks are not included because our blank water is not sterile. Specific procedures are specified in Ward, 2001. Blanks results are expected to be below reporting limits.

Instrumentation

The pH and conductivity meters are calibrated daily according to the manufacturer's directions. The pH meter is checked immediately after calibration, at midday, and at the end of the day by recording the measurement of a low ionic strength pH 7 buffer. It is also checked whenever a measurement exceeds water quality standards criteria (WAC 173-201A). If the difference between the meter measurement and the expected pH exceeds 0.10, the instrument is recalibrated and the sample re-measured (see MQOs, table 3). The conductivity and temperature meters are relatively stable; the conductivity calibration is generally checked only at the end of the day (100 µS calibration and check standards), and temperature calibration at the beginning of the sampling trip.

Table 13. Field quality control samples required annually (lab QC is specified in MEL guidance documents). QC samples are divided equally among the four sampling regions.

QC-type	Field-processed constituents [a]	Constituents without field processing [b]
Field duplicate	40	40
Field split (of duplicate)	40	0
Field blank [c]	8	8

[a] Conductivity, total phosphorus, oxygen, turbidity, total ammonia, total nitrite plus nitrate, soluble reactive phosphorus, total nitrogen.

[b] Suspended solids, fecal coliform bacteria, temperature, and pH. (Although pH measurements involve field processing, they are included in this category because samples cannot be split and measured consecutively without introducing error.)

[c] Blanks are not measured for oxygen, pH, fecal coliform bacteria, or temperature.

All meters are maintained in accordance with the user's manuals. Critical equipment and supplies are listed on a check-sheet and are the responsibility of the field personnel.

Corrective Action
The laboratory continually monitors their results for quality control sample determinations and takes appropriate action to correct problems. Frequently, samples may be re-analyzed after an analytical problem is corrected. This is also the case for field measurements with respect to check standard results. Due to sample holding time limitations, re-analysis is usually not possible if problems are discovered in field QC data. Corrective courses applying to subsequent data collection are possible, however.

If data are compromised due to poor precision, the source of the variability will determine the course of action that is required. Possible actions include 1) changing the standard operating procedures or instrumentation for field personnel; 2) informing the laboratory when lab error appears to be the source (and possibly changing analytical methods); and 3) re-evaluating the required precision, when it appears that the required $s_{error(mp)}$ is unattainable.

A persistent, consistent bias in the data may warrant adjusting the values, otherwise the corrective action for bias will be to inform the lab, which will be expected to address the problem. Significant changes in methods, instrumentation, or protocols will be made only after it has been documented that these changes will not bias the data.

Data Management Procedures

Our data are managed in an Access® database. Constituents measured in the field are recorded manually on a standard form and entered by the sampler into a temporary Access table upon return to the office. Rough validation rules prohibit obviously incorrect data from being entered. A hardcopy of the temporary table is printed and the sampler reviews the data prior to moving it into the final results table (also see the section “Data Review”).

Lab data, once available, are exported from the interface portion of Ecology's EIM system and loaded into our database. Station, date, and time must match the field data entered earlier. Characteristic, method, units, and sample fraction must match entries in a "parameters" table.

After field and laboratory data are entered, a two-tiered evaluation of results is performed (see the "Data Review, Verification, and Validation" section). Data deemed of acceptable quality are uploaded monthly as "preliminary" data to the web and quarterly to the EIM system. After the annual comprehensive QC review (published in our Annual Reports), data on the web are moved from preliminary to final.

Reports

Ecology management requires that fecal coliform bacteria results exceeding 200 colonies/100mL be reported via email to various regional and headquarters staff as soon as results are available. Regional personnel can then notify the appropriate local health agency, should they deem that appropriate. These reports are identified as being based on preliminary data.

After a full month's data are available, all results are reported to our web site, and all results exceeding water quality criteria or the usual range of results from a particular station and season

are reported to Ecology management and to our web site. These reports are identified as being based on preliminary data.

Upon completion of the WY's data collection activities, the previous year's program is summarized in an annual report (*e.g.*, Hallock, 2002). This report includes an analysis of field and some lab QC data collected during the year as well as an appendix listing known changes to the monitoring program that could potentially affect the data.

Data Review, Verification, and Validation

The laboratory verifies the data prior to reporting them to us. This includes an on-going evaluation of their QC results (using control charts, etc.). "Case Narrative" reports are included as part of each data package. Once a full month's data are received from the lab, we conduct a two-tiered validation process. The first tier consists of a computer assessment of the data and associated field QC data:

1) Each result is compared to historic data from that station collected during the same season. (Four seasons are defined: January-March, April-June, July-September, and October-December.) The datum is 'flagged' if it lies more than 2.5 standard deviations from the mean.
2) The values of replicated samples are flagged if the coefficient of variation of the replicates or split samples exceeds 20%.
3) The datum is flagged if the holding time was exceeded.
4) If internal logic checks (total phosphorus greater than soluble reactive phosphorus or total nitrogen greater than nitrate/nitrite plus ammonia) are violated, then all data values involved are flagged.

The second tier is a manual inspection and evaluation of each datum flagged by the first tier evaluation. Case Narratives provided by the lab are reviewed and questionable results confirmed with laboratory personnel. Quality Codes are assigned based on best professional judgment as follows:

1) No first tier checks were exceeded.
2) The datum has not been reviewed. (Used primarily for data that were entered into the database before this QC program was implemented.)
3) One or more first tier checks were exceeded but the second tier review indicated that the datum was 'OK.'
4) One or more first tier checks were exceeded and the second tier review was not conclusive.
5) One or more first tier checks were exceeded and the second tier review indicates that the datum was probably not 'OK.' Datum is usually not reported or used in subsequent statistical analyses.
6) One or more first tier checks were exceeded and the second tier review is currently pending.
7) Not currently used.
8) Not currently used.
9) Datum is very suspect and should not be used.

Data coded greater than "4" are not routinely reported or used in data analyses.

In addition to the procedures, above,

- Printouts of field data entered into the database are compared to the original field forms quarterly,
- Missing data are evaluated annually (using a standard report produced by our database) to ensure that no data are missing due to data management oversight (*i.e.*, all missing data can be explained), and
- We conduct an annual QC review. (See the next section for assessment procedures.)

These verification and validation steps are the responsibility of the data manager.

Missing Data

Missing data are rare. In a recent 5-year period at long-term stations, 98.8 percent of expected data were collected. Of the data collected, only 0.2 percent had quality problems severe enough to exclude them from routine use. The majority of "missing" data are due to mechanical breakdown, inaccessible sample sites, and samples lost or misplaced during transport by commercial carriers. The effects of sample size, *n*, on statistical trend analysis is discussed under "Quality Objectives." Strict adherence to standard operating procedures and clear communication between field and laboratory personnel are the best measures to prevent lost or misplaced samples. Because ambient monitoring is an ongoing process, the loss of a small percentage of the data from a long-term station will have little impact on the overall objectives.

Data Quality Assessment

Result-level data validation procedures are conducted monthly as described in the "Data Management" section. Batch-level QA assessments a made by comparing calculated percent relative standard deviations (%RSD) (Equation 5) to those specified in our MQOs (Table 3).

$$\%RSD = 100\frac{s}{\bar{x}} = 100\frac{\sqrt{(r_1 - r_2)^2 / 2}}{(r_1 + r_2) / 2} \qquad 5)$$

Where s is the standard deviation, $\bar{x}$ is the mean, and r_1 and r_2 are paired results, typically a known value (e.g., of a check standard) and the analytical result or measurement of the known value. Duplicate measurements of environmental samples may also be used to estimate precision of the analytical method, but this can include error due to matrix effects. (RSD is also known as the coefficient of variation.)

The results of the analysis of blank samples and known standards will be used to determine overall bias of the results. If a consistent method bias is discovered, even one less than the levels specified in Table 3, we should be notified prior to correction because even small changes can

affect trend analysis. Bias due to time of day of collection will be addressed on a site- and variable-specific basis as described previously (see "Representativeness").

Project-level QA assessments are conducted as part of our annual reporting process. Sources of error (lab, field, short-term in-stream) are identified to the extent possible as outlined in the "Quality Objectives" section. For constituents failing our DQOs, the central tendency of the variance of sample pairs may be grouped and compared by station, season, sampler, etc., in order to identify stations, time periods, etc., that are correlated with poor precision.

The central tendencies of the variance of sample pairs are summarized by calculating the square root of the mean of the sample-pair variances (root mean square (RMS), Equation 6). This estimate provides an unbiased—and higher—estimate than other commonly used statistics (for example, mean or median of the standard deviations). Because the variability of many parameters increases with increasing mean concentration, the RMS values of some variables will be evaluated according to concentration ranges. These results ($s_{error\ (att)}$) are then compared to requirements listed in Table 5 ($s_{error\ (mp)}$).

$$RMS = (s^2_{avg})^{0.5} \qquad 6)$$

where s^2_{avg} is the average of the variances of the paired results.

Trend Power Assessment

Whether or not trends can be detected in any particular case may be estimated for individual data sets by comparing the actual s_{obs} (after removing as much explainable variability as possible—deseasonalize, detrend, etc.) and the required s_{obs} determined by re-arranging Equation 2 ($s_{obs} = \Delta\mu/\delta$). See the Caution, below, however.

If s_{obs} for a particular (normally distributed) data set is greater than the calculated s_{obs} from Equation 2, one will be unlikely to detect a trend at the given $\Delta\mu$ and δ. One may then

- Improve field or laboratory methods to reduce error. This will reduce the variability in future data not existing data, of course. Also, if the proportion of variance due to error (ϕ) is already low, reducing s_{error} may not have much affect on s_{obs}.
- Modify expectations (decrease required confidence or increase the expected $\Delta\mu$)
- Collect (or include) more data (increase *n* thereby decreasing δ).

Caution: This power analysis is an approximation based on parametric statistics. In theory, non-parametric trend techniques are nearly as powerful as parametric methods and more so if the underlying data do not meet parametric assumptions (Hirsh *et al.*, 1991). Also note that if a data set is not normally distributed, the s_{obs} of the untransformed data may appear very large and may not accurately predict attainable $\Delta\mu$. The less normally-distributed the data, the worse the prediction will be. The predicted $\Delta\mu$ for nutrient data, for example, may be high by orders of magnitude. See Hallock (2003) for more on this phenomenon and a suggestion to account for non-normality when predicting detectable trend magnitude.

Equation 2 and Figure 1 or Table 4 can be used to estimate either the size of the trend that can be detected for a given data set or the number of independent samples needed to detect a trend of a given size. An example using dissolved oxygen data from one of our stations is shown in Table 14.

Table 14. Example for estimating required trend magnitude ($\Delta\mu$) or sample size (n) to enable trend detection in a dataset (oxygen in mg/L at station 13A060, 1991/09/01-2002/09/01).

Given:

Observed standard deviations
Original: 0.96
Detrended/deseasonalized (s_{obs}): 0.66
Mean = 10.45

To estimate trend magnitude that can be detected with n = 120 (from equation 2):

$\Delta\mu = \delta * S_{obs} = 0.93 * 0.66 = 0.61$
(This is 5.9% of mean over the ten-year period.)

To estimate the required number of samples to detect a trend magnitude of 10% (from figure 2):

Desired trend magnitude ($\Delta\mu$): 0.10 * 10.45 = 1.045
$\delta = \Delta\mu / s_{obs} = 1.045 / 0.66 = 1.58$
From Figure 2, this yields approximately n=43 or 3.6 years of monthly sampling.

This analysis assumes the data are normally distributed and without significant auto-correlation. Non-normal data will overestimate $\Delta\mu$ while significant auto-correlation will underestimate it.

Characterizing Water Quality and Analyzing Trends

Specific data analysis techniques vary depending on the history of the watershed (e.g., step vs. linear trends), the specific objectives of an analysis (e.g., reporting water quality standards criteria violations, general characterization, evaluation of management activities), the spatial scope of the report (e.g., statewide, single station, watershed), and so on. Our analyses typically use graphical displays such as time series, cumulative frequency, seasonal box, and other plots, as well as statistical (often non-parametric) techniques like the seasonal Kendall trend test. The software we use most often is WQHYDRO (Aroner, 2002). See Hallock (2002c) for an example.

References

APHA, 1998. Standard Methods for the Examination of Water and Wastewater, 20th edition. American Public Health Association, Washington, D.C.

Aroner, Eric, 2002. WQHYDRO - *Water Quality/Hydrology/Graphics/Analysis Package*. Portland, OR.

Butkus, S., 2002. Washington State Water Quality Assessment, Year 2002 Section 305(b) Report. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. Publication No. 02-03-026, 32 pp.

Ecology, 2001. The Quality Assurance Manual for the Washington State Department of Ecology Manchester Environmental Laboratory. Environmental Assessment Program, Manchester, WA. 89 pp.

Ecology, 2002. Manchester Environmental Laboratory, Lab User's Manual, 6th Edition, Environmental Assessment Program, Manchester, WA.

Environmental Protection Agency. 1983. Methods for Chemical Analysis of Water and Wastes, EPA-600/ 4-79-020. U.S. Environmental Protection Agency, Cincinnati, Ohio, USA.

Environmental Protection Agency. 2002. Guidance on Environmental Data Verification and Data Validation, EPA/240/R-02/004. U.S. Environmental Protection Agency, Washington, DC, USA.

Ehinger, W.J., 1995. Freshwater Ambient Water Quality Monitoring, Final Quality Assurance Project Plan. Washington Department of Ecology, Environmental Investigations and Laboratory Services Program, Ambient Monitoring Section, Olympia, WA. 23 pp. +appendices

Hallock, D. 2002. *River and Stream Ambient Monitoring Report for Water Year 2001*. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. Publication No. 02-03-047, 10 pp. + appendices.

Hallock, D. 2002b. *A Water Quality Index for Ecology's Stream Monitoring Program*. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. Publication No. 02-03-052, 17 pp. + appendices.

Hallock, D. 2002c. *Water Quality Assessment of the Nooksack River between Brennan and North Cedarville*. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. Publication No. 02-03-037, 25pp.

Hallock, D. 2003. *River and Stream Ambient Monitoring Report for Water Year 2003*. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. In Draft.

Hirsch, R.M., R. B. Alexander, and R.A. Smith. 1982. "Selection of methods for the detection and estimation of trends in water quality." Water Resources Research 27(5): 803-813.

Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. "Techniques of trend analysis for monthly water quality data." Water Resources Research 18(1): 107-121.

Hopkins, B., 1993. Freshwater Ambient Monitoring Report for Wateryear 1991. Washington State Department of Ecology, Ambient Monitoring Section, Publication #93-75, Olympia, WA

Hopkins, B., 1995. Ambient Metals Project Proposal; Final Quality Assurance Project Plan. Washington State Department of Ecology, Ambient Monitoring Section, Olympia, WA. 19 p. + appendices.

Lettenmaier, D.P., 1977. Detection of Trends in Stream Quality: Monitoring Network Design and Data Analysis. Technical Report No. 51, University of Washington, Department of Civil Engineering, Seattle, WA.

McDonald, M.E., *et al.*, 2002. Research Strategy, Environmental Monitoring and Assessment Program, EPA/620/R-02/002. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Smith, D.G., G.B. McBride, G.G. Bryers, R.J. Davies-Colley, J.M. Quinn, and W.N. Vant, 1989. A National Water Quality Network for New Zealand. Water Quality Centre, Department of Scientific and Industrial Research, Hamilton.

Ward, W. 2002. Stream Sampling Protocols for the Environmental Monitoring and Trends Section. Washington Department of Ecology, Environmental Assessment Program, Olympia, WA. Publication No. 01-03-036, 31 pp. + appendices.

Wrye, D. 1993. Basin Approach to Water Quality Management: Program Description. Washington State Department of Ecology, Olympia, WA.