Updated Statistical Analyses of Water Quality Data, Compliance Tools, and Changepoint Assessment for Montana Rivers and Streams

1.0 Introduction

This report presents an updated statistical analysis of water quality data from Montana rivers and streams. The purpose of the statistical analyses is to support the development by the Montana Department of Environmental Quality (MT DEQ) of nutrient and nuisance-algae criteria for flowing waters. The analysis builds upon research and analyses documented in four previous reports:

- A July 2004 report (ICF 2004a) analyzed the relationships between ambient surface water quality and selected environmental and geospatial characteristics of Montana rivers and streams;

- A November 2004 report (ICF 2004b) analyzed the relationship between the statistical distributions of water quality parameter observations at designated reference locations and in the whole population of observations from water quality monitoring locations in various classes of Montana rivers and streams;

- A June 2005 report (ICF 2005) analyzed regionally based stratification methodologies for water quality criteria determination in Montana rivers and streams. This analysis identified Omernik level III[1] ecoregions as the primary stratifying parameter; and

- A June 2006 report (ICF 2006) computed bootstrap-based[2] confidence intervals around percentiles of the reference reach distribution that were likely to be of regulatory interest.

In the current analysis, selected elements of the previous analyses have been revised and improved with an updated and expanded database. In addition, the bootstrap confidence interval generation methodology of June 2006 has been extended to level IV ecoregions in cases where specific Omernik level IV ecoregions are statistically distinct from their parent level III ecoregions.

Also provided with this report is an updated geospatial nutrient dataset that was created according to MT DEQ specifications.

[1] Omernik (1987) developed a system to classify geographic areas encompassing generally similar ecological features. In particular, Omernik ecoregions are defined based on geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology. This classification system includes a hierarchical framework designated by roman numerals. For example, North America is divided into 15 level I ecoregions (i.e., the coarsest scale in the hierarchy) and 52 level II ecoregions. Based on empirical analyses conducted in previous projects, MT DEQ selected level III and level IV Omernik ecoregions in Montana as the primary stratifying parameters for this project.
[2] Bootstrapping, a statistical technique for computing confidence intervals, is defined further in Section 4.2.2.

The report also documents two new tools developed in the current phase of this project: (1) a spreadsheet-based tool to support the assessment of compliance and non-compliance with nutrient criteria in Montana rivers and streams[3], and (2) a changepoint assessment tool to identify data points that effectively partition a dataset into two distinct groups.

The remainder of this report is structured as follows:

- Section 2 of this report describes the scope and objectives of the statistical analysis.
- Section 3 identifies the sources of data used for the statistical analysis and explains how the data were compiled into a geospatial dataset.
- Section 4 describes the statistical methods used in the analysis.
- Section 5 presents the results and conclusions of the statistical analysis.
- Section 6 describes the development of the compliance and non-compliance tools.
- Section 7 documents the technical basis for the development of the changepoint assessment tool.
- Section 8 discusses compliance with nutrient criteria by permitted point source.
- Section 9 identifies the literature cited in this report and associated appendices.
- Appendix A documents the methodologies used to group water quality parameters.
- Appendix B documents the methodologies used to group data by seasons.
- Appendix C describes the lithologic grouping methodology.
- Appendix D provides the detailed results log files of the statistical analysis.
- Appendix E provides documentation for the datasets used for this project.
- Appendix F presents a nutrient criteria developer manual with all programming language used in the current phase of this project.
- Appendix G presents a nutrient criteria user guide to facilitate the generation of new results in the future.
- Appendix H describes statistical considerations involved in the application of numeric nutrient criteria and presents decision rules for determining compliance and non-compliance with the proposed numeric criteria.
- Appendix I provides a further discussion of point source compliance with numeric nutrient criteria.

A CD accompanying this report contains the updated datasets, computer generated output or log files that form part of Appendix D, computer program files that constitute part of Appendix F, the compliance assessment tools, and the changepoint assessment tool. The following sections provide more detail on these various elements.

2.0 Scope and Objectives of the Statistical Analyses

The U.S. Environmental Protection Agency's (EPA's) guidance on nutrient and algal criteria development (EPA 2000) recognizes the need for identifying stream groups with comparable biological, ecological, physical, and chemical features, because natural levels of nutrient

[3] This tool was in fact developed as two separate tools for the convenience of users. The first is for the purpose of establishing compliance, while the second is for establishing non-compliance with standards.

concentrations in streams are likely to be related to these factors. Spatially defined nutrient zones developed by grouping streams with similar characteristics are therefore an appropriate mechanism for setting nutrient criteria that reflect natural variability. A previous phase of this project (ICF 2005) provided an empirical basis for concluding that Omernik level III and level IV ecoregions constituted statistically significant stratifying parameters and were best suited for stratifying the state of Montana into nutrient zones.

EPA guidance further recommends that the 75th percentile of the frequency distribution of water quality measurements at reference stream reaches within a nutrient zone be used to develop nutrient criteria for that zone. In the absence of data for reference stream reaches, EPA guidance recommends choosing from the 5th to 25th percentile of the frequency distribution of the whole population of a class of streams within a nutrient zone to develop the criteria.[4] Distributional analyses in a previous phase of this project (ICF 2004b, ICF 2005) demonstrated that the relationship between the reference and general population distributions is not predictable. It was therefore concluded that a criteria based on the reference population is a better choice than a criteria derived from the general population. Nutrient concentrations derived from regional, scientific impact-to-use case studies were then compared to their corresponding reference distributions and MT DEQ concluded that the 75th percentile of the reference population may be an excessively cautious standard for nutrients. MT DEQ's analysis indicated that a slightly higher percentile of the reference distribution may be appropriate to set nutrient standards (Suplee et al. 2007), for example the 90th percentile of reference. The most recent previous phase of the project (ICF 2006) computed the 80th, 85th, and 90th percentiles of the reference distributions within each nutrient zone for each nutrient of interest and derived statistical confidence intervals around each percentile using a bootstrap approach as well as a station-balanced approach.

The current statistical analysis went a step further than the previous round and used a more refined stratifying parameter — when it is statistically meaningful — in defining nutrient zones. Specific level IV ecoregions within a level III ecoregion were tested to determine if they were statistically distinct from their parent level III ecoregion. If yes, the analysis computed the percentiles of regulatory interest for those level IV ecoregions; the analysis also generated bootstrap confidence intervals around those percentiles of interest for those level IV ecoregions. This analysis was conducted for reference reaches only using station balanced median data; the analysis was performed for all nutrients and for all seasons.

The current analysis also regenerated summary statistics at the coarse and fine-scale for a range of stratifying parameters using the updated database; these summary statistics are also presented for all nutrients by season.

[4] Hereafter, this report refers to data for the reference stream reaches within a nutrient zone as the "reference population." We refer to data for non-reference stream reaches as the "non-reference population." The population consisting of both reference and non-reference stream reaches within a nutrient zone is referred to as the "whole population" or "general population".

3.0 Data Sources and Database Development

The first step of this analysis was to compile sources of water quality data and selected environmental and geospatial characteristics into a single relational database. Most of the water quality data were already available in a database compiled for the previous phases of the project (ICF 2006).[5] The 2006 version of the database (referred to hereafter as the "2006 database") contained 143,983 observations. For the current phase of the project, MT DEQ had gathered new data and developed a list of corrections/improvements to be made to the 2006 database. ICF undertook the following steps to construct the updated database:

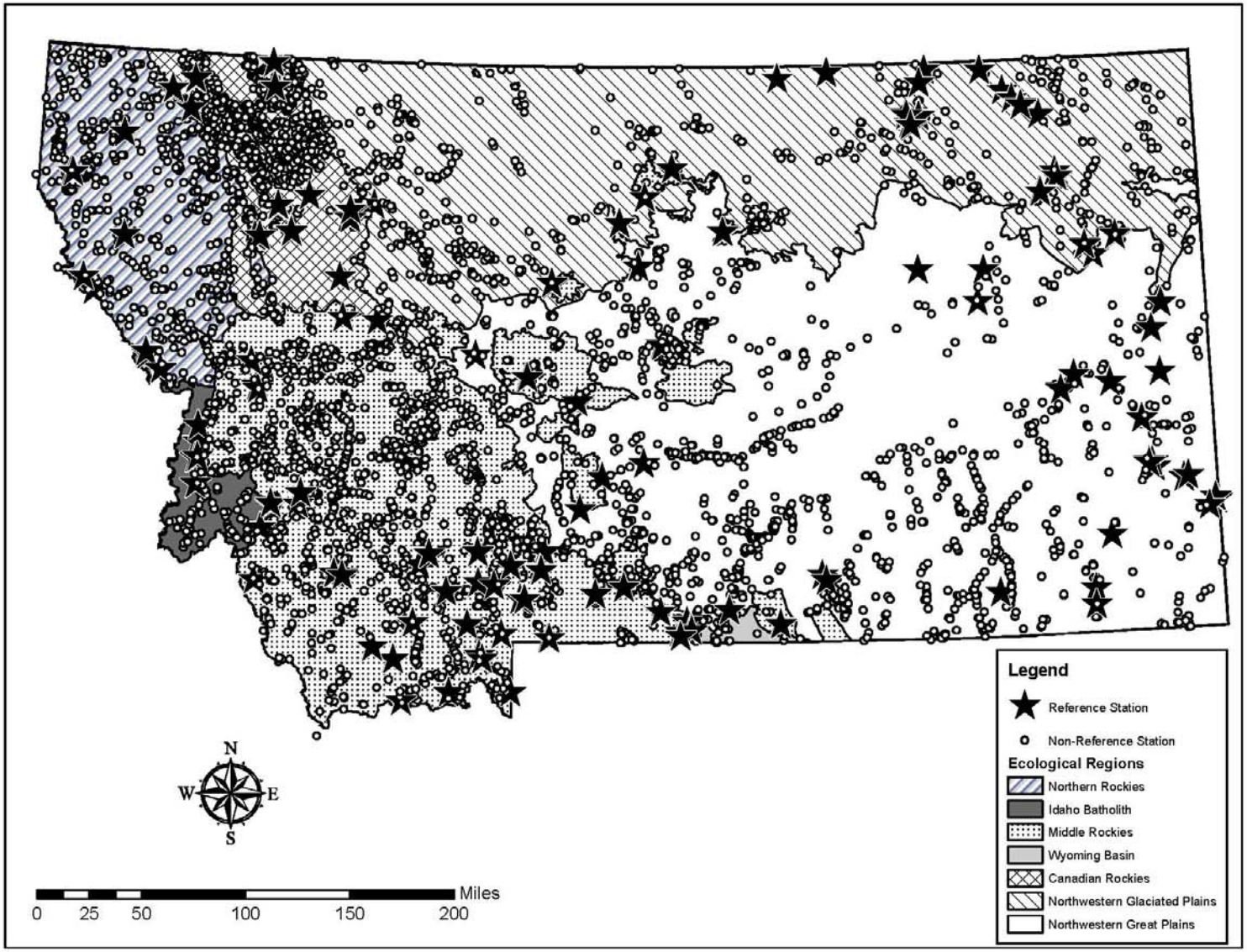
- A total of 465 observations in the 2006 database contained metadata (on spatial attributes such as ecoregions and geology) but no latitudes and longitudes. This was the result of a data upload error in the previous round, and was corrected by referencing the 2005 database.
- MT DEQ had found a total of 428 benthic chlorophyll observations all originating in the year 2004 to be defective. MT DEQ provided new data to replace these defective data.
- A total of 13,471 observations in the 2006 database lacked latitude and longitude information. None of these data were reference data. As a result, these observations did not contribute towards many of the statistical analyses. In this round of the analysis, MT DEQ provided spatial coordinates for several of the locations previously lacking these data. After adding coordinate data provided by MT DEQ, a total of 7,431 observations still did not have spatial coordinates. These 7,431 observations were removed from the database per MT DEQ instructions.
- MT DEQ provided a total of 2,815 new observations (including the 428 benthic chlorophyll overwrites noted in the second bullet above.) These data were spatially joined to the appropriate GIS layers to incorporate data on their Omernik level III and level IV classifications, as well as information on underlying geological formations and altitude.
- A limited number of corrections were made to the reference status of some field sampling stations.
- A small number of duplicate observations were dropped from the database.

After incorporating the new data described above, the new nutrient data were grouped following MT DEQ instructions. The new data received geological groupings in a manner identical to development of the 2006 dataset. A new set of definitions developed by MT DEQ was applied to define seasons for all data. More details on these operations is provided below.

Figure 3.1 displays the location of the non-reference population and reference population stations in the new database.

[5] See the four previous reports (ICF 2004a, 2004b, 2005, and 2006) for details about the sources of data used to create the database.

Figure 3.1 Reference and Non-Reference Sampling Stations



Salient information about spatial joins and data groupings used in the database is provided below:

- New water quality data, which included latitude and longitude coordinates for each observation, were spatially joined to GIS layers containing information on the stratifying parameters of interest. Using a spatial join, each observation was assigned appropriate values for each stratifying parameter depending on the location of the sampling site. The stratifying parameters thus incorporated into the database include level III ecoregions, level IV ecoregions, geologic formation, and elevation. Elevation does not play a role in this analysis but was included in the database for potential later use by MT DEQ.

- The previously compiled water quality data, which were obtained from Legacy STORET, were collected by various agencies using various methods. ICF used a methodology developed by MT DEQ to group related water quality parameters together. The objective of the water quality parameter grouping methodology was to group fundamentally equivalent analytic measurements, while avoiding double counts. A detailed explanation of the parameter grouping methodology is provided in Appendix A.

- The water quality data were seasonally defined as Winter, Runoff, or Growing using a methodology developed by MT DEQ. As noted above, the definitions of the seasons are slightly different from the previous phases of the analysis. A detailed explanation of the new seasonal grouping methodology is provided in Appendix B.

- MT DEQ had previously provided data on lithologic classifications (e.g., rock types) within Montana (ICF 2004b). MT DEQ supplied the data in a GIS shapefile on a CD; the shapefile was originally developed by Raines and Johnson (1996). The lithological classification shapefile contained over 80 groupings (i.e., rock types). After completing the GIS-based spatial join, a methodology developed by MT DEQ was used to combine the rock-type groupings into 19 major lithologic categories. The methodology for this step is described in Appendix C. Lithologic classifications are also referred to as geologic formations in this report.

- To eliminate potentially erroneous or highly uncertain data from the statistical analysis, water quality data associated with certain comment codes[6] in the Legacy STORET data were excluded from the database. For example, these codes denoted estimated values or values from analyses known to be in error.

- Legacy STORET data bearing comment codes[7] denoting non-detects were replaced with values equal to 50 percent of the reported detection limits. This approach is consistent with regulatory guidance (EPA 2006).

[6] Specifically, the data eliminated were Legacy STORET data bearing comment-codes *H, O, Q, L, Y* and *"*"*; USGS data bearing comment code *E*. Note: In previous rounds of the analysis, estimated data bearing comment code J (with the exception of benthic chlorophyll-A data bearing comment code J) were also eliminated from the analysis. No effort was made to recover these eliminated J data in this analysis because they were determined to be few in number. See details concerning the J flag on page A-2, Appendix A.

[7] The specific legacy STORET comment codes included *T, M, W, K, U, ND,"<" and "Non-detect."*

- Also eliminated from the database were all water quality parameters that had reported values of zero. Most analytical results in the database provided a result value and a detection limit, or an indication that the analyte was below the detection limit. True analytical values of zero are very unlikely (NWQL 1994), and therefore these data (most of which are old) were not used.

The resulting database, referred to hereafter as the "all-observations database," contains 138,914 observations.

The all-observations database was then processed to develop a "median database" as an alternative basis for some analyses. The median database contains a unique observation for each combination of nutrient, field sampling station, and season. These observations are the median observed values for a nutrient, at a particular station, in a particular season. The median database is less likely to be influenced by outliers and is more amenable to parametric statistical analysis. The median database contains 43,386 observations.

4.0 Statistical Methods

This section discusses the statistical methods employed in the following components of the analysis:

- Generating summary statistics; and

- Estimating percentiles of regulatory interest and generating confidence intervals around the percentiles of interest in the reference population for level III ecoregions and also for those level IV ecoregions that are distinct from their parent level III ecoregions.

These analyses were conducted by means of computer programs that performed a series of statistical calculations. The programs were run for each water quality parameter for various combinations of stratifying parameters by season. Coarse-scale stratifications refer to stratifications based on a single stratifying parameter. Fine-scale stratifications refer to stratifications based on two stratifying parameters, one embedded or sub-stratified within the other. The results of the program runs are presented in Appendix D, along with a guide to their interpretation.

Sections 4.1 and 4.2 provide further explanation of the two components of the statistical analysis of the rivers and streams database. Section 4.3 describes quality assurance measures for the statistical results.

4.1 Summary Statistics for the Reference, Non-Reference, and Whole Populations

Summary statistics in the computer outputs (referred to as log files) in Appendix D characterize the reference, non-reference, and whole population statistical distributions for the selected water quality parameter for each category of the stratifying or sub-stratifying parameter. Statistics are

reported separately for three databases: the all-observations database; the all-observations database excluding stream orders 7 and 8; and the median database. Specific summary statistics include the total number of observations, minimum, maximum, mean, standard deviation and skewness for reference observations, non-reference observations, and whole population observations. In addition, the summary statistics include the values of the 10th, 25th, 50th, 75th, and 90th percentiles for reference observations, non-reference observations, and whole population observations. These statistics are presented for all nutrients in turn by season for various combinations of stratifying and sub-stratifying parameters. In order to improve the accessibility of these statistics compared with previous phases of this project, the summary statistics are reported in separate log files (see Appendix D) and are not combined with the results of any other tests.

4.2 Observed Values and Confidence Intervals for Percentiles of Interest

As described in Section 2, percentiles of regulatory interest to MT DEQ were revised to include the 75th, 80th, 85th, and 90th percentiles of the reference population. As a first step, the analysis determined the observed values of these percentiles in the reference population in coarse-scale nutrient zones based on level III ecoregions using the median database. In order to determine confidence limits around these observed values, the analysis then used a bootstrapping technique, similar to that used in the previous phase of this project. The analysis proceeded by season within each level of the coarse-scale stratifying parameter (level III ecoregions, in this case). In addition to level III ecoregions, those level IV ecoregions that were determined to be statistically distinct from their parent level III ecoregions were also analyzed using the same methods. The analysis was performed for all nutrients, by season, for the reference population, using the median database only. Each step of the procedure is described separately below.

4.2.1 Observed Values of Percentiles of Interest

Due to a limitation in the pre-defined statistical functions available in the software application used for the analysis, it was necessary to write a user-defined function to compute the 80th and 85th percentiles. For consistency, this function was also used to compute the 75th and 90th percentile, although these percentiles may be computed using the software's pre-defined functions. There is no universally accepted definition of a percentile, and different algorithms are available for percentile estimation. The user-defined function used the following relationship (Lenter 1982):

$$R(p) = 0.5 + (p*n/100)$$

Where,

p = the pth percentile, for instance 85
n = the number of observations in the dataset
R(p) = the observation rank corresponding to the pth percentile, when the observations have been arranged in ascending order

If R(p) is an integer, then the pth percentile is simply the value of the $R(p)^{th}$ ranked observation. If R(p) includes a fraction, the pth percentile lies between the values with ranks adjacent to R(p) and may be calculated by linear interpolation as follows:

$$p^{th} \text{ Percentile Value} = \text{Obs}(\text{Int}(R(p))) + [R(p)-\text{Int}(R(p)] * [\text{Obs}(\text{Int}(R(p))+1) - \text{Obs}(\text{Int}(R(p)))]$$

Where,

Obs(x) = the value of the observation occupying the xth rank, when the observations have been arranged in ascending order
Int(y) = the integer portion of the number y.

The user-defined function was tested and verified with a second software application (for all percentiles of interest) and was found to produce identical results in the two applications in the vast majority of cases. In datasets with very few observations (fewer than 6), the results of the function differed slightly, but the differences were extremely minor and attributable to differences in the underlying algorithms.

4.2.2 Bootstrap Confidence Intervals for Percentiles of Interest

Observed values of the percentiles of interest provide a single estimate based on the available sample in a nutrient zone. Further sampling in that zone would likely result in different estimates of those percentiles; comprehensive sampling to determine the true population parameters is infeasible. Confidence limits for a statistical parameter are an interval estimate for that parameter. Interval estimates are desirable because the estimate of a statistical parameter can vary from sample to sample. Instead of a single estimate for a statistical parameter, a confidence interval generates a lower and upper limit for the parameter. The interval estimate gives an indication of how much uncertainty there is in the estimate of the true value of the parameter. The narrower the interval, the more precise is the estimate. In this analysis, the 90 percent confidence interval is reported for each percentile of interest. According to a branch of statistical science called frequentist statistics, the interpretation of a 90 percent confidence interval is that the confidence interval represents a range of values for the population parameter for which the difference between the parameter and the observed estimate is not statistically significant at the 10 percent level. In another branch of statistical science, Bayesian statistics, the interpretation of a 90 percent confidence interval is that there is 90 percent confidence that the true population parameter lies in the confidence interval. This analysis uses the Bayesian interpretation when referring to confidence intervals.

Many methods exist for computing confidence intervals for statistical parameters with mathematically uncomplicated distributions. For instance, the sampling mean of a population can be described by a mathematical probability density function known as a t distribution with n-1 degrees of freedom. Computing confidence intervals analytically for descriptive statistics with more complicated distributions – such as the median and specific percentiles – is considerably more challenging (Haukoos 2005).

Bootstrapping is a computationally intensive statistical technique that allows the computation of confidence intervals for statistics with "difficult distributions." Bootstrapping uses resampling with replacement (defined in the first bullet below) to estimate the parameter's sampling distribution. The sampling distribution is then used to estimate confidence intervals for the parameter. Estimating confidence intervals using the bootstrap involves the following steps:

- Create *m* resampled data sets that contain the same number of observations (*n*) as the original dataset. Resampling is performed by selecting an observation at random from the original dataset and copying it into the resampled dataset being created. Although that observation has been "used", it is not deleted from the original dataset and may be reselected one or more times. It is possible for some observations in the original dataset not to be selected at all.

- The statistical parameters of interest – in this case the 75th, 80th, 85th, and 90th percentiles – are calculated for each of the *m* resampled datasets, resulting in a collection of *m* values for the parameters of interest, which is referred to as the sampling distribution from the bootstrap resamples.

- A confidence interval for the parameter is calculated using the collection of values obtained for this parameter in the previous step. This analysis used three different methods for computing confidence intervals:

    1. The **normal approximation method** computes an approximate standard error using the sampling distribution from the bootstrap resamples. The confidence interval is then computed using the t-distribution, as described below.

    2. The **percentile method** uses the frequency distribution of the sampling distribution to compute the confidence intervals. The 5th and 95th percentiles of the sampling distribution define the bounds of the 90 percent confidence interval for the parameter of interest, as described below.

    3. The **bias-correction method** adjusts for bias in the bootstrapped sampling distribution compared to the actual sampling distribution.

The details of the confidence interval estimation approach using all three methods are described in mathematical terms below.

Let θ_{obs} = Observed value of the statistic;

Let $i = 1,2,\ldots.m$ denote the bootstrap samples;

Let θ_i^* = the value of the statistic computed using the ith bootstrap sample;

The standard error is estimated as:

$$\hat{se} = \{\frac{1}{m-1}\sum_{i=1}^{m}(\theta_i^* - \bar{\theta}^*)^2\}^{1/2}$$

Where,

$$\bar{\theta}^* = \frac{1}{m}\sum_{i=1}^{m}\theta_i^*$$

The bias is estimated as:

$$\hat{b} = \bar{\theta}^* - \theta_{obs}$$

Confidence intervals covering the 1-ά confidence range are computed according to the following formulas:

The normal approximation method yields confidence intervals:

$$[\theta_{obs} - t_{1-\alpha/2,m-1} * \hat{se}, \theta_{obs} + t_{1-\alpha/2,m-1} * \hat{se}]$$

The percentile methods yields confidence intervals:

$$[\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2}]$$

Where θ_p^* is the 100p^{th} percentile of the bootstrap distribution (θ_1^*,….. θ_m^*).

The confidence intervals for the bias corrected method are calculated as follows:

$$z_o = \Phi^{-1}\{\#(\theta_i^* \le \theta_{obs})/m\}$$

Where $\#(\theta_i^* \le \theta_{obs})$ is the number of elements of the bootstrap distribution that are less than or equal to the observed statistic, and Φ is the standard cumulative normal.

Let $p_1 = \Phi(2z_o - z_{1-\alpha/2})$ and $p_2 = \Phi(2z_o + z_{1-\alpha/2})$, where z represents points on the standard normal distribution corresponding to the associated cumulative probability.

Using the corrected percentile values estimated above, the bias-corrected confidence intervals may be computed as:

$$[\theta^*_{p1}, \theta^*_{p2}]$$

Where θ_p^* is the 100p^{th} percentile of the bootstrap distribution (θ_1^*,….. θ_m^*) (Efron 1993)

Although confidence intervals were computed using all three methods, the percentile method is the preferred default method. The results of the bias-corrected approach may be used to identify and correct problems arising from skewed distributions, but in some cases the bias corrected approach produces missing values. The normal approximation method is only appropriate when the underlying bootstrap distribution is normal. In most cases, the confidence intervals produced by the three methods are similar. A resampling size of 1,000 replications was chosen for this analysis, which is considered adequate for the methods implemented here. The bootstrap analysis was performed using the all observation database in the reference population only.

4.2.3 Identifying Distinct Level IV Ecoregions within Level III Ecoregions

A previous phase of the project (ICF 2005) established that the use of level IV ecoregions as a sub-stratifying parameter within level III ecoregions would result in nutrient zones that further minimize intra-zone variability and maximize inter-zone differences compared to nutrient zones based on level III ecoregions alone. The use of level IV ecoregions as the basis for creating nutrient zones would, however, create problems of sample size and regulatory complexity. MT DEQ instead decided to identify only those level IV ecoregions that were statistically distinct from their parent level III ecoregions. These distinct level IV ecoregions could potentially be used to constitute distinct nutrient zones within the parent level III ecoregion.

The two-tailed t-test was used as the basis for deciding whether the nutrient/chlorophyll concentrations within a level IV ecoregion was equal to its parent level III ecoregion. The analysis applied Welch's correction for unequal variances, which compensates for unequal variances in the two groups being compared by adjusting the degrees of freedom in computing the applicable t-statistics. In computing the concentrations in the parent level III ecoregion for this analysis, the observations from the level IV ecoregion being considered were excluded. If the concentrations were found to be different (at the 90% confidence level), the analysis declares the level IV ecoregion being considered as a distinct zone. The analysis then computes percentiles of interest for that level IV ecoregion as well as confidence intervals using the bootstrap method described in section 4.2.2. This analysis was conducted for each level IV ecoregion within each level III ecoregion by nutrient and by season. The analysis was conducted for reference data using the median database only. The analysis was only conducted when a minimum sample size of four observations were available in each group.

4.3 Quality Assurance

The quality assurance methodology adopted for this analysis involved replication of randomly selected cases from the computer log files in independent database management and statistical software applications. Because the results generated by two independent applications were in agreement for all QA cases, it was concluded that the quality assurance analysis validated the querying system, the program, and the software used for the analysis.

5.0 Results and Conclusions of the Statistical Analysis

The suite of statistical analyses described in Section 4 was conducted for ten water quality parameters and several stratification and sub-stratification methodologies. As described in

Section 4, some analyses were performed on the all-observations database, some analyses on the median database, and some analyses were performed on the percentile databases. All analyses were seasonally stratified.

Because the suite of statistical analyses produced a very large amount of results, this section presents only the most pertinent trends as well as summary tables for each type of analysis. The full results of the statistical analyses were saved in easily readable, self-explanatory, computer-generated text files, referred to as log files, which are included in Appendix D. A guide to locating and reading the log files also is included in Appendix D.

In the tables presented in this section, water quality parameters are represented by a set of abbreviations. A key to these abbreviations is presented below.[8] The tables below focus on some or all of the seven nutrient groups; statistics have also been computed for benthic chlorophyll a and for two turbidity parameters in the log files.

5.1 Summary Statistics for the Reference, Non-Reference, and Whole Populations

As discussed in section 4.1, summary statistics in the results characterize the reference, non-reference, and whole-population statistical distributions for the selected water quality parameter for each category of the stratifying or sub-stratifying parameter. Statistics are reported separately using three databases: the all-observations database; the all-observations database excluding stream orders 7 and 8; and the median database. In the log files of Appendix D, these statistics are presented for all nutrients in turn by season for various combinations of stratifying and sub-stratifying parameters. Specific summary statistics include the total number of observations, minimum, maximum, mean, standard deviation and skewness for reference observations, non-reference observations, and whole population observations. In addition, the summary statistics include the values of the 10th, 25th, 50th, 75th, and 90th percentiles for reference observations, non-reference observations, and whole population observations.

Tables 5.1.1 and 5.1.2 present nutrient concentrations corresponding to the 25th, 50th, and 75th percentile for level III ecoregions for all seasons combined. Table 5.1.1 is based on the all-observations database; Table 5.1.2 is based on the median database.[9]

5.2 Percentiles of Interest and Distinct Zone Assessments

Tables 5.2.1 through 5.2.4 present observed values of the percentiles of interest stratified by level III ecoregion for the nutrient groupings Total Phosphorus, Total Nitrogen, Nitrate plus Nitrite, and Total Kjeldahl Nitrogen using the median database for reference data for all seasons. In addition, these tables present the observed percentiles of interest for those level IV ecoregions

[8] BNCHLOR-A = Benthic Chlorophyll-A; NO3+NO2 = Nitrates and Nitrites; SRP = Soluble Reactive Phosphates; TDP=Total Dissolved Phosphorus; TKN = Total Kjeldahl Nitrogen; TOTALN = Total Nitrogen; TOTALP = Total Phosphorus; TURB-JTU = Turbidity (JTUs); TURB-NTU = Turbidity (NTUs)

[9] Summary statistics for the all observations database excluding stream orders 7 and 8 are presented in the log files of Appendix D but are not tabulated here. These log files also include the entire range of summary statistics computed for each database.

that were proven to be statistically distinct from their parent level III ecoregions. Tables 5.2.5 and 5.2.6 present similar results for the growing season for Total Phosphorus and Total Nitrogen, respectively. Bootstrap confidence intervals around the percentiles of interest were also computed for level III ecoregions as well as for those level IV ecoregions that constituted distinct nutrient zones for all nutrients for all seasons; the confidence intervals are presented in the log files of Appendix D.[10]

5.3 General Conclusions from the Statistical Analysis

The preceding statistical analyses suggest the following conclusions:

- The number of individual level IV ecoregions that are statistically distinct from their parent level III ecoregions is not excessive for data from all seasons. There are rarely more than three or four distinct level IV ecoregions with a parent level III ecoregion.

- There is not a high degree of cross-nutrient consistency in the identification of distinct level IV ecoregions, even for closely related nutrients. In other words, a level IV ecoregion that appears to be distinct from its parent level III ecoregion for a particular nutrient is often not found to be distinct for another nutrient grouping. This suggests that if nutrient zones are to be delineated using distinct level IV ecoregions (in addition to the parent level III ecoregions), the nutrient zones will vary from one nutrient grouping to another.

- There is a strong degree of correspondence between the individual level IV ecoregions that are statistically distinct from their parent level III ecoregions when one considers data taken from all seasons and data taken from the growing season alone. This suggests that intra-level III ecoregion distinctions manifest themselves most strongly in the growing season.

- Bootstrap confidence intervals around the 85th and 90th percentiles are generally narrow, implying a fair degree of confidence in the observed values of those percentiles in the reference population.

[10] The entire set of results for all percentiles of interest, as well as their bootstrap confidence intervals for level III ecoregions and distinct level IV ecoregions, are presented in the log files of Appendix D for all combinations of nutrients and seasons. These log files also include the results of the two-tailed t-tests that were used to identify level IV ecoregions that are distinct from their parent level III ecoregions.

TABLE 5.1.1. Nutrient Concentrations (mg/L) in the Reference and Non-Reference Populations (All Observations Database, All Seasons Combined)

Selected percentiles of the frequency distributions are reported for all seasons combined. Data are organized by level-III ecoregions.*

	Northern Rockies						Middle Rockies						Canadian Rockies						Northwestern Glaciated Plains						Northwestern Great Plains					
	Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference		
Nutrient Group	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th
Ammonia	0.002	0.005	0.006	0.005	0.010	0.020	0.002	0.002	0.010	0.005	0.020	0.050	0.002	0.005	0.008	0.010	0.010	0.030	0.005	0.035	0.085	0.010	0.040	0.090	0.006	0.010	0.015	0.010	0.040	0.090
$NO_3 + NO_2$	0.008	0.010	0.020	0.020	0.050	0.100	0.005	0.015	0.040	0.020	0.080	0.230	0.004	0.007	0.010	0.040	0.050	0.100	0.001	0.005	0.050	0.020	0.050	0.200	0.004	0.050	0.160	0.030	0.080	0.270
SRP	0.006	0.007	0.007	0.010	0.010	0.020	0.002	0.005	0.010	0.010	0.020	0.060	0.001	0.002	0.005	0.005	0.010	0.010	0.005	0.008	0.020	0.005	0.010	0.030	0.003	0.008	0.028	0.005	0.010	0.030
TKN	0.050	0.050	0.050	0.100	0.200	0.400	0.050	0.175	0.200	0.130	0.300	0.500	0.050	0.050	0.050	0.080	0.200	0.600	0.800	1.100	1.500	0.480	0.800	1.200	0.360	0.770	1.100	0.400	0.640	1.000
Total N	0.031	0.049	0.120	0.130	0.220	0.360	0.070	0.110	0.180	0.060	0.135	0.490	0.050	0.080	0.170	0.060	0.080	0.240	0.630	0.850	1.389	0.620	0.910	1.400	0.480	0.790	1.388	0.580	0.850	1.300
Total P	0.003	0.003	0.003	0.010	0.020	0.040	0.008	0.010	0.020	0.020	0.040	0.080	0.001	0.002	0.003	0.010	0.010	0.020	0.040	0.070	0.140	0.020	0.050	0.100	0.030	0.068	0.160	0.020	0.040	0.090
TDP	-	-	-	-	-	-	0.010	0.020	0.020	0.010	0.030	0.040	-	-	-	-	-	-	0.010	0.030	0.050	0.010	0.020	0.030	0.030	0.050	0.090	0.010	0.020	0.030

* Results for the Wyoming Basin and the Idaho Batholyth ecoregions not shown; too few reference observations (n < 4). Dashes in the table indicate too few observations (n < 4) to generate distribution.

TABLE 5.1.2. Nutrient Concentrations (mg/L) in the Reference and Non-Reference Populations (Median Database, All Seasons Combined)

Selected percentiles of the frequency distributions are reported for all seasons combined. Data are organized by level-III ecoregions.*

	Northern Rockies						Middle Rockies						Canadian Rockies						Northwestern Glaciated Plains						Northwestern Great Plains					
	Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference			Reference			Non-Reference		
Nutrient Group	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th
Ammonia	0.003	0.005	0.006	0.005	0.010	0.020	0.005	0.007	0.010	0.005	0.020	0.050	0.004	0.005	0.007	0.005	0.010	0.020	0.002	0.005	0.022	0.010	0.030	0.060	0.006	0.010	0.012	0.010	0.020	0.050
$NO_3 + NO_2$	0.005	0.020	0.094	0.015	0.045	0.130	0.008	0.015	0.070	0.020	0.050	0.130	0.005	0.010	0.020	0.035	0.050	0.105	0.004	0.010	0.090	0.050	0.180	0.700	0.002	0.007	0.100	0.015	0.055	0.200
SRP	0.002	0.003	0.007	0.010	0.010	0.020	0.003	0.007	0.015	0.010	0.020	0.050	0.001	0.002	0.003	0.005	0.010	0.010	0.004	0.009	0.021	0.010	0.030	0.070	0.002	0.006	0.029	0.010	0.020	0.030
TKN	0.050	0.050	0.050	0.100	0.200	0.500	0.050	0.110	0.200	0.090	0.240	0.450	0.050	0.050	0.088	0.050	0.110	0.300	0.618	1.050	1.750	0.340	0.707	1.160	0.420	0.800	1.270	0.285	0.520	0.990
Total N	0.037	0.053	0.167	0.100	0.150	0.230	0.070	0.119	0.208	0.070	0.130	0.320	0.060	0.090	0.139	0.060	0.080	0.090	0.638	0.740	1.335	0.680	0.975	1.370	0.423	0.736	1.290	0.500	0.810	1.200
Total P	0.003	0.006	0.010	0.010	0.010	0.030	0.006	0.010	0.025	0.010	0.030	0.060	0.001	0.004	0.005	0.009	0.010	0.020	0.020	0.060	0.110	0.020	0.050	0.120	0.030	0.065	0.136	0.020	0.040	0.090
TDP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	0.025	0.175	0.020	0.020	0.050	-	-	-	-	-	-

* Results for the Wyoming Basin and the Idaho Batholyth ecoregions not shown; too few reference observations (n < 4). Dashes in the table indicate too few observations (n < 4) to generate distribution.

Table 5.2.1 Total Phosphorus: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are from All Seasons, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.0100	0.0115	0.0132	0.0150
Idaho Batholith		0.0078	0.0088	0.0097	0.0104
Middle Rockies		0.0250	0.0308	0.0492	0.0852
	Pioneer-Anaconda Ranges (17ag)	*0.0093*	*0.0096*	*0.0098*	*0.0100*
	Absaroka-Gallatin Volcanic Mountains (17i)	*0.0993*	*0.1041*	*0.1184*	*0.1327*
	Big Snowy-Little Belt Carbonate Mountains (17q)	*0.0200*	*0.0260*	*0.0320*	*0.0380*
	Rattlesnake-Blackfoot-South Swan-Northern Garnet-Sapphire Mountains (17x)	*0.0200*	*0.0240*	*0.0280*	*0.0300*
Canadian Rockies		0.0050	0.0051	0.0056	0.0063
Northwestern Glaciated Plains		0.1075	0.1500	0.2013	0.2200
	Glaciated Dark Brown Prairie (42i)	*0.0806*	*0.0853*	*0.0900*	*0.1450*
Northwestern Great Plains		0.1360	0.1794	0.2190	0.2925
	Non-calcareous foothill grassland (43s)	*0.0160*	*0.0240*	*0.0320*	*0.0400*

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

Table 5.2.2 Total Nitrogen: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are from All Seasons, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.1678	0.1697	0.1810	0.2144
	St. Joe Schist-Gneiss Zone (15p)	*0.0555*	*0.0580*	*0.0605*	*0.0630*
Idaho Batholith		0.1120	0.1184	0.1248	0.1312
Middle Rockies		0.2125	0.2471	0.3168	0.3304
Canadian Rockies		0.1433	0.1543	0.1637	0.1776
Northwestern Glaciated Plains		1.3350	1.5220	1.7224	2.0181
	Foothill Grassland (42p)	*0.6705*	*0.6967*	*0.7229*	*0.7360*
Northwestern Great Plains		1.2900	1.4055	2.1593	2.5272
	River Breaks (43c)	*2.1704*	*2.2701*	*2.3523*	*2.3995*
	Missouri Breaks Woodland-Scrubland (43l)	*0.7813*	*0.7970*	*0.8127*	*0.8205*
	Non-calcareous foothill grassland (43s)	*0.1400*	*0.1416*	*0.1432*	*0.1440*

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

Table 5.2.3 Nitrate and Nitrite: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are from All Seasons, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.0933	0.0983	0.1192	0.1298
	St. Joe Schist-Gneiss Zone (15p)	*0.0237*	*0.0250*	*0.0256*	*0.0263*
	Purcell-Cabinet-North Bitterroot Mountains (15q)	*0.1400*	*0.1440*	*0.1480*	*0.1500*
Idaho Batholith		0.0449	0.0479	0.0486	0.0493
Middle Rockies		0.0700	0.0829	0.1000	0.1200
	Pioneer-Anaconda Ranges (17ag)	*0.0131*	*0.0163*	*0.0194*	*0.0225*
	Mid-Elevation Sedimentary Mountains (17g)	*0.0363*	*0.0420*	*0.0490*	*0.0560*
	Gneissic-Schistose Forested Mountains (17l)	*0.1000*	*0.1030*	*0.1300*	*0.1312*
	Big Snowy-Little Belt Carbonate Mountains (17q)	*0.0075*	*0.0085*	*0.0095*	*0.0100*
Canadian Rockies		0.0200	0.0410	0.0500	0.0550
Northwestern Glaciated Plains		0.0100	0.0200	0.0500	0.1186
Northwestern Great Plains		0.1000	0.1110	0.1870	0.3742
	River Breaks (43c)	*0.5994*	*0.7498*	*0.9002*	*1.0339*
	Missouri Breaks Woodland-Scrubland (43l)	*0.0200*	*0.0200*	*0.0200*	*0.0200*
	Central Grassland (43n)	*0.0427*	*0.0831*	*0.1000*	*0.1240*

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

Table 5.2.4 Total Kjeldahl Nitrogen: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are from All Seasons, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.0500	0.0500	0.0625	0.1000
Idaho Batholith		0.0500	0.7500	0.1625	0.2500
Middle Rockies		0.2000	0.2255	0.2680	0.3300
Canadian Rockies		0.0875	0.1100	0.1700	0.2075
	Southern Carbonate Front (41d)	*0.2063*	*0.2125*	*0.2188*	*0.2250*
Northwestern Glaciated Plains		1.7500	1.8000	2.3770	3.2380
	Foothill Grassland (42p)	*0.6700*	*0.6700*	*0.6700*	*0.6700*
Northwestern Great Plains		1.2775	1.4400	1.7240	1.8960

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

Table 5.2.5 Total Phosphorus, Growing Season: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are for the Growing Season, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.0100	0.0115	0.0132	0.0150
Idaho Batholith		0.0078	0.0088	0.0097	0.0104
Middle Rockies		0.0278	0.0356	0.0500	0.0905
	Pioneer-Anaconda Ranges (17ag)	*0.0093*	*0.0096*	*0.0098*	*0.0100*
	Absaroka-Gallatin Volcanic Mountains (17i)	*0.1000*	*0.1123*	*0.1245*	*0.1368*
	Big Snowy-Little Belt Carbonate Mountains (17q)	*0.0200*	*0.0260*	*0.0320*	*0.0380*
	Rattlesnake-Blackfoot-South Swan-Northern Garnet-Sapphire Mountains (17x)	*0.0090*	*0.0094*	*0.0098*	*0.0100*
Canadian Rockies		0.0050	0.0055	0.0057	0.0070
Northwestern Glaciated Plains		0.0913	0.0995	0.2065	0.3600
	Glaciated Dark Brown Prairie (42i)	*0.0552*	*0.0704*	*0.0820*	*0.0862*
Northwestern Great Plains		0.1632	0.2008	0.2263	0.3626
	Non-calcareous foothill grassland (43s)	*0.0160*	*0.0240*	*0.0320*	*0.0400*

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

Table 5.2.6 Total Nitrogen,Growing Season: Observed Values of Percentiles of Interest for Level III Ecoregions and Distinct Level IV Ecoregions
(Data are for the Growing Season, Reference Sites, Median Database)

Level III Ecoregion	Distinct Level IV Ecoregion	Observed Values of Percentiles of Interest			
		p75	p80	p85	p90
Northern Rockies		0.1678	0.1697	0.1810	0.2144
	St. Joe Schist-Gneiss Zone (15p)	*0.0555*	*0.0580*	*0.0605*	*0.0630*
Idaho Batholith		0.1120	0.1184	0.1248	0.1312
Middle Rockies		0.2125	0.2471	0.3168	0.3304
Canadian Rockies		0.1560	0.1630	0.1700	0.1890
Northwestern Glaciated Plains		1.3375	1.6964	1.8195	2.3782
Northwestern Great Plains		1.2900	1.4080	2.2749	2.6414
	River Breaks (43c)	*2.1056*	*2.2441*	*2.3523*	*2.3995*
	Non-calcareous foothill grassland (43s)	*0.1400*	*0.1416*	*0.1432*	*0.1440*

Note: Inadequate data were available for the Wyoming Basin ecoregion. A minimum sample size of 4 is required to report results.

6.0 Development of Compliance and Non-Compliance Tools

The adoption of a nutrient standard must be accompanied by the development of a set of decision rules that determine which situations constitute a violation of the standard and which are in compliance with the adopted standard. Data gathered in the field will provide the basis for supporting or disproving hypotheses about the compliance or lack of compliance of a particular reach. In this phase of the project, spreadsheet-based tools were developed that can be used to interpret field data to determine whether a reach is compliant or non-compliant. These tools are based on the exact binomial test and the t-test.

This section describes the statistical basis for the development of these tools. Section 6.1 explains the two types of decision error that are unavoidable during hypothesis testing. Sections 6.2 and 6.3 focus on the implications of decision errors when testing for compliance and non-compliance, respectively. Sections 6.4 and 6.5 describe how decision rules are derived for user specified error levels using the exact binomial test for compliance and non-compliance, respectively. Section 6.6 and 6.7 describe how the t-test may be applied to field data at user-specified levels of decision error to establish compliance and non-compliance, respectively.

6.1 Decision Errors During Hypothesis Testing

Data gathered in the field form the basis for supporting or disproving hypotheses about the compliance or lack of compliance of a stream reach. Due to the uncertainties associated with sampling variation, decisions made using hypothesis tests will be subject to errors.

Two types of decision errors may occur:

- If the null hypothesis is rejected when in fact it is true, a Type I error is said to have occurred. A Type I error is sometimes referred to as a “false positive” and the probability of its occurrence is represented by the symbol α.
- If the null hypothesis is not rejected when in fact it is false, a Type II error has occurred. A Type II error is also known as a “false negative” and the probability of its occurrence is represented by the symbol β.

Depending on policy objectives, decision makers could be concerned with minimizing the probability of either Type I or Type II error, or both (Mapstone 1995).

In this analysis, the null hypothesis and the alternative hypothesis are differently expressed when testing for compliance *vs*. testing for non-compliance. Different levels of Type I and Type II decision errors may be tolerable in the two situations, depending on policy preferences as described below.

6.2 Hypothesis Tests to Establish Compliance

When attempting to establish compliance, the null hypothesis is that the reach is non-compliant; the alternative hypothesis is that the reach is compliant. The types of possible decision error associated with these hypotheses are illustrated in Table 6-1.

Table 6-1: Decision Errors When Establishing Compliance

Null Hypothesis: ***The reach is non-compliant***		
Actual Reach Condition	**Decision Based On Sample Data**	
	Non-Compliant	**Compliant**
Non-Compliant	Correct Decision (1- α)	Type I Error, (α)
Compliant	Type II Error, (β)	Correct Decision (1- β)

If Type I errors occur, non-compliant reaches would be declared compliant. It may be important to reduce the probability of Type I errors to a low level because of the potential for adverse environmental effects.

If Type II errors occur, compliant reaches would be declared non-compliant. It is possible a higher degree of tolerance may be allowed for this type of error because there are no potential adverse environmental effects associated with such an error. To avoid the clean-up of compliant areas, however, it may be important to restrict the probability of Type II error to a reasonable level. Reducing the probability of Type II error at a given probability level of Type I error may require increasing the sample size; it may therefore be necessary to optimize between the costs of unnecessary clean-ups/effluent reduction measures and the costs of extensive sampling to establish compliance.

6.3 Hypothesis Testing to Establish Non-Compliance

When attempting to establish non-compliance, the null hypothesis is that the reach is compliant; the alternative hypothesis is that the reach is non-compliant. The types of possible decision error associated with these hypotheses are illustrated in Table 6-2.

Table 6-2: Decision Errors When Establishing Non-Compliance

Null Hypothesis: ***The reach is compliant***		
Actual Site Condition	**Decision Based On Sample Data**	
	Compliant	**Non-Compliant**
Compliant	Correct Decision (1- α)	Type I Error, (α)
Non-Compliant	Type II Error, (β)	Correct Decision (1- β)

If Type I errors occur, then reaches which are compliant will be declared non-compliant. There may not be any adverse environmental effects associated with such an error, although unnecessary clean-up actions/emission reduction measures may be initiated as a result.

If Type II errors occur, then areas which are in fact non-compliant will be declared as compliant. At a given level of Type I error, reducing Type II error may require higher sample sizes, as explained above.

6.4 Decision Rules for Compliance Using the Exact Binomial Test

Within a reach, all possible samples are assumed to be equally likely to be non-compliant. If the volume of the reach is sufficiently large, the number of possible samples from that area may be considered to be infinite[11]. In such a situation, the binomial distribution may be used to estimate the probability of finding *x* non-compliant samples from a total of *n* samples when there is a specified proportion, *p*, of non-compliant samples in the whole population. This probability is mathematically expressed as:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where, *x* is the number of non-compliant samples
f(x) is the probability that x samples will be non-compliant
n is the total number of samples taken
p is the fraction of non-compliant samples in the population representing the compliance threshold

and $$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

This probability mass function is used to formulate a relationship between the total number of samples required to be taken (*N*) and the maximum number of samples that may be found non-compliant (*y*) to support the conclusion with a specified level of confidence (*CL)* that the proportion of non-compliant samples in the whole population is no greater than a specified level (*p)*, which represents the compliance threshold.

N may be calculated for different values of *y* (or *y* may be calculated at different values of *N*) at specified levels of *CL* and *p* using the formula:

$$(1-CL) = \sum_{x=0}^{y} \binom{N}{x} p^x (1-p)^{(N-x)}$$

However, because *p* is fixed and *N* and *y* are whole numbers, it is usually not possible to exactly meet this formula. Instead, the values of *N* and *y* are chosen so that the inequality

$$(1-CL) \geq \sum_{x=0}^{y} \binom{N}{x} p^x (1-p)^{(N-x)}$$

[11] An individual sample is said to be non-compliant if it is in excess of the criterion.

is satisfied. Thus, the actual confidence level may be greater than the chosen value of *CL*.

When testing for compliance, the term (*1-CL*) in the formula above must be set to the required Type I error probability level.

After computing *N* and y, the power of the test may be estimated by performing the summation using the computed value of *N* and an appropriate assumed value of *p;* once the power is known, the probability of Type II error may be computed since the power equals (1-Type II error).

These principles have been used to develop the tool named MT-ComplianceTool.xls, which is available on the CD accompanying this report. The text within the tool provides further guidance on Type I and Type II error in the context of compliance testing.

6.5 Decision Rules for Non-Compliance Using the Exact Binomial Test

When testing to establish non-compliance, the hypotheses are stated differently than when testing for compliance, as explained in Sections 6.2 and 6.3. In this case, the relationship between *N* and *y* answers the question: "If *N* samples are taken from a homogeneous survey area, what is the minimum number of "*y*" samples that must be individually non-compliant to support the conclusion with 100 × *CL*% confidence that the fraction of non-compliant samples in the population is greater than *p*?" The same mathematical principles apply for sample size prediction, but are expressed by the formula:

$$(1-CL) \geq \sum_{x=y}^{N} \binom{N}{x} p^{x} (1-p)^{(N-x)}$$

As before, the term (*1-CL*) must be set to the Type I error probability. The compliance threshold *p* is set by the user, as before.

After computing *N*, the Type II error may be directly estimated in this case by performing the summation using the computed value of *N* and an appropriate assumed value of *p*.

These principles have been used to develop the tool named MT-NonComplianceTool.xls, which is available on the CD accompanying this report. The text within the tool provides further guidance on Type I and Type II error in the context of non-compliance testing.

6.6 Decision Rules for Compliance Using the T-Test

The binomial tests described in sections 6.4 and 6.5 depend on converting the field data to a binary form: for instance, "Yes" if the observed value is greater than the standard and "No" if the observed value is less than or equal to the standard. In this process, valuable information may be lost on the quantum by which the data point lies in excess of or below the standard. This information is particularly relevant when accounting for the phenomenon of luxury nutrient uptake in the context of water quality standards (see page H-9, Appendix H).

The t-test may be used to check if user-specified points on the distribution of the sample mean lie below the prescribed standard; if yes, this may be taken as positive confirmation that the reach is compliant. In other words, the t-test may be used to answer questions of the form: “Is there 95% confidence that the sample mean lies below the prescribed standard?” If yes, the reach would be considered compliant.

The tool allows users to specify the required Type I error level or (alpha). If the user specifies an alpha equal to 0.05, the tool will compute the upper 95% confidence limit of the sample mean using a one-tailed t-distribution. If this point lies below the standard, it is taken as confirmation that the reach is compliant.

6.7 Decision Rules for Non-Compliance Using the T-Test

For non-compliance, the t-test may be used to check if user-specified points on the distribution of the sample mean lie above the prescribed standard; if yes, this may be taken as positive confirmation that the reach is non-compliant. In other words, the t-test may be used to answer questions of the form: “Is there 95% confidence that the sample mean lies above the prescribed standard?” If yes, the reach would be considered non-compliant.

The tool allows users to specify the required Type I error level or (alpha). If the user specifies an alpha equal to 0.05, the tool will compute the lower 95% confidence level of the sample mean using a one-tailed t-distribution. If this point lies above the standard, it is taken as confirmation that the reach is non-compliant.

6.8 Decision Tree to Reconcile Multiple Tests

In certain circumstances, it is possible that the binomial test and the t-test will produce conflicting results. In such situations, users may need to perform a subjective assessment that reconciles the results of the two tests to determine the appropriate action in the circumstances. Appendix H (page H-15)includes a decision tree that focuses the results of multiple tests into a single decision; MT DEQ developed this decision process based on expert judgment for application in water quality compliance and non-compliance determinations across the state. The decision tree includes not only the results for the nutrient data from the binomial and t-tests, but also data that directly measure impacts to stream resources, for example excess algae levels.

7.0 Development of the Changepoint Assessment Tool

A common approach to establishing environmental criteria is to examine changes of selected population, community, or ecosystem attributes along a gradient of environmental conditions. Ecological attributes often show little change until a critical environmental value, or threshold, is reached. Thus, quantitative descriptions of such exposure–response relationships can be very useful in support of the development of numerical environmental criteria (Qian 2003).

The changepoint assessment tool was developed to find statistically significant breakpoints in sets of continuous data. These changepoints could be used to split these data into two distinct

groups. The tool uses the concept of deviance reduction as a method for environmental threshold estimation. The tool also uses a bootstrap technique to quantify uncertainty about the threshold.

The deviance, a measure of homogeneity, is defined for a continuous variable, as:

$$D = \sum_{k=1}^{n} (y_k - \mu)^2$$

where D is the deviance, n is the sample size, and μ is the mean of the n observations y_k.

When the response data are divided into two groups, the sum of the deviance for the two subgroup is always less than or equal to the deviance of the combined data. Each possible changepoint is associated with a deviance reduction:

$$\Delta_i = D - (D_{\leq i} + D_{>i})$$

where D is the deviance of the combined data set, $y_1, \ldots, y_n$, $D_{\leq i}$ is the deviance of the sequence$y_1, \ldots, y_i$, and $D_{>i}$ is the deviance of the sequence $y_{i+1}, \ldots, y_n$, where i = 1, . . . , n. The changepoint r is the i value that maximizes Δ_i : $r = \max_i \Delta_i$.

After computing the changepoint using the available dataset, the changepoint tool estimates uncertainty about the changepoint by using a bootstrap simulation (Efron 1993) and expresses the changepoint as a 90% confidence interval.

To estimate the statistical significance of the changepoint, the tool employs the property that the deviance reduction divided by the scale parameter describes an F distribution with one degree of freedom in the numerator and (n-1) degrees of freedom in the denominator (Venables and Ripley 1994).

Based on the above principles, the changepoint tool reports the changepoint, 90% confidence interval around the changepoint, and the statistical significance of the changepoint for any user defined dataset which specifies the response variable and the independent variable in a prescribed format.

8.0 Point Source Compliance with Numeric Nutrient Criteria

Section 6 and Appendix H present compliance and non-compliance decision-making tools for use when numeric nutrient criteria are applied to ambient water quality data. This section discusses the relationships between the numeric nutrient criteria and nutrient discharges from permitted point source. In particular, this section sumarizes relevant aspects of point source regulation, including:

- Methods used by permit writers to set effluent limits (including limits derived from numeric water quality standards);
- Effluent monitoring and reporting requirements; and

- Compliance decision-making criteria.

Appendix I discusses these subjects in greater detail and also presents potential approaches MT DEQ permit writers could use to establish nutrient effluent limits supportive of surface water quality standards for receiving streams, and to evaluate compliance. Recommended approaches are summarized in Section 8.3.

8.1 Point Source Effluent Limitations

Point source releases to surface water are regulated under the National Pollution Discharge Elimination System (NPDES) established by the Federal Water Pollution Act (Clean Water Act) and its amendments (CWA 1972). Among other things, a NPDES permit must contain effluent limitations and monitoring and reporting requirements. Effluent limitations are determined on a facility-specific basis and may be either numeric or narrative. When setting numeric effluent limits, a permit writer first identifies applicable technology-based limits and compares these to levels back-calculated with models from water quality criteria for the receiving surface water. The more stringent of these (i.e., technology-based or water quality standards based) is used to set the permitted effluent limits (EPA 1996). Technology-based effluent limits are based either on national effluent limitation guidelines (ELGs) published by EPA, or on best professional judgment. Numeric effluent limits usually include daily maximum and monthly average concentration limits (e.g., mg/L). In some cases, numeric limits are expressed as daily or cumulative pollutant mass discharge limits (e.g., kg/day).

The permitting procedures described above generally apply to all pollutants, including nonconventional pollutants such as nutrients. Technology-based effluent limits generally are not available as national ELGs for nutrients. Thus, point source effluent limits for nutrients are usually based on narrative water-quality standards for receiving waters, where applicable.

8.2 Monitoring and Compliance

States permitting agencies are required to collect monitoring data from permit holders, and these data are reported to EPA on a quarterly basis (EPA 1996). Any permitted facility must report the number of daily maximum and monthly average exceedances that have occurred during the quarterly reporting period.[12] Based on this information, states permitting agencies and EPA may determine which facilities are in compliance and pursue enforcement actions if necessary.

For nutrient discharges, a point source with a numeric, monthly-average effluent limit is in violation if the monthly average concentration exceeds or equals the effluent limit times a "technical review criterion" (TRC) of 1.4 for two months during a six month period; or if the monthly average effluent concentration exceeds the limit (not multiplied by the TRC) for four out of six months (EPA 1996).

[12] Temporal distribution or clustering of exceedances is not generally reported as permit holders only report the total number of exceedances without designation of when the exceedances occurred.

8.3 Recommended Approach for Establishing Effluent Limits and Determining Compliance

Appendix I presents six options MT DEQ could use to establish point source effluent limits for nutrients that would be supportive of the draft surface water quality standards. These options include narrative and technology-based limits, as well as potential approaches for deriving effluent limits from the receiving stream's nutrient standards. Appendix I also presents options for evaluating compliance with potential effluent limits for nutrients.

With regard to setting point-source effluent limits for nutrients, further investigation is recommended for options in which effluent limits would be derived from the draft water quality standards for nutrients. Compared with other options, these options would provide a technically-defensible and highly location-specific starting point for establishing the effluent limits. In addition, these options would employ existing, EPA-accepted models. These options are described further in Appendix I.

With regard to compliance decision-making, the recommended approach is to compare results from the both the TRC methodology (described above) and the binomial compliance/non-compliance tool discussed in Section 6 and Appendix H. If these compliance assessment methodologies provide comparable results, MT DEQ should choose to use the most cost-efficient tool. If the methodologies yield different results, the binomial compliance/non-compliance tool is recommended because it allows the user greater flexibility in defining the terms of compliance. Please refer to Appendix I for further discussion of these options.

9.0 Literature Cited

CBW 2007. Chesapeake Bay Watershed General Permit Case Study. Permit Effective Date: January 1, 2007. Available at: http://www.theexchangeassociation.org/Xchangewebdocs/News_Articles/5.%20EPA%20Case%20Study%20-%20Watershed%20Based%20Permitting%20in%20Chesapeake%20Bay.pdf.

CWA 1972. Clean Water Act, a.k.a. Federal Water Pollution Control Amendments of 1972. 33 U.S.C. 1251 - 1376. Available at: http://epw.senate.gov/water.pdf.

Efron, B. and R. Tibshirani. 1993. "An Introduction to the Bootstrap". New York: Chapman & Hall.

Haukoos, J.S. and Lewis, R.J., 2005. "Bootstrapping Confidence Intervals for Statistics with "Difficult" Distributions". Academic Emergency Medicine, April 2005, Vol.12, No. 4

ICF, 2004a. "Stratification of Algae and Nutrient Data of Montana Rivers and Streams to Assist in the Development of Numeric Standards – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, July 13, 2004.

ICF, 2004b. "Montana Rivers and Streams Reference-to-Whole-Population Percentile Mapping for Selected Water Quality Parameters – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, November 22, 2004.

ICF, 2005. "Seasonally Stratified Water Quality Analysis for Montana Rivers and Streams – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, June 29, 2005.

ICF, 2006. "Bootstrap Confidence Intervals and Updated Statistical Analyses of Water Quality Data for Montana Rivers and Streams," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, June 27, 2006.

Lentner, C. 1982. "An Introduction to Applied Statistics". New York. Prindle Weber & Schmidt Inc

Mapstone, B.D. 1995. Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors. Ecological Applications 5: 401-410.

MT DEQ, 2005. "Identification and Assessment of Montana Reference Streams: A Follow-up and Expansion of the 1992 Benchmark Biology Study", Suplee, M., Sada de Suplee R., Feldman, D., Laidlaw, T.

NWQL, 1994. Technical memorandum 94-12 from National Water Quality Laboratory to NAWQA study-unit chiefs, July 8, 1994. *Available at* http://nwql.usgs.gov/Public/tech_memos/nwql.94-12.html)

Omernik, J.M., 1987, Ecoregions of the conterminous United States (map supplement): Annals of the Association of American Geographers, v. 77, no. 1, p. 118-125, scale 1:7,500,000.

Qian, S.S., King, R.S., et al. Two statistical methods for the detection of environmental thresholds. Ecological Modelling 166 (2003) 87–97.

Raines, G.L. and Johnson, B.R., 1996. Digital representation of the Montana state geologic map: a contribution to the Interior Columbia River Basin Ecosystem Management Project, USGS Open File Report 95-691, ArcView Shapefile for Montana.

Suplee, M.W., A. Varghese and J. Cleland, 2007. Developing nutrient criteria for streams: An evaluation of the frequency distribution method. *Journal of the American Water Resources Association* 43: 453-472.

U.S. EPA 1996. U.S. EPA NPDES Permit Writers' Manual. United States Environmental Protection Agency, Office of Water. EPA-833-B-96-003. Washington, D.C.

U.S. EPA, 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams. United States Environmental Protection Agency, EPA-822-B00-002. Washington, D.C. *Available at* http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/index.html

U.S. EPA, 2006. Data Quality Assessment: Statistical Methods for Practitioners. United States Environmental Protection Agency, Office of Environmental Information, EPA/240/B-06/003. Washington, D.C. *Available at* http://www.epa.gov/quality1/qs-docs/g9s-final.pdf

U.S. EPA 2008. Nutrient Criteria Implementation. United States Environmental Protection Agency, Office of Water. Current as of 2008. Available at: http://cfpub.epa.gov/npdes/wqbasedpermitting/wspermitting.cfm#casestudy.

Venables, W.N., Ripley, B.D., 1994. Modern Applied Statistics with S-Plus. Springer, New York.

Appendix A. Water Quality Data Grouping Methodology

Prepared by: Michael Suplee, Ph.D., MT DEQ, Updated June 10, 2008

Organization of Nutrient, Benthic Chlorophyll *a* and Turbidity data from NEW, Legacy, and non-STORET Datasets

The organizational objective outlined herein was to group analytical measurements together that are fundamentally equivalent, while at the same time avoiding double-counts in cases where an agency may have measured two or more closely related parameters from the same sample. The approach was undertaken in a series of steps as follows:

1. The different analytical measurements (e.g. total nitrate, dissolved nitrate, total phosphorus) were identified in the database, checked against STORET and other records to determine what they actually measured, and then organized into groups. The groups (e.g., total N group, soluble reactive phosphorus group) were composed of similar measurements that fundamentally measured the same thing.

2. For each group, a series of exploratory queries were made in the databases, by agency, to ascertain if the various analytical measurements within the group were derived from the same sample. In cases where this occurred, only one of the analytical measurements was retained for that particular agency. Entire analytical measurements (regardless of agency) were eliminated if a clear definition for the measurement could not be located or if its inclusion resulted in other data-collation problems (these will be detailed below for each group).

This methodological approach and its refinements over time are documented in earlier appendices and articles (Appendix C, Varghese and Cleland, 2004a; Appendix A, Varghese and Cleland, 2004b; Appendix A, Varghese and Cleland, 2005; Appendix A, Varghese and Cleland, 2006, Suplee et al., 2007). The present document describes new data sources that were added to the database for the 2008 analyses, how the database was screened to assure quality, and (as in earlier versions of this document) which water quality measurements were grouped together for common analysis.

For 2008, the database was enhanced by adding about 1,300 new nutrient observations from across Montana, and over 1,000 new benthic Chl *a* observations. These data were collected as part of reference-stream sampling that occurred in summer 2006, and as part of a special 2007 nutrient-sampling project that targeted reference sites that were underrepresented in the database (MT DEQ, 2007). All of the data were from wadeable rivers and streams.

In addition, 6,033 observations (USGS data) already in the database that were accidentally missing latitude and longitude were updated to include their coordinates. None of these data were reference data. A few other minor data loading errors from the previous (2006) work were also identified and corrected.

We also dropped the Stata command which eliminated from the database data flagged with a "J". This had been carried out on all earlier versions of the database. One EPA document pertaining to nutrients indicates that J-flagged data should be eliminated because it means "Estimated Value. Value shown is not a result of an analytical measurement" (Appendix A, US EPA, 2000). However, all observations in our database *are* the result of an analytical measurement, and the more conventional definition for "J" used by MT DEQ and other agencies (i.e., USGS) is "Estimated Value. The analyte was positively identified and the associated numerical value is the approximate concentration of the analyte in the sample." That is, its value fell between the detection limit (lower bound) and the practical limit of quantitation (upper bound). For purposes of this database, then, J flagged data represent low but nevertheless quantitfiable results *above* detection limit and, so, will be retained. The correction affected just 73 observations already in the database, and will affect all data added in the future.

Table 1.0 (below) shows the thirty-five nutrient and turbidity analytical measurements that were identified in the databases. These have been organized into groups. This grouping methodology is similar to that used by Mueller et al. (1995) for an analysis of surface and ground water in the United States, and is also described in Suplee et al (2007). Individual analytical measurements that were completely eliminated from use in the present analysis are shown in the table in gray. The development of each group is detailed below, with explanations as to why the measurements were grouped together, why certain analytical measurements were eliminated, and why some were restricted to certain agencies. For simplicity, analytical measurements will be referred to by their STORET parameter code numbers (also used by the USGS in their database, the National Water Information System [NWIS]).

With the development of the NEW STORET system in the late 1990s, the use of STORET parameter code numbers was no longer continued as a means to identify STORET analytical measurements. Instead, analytical measurements are identified by information contained in several NEW STORET fields, the key fields being 'Characteristic Name', 'CharUnit', 'Sample Fraction' and 'Analysis Procedure Name'. Because the programming developed in Stata (Version 7) to identify sample types and undertake database analyses for the present project was already based on parameter codes, NEW STORET data were mapped to their corresponding parameter codes as found in Legacy STORET (Table 2). This process was also applied to non-STORET data.

Agencies not found in Legacy or NEW STORET may be included in the queries that will be detailed below. These are EMAP (the U.S. EPA's Western Environmental Monitoring and Assessment Program), MT DNRC (Montana Department of Natural Resources and Conservation), UM (the University of Montana), and USU (Utah State University).

Table 1.0 Nutrient & turbidity measurements from MT rivers & streams, identified in Legacy & NEW STORET and other data sources. Measurements are organized into functional groups. Gray areas contain measurements that were excluded from use.

AGENCY	STORET PARAM 1*	STORET PARAM 2*	STORET PARAM 3	STORET CODE	Nutrient Reported as:	NOTES (STORET codes & definitions 3/96; US EPA 1979)	What is the parameter basically a measure of	Functional Groups
VARIOUS	NH3 + NH4	N-TOTAL	MG/L	00610	N	Total ammonia nitrogen, as N	Ammonia	Ammonia group
USGS & COEOMAHA	AMMONIA	TOT-NH4	MG/L	71845	NH4	Total ammonia nitrogen, as NH4	Ammonia	
VARIOUS	NH3 + NH4	N-DISS	MG/L	00608	N	Dissoved ammonia nitrogen, as N	Ammonia	
USGS	AMMONIA	DISS-NH4	MG/L	71846	NH4	Dissolved ammonia nitrogen, as NH4	Ammonia	
VARIOUS	TOT KJEL	N	MG/L	00625	N	TKN (mg/L as N)	Total Kjeldahl N (dissolved and particulate)	TKN group
USGS	ORG-N	N	MG/L	00605	N	Total organic N	Total organic N (~ TKN)	
USGS	ORG-N	DISS-N	MG/L	00607	N	Dissolved organic N	Dissolved organic N	
U.S. EPA and USGS	KJELDL N	DISS	MG/L	00623	N	Dissolved Kjeldahl N	Dissolved Kjeldahl N	
MT DEQ and USGS	KJELDL N	SUSP	MG/L	00624	N	Suspended Kjeldahl N	Suspended Kjeldahl N	
VARIOUS	TOTAL N	N	MG/L	00600	N	Total N (mg/L as N)	Total N	Total N group
N. Park Service & USGS	DISS	NITROGEN	MG/L N	00602	N	Total N, water, filtered	Non-particulate total N	
USGS	TOTAL N	AS NO3	MG/L	71887	NO3	Total N (mg/L as NO3)	Total N	
VARIOUS	NO2 + NO3	N-TOTAL	MG/L	00630	N	NO2 + NO3, total, as N	Nitrate + Nitrite	Nitrate & nitrite group
VARIOUS	NO2 + NO3	N-DISS	MG/L	00631	N	NO2 + NO3, dissolved, as N	Nitrate + Nitrite	
VARIOUS	NO3-N	DISS	MG/L	00618	N	Dissolved nitrate nitrogen, as N	Nitrate	
VARIOUS	NO3-N	TOTAL	MG/L	00620	N	Total nitrate nitrogen, as N	Nitrate	
U.S. Forest Service & USGS	NITRATE	TOT-NO3	MG/L	71850	NO3	Nitrate nitrogen, total, as NO3	Nitrate	
VARIOUS	NITRATE	DISS-NO3	MG/L	71851	NO3	Dissolved nitrate nitrogen, as NO3	Nitrate	
VARIOUS	NO2-N	DISS	MG/L	00613	N	Dissolved nitrite nitrogen, as N	Nitrite	
USGS	NITRITE	DISS-NO2	MG/L	71856	NO2	Dissolved nitrite nitrogen, as NO2	Nitrite	
VARIOUS	NO2-N	TOTAL	MG/L	00615	N	Total nitrite nitrogen, as N	Nitrite	
VARIOUS	PHOS TOT		MG/L P	00665	P	Total phosphorus	Total Phosphorus	Total P group
USGS & BLM	TOTAL P	AS PO4	MG/L	71886	PO4	Total phosphorus reported as PO4	Total Phosphorus	
VARIOUS	PHOS TOT	HYDRO	MG/L P	00669	P	Phosphorus, total hydrolyzable	TAHP	
COEOMAHA	PHOS TOT	HYDRO + ORTH	MG/L P	00678	P	Phosphorus, hydrolyzable + ortho, total, autoanalyzer.	TAHP	
VARIOUS	PHOS-DIS	ORTHO	MG/L P	00671	P	Dissolved orthophosphate	SRP, 0.45 um filtered	Soluble reactive phosphate group
VARIOUS	PHOS-T	ORTHO	MG/L P	70507	P	Total orthophosphate	Unfiltered, total reactive phosphate, which is largely orthophosphate	
VARIOUS	ORTHOPO4	PO4	MG/L	00660	PO4	Dissolved orthophosphate as PO4	SRP, 0.45 µm filtered	
VARIOUS	T PO4	PO4	MG/L	00650	PO4	Total orthophosphate	Total reactive phosphate,probably with a more aggressive digestion-may contain organic phosphates	
VARIOUS	PHOS-DIS		MG/L P	00666	P	Dissolved phosphorus	TDP	Total Dissolved P group
VARIOUS	TURB	JKSN	JTU	00070	na	Jackson candle units. Meaningful only above 25	Turbidity	Turbidity group
VARIOUS	TURB	TRBIDMTR	HATCH FTU	00076	na	Hach kit, formazin turbidity units (meaningful below and above 25)	Turbidity	
VARIOUS	TURBIDITY	LAB	NTU	82079	na	Laboratory measured turbidity in NTU's.	Turbidity-formazin used to calibrate, results may be comparable to 00076.	
VARIOUS	TURB	TRANS	%	00074	na	Transmissometer, percent transmission.	Turbidity	
VARIOUS	TURB	HLGE	PPM SIO2	00075	na	Hellige, ppm as silicon dioxide.	Turbidity	

***USGS (Alexander et al. 1996) defines these terms as follows:**

Total: (as applicable to N, P, TKN, N) means all the constituent that was present in a water-suspended sediment sample.

"Total": (as applicable to NO_3, NO_2, NH_3 NH_4 and orthophosphate) means that part of the constituent that was in the dissolved phase of a water + suspended-sediment sample, after the sample water was decanted off (solids where allowed to settle out) and is statistically the same as filtering with a 0.45 um filter.

Dissolved: means that the sample was filtered through a 0.45 um filter. (This is part of Standard Methods [1992] and applies to all 'dissolved' measurements.)

Table 2. Translation of NEW STORET analytical measurements to equivalent Legacy STORET parameter codes. NEW STORET fields shown were those used to undertake the translation.

NEW STORET Field Descriptors					*Legacy STORET Parameter Code*			
Characteristic	Sample Fraction	Analytical Proceedure	Analytical Proceedure Name	Other data descriptors	STORET PARAM CODE	Param 1	Param 2	Param 3
"Nitrogen, ammonia as N"	Total				00610	NH3 +NH4	N-TOTAL	MG/L
"Nitrogen, Kjeldahl"					00625	TOT KJEL	N	MG/L
"Nitrogen, Nitrite (NO2) + Nitrate (NO3) as N"					00630	NO2 + NO3	N-TOTAL	MG/L
"Nitrogen, mixed forms (NH3)+(NH4)+organic+(NO2)+(NO3)"	Total				00600	TOTAL N	N	MG/L
Phosphorus	Total				00665	PHOS-TOT		MG/L P
Phosphorus as P	Total				00665	PHOS-TOT		MG/L P
"Phosphorus, hydrolyzable plus orthophosphate as P"	Dissolved			Dissolved fraction	00666	PHOS-DIS		MG/L P
"Phosphorus, orthophosphate as P"	Total	365.1	Phosphorus by Colorimetry	Total fraction, Method 365.1, Phosphorus by Colorimetry	70507	PHOS-T	ORTHO	MG/L P
"Phosphorus, orthophosphate as P"	Dissolved			Dissolved fraction	00671	PHOS-DIS	ORTHO	MG/L P
Turbidity					82079	TURBIDITY	LAB	NTU

Table 3.0 Benthic chlorophyll *a* measurements from MT rivers & streams.

AGENCY	STORET PARAM 1	STORET PARAM 2	STORET PARAM 3	STORET CODE*	Values Reported as:	NOTES	What the parameter is a measure of:
MT-DEQ and UM	CHLRPHYL	A MG/M2	CORRECTD	32223	mg Chl *a* /m^2	In "Medium" field of NEW STORET, recorded as "Rock".	Algal chlorophyll *a* from the surface of stream-bottom rocks, collected with a template; values not estimated.
MT-DEQ and UM	CHLRPHYL	A MG/M2	CORRECTD	32223	mg Chl *a* /m^2	In "Medium" field of NEW STORET, recorded as "Rock" or "Other".	Algal chlorophyll *a* from the surface of stream-bottom rocks. Six-rocks composited. Values estimated (flagged with "J").
MT-DEQ and UM	CHLRPHYL	A MG/M2	CORRECTD	32223	mg Chl *a* /m^2	In "Medium" field of NEW STORET, recorded as "Other".	Algal chlorophyll *a* collected using the Hoop Method.
MT-DEQ and UM	CHLRPHYL	A MG/M2	CORRECTD	32223	mg Chl *a* /m^2	In "Medium" field of NEW STORET, recorded as "Sediment".	Algal chlorophyll *a* collected using the Sediment Core Method.

* Although none of these data are from Legacy STORET, the parameter code and parameter names assigned are the equivelant values as found in Legacy STORET.

In a report detailing methods used by the USGS for its stream water-quality monitoring network over the period 1962-1995, Alexander et al. (1996) indicated that the term "total", when applied to measurements of dissolved solutes, is simply a methodological term and refers to a method in which the heavier solids of a sample are allowed to settle out and then the sample water is decanted off the top for analysis. The method was statistically indistinguishable from methods that filtered the sample through a 0.45µm filter (Alexander et al., 1996). A 0.45 µm filter is by convention the separation point between dissolved and suspended forms (APHA, 1992). So when dealing with *dissolved* solute measurements in the Legacy STORET dataset, measurements labeled "diss" (dissolved) and those labeled "total" are, for practical purposes, the same. This applies to all the groups that will be discussed.

A. Ammonia Group

The group was composed of four analytical measurements, all which measure ammonia and/or ammonium; 00610, 00608, 71845 and 71846 (Table 1.0). Only the USGS used 71846 and this parameter was eliminated, as it is similar to 71845 and the USGS frequently reported both 71845 and 71846 from the same sample. The USGS also frequently reported 00610, 00608, and 71845 from a single sample; so only one (00610) was retained to avoid same-sample duplication. All other agencies used 00610, 00608, and 71845, but did not report them from the same sample; so all values from those agencies could be retained. **71845 is reported as NH_4 and, in order that the data be compatible with 00608 and 00610, all sample results for 71845 need to be multiplied by 0.7765 to convert them to 'as N'.**

The database query has two important fields that are required to be correctly queried to produce the Ammonia Group: 'Agency Code' and 'Param Code'. The query for the Ammonia Group should be written as follows:

AGENCY CODE	PARAM CODE
Not "112WRD"	71845
Not "112WRD"	00608
	00610

(Microsoft Access queries described in the remainder of this report will follow the general pattern shown above, and are applicable to the database comprising data from 'tblLEGSTORETGRAB(mostly nutrients)' [see Appendix B, Varghese and Cleland, 2004a] *and* data from NEW STORET, EMAP and the universities.)

B. TKN /Organic N Group

The database contained five analytical measurements in this group (00625, 00605, 00607, 00623, and 00624). The USGS reported from individual samples both 00625 (total Kjeldahl N) and 00605 (total organic N), and since 00605 is unique to the USGS it was eliminated. Many other different agencies reported 00625. 00607 (dissolved organic N)

is unique to the USGS but is not useful in this effort because it excludes particulate organic N. Similarly, 00623 and 00624 measure either dissolved or suspended TKN, respectively, and therefore they were eliminated. The query was written to capture only 00625 and is shown here:

AGENCY CODE	PARAM CODE
	00625

C. Total N Group

Three TN analytical measurements were located in the database (00600, 00602 and 71887). 71887 is unique to the USGS and is frequently reported along with 00600 from the same samples (71887 is total N reported at NO_3), and was eliminated from use. 00602 only measured total N from the dissolved fraction and therefore is not comparable to 00600. The query for total N is as follows:

AGENCY CODE	PARAM CODE
	00600

D. Nitrate + Nitrite Group

This group contained the largest number of analytical measurement variations and presented the most complicated query challenge. At first, the possibility of summing the individual nitrate value and individual nitrite value (e.g., 71851 and 71856) from the same sample was considered, but this presented a major database effort in order to correctly match stations, dates and times for numerous different agencies. In most natural water exposed to oxygen, nitrite is only present in trace quantities and most dissolved inorganic N is nitrate (Horne and Goldman, 1994). A review of this database showed that most nitrite measurements were very low or below the detection limit. Further, analytical measurements that analyzed for both NO_3 + NO_2 simultaneously (both 00630 and 00631 do this) were the largest contributors to the group in terms of sample numbers, and therefore nitrite is included in the bulk of the samples. For simplicity, it was decided to eliminate all nitrite-only measurements (00615, 00613, and 71856) from use (Table 1.0).

Large numbers of USGS samples reported both 71851 (nitrate as NO_3) and 00618 (nitrate dissolved, as N) from the same sample. But there were only 114 non-USGS records for 71851 (for streams & rivers) in the database. Therefore, 71851 was completely eliminated (Table 1.0). The USGS and the U.S. Forest Service were the only agencies to measure 71850 (“total” nitrate as NO3), and sometimes reported both 00618 and 71850 from the same samples; therefore, 71850 (the measurement with fewest values) was eliminated.

The remaining four ‘nitrate’ or ‘nitrate + nitrite’ measurements (00630, 00631, 00618 and 00620) were queried on an agency-by-agency basis to determine which agencies reported more than one of these measurements from a single sample. It resulted that the USGS

frequently reported all four, and so their samples are restricted in the query only to 00630 (NO3 + NO2 N-TOTAL). MT DEQ (Legacy STORET agency code 21MTHDWQ) did, on very rare occasions, report 00630 and 00620 from the same sample, but in the vast majority of cases reported values were unique and so all four measurements are retained. EPA (agency code 11EPALES) usually reported 00630 paired with 00620, and so 00620 was excluded. EMAP reported both 00618 and 00631, but there were only a few 00631 samples so that parameter was eliminated for EMAP and 00618 was retained. None of the remaining agencies reported more than one of the four measurements from any single sample. 11NPSWRD was excluded because the few sample shown in the database appeared to be end-of-pipe data, even though such data should not have been part of assembled dataset. The final query is shown below:

Agency Code	Param Code
Not "112WRD" and Not "11NPSWRD" and Not "EMAP"	00631
Not "11NPSWRD"	00630
Not "112WRD" and Not "11NPSWRD"	00618
Not "112WRD" and Not "11EPALES" and Not "11NPSWRD"	00620

E. Total P Group

AGENCY CODE	PARAM CODE
Not “112WRD”	71886
	00665

This group was comprised of four measurements (00665, 71886, 00669, and 00678).

00669 and 00678 were completely excluded because there was some uncertainty in the comparability of these samples to 00665 and 71886 and, together, they represented < 1% of the data of this group. The two remaining measurements (00665 and 71886) were both found to be frequently reported on from a single sample by the USGS, and so the query restricts USGS data to 00665. All other agencies reported one or the other from any given sample. **Values from 71886 are reported as PO_4 and will need to be converted to units of P to combine with 00665, therefore all samples under the 71886 code need to be multiplied by 0.3261.** The query is shown below:

F. Soluble Reactive Phosphate (SRP) Group

This group consists of four analytical measurements (00671, 70507, 00660, and 00650). 00650 was eliminated completely because of uncertainty in the comparability of this analytical measurement to the other three. The USGS frequently reported the remaining three (00671, 70507, 00660) from a single sample, but one of them (00660) was simply a reporting unit change (00660 is reported as PO_4) and of the remaining two, there were six times more records for 00671 than for 70507. Therefore, for the USGS only 00671 is retained. EPA (agency code 1119C050) often reported 70507 and 00671 from a single sample and therefore the query restricts that agency to just 70507. The other agencies did not report more than one measurement from a single sample. The query is shown below:

AGENCY CODE	PARAM CODE
Not “112WRD”	70507
Not “112WRD”	00660
Not “1119C050”	00671

G. Total Dissolved Phosphorus (TDP) Group

Only one analytical measurement was found in the database (00666). No restrictions are placed on any agency for this query, which is shown below:

AGENCY CODE	PARAM CODE
	00666

H. Turbidity Group

The turbidity group consists of five analytical turbidity measurements (00070, 00076, 82079, 00074, and 00075). 00075 was eliminated as no record of what it actually measures could be found. 00074 was eliminated after a conversation with the USGS revealed that 00074’s units (percent light light transmissivity) could not be directly converted to the other commonly used turbidity unit, JTU or NTU (John Lambing, personal communication). The remaining three measurements (00070, 00076, and 82079) are based on differing measurement methodologies relying on visible light and infrared light respectively. There is no consistent conversion mechanism that can be applied to standardize 00070 and 00076 measurements; the measurements (JTU and FTU) were therefore analyzed separately. Further, 00070 (JTU) values are only meaningful when they are higher than 25 (Hach Chemical Co., 1975). 00076 (FTU) and 82079 (NTU) values are evidently comparable because 82079 values were calibrated using formazin; 00076 values are meaningful down to a value of 1.0 (Hach Chemical Co., 1975; APHA, 1992), whereas 82079 values are meaningful down to 0.02 (APHA, 1992). No agency reported all three, or pairs, from the same sample. **After the query is run, all values associated with 00070 that are less than 25 should be eliminated, and in any analyses the values from 00070 should be kept separate from those of 00076 and 82079**. The query is shown below:

AGENCY CODE	PARAM CODE
	00070
	00076
	82079

I. Benthic Chlorophyll *a* Group

There is a single analytical measurement for this group (32223), although there are several ways in which benthic Chl *a* were collected (Table 3; see Suplee 2004 or MT

DEQ SOPs for more explanation). All of the samples have been adjusted for phaeophytins (monochromatic method; APHA, 1992), and are only measures of *algal* Chl *a*. By definition one of the rock methods, wherein 6 stream bottom rocks are collected and the total attached Chl *a* is extracted (see Notes, table 3.0), is considered an estimate ("J" flag in the Remark Code field). This is because it is assumed that the Chl *a* was recovered from 50% of the rocks' total surface area, which may or may not be true. The query for these data is shown below.

AGENCY CODE	PARAM CODE
	32223

References

Alexander, R. B., Ludtke, A. S., Fitzgerald, K. K., and T. L. Schertz. 1996. Data from selected U.S. geological survey national stream water-quality monitoring networks (WQN) on CD-ROM. United States Geological Survey, Open-File Report 96-337.

American Public Health Association (APHA). 1992. Standards methods for the examination of water and wastewater, 18th edition. American Public Health Association, Washington, D.C.

Hach Chemical Co. 1975. Procedures, chemical lists and glassware for water and wastewater analysis with calibrations for the Bausch and Lomb Spectronic 20, Mini Spec 20 and other popular spectrophotometers. 2nd edition. Hach Chemical Company, Ames, Iowa.

Horne, A. J., and C. R. Goldman. 1994. Limnology, 2nd edition. McGraw-Hill.

Lambing, J. 2004. USGS Water Quality Specialist. Personal communication. Helena, MT, May 10th, 2004.

MT DEQ. 2005. Sample collection and laboratory analysis of chlorophyll-*a*, standard operation procedure. QPBWQM-011, review draft of 5/16/2005. Available from the MT Department of Environmental Quality, 1520 E. 6th Ave, Helena, MT 59620.

MT DEQ. 2007. Sampling and analysis plan: Sampling reference sites to produce a more uniform nutrient dataset and improve Montana's wadeable stream nutrient criteria. Montana Department of Environmental Quality, Water Quality Planning Bureau, May 21, 2007.

Mueller, D.K., Hamilton, P.A., Helsel, D.R., Hitt, K.J., and B.C. Ruddy. 1995. Nutrients in ground water and surface water of the United States —An analysis of

data through 1992. U.S. Geological Survey Water-Resources Investigation Report 95-4031. Denver, CO, 74 p.

STORET. 1996. Printout of STORET parameter numbers and their associated descriptions. Printed from the EPA STORET system on September 11, 1996 (on file at the MT Department of Environmental Quality, Helena, MT).

Suplee, M. 2004. Wadeable streams of Montana's Hi-line region: an analysis of their nature and condition, with an emphasis on factors affecting aquatic plant communities *and* recommendations to prevent nuisance algae conditions. Montana Department of Environmental Quality, May 2004. Available at: http://www.deq.state.mt.us/wqinfo/Standards/Master_Doc_DII.pdf

Suplee, M.W., Varghese, A., and Cleland, J. 2007. Developing nutrient criteria for streams: an evaluation of the frequency distribution method. Journal of the American Water Resources Association 43: 453-472.

U.S EPA. 1979. Methods for chemical analysis of water and wastes. United Stated Environmental Protection Agency, publication No. EPA-600/4-79-020. March 1979.

U.S. EPA. 2000. Ambient water quality criteria recommendations, information supporting the development of State and Tribal nutrient criteria for rivers and streams in nutrient ecoregion II. United States Environmental Protection Agency, Office of Water, EPA-822-B-00-015. Washington, D.C. *Available at* http://www.epa.gov/waterscience/criteria/nutrient/ecoregions/rivers/rivers_2.pdf

Varghese, A. and J. Cleland. 2004a. Stratification of algae and nutrient data of Montana rivers and streams to assist in the development of numeric standards-final report. Submitted to Dr. Michael Suplee, Montana Department of Environmental Quality, Helena, MT. July 13, 2004.

Varghese, A. and J. Cleland. 2004b. Montana rivers and streams reference-to-whole-population percentile mapping for selected water quality parameters-final report. Submitted to Dr. Michael Suplee, Montana Department of Environmental Quality, Helena, MT. November 19, 2004.

Varghese, A. and J. Cleland. 2005. Seasonally stratified water quality analysis for Montana rivers and streams-final report. Submitted to Dr. Michael Suplee, Montana Department of Environmental Quality, Helena, MT. June 29, 2005.

Varghese, A. and J. Cleland. 2006. Bootstrap confidence intervals and updated statistical analayses of water quality data for Montana rivers and streams. Submitted to Dr. Michael Suplee, Montana Department of Environmental Quality, Helena, MT. June 30, 2006.

Appendix B. Seasonal Grouping Methodology

Prepared by: Michael Suplee, Ph.D., MT DEQ, June 10, 2008

Determination of Seasonal Periods for the Purpose of Segregating Nutrient and Chlorophyll *a* Data by Time-of-Year.

Results from earlier analysis (Varghese and Cleland, 2005) demonstrate that level III and level IV ecoregions (Omernick, 1987) are meaningful tools for the purpose of stratifying nutrient data in Montana. At the time this work was undertaken, it was also hypothesized that ecoregions might show distinct inter-seasonal nutrient concentrations and, therefore, stratifying the nutrient data by seasons would further improve our ability to characterize regional nutrient concentrations. Flowing waters often demonstrate distinct seasonal patterns (Lohman and Priscu, 1992). For example, phosphorus is frequently associated with total suspended sediment (TSS)(Froelich, 1988), and during spring runoff in streams both TSS and total phosphorus can be orders-of-magnitude higher than at other times (Horne and Goldman, 1994). Seasonal variation in stream nutrient concentrations are not only influenced by abiotic factors such as runoff patterns, they are also influenced by biological uptake and release by organisms such as aquatic plants. Aquatic plant growth — including algal growth — is influenced by (among other things) light availability and temperature, which are themselves climatically driven. Therefore, the development of seasons to define nutrient concentrations had to consider not only hydrologic patterns, but also climatic factors such as the onset of cold winter temperatures.

Work was undertaken in 2004 and 2005 to define three seasons for each ecoregion: a growing season, which would roughly correspond to the summer months; winter, which would follow the growing season; and runoff, which would terminate the winter period and comprise the yearly high flow period. A detailed description of the techniques used to determine the growing, winter, and runoff seasons was originally provided in Appendix B of Varghese and Cleland (2005). These methods were also presented on pages 457-459 of Suplee et al. (2007). The purpose of this appendix is to describe some minor changes to the start and end dates that have resulted from work carried out since the publication of the aforementioned documents.

Updated Cutoff Dates for Winter, Run-off and Growing Seasons

The final, updated cutoff dates for the three seasons in each ecoregion are shown in Table 1.0 below. They were derived using a combination of the methods and considerations discussed in Varghese and Cleland (2005) and Suplee et al. (2007). The changes affect the starting period for the growing season of the Middle Rockies and Idaho Batholyth ecoregions. The growing season start date has been moved back to July 1, from July 16. This was done after a careful re-review of the ecoregion hydrograph dataset. The review showed that an end-of-June endpoint for runoff was, overall, more representative of the region and, given the general warming trend observed in Montana, setting the date two

weeks earlier makes good sense. For simplicity, all dates are still rounded to the nearest mid- or end-of-month date.

Table 1. Updated Start and Ending Dates for Three Seasons (Winter, Runoff and Growing), by Level III Ecoregion.

Ecoregion Name	Start of Winter	End of Winter	Start of Runoff	End of Runoff	Start of Growing Season	End of Growing Season
Canadian Rockies	Oct.1	April 14	April 15	June 30	July 1	Sept. 30
Northern Rockies	Oct.1	March 31	April 1	June 30	July 1	Sept. 30
Idaho Batholith	Oct.1	April 14	April 15	June 30	July 1	Sept. 30
Middle Rockies	Oct.1	April 14	April 15	June 30	July 1	Sept. 30
Northwestern Glaciated Plains	Oct.1	March 14	March 15	June 15	June 16	Sept. 30
Northwestern Great Plains	Oct.1	Feb. 29	March 1	June 30	July 1	Sept. 30
Wyoming Basin	Oct.1	April 14	April 15	June 30	July 1	Sept. 30

References

Froelich, P. N. 1988. Kinetic control of dissolved phosphate in natural rivers and estuaries: A primer on the phosphate buffer mechanism. Limnology and Oceanography 33: 649-668.

Horne, A. J., and C. R. Goldman. 1994. Limnology, 2nd edition. McGraw-Hill Inc.

Lohman, K., and J. C. Priscu. 1992. Physiological indicators of nutrient deficiency in *Cladophora* (chlorophyta) in the Clark Fork of the Columbia River, Montana. J. Phycol. 28: 443-448.

Omernik, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers 77: 118-125.

Suplee, M.W., Varghese, A., and Cleland, J. 2007. Developing nutrient criteria for streams: an evaluation of the frequency distribution method. Journal of the American Water Resources Association 43: 453-472.

Varghese, A. and J. Cleland. 2005. Seasonally stratified water quality analysis for Montana rivers and streams-final report. Submitted to Dr. Michael Suplee, Montana Department of Environmental Quality, Helena, MT. June 29, 2005.

Appendix C. Lithographic Grouping Methodology, Nutrient – Lithology – Geology Layer

Prepared by Jim Stimson, Montana Department of Environmental Quality, Source Water Protection Program. June 7, 2004

Source of Geologic Information: Gary L. Raines and Bruce R. Johnson, 1996, Digital representation of the Montana state geologic map: a contribution to the Interior Columbia River Basin Ecosystem Management Project, USGS Open File Report 95-691, ArcView Shapefile for Montana.

Source of information on geologic sources of Nitrogen: Holloway, J.M., Dahlgren, R.A., and Casey, W. H., 1998, Contribution of bedrock nitrogen to high nitrate concentrations in stream water, Nature: Vol 395 October 1998, p. 785 – 788.

Reason for lithology groupings: Based on information from the Holloway et al. study forms of nitrate (ammonium) can substitute into the mica crystal structure. In addition, several other sources identify marine shale rocks as containing significant amounts of nitrate. Therefore, sedimentary shale rocks and metamorphic that contain significant volumes of mica minerals, like biotite and muscovite, were two of the lithology categories culled out of the digital version of the 1955 Geologic map of Montana.

How the lithology groupings were made: The digital version of the 1955 Geologic map of Montana comes with an attribute data table that includes multiple fields including eight lithology fields. In transferring the original lithologic descriptions from the 1955 geologic map, the dominant lithology type was placed into the Lith1 field, the next abundant lithology was placed into the Lith2 field, and so on. These lithology fields were used to isolate 17 lithology types. The dominant lithology names taken from the Lith1 and Lith2 fields are included in the lithology shapefile names. A quick review of the attribute table of each shapefile shows which specific lithologies are included in each group. Once a group of lithology types was selected from the original geologic shapefile, the selected records were exported into a new shapefile bearing the name of new lithologic group.

Lithologic Groups: The following lithologic shapefiles have been created:

1) **Igneous Intrusives and extrusives but not volcanics**. Major types of igneous rock. Volcanics excluded.

2) **Igneous volcanics**. Rocks identified specifically as volcanics.

3) **Metamorphic Schist.** Schist is a name applied a metamorphic rock dominated by aligned mica minerals usually biotite and muscovite but there is a host of other mica minerals. Because the schist rocks contain significant volumes of mica minerals they are thought to be a possible source of geologic nitrate.

4) **Metamorphic Schist and Argillite**. Argillite is a name usually applied a fine-grained rock like a shale or mudstone that is quite hard and breaks into angular "chucks". Argillites are thought to have been metamorphosed so that the original minerals have been re-crystallized and have grown, and exhibit alignment with the general bedding direction. Because the argillite rocks contain significant volumes of mica minerals they are thought to be a possible source of geologic nitrate.

5) **Sedimentary Sandstones**. Thought to not be a source of geologic nitrate.

6) **Sedimentary Quartzite**. A metamorphosed sandstone composed almost entirely of quartz grains. Thought to not be a source of geologic nitrate.

7) **Sedimentary Gravel, Boulders, and Conglomerates**. Range front gravel deposits and the Flaxville gravels. Flaxville gravels host aquifers that can have elevated nitrate levels related to agricultural practices. Otherwise, thought to not be a source of geologic nitrate.

8) **Sedimentary Limestone and Dolomite**. Thought to not be a source of geologic nitrate.

9) **Sedimentary Argillaceous Limestone**. Limestones with a significant volume of fine grained material (argillite) and could be a source of geologic nitrate.

10) **Sedimentary Shaly Sandstone**. Sandstones with a significant volume of shale and could be a source of geologic nitrate.

11) **Sedimentary Sandy Shale**. Sandstones with a significant volume of shale and could be a source of geologic nitrate.

12) **Sedimentary Shale and Clay**. A grouping of several geologic formations where the Lith1 and Lith2 fields listed shale and clay. Could be a source of geologic nitrate.

13) **Sedimentary Carbonaceous Shale**. Shales with high organic (carbonaceous) content. Could be a source of geologic nitrate.

14) **Sedimentary Siliceous Shale**. Shales with relatively high silica content. Could be a source of geologic nitrate.

15) **Sedimentary Calcareous Shale**. Shales with relatively high carbonate (calcite or dolomite, or other carbonates) content. Could be a source of geologic nitrate.

16) **Sedimentary Argillaceous Shale**. I believe these are shales with relatively high mica content. The could also be shales that are more lithified and hard but I think the term is used to express that the rock has identifiably large mica crystals present. Could be a source of geologic nitrate.

17) **Sedimentary Silt Dominated**. Since there can be a fine line between shales and silts, I pulled out formations that listed silt as the dominant lithology. Usually the term silt is

used to describe rocks that are more gritty than shales or clays because the silt rock contains more grains of quartz. If they are associated with shale beds, they could be a source of geologic nitrate.

18) **Sedimentary Silt Shale – Hell Creek Formation.** The Hell Creek is part of the Fort Union Formation and is usually described as containing more shale beds than the Fox Hills member (located below the Hell Creek) and the Tongue River member (located above the Hell Creek. Could be a source of geologic nitrate.

Appendix D. Log Files of *Stata* Statistical Analyses

Enclosed with this report is a CD-ROM that contains the detailed output of the statistical analyses conducted for this project. The CD contains several hundred statistical analyses (i.e., *Stata* log files) for ten water quality parameters evaluated in various combinations of coarse- and fine-scale stratifying parameters for different seasons, as described in Section 4 of the report. Below, we provide guidance on locating the files on the CD and finding specific statistical results within the log files.

1.0 Overview of Folders

1.1 Contents of Folders

The RUNS folder on the CD-ROM contains two subfolders. Table D-1 describes the contents of each subfolder and the databases analyzed.

Table D-1: Folder Contents

Folder Name	Contents	Databases Analyzed
SUMSTATS	Summary Statistics	All-Observation; Median; All Observation Excl. 7&8 Databases
PTILE	Percentiles of Interest and Distinct Zone Assessments	Median

Detailed guides to the contents of each folder are provided in Section 2 of this appendix.

1.2 Guide to Terminology

The log files contain a number of symbols, or codes, used as short-hand for water quality parameters and stratifying parameters in the database and in the statistical programs. In many cases, the log files provide comments that translate these symbols or codes into plain English equivalents. However, such translation was not always possible, given the complexity of some of the procedures. Therefore, the tables below provide a key to the codes used in this analysis.

Tables D-3, D-4, and D-5 list the codes used for water quality parameters, stratifying parameters, and geologic formations, respectively. Figure D-1 (page D-3) presents the codes used to represent Omernik Levels III and IV ecoregions.

Table D-3: Water Quality Parameter Codes

Code	Water Quality Parameter
AMMONIA	Ammonia (mg/l)
BNCHLOR-A	Benthic Chlorophyll-*a* (mg/m^2)
NO3+NO2	Nitrate and Nitrite group (mg/l)
SRP	Soluble Reactive Phosphate (mg/l)
TDP	Total Dissolved Phosphorus (mg/l)
TKN	Total Kjeldahl Nitrogen (mg/l)
TOTALN	Total Nitrogen (mg/l)
TOTALP	Total Phosphorus (mg/l)
TURB-NTU	Turbidity (NTU)
TURB-JTU	Turbidity (JTU)

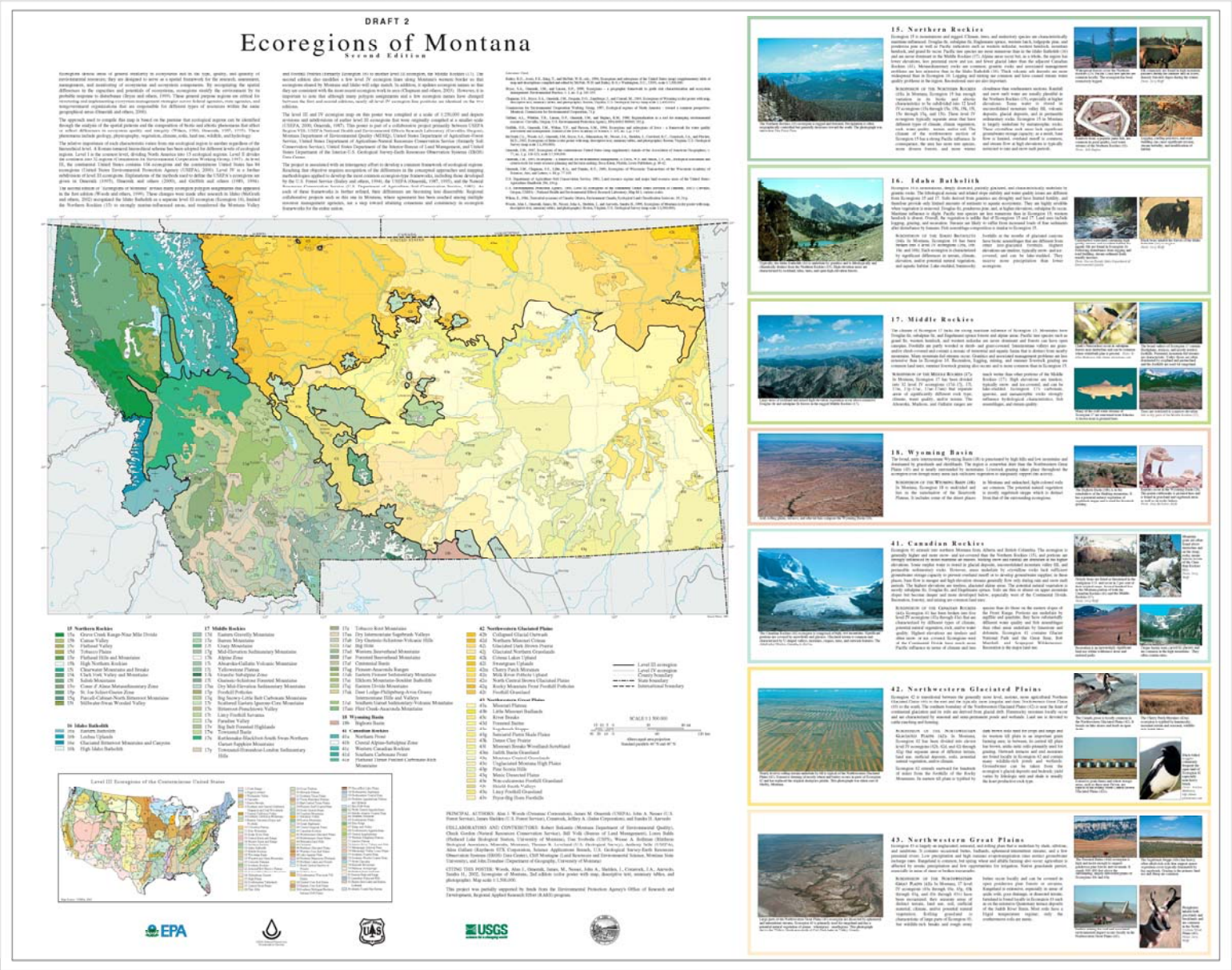
Table D-4: Stratifying Parameter Codes

Code	Stratifying Parameter
eco3	Omernik level III ecoregions
eco4	Omernik level IV ecoregions
sfname	Geologic formation
geco	Grouped level III ecoregions (Mountain or Prairie)
strahgroup	Grouped Strahler Stream Order (1&2, 3&4, 5&6 and 7&8)

Table D-5: Geologic Formation Codes

Code	Geologic Formation
1	Metamorphic Argillite
2	Argillaceous Shale
3	Argillaceous Limestone
4	Calcareous Shale
5	Carbonaceous Shale
6	Gravel, Boulders, Conglomerate
7	Hell Creek Silty Shale
8	Chert
9	Igneous but not Volcanic
10	Limestone and Dolomite
11	Quartzite
12	Sedimentary Sandstone
13	Sandy Shale
14	Metamorphic Schist
15	Shale and Clay
16	Shaly Sandstone
17	Siliceous Shale
18	Sedimentary Silt Dominated
19	Cretaceous and Tertiary Volcanics

Figure D-1: Omernik Ecoregions in Montana



2. Detailed Guide to Folders

2.1 SUMSTATS

The SUMSTATS folder contains the summary statistics generated as part of the current task (Task 5). The folder contains separate sub-folders for the all observations database, median database, and the all observations database excluding stream orders 7&8. Within each of these sub-folders are further sub-folders leading to coarse- and fine-scale analyses. Within the coarse or fine folders are ten sub-folders for each nutrient grouping analyzed. Within the nutrient folders, users should follow the path leading to the stratification methodology of interest. For instance to locate summary statistics for level IV ecoregions within level III ecoregions for total phosphorus using the all observations database, follow the path:

SUMSTATS/ALLOBS/FINE/TOTALP/ECO3/ECO4/*TOTALP.log*

Each summary statistics log file is self-explanatory with several comments to guide the user. Descriptive statistics are sequentially reported for each season within each level of the fine-scale parameter within each level of the coarse-scale parameter. Units are reported in the header.

The summary statistics log files are similar in structure to those contained in previous analyses. In order to improve ease of access to the descriptive statistics, the log files generated in the current analysis report summary statistics exclusively and do not contain the results of any other statistical tests. The results of statistical tests have been separately reported in other folders.

2.2 PTILE

The PTILE folder contains ten subfolders, one for each water quality parameter listed above in Table D-3. Within the water quality parameter subfolder, is a single log file.

For example, to access the percentile mapping results for TKN, follow the path:

Ptile/TKN and open the log file TKN.log

To access the percentile mapping results for SRP, follow the path:

Ptile/SRP and open the log file SRP.log.

The log files have the following general structure. For each category of the level III ecoregion stratifying variable, each season is considered in turn, beginning with the results for all seasons combined. For each season, the log file computes the observed values of the percentiles of interest for the current level III ecoregion as well as the 90% bootstrap confidence intervals around those percentiles of interest. Thereafter, the distinct zone assessment commences to identify level IV ecoregions within the current level III ecoregion for the current season that may be considered to be statistically distinct from their parent level III ecoregion. If a particular level IV ecoregion is found to be distinct, then the log file will report the observed values of the percentiles of interest for that distinct level IV ecoregion as well as the bootstrap confidence

intervals. After considering all the level IV ecoregions within the current level III ecoregion for the current season, the analysis proceeds to the next season. After all the seasons have been analyzed, the log file proceeds to report the results for the next level III ecoregion. The log file is supplied with several explanatory comments that will make evident to the reader at every point the nutrient, season, and ecoregion being analyzed. The results of the statistical analysis are presented both in statistical terms as well as with interpretative comments in plain language.

It may be noted that the analysis offers three types of bootstrap estimates of the percentiles of interest: (i) based on the normal approximation, (ii) based on observed bootstrap sample percentiles (iii) based on a system of bias correction. We recommend the use of the percentile-based bootstrap confidence interval estimates because these estimates are free of distributional assumptions. The alternative bootstrap estimates and the recommended option are illustrated below.

```
OBSERVED VALUES AND BOOTSTRAP 90% CI FOR PERCENTILES OF INTEREST
FOR NUTRIENT = AMMONIA, LEVEL 3 ECOREGION= Northern Rockies & SEASON = ALL SEASONS

(obs=12)

Bootstrap statistics

Variable |   Reps   Observed       Bias   Std. Err.   [90% Conf. Interval]
---------+-------------------------------------------------------------------
  Refp75 |   1000    .005875   .0001769   .0011067    .0040529  .0076971  (N)
         |                                                .005   .008875  (P)
         |                                            .0053014   .008875 (BC)
---------+-------------------------------------------------------------------
  Refp80 |   1000    .006025   .0004218   .0015071    .0035437  .0085063  (N)
         |                                               .0051     .0115  (P)
         |                                            .0056175     .0115 (BC)
---------+-------------------------------------------------------------------
  Refp85 |   1000    .006175    .001069   .0020203    .0028487  .0095013  (N)
         |                                            .0056028     .0115  (P)
         |                                            .0057058     .0115 (BC)
---------+-------------------------------------------------------------------
  Refp90 |   1000    .007825   .0004408   .0022163    .0041761  .0114739  (N)
         |                                             .005825     .0115  (P)
         |                                              .00625     .0115 (BC)
-------------------------------------------------------------------------------
                              N = normal, P = percentile, BC = bias-corrected
```

Refp75, *Refp80*, *Refp85* and *Refp90* in the extract from the log file above refer to the 75th, 80th, 85th, and 90th percentile from the reference population, respectively. The column titled "*Observed*" presents the observed values of the percentiles of interest for the data in question. For bootstrap 90% confidence intervals, we recommend the use of the highlighted bootstrap confidence intervals, which represent the percentile-based estimates, as described above.

Appendix E. Dataset Documentation

This appendix lists the names and contents of the datasets created and delivered as part of this project. It also documents the sources of these data.

1.0 Dataset Deliverables

1. The *Microsoft Access* file *Task5DataTable.mdb* contains a single data table named *Task5DataTable*. This dataset represents the "All Observations" database described in Section 3 of this report. The dataset contains a total of 138,914 observations with 32 variables. Viewing the dataset in "design view" in *Microsoft Access* will display labels that explain what each variable represents. Appendix D provides further detail on specific variable codes. This dataset has been processed so that concentration values are adjusted for non-detects. Seasonal, geological and nutrient groupings have also been performed per the descriptions in Appendices A, B and C.

2. The *Stata* data file *t5prmtdb.dta* represents the "All Observations" database described in Section 3 of this report. The dataset contains a total of 138,914 observations with 32 variables. This dataset has been processed so that concentration values are adjusted for non-detects. Seasonal, geological, and nutrient groupings have also been performed per the descriptions in Appendices A, B and C.

3. The *Stata* data file *t5medmtdb.dta* represents the "Median" database described in Section 3 of this report. The median database contains a unique observation for each nutrient, for each station, for each season. This unique observation is the median observed value for a nutrient, at a particular station, in a particular season. This database contains a total of 43,386 observations. This dataset has also been processed so that concentration values are adjusted for non-detects. Seasonal, geological, and nutrient groupings have also been performed per the descriptions in Appendices A, B and C.

2.0 Data Sources

The datasets described above have evolved over several phases of analysis. The sources of the data have been described in previous reports (ICF 2004a, 2004b, 2005, and 2006) and the current updates are described in Section 3 and also in this Appendix (see below). Many of the data gathering and processing steps were performed by MT DEQ, while others were performed by ICF in accordance with MT DEQ directions. A brief summary of the sources and construction of the dataset is provided below.

The primary data source for this dataset was the U.S. EPA's Storage and Retrieval (STORET) database. In March 2001, MT DEQ placed a request with the then-functioning mainframe STORET database for all ambient surface water-quality data for Montana, excluding data from pipes, wells and springs. The delimited text file received was then transferred to a Microsoft Access relational database. The STORET data (also referred to as Legacy STORET) contained

data collected by 33 agencies or entities, and held nutrient data from the early 1960s to 1998. A query was run in the "Type" field (a field indicating the waterbody type) to remove lake data. The database was supplemented with all river and stream nutrient data from MT DEQ found in modernized STORET, which were collected from 2000-2004. Also added to the relational database were Montana river and stream data collected by the University of Montana, Utah State University, and the U.S. EPA's Environmental Monitoring and Assessment Program (EMAP; Lazorchek et al., 1998).

Each analytical measurement in Legacy STORET was uniquely identified by a parameter code (e.g., 00665; total P). Other data that were incorporated into the relational database, including those from modernized STORET, did not use these codes. To assure consistency and to facilitate the grouping of data, the appropriate parameter code was assigned to each observation lacking a code. The water quality data in the assembled database, which included latitude and longitude coordinates for each observation, were then spatially joined to Geographic Information System (GIS) layers containing information on level III and level IV ecoregions (Woods et al., 2002). The final database was transferred to *Stata* (version 7), which was more amenable to statistical analysis programming,

Examination of the Legacy STORET dataset confirmed that it did not contain water quality data from pipes, wells or springs. This was established by determining if pipe, well or spring sampling stations, which were included in a Legacy STORET metadata (station-information) file, could be joined with any of the data in the database. To eliminate potentially erroneous or highly uncertain data from the analyses, data bearing certain comments codes were excluded, as described in Section 3. Further, observations in the database bearing comment codes indicating that the analysis was below detection levels were replaced with values equal to 50% of the reported detection limits. For datasets skewed to the right, the DL/2 method is reported to be sufficiently accurate for determining descriptive statistics like the mean and standard deviation (Hornung and Reed, 1990). Less than 15% of the total dataset was adjusted in this manner, which should not have significantly influenced the results and is consistent with U.S. EPA guidance (U.S. EPA, 2006). Also eliminated from use were water quality observations that had reported values of zero. Most analytical results in the database provided a result value and a detection limit, or an indication that the measurement was below the detection limit. True analytical values of zero are very unlikely and therefore these data were not used in the analyses. (NWQL, 1994)

The identification and assessment of Montana reference stream sites is discussed in detail in Suplee et al. (2005); please see that document for details. All data associated with reference sites were flagged in the *Stata* database to distinguish them from non-reference data. The locations of reference and non-reference sites and their corresponding level III ecoregions are shown in Figure 3.1 in the main report.

Finally, for the current database, MT DEQ provided a total of 2,815 new observations (including 428 benthic chlorophyll overwrites for existing data that were found to have had an analytical error). These new data were spatially joined to the appropriate GIS layers to incorporate data on their Omernik level III and level IV classifications, geological formations, and altitude. The vast majority of these new data were from MT DEQ reference sites. The data had been collected in a

manner intended to make each reference site in the nutrient database more equitably representative. A database review in 2006 showed that some reference sites contributed an unusually large proportion of the nutrient observations to a given level III ecoregion. In 2007, MT DEQ used an evenness index and then targeted under-represented reference sites for nutrient sampling (TP, SRP, TN, TKN, NO_{2+3}) (MT DEQ 2007). After having incorporated these new data into the database, each reference site now contributes more equitably to the whole in any given ecoregion.

3.0 References

Hornung, R.W., and L.D. Reed, 1990. Estimation of Average Concentration in the Presence of Nondetectable Values. Applied Occupational and Environmental Hygiene 5: 46-51.

ICF, 2004a. "Stratification of Algae and Nutrient Data of Montana Rivers and Streams to Assist in the Development of Numeric Standards – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, July 13, 2004.

ICF, 2004b. "Montana Rivers and Streams Reference-to-Whole-Population Percentile Mapping for Selected Water Quality Parameters – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, November 22, 2004.

ICF, 2005. "Seasonally Stratified Water Quality Analysis for Montana Rivers and Streams – Final Report," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, June 29, 2005.

ICF, 2006. "Bootstrap Confidence Intervals and Updated Statistical Analyses of Water Quality Data for Montana Rivers and Streams," Memorandum to Dr. Michael Suplee, Montana Department of Environmental Quality, from Arun Varghese and Joshua Cleland, ICF Consulting, June 27, 2006.

Lazorchak, J.M., D.J. Klemm, and D.V. Peck, 1998. Environmental Monitoring and Assessment Program-Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams. United States Environmental Protection Agency, EPA/620/R-94/004F, Washington, D.C. Available at:
http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwatr/field/ws_abs.html

MT DEQ (Montana Department of Environmental Quality), 2007. Sampling and Analysis Plan, and Associated White Paper: Sampling Reference Sites to Produce a More Uniform Nutrient Dataset and Improve Montana's Wadeable Stream Nutrient Criteria. Finalized May 21, 2007.

NWQL, 1994. Technical memorandum 94-12 from National Water Quality Laboratory to NAWQA study-unit chiefs, July 8, 1994. *Available at*
http://nwql.usgs.gov/Public/tech_memos/nwql.94-12.html)

Seaber, P.R., F.P. Kapinos, and G.L. Knapp, 1987. Hydrologic Unit Maps. U.S. Geological Survey Water-Supply Paper 2294. 63 p.

U.S. EPA, 2006. Data Quality Assessment: Statistical Methods for Practitioners. United States Environmental Protection Agency, Office of Environmental Information, EPA/240/B-06/003. Washington, D.C. *Available at* http://www.epa.gov/quality1/qs-docs/g9s-final.pdf

Woods, A.J., J.M. Omernik, J.A. Nesser, J. Shelden, J.A. Comstock, and S. J. Azevedo, 2002. Ecoregions of Montana, 2nd edition. (Color Poster with Map, Descriptive Text, Summary Tables, and Photographs): Reston, Virginia, U.S. Geological Survey (map scale 1:1,500,000).

Appendix F. Nutrient Criteria Developer Manual

All the statistical analyses included in the report, with the exception of the compliance and non-compliance tools described in Section 6 of the main report, are generated by means of computer programs written in the *Stata* programming language. These programs are contained in text files, known as "do files" in *Stata* terminology.

This manual documents all the programs used to generate statistical analyses in the current project, as well as in the previous phase of the project (ICF, 2006). The programs, or do files, are stored on a CD accompanying this report. The programs may be accessed using any text editor or with the *Stata* do file editor. Table F-1 presents a guide to the names, locations, and purposes of all the do files included on the CD.

Although all the do files listed in Table F-1 were written in *Stata* version 7, they have been rendered backward version compatible so that users operating with more advanced versions of *Stata* will be able to run these files. Users operating with older versions of *Stata* (i.e., earlier than version 7) may not be able to run the files.

In addition to the do files described above, the analysis also relies upon a limited number of user-defined program files, which are known as "ado files" in *Stata* terminology. These ado files may be accessed by means of any text editor or by using the *Stata* do file editor; it is recommended that users not make any edits to the provided ado files. These ado files must be installed on the user's hard drive at c:\ado\personal before the do files can be run. Table F-2 presents a guide to the name, location and function of ado files necessary to run these analyses.

Apart from the *Stata* program files described above, this project also required the creation of certain tools in *Microsoft-Excel*, as described in Section 6 of the main report. Table F-3 provides a guide to the location and purpose of these files. These tools contain macros that are governed by computer code written in the *Visual Basic for Applications (VBA)* programming language. The macros are locked to prevent unintentional modifications but the underlying code may be examined and modified by using the password separately provided as part of the project deliverables. The content of these macros is not included here because it may be confusing to view the macro outside of the context of the particular spreadsheet for which it was written.

Table F-1: Developer's Guide to *Stata* Do Files

Name of Do File	Location of File on CD	Purpose of Do File
changepoint.do	AppendixF_DoFiles\Changepoint_DoFile	Runs the changepoint analysis
ptile.do	AppendixF_DoFiles\Task5_DoFiles\PTILE	Computes the observed values of the percentiles of interest and also runs the distinct zone assessment
t5sumstatsallobs.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the all observation database for all nutrients[1] at the coarse scale for various stratifying parameters
t5sumstatsallobs-fine.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the all observation database for all nutrients at the fine scale for various stratifying and sub-stratifying parameters
t5sumstatsallobsno7&8.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the all observation database excluding data from stream orders 7&8 for all nutrients at the coarse scale for various stratifying parameters
t5sumstatsallobsno7&8-fine.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the all observation database excluding data from stream orders 7&8 for all nutrients at the fine scale for various stratifying and sub-stratifying parameters
t5sumstatsmedian.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the median database for all nutrients at the coarse scale for various stratifying parameters
t5sumstatsmedian-fine.do	AppendixF_DoFiles\Task5_DoFiles\SUMSTATS	Presents summary statistics by season for the median database for all nutrients at the fine scale for various stratifying and sub-stratifying parameters
t5mtmeddb.do	AppendixF_DoFiles\Task5_DoFiles\MEDIANDB	Converts the all observation database to a median database
bsallobs.do	AppendixF_DoFiles\Task3_DoFiles\BS	Computes observed values of percentiles of interest and bootstrap confidence intervals by season for the all observation database for all nutrients for various stratifying parameters
cip80.do	AppendixF_DoFiles\Task3_DoFiles\BS	Computes observed values of percentiles of interest and bootstrap confidence intervals by season for the p80 database for all nutrients for various stratifying parameters
cip85.do	AppendixF_DoFiles\Task3_DoFiles\BS	Computes observed values of percentiles of interest and bootstrap confidence intervals by season for the p85 database for all nutrients for various stratifying parameters
cip90.do	AppendixF_DoFiles\Task3_DoFiles\BS	Computes observed values of percentiles of interest and bootstrap confidence intervals by season for the p90 database for all nutrients for various stratifying parameters
kskweco3.do	AppendixF_DoFiles\Task3_DoFiles\KSKWFORPAPER	Compares reference and non-reference distributions by season for various stratifying parameters using the all observation database

[1] In this table, the phrase "for all nutrients' should be interpreted to mean all nutrient, turbidity and benthic chlorophyll-a groupings.

Name of Do File	Location of File on CD	Purpose of Do File
kskweco3_med.do	AppendixF_DoFiles\Task3_DoFiles\KSKWFORPAPER	Compares reference and non-reference distributions by season for various stratifying parameters using the median database
kskweco3_no7&8.do	AppendixF_DoFiles\Task3_DoFiles\KSKWFORPAPER	Compares reference and non-reference distributions by season for various stratifying parameters using the all observation database excluding stream orders 7&8
ptilefun.do	AppendixF_DoFiles\Task3_DoFiles\PTILE	Performs the percentile mapping exercise between reference and non-reference populations by season at the coarse scale for various stratifying parameters using the all observation database
ptilefun-fine.do	AppendixF_DoFiles\Task3_DoFiles\PTILE	Performs the percentile mapping exercise between reference and non-reference populations by season at the fine scale for various stratifying and sub-stratifying parameters using the all observation database
strahcomp.do	AppendixF_DoFiles\Task3_DoFiles\STRAHCOMP	Compares nutrient concentrations across select stream order groupings by season for all nutrients at the coarse scale for various stratifying parameters
t3sumstatsallobs.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 all observation database for all nutrients at the coarse scale for various stratifying parameters
t3sumstatsallobs-fine.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 all observation database for all nutrients at the fine scale for various stratifying and sub-stratifying parameters
t3sumstatsallobsno7&8.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 all observation database excluding data from stream orders 7&8 for all nutrients at the coarse scale for various stratifying parameters
t3sumstatsallobsno7&8-fine.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 all observation database excluding data from stream orders 7&8 for all nutrients at the fine scale for various stratifying and sub-stratifying parameters
t3sumstatsmedian.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 median database for all nutrients at the coarse scale for various stratifying parameters
t3sumstatsmedian-fine.do	AppendixF_DoFiles\Task3_DoFiles\T3SUMSTATS	Presents summary statistics by season for the Task 3 median database for all nutrients at the fine scale for various stratifying and sub-stratifying parameters

Table F-2: Developer's Guide to *Stata* Ado Files

Name of Ado File	Location of File on CD	Purpose of Ado File	To be copied to:
poi.ado	AppendixF_DoFiles\Ado_Files	To compute percentiles of interest	C:\ado\personal[2]
poi2.ado	AppendixF_DoFiles\Ado_Files	To compute percentiles of interest – version 2	C:\ado\personal
distplot7.ado	AppendixF_DoFiles\Ado_Files	To plot distributions	C:\ado\personal
kwallis2.ado	AppendixF_DoFiles\Ado_Files	Modified Kruskal Wallis Test	C:\ado\personal
tostring.ado	AppendixF_DoFiles\Ado_Files	String function	C:\ado\personal

Table F-3: Developer's Guide to *Microsoft Excel* Tools

Name of Excel File	Location of File on CD	Purpose of Excel File
MT-ComplianceTool.xls	ComplianceTools	To confirm compliance with standards based on the t-test and the exact binomial test
MT-NonComplianceTool.xls	ComplianceTools	To confirm non-compliance with standards based on the t-test and the exact binomial test

[2] If the pathway c:\ado\personal does not already exist on the user's hard drive, the user should create the same.

Appendix G. User Guide for *Stata* Program Modification

Table of Contents

1.0. Introduction.....G-2
2.0. Creating a *Stata* Dataset.....G-3
3.0. Modifying Do Files to Work Off New Datasets.....G-5
4.0. Copying Ado Files to the User's Hard Drive.....G-9
5.0. Examples of Do File Modification G-10
5.1 How to modify *ptile.do*……...G-11
5.2 How to modify *t5sumstatsallobs.do*…….G-17
5.3 How to modify *t5sumstatsmedian-fine.do*……G-22
5.4 How to modify *t5mtmeddb.do*…….G-29

1.0 Introduction

The statistical analyses generated in this and previous phases of the project are based on output produced by computer programs written in the *Stata (Version 7)* programming language.[1] *Stata* applications are widely used in academia and industry for statistical and numerical analyses.

The programs developed in this project are stored in text files, which are referred to as "do files" in *Stata* terminology. When run or executed within the Stata operating environment, these do files produce output files which are referred to as "log files" in *Stata* terminology.

A significant advantage of the *Stata* environment, compared to applications that employ menu-based interfaces, is that *Stata* do files may be re-run on updated databases in the future. Minor modifications would, however, have to be made to the do files to enable them to read data from new databases and to write log files to new locations. The programs developed in this project, especially in the current phase, have been written with the aim of minimizing the modifications required to re-run the do files on future databases. Users will require only a minimal knowledge of *Stata* syntax to make the necessary modifications. This user guide describes the modifications needed for users of Windows-based computers to re-run the do files on new databases in future.

Users should note that *Stata* syntax is case insensitive for filenames and location pathways. When specifying file locations, the forward slash and the backward slash are considered equivalent; either direction of slash will work.

[1] See *www.stata.com* for details on the Stata statistical programming environment.

2.0 Creating a *Stata* Dataset

Users may find it easiest to append new data to the current dataset and edit existing data in the current dataset using the well-known *Microsoft Excel* or *Microsoft Access* applications. Directly making such changes to the current *Stata* dataset (provided in the accompanying CD) will require knowledge of specialized *Stata* syntax; covering the range of such syntax is out of the scope of this user guide. We therefore recommend that users rely on the *Microsoft Access* dataset (also included in the accompanying CD) to make revisions and updates.

Following the addition of new data and/or editing of old data, users will need to translate the updated *Microsoft Access* dataset to a *Stata* dataset; this section describes the procedure and commands needed to perform such a conversion. Note that when adding new data users will need to perform adjustments for non-detects, delete defective and unwanted data, and perform nutrient grouping, seasonal groupings and lithologic groupings per the methods reported in Section 3 of the main report and Appendices A, B and C. Please note that the updated *Microsoft Access* dataset must carry the same variable names (i.e., data field names) as the original; if any variable names are changed, the do files may not run.

Users should also remember to create a median database from the updated dataset for any analysis which runs on station median data. An updated median dataset may be created from the updated all observation dataset by using the do file provided in Appendix F. Detailed instructions on how to update the do file for creating the median dataset are provided in Example 4 below.

Step 1: Convert the Updated Microsoft Access Dataset Into a Tab-Separated Text File

- Load the updated *Microsoft Access* dataset in table view.
- Click on the “File” menu and scroll down to the “Export” option.
- Specify an export location and a file name in the window that pops up.
- In the “Save as Type” box, remember to choose the “Text Files (*.txt, *.csv, *.tab, *.asc)” option.
- Click on the “Export All” button.
- The following screen will offer an option of “delimited” or “fixed width”. Choose “delimited” and click on “Next”.
- At the following screen, choose the “tab” delimiter.
- Also, at the same screen as above, check the “Include Field Names on First Row” check-box, and choose “{none}” for the text qualifier.
- Click on “Next”.
- Click on “Finish”.

This procedure will cause a tab-delimited text file to be exported to the specified location. The contents of this text file will be identical to the user’s updated *Microsoft Access* dataset.

Step 2: Input the Tab-Separated Text File Into Stata

First start up the *Stata* application by double clicking on the *Stata* icon on the desktop. Alternatively, start up *Stata* from the Start/Programs path in Windows.

In the *Stata* command window enter the commands below (without the quotation marks, which are provided here to distinguish the command from the surrounding explanatory text):

(i) Enter "set mem 500m" (This allocates sufficient memory for a large dataset. For a very large dataset, enter "set mem 1g" if the computer's memory is sufficient.)

(ii) Enter "insheet using pathname\filename.txt, clear", where pathname refers to the location of the text file on the user's computer and the filename refers to the name of the text file. For example, if the tab separated text file created in Step 1 is called newdatabase.txt and it is located at c:/helena then the command should read: "insheet using c:/helena/newdatabase.txt, clear"

These commands will load the text file into *Stata*.

Step 3: Saving the Stata Dataset

After the text file has been loaded into *Stata* in Step 2, enter the following command in the command window:

"save c:/pathname/filename.dta, replace", where pathname refers to the preferred location of the new *Stata* dataset and filename refers to the preferred file name for the new *Stata* data set. For instance, if the user wishes to call the dataset "updated_data" and wishes to store the file at c:/helena, the user will need to enter the command:

"save c:/helena/updated_data.dta, replace"

Note that the suffix for data files in *Stata* is ".dta".

After saving the file, the user may look over the file by using the data browser icon on the main *Stata* menu. To prevent unintentional edits to the dataset, users should avoid using the data editor icon.

3.0 Modifying Do Files To Work Off New Datasets

The programs developed for this project have been written to minimize the modifications necessary to work with new datasets. The general steps in making the necessary modifications are to:

(i) Direct the program to retrieve the new dataset from a location of the user's choosing,
(ii) Direct the program to save any interim datasets in an appropriate location of the user's choosing, and
(iii) Direct the output log files to an appropriate location of the user's choosing.

Users wishing to minimize text edits to the do files should consider setting up pathnames for the log files and database names identical to those used with the original do files.

This section provides a more detailed description of the general steps listed above. Section 1.4 will illustrate and clarify these instructions by using four actual examples based on do files developed in this analysis.

Step 1: Load the Required Do File Into the Do File Editor

Appendix F provides a table indicating the name and purpose of each do file in the current and previous phase of the analysis. The required do file may be accessed from the accompanying CD and downloaded to a location on the user's hard drive. The user must then open the do file using *Stata*'s do file editor. The method of accessing the do file editor may vary between *Stata* versions; the *Stata* help tab or *Stata* user guides will clarify how to open the do file editor. Once within the do file editor, the user has to browse through the hard drive for the saved do file and open it.

Step 2: Modify the Do File to Reflect Location and Name of New Dataset

Every do file will call on a primary dataset that supports the analysis being performed. This could be the all observations dataset or the median dataset described in Section 3 of the main report. If the user intends to run the analysis on an updated dataset (either all observations or median), it is imperative that the user modify the do file to reflect the path location and file name of the new dataset. For the programs created in Task 5, the all observation dataset is named *t5prmtdb.dta* and the median dataset is named *t5medmtdb.dta.* Users should look within the relevant do file for all instances of these filenames; the users should then replace the filename of the old dataset with the filename of the updated dataset along with the updated path location of the new dataset.

For instance, in the do file "ptile.do" – which computes the observed values of the percentiles of interest and also performs the distinct zone assessment described in Sections 4 and 5 of the main report – the analysis calls on the median database with the following statement:

```
quietly use "C:\Montana\Task5\DB\t5medmtdb.dta", clear
```

Users should update the above statement with the path location and filename of the updated median dataset. For instance, if the new median dataset is called newmediandataset.dta and it is located at c:/Helena/Data/ then the above statement should be modified as follows:

quietly use "C:\Helena\Data\newmediandataset.dta", clear

This process should be repeated for every instance where the old dataset is called in the relevant do file. Users are advised to use the "search" tab within the do file editor to search for instances where the old dataset is called by the do file. In some do files, the dataset is loaded repeatedly and the necessary revisions may need to be made more than once within the do file.

Step 3: Modify the Do File to Reflect Pathway for Creation of Interim Datasets

Most of the programs created as part of this analysis rely on a programming construct that may be described as the creation of an interim dataset. These interim datasets are used to isolate the specific data required for the analysis being conducted instead of working with the whole dataset. For instance, an interim database may contain data only the nutrient, season, and level III ecoregion being analyzed; this construct helps perform the types of repetitive looping analyses which focus on each parameter in turn.

Users do not need to do anything to create these interim databases; the do file automatically does that.

Users do, however, need to make sure that an appropriate path location is provided for the interim datasets. The do file will already contain a path location for the creation of the interim dataset. Users will need to either (i) create an identical path location on their hard-drive or (ii) modify the do file to reflect the user's preferred path location for the interim dataset.

The first step in this process is the identification of the interim datasets created in a do file. A keyword search for the word "save" will identify instances where interim datasets are created in a do file; all interim datasets are created with statements beginning with the word "save".

For instance, in the do file "ptile.do" there are two instances where interim datasets are created:

quietly save "C:\Montana\Task5\dummydb\dbeco3.dta", replace

and

quietly save "C:\Montana\Task5\dummydb\dbeco4.dta", replace

For this do file, users will have to either:

(i) create the path **C:\Montana\Task5\dummydb** on their hard-drives. Once this pathway is created, the process is complete; no edits need be made to the do file at all. (This is a much simpler option than the second option below.)

OR

(ii) edit the do-file to make sure that the interim datasets are written to the preferred location. For instance, if the preferred location is “C:\Helena\InterimData”, users would have to edit the two statements above as follows:

quietly save "C:\Helena\InterimData\dbeco3.dta", replace

and

quietly save "C:\Helena\InterimData\dbeco4.dta", replace

respectively.

(Note: It is not recommended that the user rename the filenames of the interim datasets that were created – in this case it is not recommended that the user change the filenames *dbeco3.dta* and *dbeco4.dta*).

In addition to editing the “save” statements above, the user will need to locate all instances of “use” statements where the interim datasets are called by the do file and change the path location to reflect the updated location of the interim database. Users will need to perform a keyword search using the name of the interim dataset that was created above and then change its path location to reflect the one specified by the user. For instance, such a keyword search on *dbeco3.dta* and *dbeco4.dta* would yield the statements:

quietly use "C:\Montana\Task5\dummydb\dbeco3.dta",clear

and

quietly use "C:\Montana\Task5\dummydb\dbeco4.dta", clear

Every instance of these statements would have to be modified as follows:

quietly use "C:\Helena\InterimData\dbeco3.dta",clear

and

quietly use "C:\Helena\InterimData\dbeco4.dta",clear

respectively, to reflect the user specified location of the interim database.

Users should note that the interim datasets will not contain any meaningful data at the end of the run process; they serve a purely internal function and should not be consulted.

Step 4: Modify the Do File to Reflect the Pathway for Creation of Log Files

The *Stata* do files developed in this analysis write their output as text files known as log files. The do files currently direct the output log files to be written to location pathways specified by

the original programmers. It is likely that future users would wish to specify different pathways. The do files offer limited flexibility to direct the output log files because, like most applications, *Stata* is unable to order Windows to create a structure of folders on the user's hard drive; the required structure of folders must be created independently by the user prior to running the do file.

In the current analysis, log files were written to: C:/Montana/Task5/Runs. Within the Runs folder are a set of sub-folders whose structure and contents are further explained in Appendix D.

The do files may be easily edited to redirect the location of the Runs folder, but the sub-structure within the Runs folder may not be altered.

To illustrate by means of an example, the user may decide to locate the Runs folder at C:/Helena/Output. To do so, the user must follow the following steps:

(i) Create a folder named "Output" located at C:/Helena

(ii) Copy the Runs folder from Appendix D of this report to C:/Helena/Output. (This is the easiest way to create the required nested structure of sub-folders. Unfortunately, the copied Runs folder will also include the old log files within it.) Alternatively, users may create an identical nested structure of sub-folders as in the Runs folder but directly from Windows. This step must be completed before running the do file.

(iii) Optional: Delete all the old log files from the copied Runs folder to avoid version confusion. (Note that only the log files must be deleted and not any of the subfolders.) If all the do files created in this project are being rerun then all the old log files in the Runs folder will be rewritten -- this step may not be required in those circumstances.

(iv) Edit the do file of interest to reflect the C:/Helena/Output pathway. To take the example of the *ptile.do* do file, the log files are currently created by the *Stata* command:

```
log using c:/montana/task5/runs/ptile/`1'/`1'.log, replace
```

This statement must be edited to reflect the preferred output location as follows:

```
log using c:/Helena/Output/runs/ptile/`1'/`1'.log, replace
```

As explained above, no edits can be made to the pathway in the do file beyond the "runs" folder.

The same process may be followed for other do files. The log files are always created by the "log using" command and are usually located close to the beginning of the do file.

Step 5: Save and Run the Modified Do File

After making the modifications described in the steps above, the do file should be saved using the appropriate icon or tab in the *Stata* do file editor. It is recommended that the modified do file be saved under a new name. The do file may be saved to any location of the user's preference. The user may then run the do file using the appropriate icon on the *Stata* do file editor.

4.0 Copying Ado Files to the User's Hard Drive

The programs developed for this project make use of certain user-defined functions which are referred to as "ado" files in *Stata* terminology. These files must be loaded onto the user's computer at an appropriate location or the do files which reference these functions will not work. This is a one-time procedure. Once the ado files have been loaded to the user's computer, any number of do files can be modified or run without any further requirements with regard to the ado files.

The ado files may be found in Appendix F. The ado files included in Appendix F should be loaded onto the user's computer at: C:\ado\personal. These folders may already have been created automatically at the time of installing *Stata*. If not, the user should create the pathway C:\ado\personal. All the user then needs to do is cut and paste the ado files out of Appendix F into C:\ado\personal. This step should be performed before any do files are run.

5.0 Examples of Do File Modification

This section presents four complete examples of all the modifications needed to make a do file ready to operate on a new dataset. These examples are based on four existing do files developed as part of this project. The first three examples illustrate the modifications that must be effected to the existing do files under a set of common assumptions about the names and file locations of the new datasets. These assumptions are:

(i) The new dataset is called NewDataset.dta and it is located at C:/Helena/NewData/NewDataset.dta
(ii) The interim datasets will be written to: C:/Helena/DummyData
(iii) The log files will be written to: C:/Helena/Output

The fourth example illustrates how to modify the do file that creates a median database out of an all observation database.

Users can use these examples as a template to make modifications by substituting the names of their own dataset names and locations at the appropriate points as illustrated. Users should remember that do files intended for the all observation dataset must be run only on the all observation dataset; similarly, do files intended for the median dataset must be run on the median dataset only.

In all the examples, lines highlighted in red are no longer relevant for the new dataset and should be removed. The lines highlighted in green reflect the names and locations of the new data and should replace the lines in red.

Example 1: How to Modify *ptile.do*[2]

Name of User's New Dataset: NewDataset.dta
Location of User's New Dataset: C:/Helena/NewData/NewDataset.dta
User's Preferred Location for Interim Datasets: C:/Helena/DummyData
User's Preferred Output Location: C:/Helena/Output

```
version 7

clear
capture log close
set matsize 150
set mem 250m
set trace off
set more off

capture program drop mtpt

program define mtpt

version 7

/* OPEN LOG FILE */

log using c:/montana/task5/runs/ptile/`1'/`1'.log, replace

log using c:/helena/output/runs/ptile/`1'/`1'.log, replace


di" "
di
"XXXX
```

```
quietly use "C:\Montana\Task5\DB\t5medmtdb.dta", clear

quietly use "C:\Helena\NewData\newdataset.dta", clear


quietly drop if nutgroup ~="`1'"

quietly keep if ref=="Y"

/* UNITS STANDARDIZATION */

quietly replace param3="MG/L" if param3=="MG/L P"
quietly replace param3="mg Chla /m^2" if param3=="CORRECTD"& nut=="BNCHLOR-A"


/* ADDS LABEL*/

quietly label variable mednum "Median Nutrient Concentration"


/*CREATES LABEL VARIABLES */

local unit=param3

local label0="Level 3 Ecoregion"
local label00="eco3_name"

/* GENERATE INDICES FOR SEASON */


local seasonname0="ALL SEASONS"
local seasonname1="WINTER"
local seasonname2="RUNOFF"
local seasonname3="GROWING"

/* START MAIN PROGRAM */

quietly sort eco3

quietly egen neco3=group(eco3)

quietly sort neco3

local eco3loopmax=neco3[_N]

quietly save "C:\Montana\Task5\dummydb\dbeco3.dta", replace

quietly save "C:\Helena\DummyData\dbeco3.dta", replace

local j=1

while `j'<=`eco3loopmax' {

quietly use "C:\Montana\Task5\dummydb\dbeco3.dta",clear

quietly use "C:\Helena\DummyData\dbeco3.dta",clear
```

```
quietly keep if neco3==`j'

local nameofeco3=eco3_name


di" "
di "##################NEW ECO3 LOOP: `nameofeco3' ####################"
di" "


local seasonindex=0

while `seasonindex'<4 {

quietly use "C:\Montana\Task5\dummydb\dbeco3.dta",clear

quietly use "C:\Helena\DummyData\dbeco3.dta",clear


quietly keep if neco3==`j'

quietly drop if season~="`seasonname`seasonindex'""

di" "
di"%%%%%%%%%%%%%%%% NEW SEASON LOOP: `seasonname`seasonindex''
%%%%%%%%%%%%%%%%%%"
di" "

if _N<4 {
di "NO DATA OR INSUFFICIENT DATA AVAILABLE FOR `seasonname`seasonindex'""
di"FOR NUTRIENT = `1' & LEVEL 3 ECOREGION= `nameofeco3'"}
else {

di " "
di "*************************************************************************"
di "*************************************************************************"
di " "
di "OBSERVED VALUES AND BOOTSTRAP 90% CI FOR PERCENTILES OF INTEREST "
di"FOR NUTRIENT = `1', LEVEL 3 ECOREGION= `nameofeco3' & SEASON =
`seasonname`seasonindex'""
di " "

bstrap poi2, reps(1000) level (90)


di " "
di "*************************************************************************"
di "*************************************************************************"
di " "

di"^^^
```

```
di " "
di"^^^
```

```
local group2ss=r(N)

if `group1ss'>3&`group2ss'>3 {

di "_____"
di " "
di "DISTINCT ZONE ASSESSMENT FOR `nameofeco4' (`codeofeco4') BASED ON THE T-TEST: "
di "FOR NUTRIENT = `1', FOR  SEASON = `seasonname`seasonindex'' "
di " "

ttest mednum, by (group) unequal welch

quietly keep if group==2

di "  "
di"GROUP 1 ABOVE REFERS TO THE PARENT ECOREGION 3 = `nameofeco3'"
di "GROUP 2 REFERS TO THE LEVEL 4 ECOREGION = `nameofeco4' (`codeofeco4')"
di "  "
di "BASED ON THE T-TEST REPORTED ABOVE: "
di "  "

if r(p)<=0.10 {
di " "
di "`nameofeco4' (`codeofeco4') IS A DISTINCT ZONE WITHIN `nameofeco3' FOR NUTRIENT = `1' &
SEASON = `seasonname`seasonindex''"
di "  "
di"(Two-tailed T-test with unequal variances, Welch Correction, 90% confidence level)"
di " "
di "OBSERVED VALUES AND BOOTSTRAP 90% CI FOR PERCENTILES OF INTEREST "
di"FOR NUTRIENT = `1', SEASON = `seasonname`seasonindex'' & LEVEL 4
ECOREGION=`nameofeco4' (`codeofeco4')"
di " "

bstrap poi2, reps (1000) level (90)
di " "
di "_____"

di " "

}

else {di " "
di "`nameofeco4' (`codeofeco4') IS NOT A DISTINCT ZONE WITHIN `nameofeco3' FOR NUTRIENT =
`1' & SEASON = `seasonname`seasonindex''"
di "  "
di"(Two-tailed T-test with unequal variances, Welch Correction, 90% confidence level)"
di "  "
di "_____"
di " "}

}

else {di " "
di "_____"
di "  "
```

```
di"SAMPLE SIZE IN `nameofeco4' (`codeofeco4') IS NOT ADEQUATE FOR DISTINCT ZONE
ASSESSMENT FOR NUTRIENT = `1' & SEASON = `seasonname`seasonindex""
di "  "
di "_____________________________________________________________________________________"
di " "}

local k=`k'+1

}

}

local seasonindex=`seasonindex'+1

}

local j=`j'+1

}

di "  "
di "END OF LOG FILE "
di " "
di "_____________________________________________________________________________________"

capture log close

end

mtpt AMMONIA
mtpt BNCHLOR-A
mtpt NO3+NO2
mtpt SRP
mtpt TDP
mtpt TKN
mtpt TOTALN
mtpt TOTALP
mtpt TURB-JTU
mtpt TURB-NTU
```

Example 2: How to Modify t5sumstatsallobs.*do*[3]

Name of User's New Dataset: NewDataset.dta
Location of User's New Dataset: C:/Helena/NewData/NewDataset.dta
User's Preferred Location for Interim Datasets: C:/Helena/DummyData
User's Preferred Output Location: C:/Helena/Output

```
/*Task 5 version of Sum Stats -- Coarse -- All Obs*/

clear
capture log close
set matsize 150
set mem 50m
set trace off
set more off

capture program drop mtpt

program define mtpt

log using c:/montana/task5/runs/sumstats/allobs/coarse/`1'/`2'/`1'.log, replace

log using c:/helena/output/runs/sumstats/allobs/coarse/`1'/`2'/`1'.log, replace


use c:/montana/task5/db/t5prmtdb.dta, clear

use c:/helena/newdata/newdataset.dta, clear



/* ADDS LABEL*/

quietly label variable numericvalue "Concentration"

quietly label variable ref "Reference Status"

/* REDEFINE REF FIELD */

quietly replace ref="Reference" if ref=="Y"
quietly replace ref="Non-Reference" if ref=="N"


/* UNITS STANDARDIZATION */

quietly replace param3="MG/L" if param3=="MG/L P"
quietly replace param3="mg Chla /m^2" if param3=="CORRECTD"& nut=="BNCHLOR-A"
```

[3] (This do file is used to generate summary statistics at the coarse scale using the all observations database. See Appendix F.)

```
/* DROPS REDUNDANT NUTRIENTS*/

quietly drop if nutgroup ~="`1'"

/*CREATES LABEL VARIABLES */

local unit=param3

if "`2'"=="eco3" {
local label0="Level 3 Ecoregion"
local label00="eco3_name"
}
else if "`2'"=="eco4" {
local label0="Level 4 Ecoregion"
local label00="eco4_name"
}
else if "`2'"=="sfname" {
local label0="Geological Formation"
local label00="geo_name"
}

else if "`2'"=="geco" {
local label0="Grouped Level 3 Ecoregions"
local label00="geco"
}


di" "
di
"XXXX
```

```
/* GENERATE INDICES FOR SEASON */

local index1=1
local seasonname0="ALL SEASONS"
local seasonname1="WINTER"
local seasonname2="RUNOFF"
local seasonname3="GROWING"


while `index1'<`n0'{


di "##########################################################################"
di "###########     NEW `label0' CATEGORY      ###########"
di "##########################################################################"


local seasonindex=0

while `seasonindex'<4 {

use c:/montana/phaseiii/dbasegroup.dta, clear

use c:/helena/dummydata/dbasegroup.dta, clear

quietly drop if paramcat~=`index1'

if `seasonindex'==0 {quietly drop if
season~="WINTER"&season~="RUNOFF"&season~="GROWING"}

else {quietly drop if season~="`seasonname`seasonindex'"}

if "`2'"=="eco3" {
local label1=eco3_name
}
else if "`2'"=="eco4" {
local label1=eco4_name
}

else if "`2'"=="sfname" {
local label1=geo_name
}

else if "`2'"=="geco" {
local label1=geco
}


di" "
di "*************************************************************************"
di "************************   NEW SEASON    ************************"
di "*************************************************************************"
di " "
di "ANALYSIS COMMENCES for NUTRIENT = `1' & `label0' = `label1'"
```

```
di " "
di"for SEASON = `seasonname`seasonindex""
di " "

quietly tab ref
local n00=r(r)
quietly tab ref if ref=="Reference"
local n000=r(N)
quietly tab ref if ref=="Non-Reference"
local n0000=r(N)


if (`n00'==2&`n000'>3&`n0000'>3) {


di " "
di "_____________________________________________________________________________________"
di " "
di "SUMMARY STATISTICS for Nutrient = `1' and `label0' = `label1'"
di " "
di"for SEASON = `seasonname`seasonindex'' (in units of `unit')"
di " "
tabstat num, by (ref) stats (count, min, max, mean, sd, skewness) format (%9.3f)
tabstat num, by (ref) stats (count, p10, p25, p50, p75, p90) format (%9.3f)
di" "

}

else {di"Insufficient Reference and/or Non-Reference Data Available"
di " "
di"for Nutrient=`1', `label0' = `label1'"
di " "
di"for  SEASON = `seasonname`seasonindex''"}
di " "
di "_____________________________________________________________________________________"
di " "

local seasonindex=`seasonindex'+1
}

local index1=`index1'+1

}

di" "
di
"XXXX
```

mtpt AMMONIA geco
mtpt BNCHLOR-A geco
mtpt NO3+NO2 geco
mtpt TKN geco
mtpt TOTALN geco
mtpt TDP geco
mtpt SRP geco
mtpt TOTALP geco
mtpt TURB-NTU geco
mtpt TURB-JTU geco

mtpt AMMONIA eco3
mtpt BNCHLOR-A eco3
mtpt NO3+NO2 eco3
mtpt TKN eco3
mtpt TOTALN eco3
mtpt TDP eco3
mtpt SRP eco3
mtpt TOTALP eco3
mtpt TURB-NTU eco3
mtpt TURB-JTU eco3

mtpt AMMONIA sfname
mtpt BNCHLOR-A sfname
mtpt NO3+NO2 sfname
mtpt TKN sfname
mtpt TOTALN sfname
mtpt TDP sfname
mtpt SRP sfname
mtpt TOTALP sfname
mtpt TURB-NTU sfname
mtpt TURB-JTU sfname

Example 3: How to Modify t5sumstatsmedian-fine.*do*[4]

Name of User's New Dataset: NewDataset.dta
Location of User's New Dataset: C:/Helena/NewData/NewDataset.dta
User's Preferred Location for Interim Datasets: C:/Helena/DummyData
User's Preferred Output Location: C:/Helena/Output

```
/*Task 5 Version of Sum Stats -- Fine -- Median */

 /*FINE-SCALE FUNCTION*/

clear
capture log close
set matsize 150
set mem 50m
set trace off
set more off

capture program drop mtpt2

program define mtpt2

log using c:/montana/task5/runs/sumstats/median/fine/`1'/`2'/`3'/`1'.log, replace

log using c:/helena/output/runs/sumstats/median/fine/`1'/`2'/`3'/`1'.log, replace


use c:/montana/task5/db/t5medmtdb.dta, clear

use c:/helena/newdata/newdataset.dta, clear


/* RENAME MEDNUM AND MEDLOGNUM */

rename mednum numericvalue

rename medlognum lognum

/* ADDS LABEL*/

quietly label variable numericvalue "Concentration"

quietly label variable ref "Reference Status"

/* REDEFINE REF FIELD */

quietly replace ref="Reference" if ref=="Y"
quietly replace ref="Non-Reference" if ref=="N"

/* UNITS STANDARDIZATION */
```

[4] (This do file is used to generate summary statistics at the fine scale using the median dataset. See Appendix F.)

```
quietly replace param3="MG/L" if param3=="MG/L P"
quietly replace param3="mg Chla /m^2" if param3=="CORRECTD"& nut=="BNCHLOR-A"

/* DROP REDUNDANT NUTRIENTS*/

quietly drop if nutgroup ~="`1'"

/*CREATES LABEL VARIABLES */

local unit=param3

if "`2'"=="eco3" {
local label0="Level 3 Ecoregion"
local label00="eco3_name"
}
else if "`2'"=="eco4" {
local label0="Level 4 Ecoregion"
local label00="eco4_name"
}
else if "`2'"=="sfname" {
local label0="Geological Formation"
local label00="geo_name"
}

else if "`2'"=="geco" {
local label0="Grouped Level 3 Ecoregions"
local label00="geco"
}


if "`3'"=="eco3" {
local label000="Level 3 Ecoregion"
local label0000="eco3_name"
}
else if "`3'"=="eco4" {
local label000="Level 4 Ecoregion"
local label0000="eco4_name"
}

else if "`3'"=="sfname" {
local label000="Geological Formation"
local label0000="geo_name"
}

di" "
di
"XXXX
```

```
di
"XXXX
```

```
quietly tab `3', gen(subcat)

local iii=1
quietly gen subparamcat=.
while `iii'<`nn' {
local subcatn=substr("subcat",1,6)+string(`iii')
quietly replace subparamcat=`iii' if `subcatn'==1
local iii = `iii'+1
}

capture drop subcat*

quietly save c:/montana/phaseiii/dbasegroup2.dta, replace

quietly save c:/helena/dummydata/dbasegroup2.dta, replace



local index2=1
local seasonname0="ALL SEASONS"
local seasonname1="WINTER"
local seasonname2="RUNOFF"
local seasonname3="GROWING"



while `index2'<`nn' {


di "##########################################################################"
di "###########     NEW `label000' CATEGORY      ###########"
di "##########################################################################"

local seasonindex=0

while `seasonindex'<4 {

use c:/montana/phaseiii/dbasegroup2.dta, clear

use c:/helena/dummydata/dbasegroup2.dta, clear


quietly drop if subparamcat~=`index2'

quietly drop if season~="`seasonname`seasonindex'""

local label2 = `label0000'



di" "
di "**************************************************************************"
di "************************   NEW SEASON    ************************"
di "**************************************************************************"
di " "
di "_____________________________________________________________________________________________"
```

```
di " "
di "ANALYSIS COMMENCES for NUTRIENT = `1', `label0' = `label1'"
di " "
di"and `label000' = `label2'"
di " "
di"for SEASON = `seasonname`seasonindex''"
di " "
di "____________________________________________________________________________________________"
quietly tab ref
local n00=r(r)
quietly tab ref if ref=="Reference"
local n000=r(N)
quietly tab ref if ref=="Non-Reference"
local n0000=r(N)

if (`n00'==2 & `n000'>3 & `n0000'>3) {


di " "
di "____________________________________________________________________________________________"
di " "
di "SUMMARY STATISTICS for Nutrient = `1', `label0' = `label1'"
di " "
di"and `label000' = `label2'"
di " "
di"for SEASON = `seasonname`seasonindex'' (in units of `unit')"
di " "
tabstat num, by (ref) stats (count, min, max, mean, sd, skewness) format (%9.3f)
tabstat num, by (ref) stats (count, p10, p25, p50, p75, p90) format (%9.3f)
di" "
di "____________________________________________________________________________________________"
di " "

}

else {di"Insufficient Reference and/or Non-Reference Data Available"
di" "
di"for Nutrient = `1', `label0 ' = `label1'"
di" "
di"and `label000' = `label2'"
di " "
di"for  SEASON = `seasonname`seasonindex'' }
di " "
di "____________________________________________________________________________________________"
di " "

local seasonindex=`seasonindex'+1

}

local index2=`index2'+1

}

di" "
```

```
di
"XXXX
```

mtpt2 SRP eco3 sfname
mtpt2 TOTALP eco3 sfname
mtpt2 TURB-NTU eco3 sfname
mtpt2 TURB-JTU eco3 sfname

mtpt2 AMMONIA sfname eco3
mtpt2 BNCHLOR-A sfname eco3
mtpt2 NO3+NO2 sfname eco3
mtpt2 TKN sfname eco3
mtpt2 TOTALN sfname eco3
mtpt2 TDP sfname eco3
mtpt2 SRP sfname eco3
mtpt2 TOTALP sfname eco3
mtpt2 TURB-NTU sfname eco3
mtpt2 TURB-JTU sfname eco3

mtpt2 AMMONIA sfname eco4
mtpt2 BNCHLOR-A sfname eco4
mtpt2 NO3+NO2 sfname eco4
mtpt2 TKN sfname eco4
mtpt2 TOTALN sfname eco4
mtpt2 TDP sfname eco4
mtpt2 SRP sfname eco4
mtpt2 TOTALP sfname eco4
mtpt2 TURB-NTU sfname eco4
mtpt2 TURB-JTU sfname eco4

Example 4: How to Modify t5mtmeddb.*do*[5]

In this example, modifications must be effected under the assumption that:

- The new all observation dataset is named NewAllObsData.dta and is located at: C:/Helena/NewData/NewAllObsData.dta
- Interim datasets must be written to C:/Helena/DummyData
- The new median dataset must be name NewMedianData.dta and is to be stored at: C:/Helena/NewData/NewMedianData.dta

```
/*UPDATED FOR TASK 5 */

clear
set mem 500m

/* MEDIAN DATABASE CREATION PROGRAM */

capture program drop mediandb

program define mediandb

/* TO CREATE SEASONAL MEDIAN DB */

/* LOAD PROCESSED DATABASE */

use c:/montana/task5/db/t5prmtdb.dta, clear

use c:/helena/newdata/newallobsdata.dta, clear


drop if season==""

/* Seasons are missing when eco3 is missing. Need to drop missing obs for median program to
work. No harm done because eco3 is critical to analysis*/

sort nutgroup agency station season

by nutgroup agency station season: egen mednum =median(numericvalue)

by nutgroup agency station season: egen medlognum =median(lognumericvalue)

by nutgroup agency station season: gen dupcount= _n

drop if dupcount~=1

egen unique = group (nutgroup agency station season)

if unique[_N]~=_N {
```

[5] (This do file is used to create a median dataset from an all observation dataset. See Appendix F.)

```
di "ERROR" }
else di "OKAY"

drop paramcode paramcodetext orig dup param1 param2 numericvalue lognumericvalue
remarkcode orig dup year month day time dateandtime

save c:/montana/task5/db/mediandb/t5medmtdbseason.dta, replace

save c:/helena/dummydata/t5medmtdbseason.dta, replace


/* TO CREATE ALL SEASON MEDIAN DB */

/* LOAD PROCESSED DATABASE */

clear

use c:/montana/task5/db/t5prmtdb.dta, clear

use c:/helena/newdata/newallobsdata.dta, clear

drop if season==""

/* Seasons are missing when eco3 is missing. Need to drop missing obs for median program to
work. No harm done because eco3 is critical to analysis*/

sort nutgroup agency station

by nutgroup agency station : egen mednum =median(numericvalue)

by nutgroup agency station : egen medlognum =median(lognumericvalue)

by nutgroup agency station : gen dupcount= _n

drop if dupcount~=1

egen unique = group (nutgroup agency station)

if unique[_N]~=_N {
di "ERROR" }
else di "OKAY"

replace season="ALL SEASONS"

drop paramcode paramcodetext orig dup param1 param2 numericvalue lognumericvalue
remarkcode orig dup year month day time dateandtime

save c:/montana/task5/db/mediandb/t5medmtdballseason.dta, replace

save c:/helena/dummydata/t5medmtdballseason.dta, replace

/* TO APPEND THE TWO MEDIAN DATABASES */

append using c:/montana/task5/db/mediandb/t5medmtdbseason.dta,

append using c:/helena/dummydata/t5medmtdbseason.dta,
```

```
/* SAVE MEDIAN DB */

drop dupcount unique

save c:/montana/task5/db/t5medmtdb.dta, replace

save c:/helena/newdata/newmediandata.dta, replace

end

/* CALL MEDIAN DB PROGRAM */

mediandb
```

Appendix H. Statistical Considerations for Applying Montana's Numeric Nutrient Standards: Recommendations for 303(d) Listing and TMDLs

Prepared by Michael Suplee, Ph.D., Montana Department of Environmental Quality. June 23, 2008

Summary

The following summary should provide sufficient detail to allow users to determine compliance with the numeric nutrient standards. Details concerning the whys and wherefores of each of these recommendations are provided further on in this document.

1. Determination of compliance with these numeric nutrient standards is to be carried out on a waterbody-by-waterbody basis. A waterbody, in this case, is defined as: *a wadeable stream segment listed in the Assessment Data Base (ADB) (MT DEQ, 2006a, and updates), <u>or</u> a sub-segment of an ADB stream segment.* A sampling unit within each ADB segment (or sub-segment) is defined as a water quality sample collected independently of other water quality samples within the same segment/sub-segment. Please consider the following:
    a. The aggregate of samples collected from a segment or sub-segment should provide good overall representation of the ADB segment (or sub-segment). Individual sites within a segment or sub-segment that have known or suspected pollution problems should be sampled equitably *along with* sites where pollution problems are not suspected or are minimal or less pronounced. Do not just target the hotspots.
    b. Given the guidelines in 1a, the assessor will have to judge if further stratification of the stream (i.e., create 2 or more sub-segments) is warranted. If a relatively un-impacted upstream or downstream reach of an ADB segment can be isolated and its condition is substantially different from other part of the segment, sub-segmenting is likely justified. As a rule of thumb, it is better to lump than split to avoid excessive sub-segmentation of streams and the consequential administrative and sampling requirements.
    c. Each sub-segment will have the same data requirements (dataset minimums, tests, etc.) as the parent segment if it hadn't been divided.
    d. Samples must be collected when the criteria apply, which is expected to be during the ecoregion-specific Growing Season (appendix B, this report). Samples collected outside the Growing Season may be useful for other purposes (e.g., isolating load sources), but should not to be used for compliance determination for the Growing Season.

2. Water quality samples from an ADB segment (or sub-segment) may each be considered independent of the others if they meet the following conditions:

a. Sites (or very short reaches functionally equivalent to sites) should be located at least 1 stream mile apart.
b. Sites may be placed < 1 mile apart on an ADB stream segment/sub-segment **if** there is a tributary confluencing with the segment between the two sites.
c. Land use changes and land form changes should be considered and can be used to help define (1) breaks between sampling frames or (2) additional sampling sites within a sampling frame. See page 11-12 of MT DEQ (2005).
d. Samples collected at the same site should be collected at least 30 days apart.

3. The following are recommended to determine compliance with the nutrient criteria:
   a. For Cold-water and Transitional[1] Wadeable Streams: Measurement of the segment's benthic algal Chl *a* at one or more sites[2] (following MT DEQ SOPs, including approved low-chlorophyll visual estimation methods) and a minimum of 12 or 13[3] independent nutrient samples from the segment. The nutrient data are evaluated using the "Exact Binomial Test" *and* the "One-Sample Student's t-test for the Mean". (Each statistical test evaluates somewhat different aspects of the nutrient dataset.) Compliance with the standards is determined via decision rules, which consider together the results from the algae sampling and the two statistical tests (see No. 5 below, Table 1).
   b. For Warm-water Wadeable Streams: Measurement of nighttime dissolved oxygen (DO) levels at one or more sites, either measured directly by instrument or via diatom-inferred DO (van Dam et al., 1994). Sample size minima and statistical tests are the same as for cold water segments. Compliance with the standards is then determined via decision rules, which consider together the results from the DO sampling and the two statistical tests (see No. 5 below, Table 2).

4. To use the statistical tests, make sure to do the following:
   a. For new, un-listed stream segments, use the Excel spreadsheet tool named "MT-NoncomplianceTool.xls".
   b. For already-listed stream segments, use the Excel spreadsheet tool named "MT-ComplianceTool.xls".
   c. In both tools, for either test, set alpha to 0.25 (25%), effect size (gray zone) to 0.15 (15%), and the critical exceedence rate to 0.2 (20%). Note in the Exact Binomial Test you must add or subtract the 0.15 effect size from the critical exceedence in order to enter the correct "p2" value; please carefully read the spreadsheet instructions on this.

[1] Transitional streams are streams in ecoregion 42 or 43 whose headwaters are in a mountainous ecoregion and whose water quality still largely reflects mountain conditions. Specific level IV ecoregions will be identified as containing transitional streams in the nutrient criteria circular under development. See footnote 7 for details on the terms "cold-water" and "warm-water" as used in this document.
[2] Treat each Chl *a* site average as an independent evaluation of use support. Do not average together results from different sites within the same ADB segment or sub-segment. If Chl *a* is measured more than once at the same site, treat each sampling event as unique (NOT as temporal repeat measures).
[3] Twelve independent nutrient samples for new, unlisted streams, and thirteen for streams already listed for nutrients on the 303(d) list. A "nutrient" sample is a type of nutrient (or group of equivalent nutrient measurements; see page A-3, Appendix A) such as TP, NO_{2+3}, or TN. Sample minimums apply to *each* nutrient type.

5. Use the following two tables to determine the stream segment's compliance:

Table 1 (Summary). Decision rules for Determining Compliance with Nutrient Standards, Cold-Water & Transitional* Streams. For each Statistical Test, Regardless of the Form of the Null Hypothesis, PASS Means the Stream Segment Complies with the Nutrient Standard, FAIL Means the Stream Segment Does Not Comply with the Nutrient Standard.

Scenario	Binomial Test	T-test	Benthic Algae Chl*a* (mg /m^2)†	**Resulting Decision**	Notes on Decision
1	PASS	PASS	≤ 150 mg Chl*a* /m^2	In Compliance	All indications show that the stream is in compliance.
2	PASS	FAIL	≤ 150 mg Chl*a* /m^2	In Compliance	Suggests pulsed nutrient loads occur but are not resulting in elevated benthic algae biomass.
3	FAIL	PASS	≤ 150 mg Chl*a* /m2	Borderline	Likely that segment sometimes has high benthic algae biomass, but the timing of the algae sampling may have missed high levels. Borderline case; compliance or non-compliance may be equally justifiable; use other accompanying data (e.g., biometrics) to confirm decision. **OR, further sample algae & nutrients.**
4	FAIL	FAIL	≤ 150 mg Chl*a* /m2	Not in Compliance	Likely that segment has high benthic algae biomass, but the timing of the algae sampling may have missed the high levels. **Further algae sampling may be justified.**
5	PASS	PASS	> 150 mg Chl*a* /m2	Not in Compliance	Algae may be taking up nutrients and leading to low instream nutrient concentrations with concurrent high benthic algae biomass.
6	PASS	FAIL	> 150 mg Chl*a* /m2	Not in Compliance	Non-compliance with the T-test suggests that pulsed nutrient loads are allowing high algae biomass to be maintained via luxury uptake.
7	FAIL	PASS	> 150 mg Chl*a* /m2	Not in Compliance	Suggests sustained nutrient values near the standard but not necessarily pulsed nutrient loading.
8	FAIL	FAIL	> 150 mg Chl*a* /m2	Not in Compliance	All indicators show that the stream is not in compliance.

* Transitional streams are streams in ecoregion 42 or 43 but whose headwaters are in a mountainous ecoregion and whose water quality still largely reflects mountain conditions. Specific level IV ecoregions will be identified as containing transitional streams in the numeric nutrient criteria circular under development.

† Algae biomass collected and summarized as per DEQ SOPs, for a single site (short reach), during any given sampling event. Unlike nutrient samples, *do not* average together algae biomass results from different sites (short reaches) across the stream segment. Consider each site (short reach) on its own merits.

Table 2 (Summary). Decision rules for Determining Compliance with Nutrient Standards, Warm-Water Streams. For each Statistical Test, Regardless of the Form of the Null Hypothesis, **PASS** Means the Stream Segment Complies with the Nutrient Standard, **FAIL** Means the Stream Segment Does Not Comply with the Nutrient Standard.

Scenario	Binomial Test	T-test	Dissolved Oxygen (DO)†	**Resulting Decision**	Notes on Decision
1	PASS	PASS	Meets standards	In Compliance	All indications show that the stream is in compliance.
2	PASS	FAIL	Meets standards	In Compliance	Suggests pulsed nutrient loads occur but are not resulting in large nighttime DO sags.
3	FAIL	PASS	Meets standards	Borderline	Likely that segment periodically does not meet DO standard, but the timing of the sampling may have missed the problem; conclusion that the stream segment *is* or *is-not* in compliance may be equally justified; use other data (e.g., biometrics) to decide. **Further DO & nutrient sampling justified.**
4	FAIL	FAIL	Meets standards	Not in Compliance	Likely that segment periodically does not meet DO standard, but the timing of the sampling may have missed the problem. **Further DO sampling is justified.**
5	PASS	PASS	Does not meet standards	Not in Compliance	Algae may be taking up nutrients and leading to low instream nutrient concentrations with concurrent high productivity and low nightime DO.
6	PASS	FAIL	Does not meet standards	Not in Compliance	Non-compliance with the T-test suggests that pulsed nutrient loads are allowing high productivity to be maintained via luxury uptake, leading to low nighttime DO.
7	FAIL	PASS	Does not meet standards	Not in Compliance	Suggests sustained nutrient values near the standard but not necessarily pulsed nutrient loading.
8	FAIL	FAIL	Does not meet standards	Not in Compliance	All indicators show that the stream is not in compliance.

† Nightime sampling of DO for at least several days should be measured directly via a YSI sonde or similar; or, diatom-inferred DO can be used. Diatom-inferred DO concentrations should be compared to the 7 day mean minimum in DEQ-7 (MTDEQ, 2006b).

1.0 Introduction

The numeric nutrient criteria addressed in this document are not intended to be ideal standards, i.e., "no sample shall exceed" values. As such, appropriate inferential statistical tests, assumptions about stream sampling frames, etc. must be developed so that the criteria can be correctly applied. This appendix outlines these statistical considerations and provides rationales for the various approaches used. It also provides precautionary points where certain assumptions depart from more conservative statistical thinking, and discusses how improper sampling design has the potential to mislead a conclusion made about a stream's condition. The key issues addressed herein are:

- Sampling frames, populations and sampling units for streams, and associated assumptions and precautions
- Consideration of what constitutes sample independence in streams (spatial and temporal), and associated assumptions and precautions
- Determination of appropriate critical exceedence rates for nutrients (nitrogen and phosphorus)
- Statistical testing procedures and accompanying decision rules
- Minimum sample sizes

2.0 Sample Frame, Population, and Sampling Units

All studies involving statistical evaluations of data require that a sample frame, population, and sampling unit be defined. Streams are particularly poor entities for establishing these parameters because streams are an interconnected network rather than discreet entities. Nevertheless, streams are the entities to be sampled so some effort must be made to segregate them into definable units. For the purposes of determining compliance with numeric nutrient criteria, I define the following:

- **Sample Frame**: *A wadeable[4] stream segment listed in the Assessment Data Base (ADB) (MT DEQ, 2006a, and updates)* ***OR*** *a sub-segment of an ADB stream segment.*
- **Population**: *All the water flowing through the ADB stream segment (or sub-segment) during the time period when the numeric nutrient criteria apply.*
- **Sampling Unit**: *A water quality sample collected from the ADB segment (or sub-segment) that is independent of other water quality samples collected from the same segment (or sub-segment) and collected during the time the numeric nutrient criteria apply.*

[4] Wadeable streams are perennial and intermittent streams whose Strahler order (Strahler, 1964) is one through (at most) six (1:100,000 map scale) in which large portions of the channel are wadeable during baseflow conditions.

Assumptions: Each sampling frame (ADB segment or sub-segment) will be made up of a series of sampling sites, or often a series of very short study reaches that are essentially sites (Figure 1). The minimum number of sites on an ADB segment is established in MT DEQ SOPs (MT DEQ, 2005). Figure 1 illustrates the variety of ADB segments that may be found; lengths can vary tremendously. For purposes of determining compliance with numeric nutrient standards using statistical methods, it is assumed that (1) pollution sources are evenly dispersed along the reach, (2) sampling sites are randomly located along the reach, and (3) each sample is independent of the others. (Spatial and temporal independence criteria for sites will be addressed later.)

It some cases, ADB segments may have pollution problems (hotspots) concentrated only in a particular part of the stream, say, the last 5 stream miles. In such cases, it may not make sense to view the original ADB segment as the best possible sampling frame. That is, it would be better to further stratify the sample frame and, thus, the population of interest. This will prevent distortion of results caused by mixing together, for common analysis, data from the relatively un-impacted sub-segment with data from the impacted sub-segment. For example, in Figure 1 it might be prudent to consider the sub-segment upstream of the Star Mine as a sampling frame apart from the sub-segment below the mine. Stratification is common in studies employing purely random sampling, where it is referred to as stratified random sampling (Cochran, 1977). Stratification allows maximal precision of estimates for minimal sampling effort (Norris et al., 1992). **The assessor carrying out the analysis on an ADB segment will have to judge if further stratification is warranted**. If it is warranted, then sampling requirements, described above and further detailed below, would apply to *each* of the new sub-segments, individually.

Precautionary Considerations: Pollution sources are rarely evenly-dispersed along stream segments, violating assumption 1 above. And purely random sampling is usually not practical due to stream access issues, etc. Targeting only the known or potential pollution "hotspots" — even within a sub-segment has been broken out from a larger ADB segment — may over represent the hotspots and distort the statistical tests. Sampling and analysis plans (SAPs) should proceed with goal-oriented sampling (USEPA, 2000) that works towards striking an equitable balance between the number of hotspot sites and the number of un- or minimally impacted sites *within the defined sampling frame*. That is, the aggregate of collected samples should be representative (USEPA, 2002) of the ADB stream segment (or sub-segment) as a whole. Advanced knowledge and expertise of the field will be needed to accomplish this (Norris et al., 1992), and modifications to the sampling frame can be made on-the-fly during field work, if deemed necessary. It is in theory possible to sub-segment a stream reach to the point where, for a particular sub-segment, there really is nothing left but hotspots; if this is the case, then the hotspots *are* representative of the sub-segment. As a general rule, it is better to lump than split to avoid unnecessary sampling and administrative work.

Although this quasi-systematic approach is not a substitute for truly random sampling it will, if carried out properly, achieve good sample interspersion and representativeness. For further discussion of randomization vs. interspersion approaches, see page 196 of Hurlbert (1984).

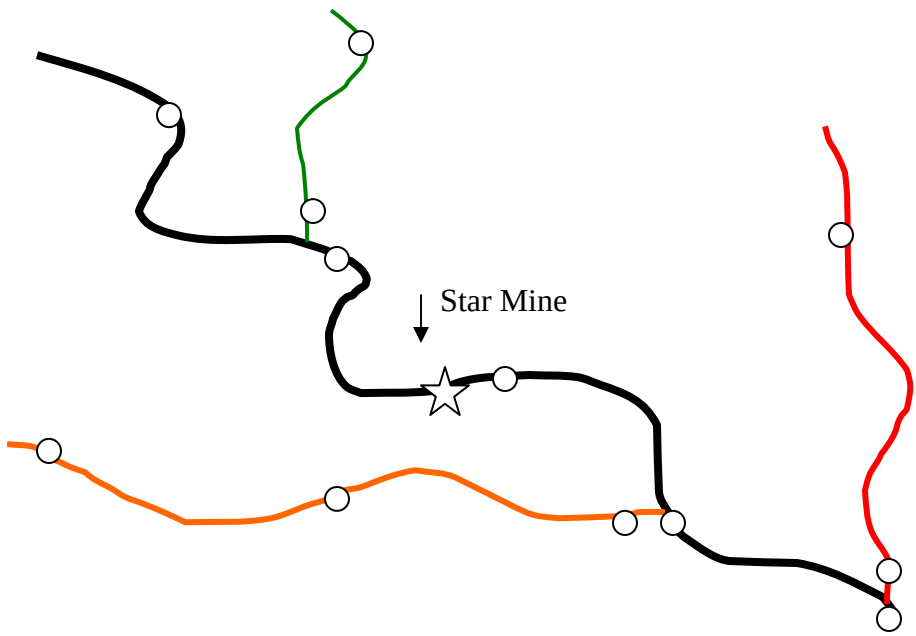


Figure 1. Four different stream reaches (shown by different colors), each representing 1 sampling frame (ADB stream segment). Example sampling sites (hollow dots) are shown along each segment.

3.0 Determining Sample Independence

According to definitions in Hurlbert (1984), much sampling carried out by MT DEQ on individual streams tends to violate spatial and temporal independence assumptions and results in pseudoreplication. For example, samples collected over time at a site can be serially correlated, which precludes temporal independence (Hurlbert, 1984). However, the conservative statistical view advocated by Hurlbert is not universally supported; contrary opinions on the matter can be found in the literature (Stewart-Oaten et al., 1986; Stewart-Oaten et al., 1992; Osenberg et al, 1994) and have led to what one journal referred to as a "healthy debate" (*Ecological Applications*, volume 4, No. 1, 1994). In general, more needs to be known about detection of non-independence and the frequency with which temporally independent samples can be collected (Underwood, 1994).

Time-series collected samples from a site may be used in inferential statistical testing, if used cautiously; this requires that one assumes that actual trends in time are identical in magnitude and direction for all the sites across the study (Norris and Georges, 1993). Osenberg et al. (1994) examine time-series serial correlation of physical and biological measurements in a BACI (Before-After-Control-Impact) study and conclude that, in the marine environment they study, sampling can occur at a site every 60 days without yielding substantial serial correlation.

MT DEQ recognizes the issue of temporal pseudoreplication, but also needs to be practical about the reality of sampling streams which, by their very nature, make collection of independent samples difficult. In MT DEQ's reference project (Suplee et al., 2005), 30 days has generally been used as a minimum time span between sampling events at a site to infer temporal

independence of water samples. This time span was based on the experiential observation that, during the brief Montana summer, substantial changes in flow, temperature, and vegetation (riparian and instream) occur from month to month, changes that would likely effect water quality. But Stewart-Oaten et al. (1986) recommend that the assumption of temporal independence be tested, rather than assumed. The Durbin-Watson test statistic is widely used to check for time-series serial correlation. Stream sites with monthly nutrient sampling during the summer were available in Montana, and some of these sites were tested using the Durbin-Watson statistic. Results are shown in Table 3.1 below.

Table 3.1. Durbin-Watson Values for Time-series Collected Nutrient Samples at Selected Sites. All Samples were Collected Approximately 30 Days Apart. Nutrients Showing Probable Time-series Serial Correlation (95% Confidence Level) are Highlighted.

Stream Site	Months Sampled	Time Range	n	Nutrient		
				Total N	Total P	NO_{2+3}
Rock Creek Site 2	June, July, Aug, Sept	2001-2004	12	1.18	1.43	2.3
Clark Fork R. at Deer Lodge (site 9)	July, Aug, Sept	1998-2006	25	1.81	1.78	1.68
Clark Fork R. above Little Blackfoot R. (site 10)	July, Aug, Sept	1998-2006	26	2.01	1.57	1.46
Clark Fork R. above Flathead R. (site 25)	July, Aug, Sept	1998-2006	26	1.76	1.21	2.08

In general, Durbin-Watson values around 2 mean there is no serial correlation, whereas values greater than approximately 2.5 or less than about 1.5 lead one to suspect negative or positive serial correlation, respectively (Neter et al., 1989; Ott, 1993). What can be concluded from this limited analysis? Most nutrients did not show serial correlation, and one of the three that did is borderline cases (statistic =1.43, but power of test very low). Overall, it appears that serial correlation is present in nutrient samples collected a month apart, but the effect is weak. It is evident that 30-day separated water samples can provide a fairly high degree of independence for nutrients.

MT DEQ is aware that spatial independence is also a concern. Water flows from upstream to downstream, consequently influencing the spatial independence of downstream sampling sites. No generally applicable spatial minimums were found as of this writing. U.S.EPA guidance (USEPA, 2002) generally glosses over the topic of spatial independence in streams. Spatial independence has been established (albeit only as rules of thumb) as a minimum of 1 mile between two sites (or short reaches), or a tributary confluences on a stream in between them, or if major land form or land use changes occur along the reach (MT DEQ, 2005; MT DEQ, 2007).

Given all these considerations, below are defined guideline recommendations for establishing independence of samples collected within an ADB segment or sub-segment:

- Sites (or short reaches equivalent to sites) should be located a minimum of 1 stream mile apart along an ADB stream segment or sub-segment.
- Sites may be placed < 1 mile apart on an ADB stream segment **if** there is a tributary confluencing with the segment between the two sites.
- Land use changes and land form changes should be considered and can be used to help define (1) breaks between sampling frames or (2) additional sampling sites within a sampling frame. See page 11-12 of MT DEQ (2005).

- Samples collected at the same site (or short reach) should be collected a minimum of 30 days apart.

Samples that meet the above conditions may be considered spatially and temporally independent for the purposes of determining compliance with the nutrient criteria. As such, they may be used in inferential statistical analyses and to make conclusions about the segment (or sub-segment) in question.

Precautionary Considerations. The last bullet above (temporal independence resulting from 30-day timespans) is not applicable for bioassemblage samples (i.e., macroinvertebrates, diatoms, and fish). These types of samples are based on organism populations that operate on different time scales from water samples and may show considerable year-to-year stability. Please see Bollman (2003), Bramblett et al. (2005), and Teply and Bahls (2007) for more details on temporal patterns of these biological assemblages.

4.0 Selection of Inferential Statistical Tests, Confidence Levels, and Associated Decision Rules

4.1 Rationale for Using Two Inferential Statistical Tests to Help Determine Compliance with Nutrient Standards

Exhaustive reviews of the pros and cons of statistical tests available for determining compliance with numeric standards have already been published (USEPA 2002; CA SWRCB, 2004). For brevity, rather than revisit all the detailed considerations put forward in those documents, recommendations are provided herein concerning what where judged to be the most applicable tests. These recommendations are then followed by a series of decision rules that allow the user to apply the tests in tandem. For purposes of compliance with numeric nutrient criteria, two tests should be used; the Exact Binomial Test and the One-Sample Student's t-test for the Mean.

- Exact Binomial Test: This test assumes data are dichotomous in nature (i.e., only two possible answers). For compliance with a criterion this reduces to (1) samples that exceed the criterion and (2) samples that do not exceed the criterion. If confidence levels, power, and exceedence rates (more on these below) are established upfront, minimum sample sizes can also be determined. The main disadvantage of the test is that it is blind to exceedence magnitude; that is, it takes no account of whether a sample exceeds the criterion by 1% or 1,000%.

- One-Sample Student's t-test for the Mean: This test does not assume the data take on a dichotomous relationship relative to the criterion. The test compares the mean of the samples in question to the criterion. The desired confidence levels in the test are established upfront. But unlike the Exact Binomial Test, it *is* greatly influenced by high values and outliers which can skew the dataset mean relative to the bulk of the other

samples in the dataset. It is also influenced by the proportion of non-detects in the dataset[5].

The Exact Binomial Test is useful for determining sample sizes, and is not influenced by large numbers of non-detects in the dataset. In fact, if the magnitude of nutrient criterion exceedences was irrelevant, then the Exact Binomial Test could probably be used by itself. But this is not the case; one must consider the issue of luxury nutrient uptake by algae.

One of the main purposes of establishing nutrient criteria is to control excess algae growth and its effects on water quality. Many benthic and water-column algae have the ability to take up the non-limiting nutrient, be that N or P, in excess of immediate need and utilize it for growth later (luxury nutrient uptake; Stevenson and Stoermer, 1982; Elrifi and Turpin, 1985; Portielje and Lijklema, 1994). If extracellular nutrient concentrations then decline in the water, growth can still be maintained on intracellular stores (Rhee, 1973; Droop, 1974). Therefore, pulsed loading events of nutrients to streams may allow algae to carry out luxury nutrient uptake which may sustain growth after the pulse has ended.

Luxury nutrients uptake is a kinetics phenomenon dependent on the physiological condition of the algae, duration and magnitude of the nutrient pulse, etc.; complex factors not easily addressed by a simple t-test. But the t-test can help assess the *potential* for luxury nutrient uptake because pulsed loads of elevated nutrient concentrations, if captured during sampling, would increase the dataset mean and would show that mean water quality has exceeded the criterion; this is useful information not provided by the Exact Binomial Test.

Each test possesses strengths the other does not. Therefore, it is recommended that the t-test be used in tandem with the Exact Binomial Test via a series of decision rules; these will be outlined in Section 4.3.

4.2 Form of the Null Hypothesis, Alpha, Beta, Effect Size, and Critical Exceedence Rate

All of these factors listed in Section 4.2's title are interrelated and influence one another in statistical hypothesis testing. Again, rather than reiterate here the mass of discussion devoted to these topics already covered elsewhere (EPA, 2002; CA SWRCB, 2004), I will simply state what I concluded to be the best statistical parameters (form of null hypothesis, alpha, beta, etc.) associated with the two tests (Exact Binomial and One-Sample Student's t-test for the Mean), and provide further explanation where warranted.

[5] For the purposes of using the t-test, users should initially convert all non-detects to 50% of the reported detection limit (USEPA, 2006). If >> than 15% of the dataset will be affected, consult Standards Section.

4.2.1 Form of the Null Hypothesis for the Statistical Tests

For Streams Already on the 303(d) List:

- Null Hypothesis: Waterbody is not in compliance with numeric nutrient standards
- Alternative Hypothesis: Waterbody is in compliance with numeric nutrient standards

For Streams Not on the 303(d) List:

- Null Hypothesis: Waterbody is in compliance with numeric nutrient standards
- Alternative Hypothesis: Waterbody is not in compliance with numeric nutrient standards

In effect, this is a "guilty until proven innocent" approach for streams already considered to have water quality problems, and an "innocent until proven guilty" approach for newly assessed streams. California uses the same approach (CA SWRCB, 2004).

4.2.2 Alpha, Beta, Effect Size

In statistical testing alpha, beta, effect size, and critical exceedence rate interact, and changes in one affect changes in the others. In environmental compliance work, there are strong arguments for attempting to balance type I (alpha) and type II (beta) errors; in doing so, it is important to consider the form of the null hypothesis and the implications for making one error or the other. Basically, each type of error has ramifications; one type of error leads to degradation of the environment, the other type of error leads to unnecessary expenditures on the part of the regulated entity. Working towards balancing type I and II errors is a process which inherently recognizes the consequences of each error type (Schroeter et al., 1993; Mapstone, 1995; page 178, CA SWRCB, 2004; page 52, Appendix C, EPA, 2002). *Given that* working towards balancing type I and type II errors is a valuable endeavor, here are general recommendations for the parameters to be input into statistical tests for nutrients in wadeable streams:

- Alpha should be about 0.25 or less (equates to ≥ 75% confidence level), depending on the form of the null hypothesis and its implications.
- Beta should be about 0.3 or less (equates to ≥ 70% power), and will vary according to the samples size (more on sample size minimum in Section 5.0).
- Effect size (Gray Zone) should be set at 0.15, per USEPA (2002).

In the statistical spreadsheet tools that accompany this technical appendix ("MT-NoncomplianceTool.xls" and "MT-ComplianceTool.xls"), one or the other file is used depending upon whether you are dealing with a new, unlisted stream (use "MT-NoncomplianceTool.xls") *vs.* a 303(d) listed stream (use "MT-ComplianceTool.xls"). You will be able to set alpha, critical exceedence rate (p), and effect size (p2) in the Exact Binomial Test in both of the files. The program will then return various sample sizes, their associated beta

values, and the maximum number of exceedences allowed while still remaining in compliance with the criterion.

For the One-Sample Student's T-test, you must input alpha and the nutrient criterion in mg/L. The One-Sample Student's T-tests will then provide a result indicating if the statistical test can or cannot confirm the alternative hypothesis. (The alternative hypothesis will reverse, according to whether you are using the tool for a listed or for a new, unlisted stream.)

4.2.3 Critical Exceedence Rate

Critical Exceedence Rate: *An estimate of the actual proportion of samples that exceed an applicable water quality criterion. When more than this proportion exceeds the criterion, the standard is not attained (i.e., stream is not in compliance with standard).*

Among the four statistical parameters critical to the Exact Binomial Test — alpha, beta, effect size, and exceedence rate — exceedence rate needs some kind of empirical ground-truthing to assure its validity. The implications of different alpha and beta errors can be understood relative to the form of the null hypothesis, and the effect size (gray zone) is not knowable *a priori*, therefore it is assumed; I recommend an effect size of 0.15 per EPA (USEPA, 2002). An exceedence rate, on the other hand, can be estimated using lines of reasoning and empirical evidence. The considerations used to estimate an exceedence rate for numeric nutrient standards were (1) recommended exceedence rates from EPA (USEPA, 2002) and (2) long-term benthic algae and nutrient relationships on the Clark Fork River, MT. Based on 1 and 2 (which will be further detailed):

- It is recommended that a critical exceedence rate for compliance with numeric nutrient standards be set at 0.2 (20%). Consideration 2 is detailed below using the scientific IMRAD reporting format for better clarity as to how the conclusion was reached.

(1) EPA recommends that, for a number of different polluting substances (e.g., fecal bacteria, conventional pollutants, toxic trace metals, etc.), criteria exceedence rates be set between 0.1 and 0.25 (10 to 25%) to protect beneficial uses (USEPA, 1997; USEPA, 2002).

(2) **Introduction.** Numeric nutrient (TN and TP) and benthic algae (mg Chl *a*/m^2) standards have been in place on most of the Clark Fork River in Montana for about 6 years. A systematic collection of nutrient and algae data has been ongoing since 1998. At a number of sites both algae and nutrient data have been collected multiple times each year for nearly 10 years. These data lent themselves well to empirically deriving a numeric nutrient exceedence rate because some river sites almost always exceed the algae standards, while others do not. The question became:

> *Do sites on the Clark Fork River that routinely exceed the numeric algae standards exceed the river's established numeric nutrient (TN and TP) standards more frequently than sites that do not exceed the numeric algae standards?*

Benthic algae levels in excess of 150 mg Chl *a*/m^2 (maximum) are not to be exceeded during the summer (ARM 17.30.631). Maximum in this case does not refer to a single high repeat measure from a Clark Fork River site; it refers to the mean value of a series of repeat measures (n = 15 to 20) that are collected at a site *during a particular sampling event*. Clark Fork River sites are usually sampled several times throughout the summer. It has been noted for some years that, during the summer, some sites are usually above the algae standards, while others are not. TN and TP standards were established on the Clark Fork River (ARM 17.30.631) and, if ultimately met, should keep benthic algae below the nuisance threshold described above. However, an exceedence rate was never explicitly established in the regulations. In carrying out the exceedence rate determination described herein, it is assumed that the magnitude of the TN and TP criteria on the Clark Fork River were accurately determined, and therefore any exceedence rate drawn from this analysis is meaningful.

Methods. Benthic algae and TN and TP concentration data where concurrently available for seven Clark Fork River sites from 1998-2006. Data were restricted to the time period June 30th to October 1st to generally comply with the summer growing season for this ecoregion (Suplee et al., 2007) and the regulatory timeframe in ARM 17.30.631. Every benthic Chl *a* measurement from a site (n = 15-20 per sampling event) collected over time was treated as a repeat measure. This resulted in a grand total of 285 to 333 repeat measures of Chl *a* at each site for the period 1998-2006. A grand benthic Chl *a* mean was calculated for a site by averaging all the repeat measures collected between June 30th and Oct 1st for all available years. Nutrient data collected at the corresponding sites during the same time frames where similarly compiled. At each site nutrients were collected as a single grab sample and, as a consequence, there were fewer data (43 to 78 N or P samples per site). Total N data were not collected; however, TKN and NO_{2+3} were. Therefore, for each site, individual Total N concentrations were calculated by summing the TKN and NO_{2+3} sample results collected simultaneously during a sampling event.

Next, the Clark Fork River TN and TP criteria concentrations were matched to their corresponding values in the nutrient cumulative frequency distributions for each site, and the associated percentile was recorded. For example, the TN criterion for the Clark Fork River is 0.3 mg/L, and it resulted that at site 9.0 (Clark Fork at Deer Lodge) 0.3 mg TN/L corresponded to the 23rd percentile of site 9.0's cumulative TN frequency distribution. This process was carried out for all 7 sites for both TN and TP. There is a break at the Blackfoot River confluence where the Clark Fork's upstream TP criterion (0.02 mg/L) differs from that below (0.039 mg/L); each TP criterion was applied as appropriate for a site's location along the river.

Results. Table 4.1 shows the results for 3 sites that, over the 1998-2006 time period, did not exceed the Clark Fork River's benthic algal biomass criteria. For this group of sites the nutrient criteria exceedence rate (both TN and TP) was, on average, about 8%. That is, nutrient samples whose concentrations exceed the standards occur only about 8% of the time at these sites. Table 4.2 shows three sites that *did* exceed the benthic algae standard; for this group of sites, the nutrient criteria exceedence rate was, on average, about 58%. Sites in Table 4.1 (did not exceed algae standard) had a range of exceedence rates (TN and TP) from 0.1%-24%, and sites in Table 4.2 (exceed algae standard) had a range of exceedence rates from 27.7% to 88%. The remaining site examined (Site 12; Clark Fork River at Bonita), which is not presented in Tables 4.1 or 4.2,

had a mean algae density (144 mg Chl *a*/m^2) so close to the algae standard it was considered borderline. Site 12's exceedence rate was 30.8% for TN, 68% for TP.

Table 4.1. Sites on the Clark Fork River (CFR) Not Exceeding the Maximum Benthic Algae Standard (Growing Season, 1998-2006)

Clark Fork River Site #	Site Name	Mean and [Median] Long-term Benthic Algal Biomass (mg Chl *a* /m^2, growing season)	Percentile in Site's Nutrient Frequency Distribution Matching CFR Standard*		**Criteria Exceedance Rate (%)**	
			TN	TP	TN	TP
15.5	Clark Fork above Missoula	96 [80]	90th	95th	10.2%	5.4%
22	Clark Fork at Huson	72 [52]	76th	96th	24.0%	3.8%
25	Clark Fork above Flathead	35 [20]	100th	99th	0.1%	1.5%
				Grand Mean:	**7.5%**	
				Grand Median:	**4.6%**	
				Maximum:	**24.0%**	
				Minimum:	**0.1%**	

* Rounded to nearest whole number for clarity.

Table 4.2. Sites on the Clark Fork River (CFR) Consistently Exceeding the Maximum Benthic Algae Standard (Growing Season, 1998-2006).

Clark Fork River Site #	Site Name	Mean and [Median] Long-term Benthic Algal Biomass (mg Chl *a* /m^2, growing season)	Percentile in Site's Nutrient Frequency Distribution Matching CFR Standards*		**Criteria Exceedance Rate (%)**	
			TN	TP	TN	TP
9	Clark Fork at Deer Lodge	180 [147]	23rd	50th	77.0%	50.0%
10	Clark Fork above Little Blackfoot River	163 [117]	48th	12th	52.0%	88.0%
18	Cark Fork at Shuffields	197 [181]	50th	72nd	50.4%	27.7%
				Grand Mean:	**57.5%**	
				Grand Median:	**51.2%**	
				Maximum:	**88.0%**	
				Minimum:	**27.7%**	

* Rounded to nearest whole number for clarity.

Discussion and Conclusion. The main assumption of this analysis was that the magnitudes of the Clark Fork River nutrient criteria, which were established as standards for the river, are correct. That is, if the nutrient standards are achieved, then summertime algae levels should be kept below the established nuisance thresholds. It was assumed that, as has previously been shown, both N and P co-limit in the Clark Fork River (Loman & Priscu, 1993; Dodds et al., 1997). It was further assumed that the algae standard (150 mg Chl *a*/m^2, site mean per sampling event) will protect beneficial uses. Regarding the later, research completed since the Clark Fork River standards were adopted in 2002 show that 150 mg Chl *a*/m^2 (site mean) is identified as a nuisance threshold by the Montana public majority (Suplee et al., *In Publication*[6]). If all these

[6] Available from the author upon request.

assumptions hold true, then reasonable exceedence rates for the 9 year dataset can be derived and used as a case study. It would have been ideal to have a true population of data (rather than a subset of data for a single river over a specific time period) with which to carry out this analysis. But such data are not readily available, and the long-term dataset examined here will have to serve as a proxy.

Comparison of Clark Fork River sites 15.5, 22, and 25 (don't exceed algae standard; Table 4.1) *vs.* 9, 10, and 18 (do exceed algae standard; Table 4.2) show a clear separation in the consistency of compliance with the river's numeric nutrient standards. It is clear from Table 4.2 that if the exceedence rate is ca. 50% then nuisance algae growth will almost certainly occur. But when the exceedence rate is ca. 5-10%, nuisance algae is unlikely to occur (Table 4.1). For purposes of estimating a protective nutrient criteria exceedence rate, the range of exceedence rates from these site groups needs to be considered as well. Note that an exceedence rate *of as much as* 24% does not result in excess benthic algae in some cases (site 22; Table 4.1). On the other hand, notice that an exceedence rate *of as little as* 27.7% can result in non-compliance with the algae standard (site 18; Table 4.2). Thus, an exceedence rate around 25% probably represent a threshold; if about 25% of the dataset exceeds the nutrient criteria, then there are roughly equal odds that the site could have nuisance algae (or not). This is partially supported by the fact that the single site with borderline algae conditions (site 12, Clark Fork River at Bonita; 144 mg Chl *a*/m^2) had a TN exceedence rate of 30.8%.

In conclusion, these analyses show that over a 9 year period (1998-2006) sites on the Clark Fork River that have consistently exceeded the nuisance algae standard (150 mg Chl *a*/m^2, summertime max) have TN and TP exceedence rates with a central tendency around 54%. On the other hand, sites that did not exceed the benthic algae standards had TP and TN exceedence rates with a central tendency around 6%. Within each group (sites that do not exceed algae standards, those that do; Tables 4.1 and 4.2), individual sites had exceedence rates as high as or as low as about 25%. This suggests that 25% may be an exceedence rate threshold where the ability to assure compliance with the algae standard becomes tenuous. Given that ca. 50% is certainly too high of an exceedence rate and will not protect beneficial uses, approximately 10% is probably too restrictive, and 25% is borderline, it is recommended that a nutrient exceedence rate be set to 20%.

4.3 Decision Rules for Applying the Inferential Statistical Tests

Measurements of benthic algal Chl *a* data (cold-water and transitional streams) and nighttime dissolved oxygen (DO) concentrations (warm-water streams)[7] should be available along with nutrient concentration data. DO and Chl *a* samples can be collected less frequently than nutrients (see footnotes in Tables, below) but are better if collected more than once during a field season. In streams, either benthic Chl *a* or nighttime DO data (i.e., effect variables) need to be considered along with nutrient concentrations to assess nutrient criteria compliance because heavy algae growth can deplete water nutrients concentrations to such an extent that nutrients are

[7] See page 11 of Suplee et al. (2005) for descriptions of ecoregionally segregated cold-water and warm-water streams. The definitions therein are general categorizations that correspond to ecoregional characteristics, and in no way override stream classifications codified in ARM17.30.607 through 17.30.617 or change any beneficial uses in ARM 17.30.621 through 17.30.629.

below the criteria, while at the same time excess algae/low DO is occurring and impacting beneficial uses. The decision rules, therefore, reflect decision making involving both inferential statistical tests and benthic algae or DO data.

Table 4.3. Decision rules for Determining Compliance with Nutrient Standards, Cold-Water & Transitional* Streams. For each Statistical Test, Regardless of the Form of the Null Hypothesis, PASS Means the Stream Segment Complies with the Nutrient Standard, FAIL Means the Stream Segment Does Not Comply with the Nutrient Standard.

Scenario	Binomial Test	T-test	Benthic Algae Chl*a* (mg /m^2)†	**Resulting Decision**	Notes on Decision
1	PASS	PASS	≤ 150 mg Chl*a* /m^2	In Compliance	All indications show that the stream is in compliance.
2	PASS	FAIL	≤ 150 mg Chl*a* /m^2	In Compliance	Suggests pulsed nutrient loads occur but are not resulting in elevated benthic algae biomass.
3	FAIL	PASS	≤ 150 mg Chl*a* /m2	Borderline	Likely that segment sometimes has high benthic algae biomass, but the timing of the algae sampling may have missed high levels. Borderline case; compliance or non-compliance may be equally justifiable; use other accompanying data (e.g., biometrics) to confirm decision. **OR, further sample algae & nutrients.**
4	FAIL	FAIL	≤ 150 mg Chl*a* /m2	Not in Compliance	Likely that segment has high benthic algae biomass, but the timing of the algae sampling may have missed the high levels. **Further algae sampling may be justified.**
5	PASS	PASS	> 150 mg Chl*a* /m2	Not in Compliance	Algae may be taking up nutrients and leading to low instream nutrient concentrations with concurrent high benthic algae biomass.
6	PASS	FAIL	> 150 mg Chl*a* /m2	Not in Compliance	Non-compliance with the T-test suggests that pulsed nutrient loads are allowing high algae biomass to be maintained via luxury uptake.
7	FAIL	PASS	> 150 mg Chl*a* /m2	Not in Compliance	Suggests sustained nutrient values near the standard but not necessarily pulsed nutrient loading.
8	FAIL	FAIL	> 150 mg Chl*a* /m2	Not in Compliance	All indicators show that the stream is not in compliance.

* Transitional streams are streams in ecoregion 42 or 43 but whose headwaters are in a mountainous ecoregion and whose water quality still largely reflects mountain conditions. Specific level IV ecoregions will be identified as containing transitional streams in the numeric nutrient criteria circular under development.

† Algae biomass collected and summarized as per DEQ SOPs, for a single site (short reach), during any given sampling event. Unlike nutrient samples, *do not* average together algae biomass results from different sites (short reaches) across the stream segment. Consider each site (short reach) on its own merits.

Table 4.4. Decision rules for Determining Compliance with Nutrient Standards, Warm-Water Streams. For each Statistical Test, Regardless of the Form of the Null Hypothesis, **PASS** Means the Stream Segment Complies with the Nutrient Standard, **FAIL** Means the Stream Segment Does Not Comply with the Nutrient Standard.

Scenario	Binomial Test	T-test	Dissolved Oxygen (DO)†	**Resulting Decision**	Notes on Decision
1	PASS	PASS	Meets standards	In Compliance	All indications show that the stream is in compliance.
2	PASS	FAIL	Meets standards	In Compliance	Suggests pulsed nutrient loads occur but are not resulting in large nighttime DO sags.
3	FAIL	PASS	Meets standards	Borderline	Likely that segment periodically does not meet DO standard, but the timing of the sampling may have missed the problem; conclusion that the stream segment *is* or *is-not* in compliance may be equally justified; use other data (e.g., biometrics) to decide. **Further DO & nutrient sampling justified.**
4	FAIL	FAIL	Meets standards	Not in Compliance	Likely that segment periodically does not meet DO standard, but the timing of the sampling may have missed the problem. **Further DO sampling is justified.**
5	PASS	PASS	Does not meet standards	Not in Compliance	Algae may be taking up nutrients and leading to low instream nutrient concentrations with concurrent high productivity and low nightime DO.
6	PASS	FAIL	Does not meet standards	Not in Compliance	Non-compliance with the T-test suggests that pulsed nutrient loads are allowing high productivity to be maintained via luxury uptake, leading to low nighttime DO.
7	FAIL	PASS	Does not meet standards	Not in Compliance	Suggests sustained nutrient values near the standard but not necessarily pulsed nutrient loading.
8	FAIL	FAIL	Does not meet standards	Not in Compliance	All indicators show that the stream is not in compliance.

† Nightime sampling of DO for at least several days should be measured directly via a YSI sonde or similar; or, diatom-inferred DO can be used. Diatom-inferred DO concentrations should be compared to the 7 day mean minimum in DEQ-7 (MTDEQ, 2006b).

5.0 Minimum Sample Size

The final consideration is minimum nutrient sample size. A "nutrient" sample refers to a nutrient type such as TP, NO_{2+3}, or TN, or groups of equivalent nutrients as outlined on page A-3 of Appendix A. Sample sizes apply to each nutrient type, and not to the total number of nutrient samples collected from a stream segment. So, if 7 TN and 7 TP samples were collected from a segment they would not represent 14 samples, but rather 7 of TN and 7 of TP. There is extensive discussion of determining appropriate sample size on a study-by-study basis in USEPA (2002). However, the recommendations made here for determining compliance with the numeric nutrient standards are meant to apply generally to all Montana wadable streams, mainly for purposes of 303(d) listing/delisting. *Please note that these sample size minimums do not apply to biological samples (i.e., benthic algae or diatoms) that may be collected concurrently with the nutrient samples.*

For unlisted streams, those for which the form of the null hypothesis is "complies with standard" (Section 4.2.1), the implication for making a type II (beta) error is that a truly non-compliant stream segment would be incorrectly declared compliant. This is a scenario MT DEQ wants to minimize, and so the probability of such an outcome should be reduced well below 50%, i.e., a level the difference from which is different from a coin flip. The Exact Binomial Test in the accompanying spreadsheet tool can be used to estimate minimal sample sizes. In the test it can be seen that if alpha (type I) error is set to 0.25, exceedence rate (p) to 0.2, and effect size/gray zone (p2) to 0.15 (entered value = 0.35), then a Beta (type II) error of 0.35 is achieved with 12 samples. (Note in the spreadsheet that introducing lower and lower alpha values causes beta error to increase and, therefore, many more samples are needed to try to balance alpha and beta errors.) Twelve samples is about as low an **n** that can be used and still have roughly balanced (0.25 *vs*. 0.35) alpha and beta errors that are each well below 50%.

For listed streams, a similar approach is used. Listed streams are those for which the form of the null hypothesis is "does not comply with standard" (Section 4.2.1). In this scenario, the implication of making a type I (alpha) error is that a truly non-compliance stream segment is incorrectly declared compliant; again, this is a scenario MT DEQ wants to minimize. Setting the alpha to 0.25, exceedence rate (p) to 0.2, and effect size/gray zone (p2) to 0.15 (entered value = 0.05), a beta error of 0.14 (14%) can be achieved with 13 samples. This is a reasonable balance of type I and II errors, and provides a total sampling effort about the same as that for unlisted streams. Given these considerations, it is recommended that:

- For new, unlisted stream segments, a minimum of 12 independent samples for any given nutrient be collected for compliance determination.

- For 303(d)-listed stream segments having 1 nutrient criteria exceedence for a given nutrient, a minimum of 13 independent samples (this total can include newly collected as well as previously collected samples) should be used for compliance determination.

- For listed streams with ≤ 13 total samples that already have 2 or more exceedences for a given nutrient, the default conclusion is that the stream segment has failed the Exact Binomial Test (no further sampling required at this time). Run the t-test as well, but be

aware that low sample size will often contribute to a T-test result of "cannot confirm that the site complies with the standard" (i.e, FAIL).

- For listed streams that have > 13 samples for a given nutrient, set alpha to 0.25, exceedence rate to 0.2 and effect size to 0.15 in the Exact Binomial Test, and determine if the reach is (or is not) in compliance with the Exact Binomial Test Carry out the same for the T-test.

If a *very* large dataset (> 300 samples) is available for a particular stream, then lower type I (alpha) and type II (beta) error can be achieved with higher confidence in the results. Use the special feature of the Exact Binomial Test to help define these confidence levels. Confer with Standards Section if needed.

REFERENCES

Bollman, W., 2003. The Northeaast Montana Nutrient Pilot Study, 2001-2002. An Addendum. Submitted to the Montana Department of Environmental Quality, Helena MT.

Bramblett R.G., T.R. Johnson, A.V. Zale, and D.G. Heggem, 2005. Development and Evaluation of a Fish Assemblage Index of Biotic Integrity for Northwestern Great Plains Streams. Transactions of the American Fisheries Society 134: 624-640.

Cochran, W.G., 1977. Sampling Techniques, 3rd Edition. John Wiley and Sons, New York.

CA SWRCB (California State Water Resources Control Board), 2004. Functional Equivalent Document, Water Quality Control Policy for Developing California's Clearn Water Act Section 303(d) List — Final. September 2004.

Dodds, W.K., V.H. Smith, and B. Zander, 1997. Developing Nutrient Targets to Control Benthic Chlorophyll Levels in Streams: A Case Study of the Clark Fork River. Water Research 31: 1738-1750.

Droop, M.R., 1973. Some Thoughts on Nutrient Limitation in Algae. Journal of Phycology 9: 264-272.

Elrifi, I.R., and D.H. Turpin, 1985. Steady-state Luxury Consumption and the Concept of Optimum Nutrient Ratios: A Study with Phosphate and Nitrate Limited *Selenastrum minutum* (Chlorophyta). Journal of Phycology 21: 592-602.

Hurlbert, S. H., 1984. Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs 54: 187-211.

Larix Systems, 2006. Statistical Survey Analysis Report: Algae Perception Survey —Final. Submitted to Michael Suplee, Ph.D. of the Montana Department of Environmental Quality. December 2006.

Lohman, K., and J. C. Priscu. 1992. Physiological Indicators of Nutrient Deficiency in *Cladophora* (chlorophyta) in the Clark Fork of the Columbia River, Montana. Journal of Phycology 28: 443-448.

Mapstone, B. D., 1995. Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors. Ecological Applications 5: 401-410.

MT DEQ (Montana Department of Environmental Quality), 2005. Field Procedures Manual for Water Quality Assessment Monitoring. Water Quality Planning Bureau, Document No. WQPBWQM-020, Revision 2. April 21, 2005.

MT DEQ (Montana Department of Environmental Quality), 2006a. 2006 Integrated 303(d)/305(b) Water Quality Report for Montana. December 7, 2006.

MT DEQ (Montana Department of Environmental Quality), 2006b. Circular DEQ-7, Montana Numeric Water Quality Standards. February, 2006.

MT DEQ (Montana Department of Environmental Quality), 2007. Sampling and Analysis Plan, and Associated White Paper: Sampling Reference Sites to Produce a More Uniform Nutrient Dataset and Improve Montana's Wadeable Stream Nutrient Criteria. Finalized May 21, 2007.

Neter, J., W. Wasserman, and M.H. Kutner, 1989. Applied Linear Regression Models, 2nd Edition. Irwin, Burr Ridge, Illinois.

Norris, R.H., E.P. McElravy, and V.H. Resh, 1992. The Sampling Problem. In: *The River Handbook*, P. Calow and G.E. Petts (Editors). Blackwell Scientific Publications, Oxford, England.

Norris, R.H., and A. Georges, 1993. Analysis and Interpretation of Benthic Macroinvertebrate Surveys. In: *Freshwater Biomonitoring and Benthic Macroinverebrates*, D.M. Rosenberg and V.H. Resh (Editors). Chapman and Hall, New York, U.S.A.

Osenberg, C.W., R.J. Schmitt, S.J. Holbrook, K.E. Abu-Saba, and A.R. Flegal, 1994. Detection of Environmental Impacts: Natural Variability, Effect Size, and Power Analysis. Ecological Applications 4: 16-30.

Ott, R. L., 1993. An Introduction to Statistical Methods and Data Analysis, 4th Edition. Duxbury Press, Belmont, CA.

Portielje, R., and L. Lijklema, 1994. Kinetics of Luxury Uptake of Phosphate by Algae-Dominated Benthic Communities. Hydrobiologia 275/276: 349-358.

Rhee, G.Y., 1973. A Continuous Culture Study of Phosphate Uptake, Growth Rate and Polyphosphate in *Scenedesmus sp*. Journal of Phycology 9: 495-506.

Schroeter, S.C., J.D. Dixon, J. Kastendiek, and R.O. Smith, 1993. Detecting the Ecological Effects of Environmental Impacts: A Case Study of Kelp Forest Invertebrates. Ecological Applications 3: 331-350.

Stevenson, R.J., and E.F. Stoermer, 1982. Luxury Consumption of Phosphorus by Five *Cladophora* Epiiphytes in Lake Huron. Transactions of the American Microscopy Society 101: 151-161.

Stewart-Oaten, A., and W.W. Murdoch, 1986. Environmental Impact Assessment: "Psuedoreplication" in Time? Ecology 67: 929-940.

Stewart-Oaten, A., J.R. Bence, and C.W. Osenberg, 1992. Assessing Effects of Unreplicated Perturbations: No Simple Solutions. Ecology 73: 1396-1404.

Strahler, A. N., 1964. Quantitative Geomorphology of Drainage Basins and Channel Networks. Pages 439-476 in V. T. Chow (editor). Handbook of Applied Hydrology. McGraw-Hill, New York.

Suplee, M., R. Sada de Suplee, D. Feldman, and T. Laidlaw, 2005. Identification and Assessment of Montana Reference Streams: A Follow-up and Expansion of the 1992 Benchmark Biology Study. Montana Department of Environmental Quality, Helena, Montana, 41 p, November 3, 2005. *Available at* http://deq.mt.gov/wqinfo/Standards/Refsites_writeup_FINALPrintReady.pdf.

Suplee, M.W., A. Varghese, and J. Cleland, 2007. Developing Nutrient Criteria for Streams: An Evaluation of the Frequency Distribution Method. Journal of the American Water Resources Association 43: 453-472.

Suplee, M.W., V. Watson, M. Teply, and H. McKee. How Green is Too Green? Public Opinion of What Constitutes Undesirable Algae Levels in Streams. Journal of the American Water Resources Association, *In Publication*.

Tepley, M., and L.L. Bahls, 2007. Statistical Evaluation of Periphyton Samples from Montana Reference Streams. Technical Report, Prepared for the Montana Department of Environmental Quality. February 2007.

Underwood, A.J., 1994. On Beyond BACI: Sampling Designs that Might Reliably Detect Environmental Change. Ecological Applications 4: 3-15.

USEPA (United State Environmental Protection Agency), 1997. Guidelines for Preparation of the Comprehensive State Water Quality Assessments (305(b) Reports) and Electronic Updates: Supplement. EPA-841-B-97-002B. Washington, D.C.; Assessment and Watershed Protection Division (4503F), Office of Wetlands, Oceans, and Watersheds, and Office of Waters, U.S. Environmental Protection Agency.

USEPA (United State Environmental Protection Agency), 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams. United States Environmental Protection Agency, EPA-822-B00-002. Washington, D.C. *Available at* http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/index.html

USEPA (United States Environmental Protection Agency), 2002. Consolidated Assessment and Listing Methodology, Toward a Compendium of Best Practices, 1st Edition. Prepared by USEPA Office of Wetlands, Oceans and Watersheds. July 2002. *Available at* http://www.epa.gov/owow/monitoring/calm.html

van Dam, H., A. Mertens, and J. Sinkeldam, 1994. A Coded Checklist and Ecological Indicator Values of Freshwater Diatoms from the Netherlands. Netherlands Journal of Aquatic Ecology 28: 117-133.

Appendix I. Point Source Permitting and Compliance for Nutrients

Pursuant to the Clean Water Act (CWA), MT DEQ in the process of developing water quality standards for nutrients in Montana rivers and streams. Because nutrients regulation is an evolving area of CWA implementation and because of inherent differences between nutrients and other potential water pollutants, MT DEQ is examining whether existing compliance assessment methods and tools are appropriate to apply to the draft nutrient criteria. Section 6 and Appendix H document draft statistical tools to support regulatory decisions about whether instream nutrient concentrations should be considered compliant with the draft water quality criteria. This appendix discusses the relationships between water quality criteria and point source discharge permitting under the National Pollution Discharge Elimination System (NPDES) established by the CWA (1972). In particular, the appendix describes how NPDES effluent limits are determined and how monitoring and reporting requirements are used to determine compliance. In addition, this section describes relevant nutrient permitting and compliance mechanisms of the Chesapeake Bay Program. Finally, conclusions and recommendations are provided for an optimized NPDES nutrient standards compliance strategy.

1.0 Regulatory Overview

Point source releases to surface water are regulated under NPDES, established by the Federal Water Pollution Act (Clean Water Act) and its amendments (CWA 1972). NPDES is a permitting program that applies to all facilities that discharge pollutants from any point source to US waters. EPA may authorize states, territories, or tribes to implement CWA-compliant point source permitting programs (EPA 1996). A permit is a license for a facility (i.e., individual permit) or group of facilities (i.e., general permit) to discharge a specified amount of a pollutant into a receiving water body. A general permit may be issued if the permitting authority determines that a group or category of dischargers have similarities that warrant a general permit.

Among other things, a NPDES permit must contain effluent limitations and monitoring and reporting requirements. It is the responsibility of the permit writer to do the following:

(1) Determine technology-based effluent limits for each pollutant;
(2) Determine effluent limits protective of state water quality standards; and
(3) Compare the technology-based standards with the water quality-based standards and apply the more stringent of these in the NPDES permit (EPA, 1996).

The terms “technology-based effluent limit” and “water-quality based effluent limit” are defined in Sections 1.1 and 1.2, respectively. The methodology used by permit writers to determine each type of limit and determine which to implement in the permit is described. Once effluent limits are set, facilities must remain in compliance with these limits. If a facility exceeds its limits, it may be subject to penalty by the state or federal government. Section 1.3 describes compliance and enforcement of permit limits in greater detail.

1.1 Technology-Based Effluent Limits

Technology-based effluent limits are "end-of-the pipe" effluent limits intended to reflect pollutant levels attainable by available control technologies. Permit writers determine technology-based effluent limits based either on national effluent limitation guidelines (ELGs) published by EPA, or when ELGs are unavailable for a particular source type or pollutant, on best professional judgment. ELGs include maximum and long-term average limitations based on the demonstrated performance of a reasonable level of treatment that is within reasonable economic limits of categories of industrial facilities. Permit writers establish technology-based effluent limits for facilities using the following information: the number of production days per year, the annual and average daily effluent discharge rate, and the daily maximum and monthly average ELGs for pollutants present in the point source discharge. Technology-based effluent limits are expressed as a concentration or daily production rate (i.e., lbs per day). All limits must be expressed as either daily maximums or monthly averages. According to the NPDES Permit Writers' Manual (EPA 1996), permit holders should design facility treatment systems to meet the monthly average rather than the daily maximum limit because daily fluctuations are expected.

1.2 Water-Quality Based Effluent Limits

When technology-based effluent limits for point source discharges are not stringent enough to prevent pollutant concentrations in receiving surface waters from exceeding water quality standards, permit writers may use state water quality standards to determine effluent limits. Under section 303(c) of the CWA, individual states are required to set water quality standards for specific water bodies. These water quality standards are composed of three parts: (1) use classifications for the water body; (2) numeric and/or narrative water quality criteria; and (3) an antidegradation policy. Each of these parts is described below.

1.2.1 Use Classifications

Also under the CWA, states are required to specify appropriate uses for water bodies. Uses are determined by considering the biological, physical, and chemical properties of the water body as well as its geographical location and setting. Designated uses include public water supply and protection of fish, shellfish, and wildlife, and recreation. Water bodies may also be used for agricultural, industrial, and navigational purposes. Water bodies may have multiple uses. States determine use classifications for water bodies based on their designated uses. EPA recommends five use classifications (EPA 2006):

1. All designated uses are supported and no use is threatened.
2. Available data indicate that some but not all designated uses are supported.
3. There are insufficient data to support a use determination.
4. Available data indicate that at least one designated use is not being support or is threatened, but a TMDL[1] is not needed.

[1] TMDL stands for total maximum daily load. The TMDL is the allocated amount of a specific pollutant that may be discharged into a waterbody from a specific source. Any discharge load that exceeds the TMDL may result in a violation of the water quality standard. A TMDL may be expressed in terms of pollutant mass per unit of time or by toxicity (EPA 1996).

5. Available data indicate that at least one designated use is not being supported or is threatened and a TMDL is needed.

Each designated use for a water body has an associated water quality criterion. For example, swimming would typically be associated with a more stringent criterion than limited-contact uses like navigation. These criteria are discussed in further detail below.

1.2.2 Water Quality Criteria

Water quality criteria may be numeric or narrative. If numeric criteria are unavailable, narrative criteria may be used instead. For example, a narrative criterion could assert that no discharge causing nuisance algal blooms is acceptable. When numeric criteria are available, EPA asserts that for "conventional pollutants" (e.g., total suspended solids, pH, and fecal coliform bacteria but *not* including nutrients), a water body may not attain its designated use if more than 10 percent of the numeric measurements exceed a water quality criterion (EPA 1997). However, EPA has not encouraged this "10 percent rule" to be applied for other pollutants such as toxics (EPA 2006).

When available, numeric water quality criteria include (EPA 1996):

- Magnitude (i.e., allowable pollutant concentration in the water body);
- Duration (i.e., the period of time over which the instream concentration is averaged for comparison with criteria concentration factors); and
- Frequency (i.e., how often the criteria can be exceeded).

1.2.2.1 Magnitude

The magnitude is a pollutant-specific value that is generally expressed as a concentration. This concentration, in combination with the duration and frequency specified by the criterion, indicates an allowable concentration that is considered protective of the water body and its aquatic environment. A criteria continuous concentration (CCC) and a criteria maximum concentration (CMC) are set by EPA for many pollutants. The CCC is the water quality criteria recommendation for the highest instream concentration of a pollutant or effluent to which organisms can be exposed continuously without causing unacceptable effects. Significant excursions above the CCC will likely reduce growth and reproduction (EPA 1991). The CMC indicates the Agency's recommendation for the highest instream concentration of a pollutant or effluent to which organisms can be exposed briefly without causing unacceptable effects. Significant excursions above the CMC will likely cause death and other severe effects (EPA 1991).

1.2.2.2 Duration

EPA recommends that different duration periods be used to protect against chronic and acute effects. According to the EPA *Guidelines for Deriving Numerical National Water Quality Criteria* (EPA 1985), chronic effects in aquatic wildlife, excluding sensitive species, should not be affected adversely if the four-day average instream concentration of the pollutant does not

exceed the criteria continuous concentration. Thus, to protect against chronic effects, EPA recommends a four-day duration period. For acute effects, excluding sensitive species, aquatic wildlife should not be adversely affected if the one-hour average concentration of the pollutant does not exceed the criteria maximum concentration. Thus, to protect against acute effects, EPA recommends an hour-long duration period.

1.2.2.3 Frequency

EPA also recommends a number of times that the CCC or CMC may be exceeded over the recommended duration period without posing risk to the aquatic environment. EPA (1985) recommends that the CCC or CMC not be exceeded more than one time every three years on the average.

1.2.3 Antidegradation Policy

The third element of a water quality standard is the state's tiered antidegradation policy. States are required to show that once a water body has been established to be protective of its uses, it cannot "degrade" to the point that it fails to meet its designated uses (Hair 2008). States must have policies to evaluate degradation of water bodies. If, for example, there is an increase in discharge from one or a group of facilities, the increase must be justified by demonstrating that the increase will not degrade the water body.

1.2.4 Determination of Effluent Limit

As stated above, NPDES permit writers base effluent limits on either technology-based standards or water quality-based standards, depending on whichever are more stringent. However, the water quality-based standards apply to instream concentrations, not end-of-the-pipe effluents. Permit writers use models to back-calculate from narrative or numeric water quality criteria to determine effluent limits that are protective of the water body. In doing so, the lowest instream flow and the highest effluent design flow for the discharging facility (or facilities) are assumed to simulate the most conservative conditions. Models take into account the standard or other water quality goals for the receiving water body in addition to the water body's assimilative capacity to determine what effluent load can be discharged by the facility (or facilities) without degrading the quality of the water body. The resulting facility effluent limit is expressed as a concentration (e.g., mg/L) or daily production rate (e.g., pounds per day) and, for conventional pollutants, must be expressed as either a daily maximum or monthly average limit.

1.3 Effluent Limit Compliance and Enforcement

States authorized to implement point source permit programs are required to collect monitoring data from permit holders on a quarterly basis. Any permitted facility must report the number of daily maximum and monthly average exceedances that have occurred during the quarterly reporting period.[2] Based on this information, states determine which facilities are in compliance. One violation of a permit-defined effluent limit may be considered by the state to be a violation,

[2] Temporal distribution or clustering of exceedances is not generally reported as permit holders only report the total number of exceedances without designation of when the exceedances occurred.

but a single violation will likely not incur the same degree of enforcement as a series of violations. States may implement state program-specific enforcement systems such as fines or other penalties. For example, Montana implements a graduated enforcement system. All permit holders are required to pay an annual fee based on the type and volume of the discharge (Reid 2008). If the permit holder remains in compliance with all effluent limits, a 25% discount on the annual fee is granted. Violation of a single effluent limit results in loss of this discount. In Montana, if permit holders continue to violate effluent limits, the state takes further enforcement action. In addition, any facilities or groups of facilities that are not in compliance are reported by states to EPA in a Quarterly Noncompliance Report. EPA determines what enforcement measures are necessary to bring facilities back into compliance.

There are five types of violations that states must report including (EPA 1996):

- Violations of monthly average effluent limits – A facility is in violation of a monthly average effluent limit if the monthly average effluent concentration exceeds or equals the limit times the technical review criteria (defined further below) for two months during a six month period; or if the monthly average effluent concentration exceeds the limit (not multiplied by the technical review criteria) for four out of six months.
- Interim effluent limits set forth in a formal enforcement action – A facility must also report violations of any magnitude interim effluent limits that are established in any formal enforcement action.
- Schedule – A facility is in violation if it misses a compliance schedule milestone date by 90 days.
- Reporting – Reporting violations must be reported when a facility misses a report due date by 30 days.
- Single event – A violation occurs when there is any violation of any magnitude that is considered to have and adverse effect on water quality or public health.

With regard to the first type of violation (violation of monthly average effluent limits), EPA established different technical review criteria (TRC) for categories of pollutants for which there is some uncertainty in determining exceedances (EPA 1996). The technical review criterion for a pollutant is either 1.4 or 1.2 for Group I and Group II pollutants, respectively. Group I pollutants include inorganic phosphorus compounds, inorganic nitrogen compounds, and other nutrients (40 CFR 123.45). Using the TRC methodology, water quality data collected by permit holders may be up to 40% or 20% higher than the monthly average effluent limit for Group I and II pollutants respectively for any two months out of a six month period without violating the permit. The goal of the TRC is to distinguish true exceedances of limits from measurements that appear to be over the limit, but are in fact not over the limit. This type of measurement error may be referred to as a “false positive” or a type I (alpha) error.

The binomial tool presented in Section 6 of the main report uses a comparable, but more sophisticated methodology to distinguish between true exceedances and false negatives. When using the binomial test, the user reduces all observations to a binary format where "yes" is entered if the observation exceeds the standard and "no" is entered if otherwise. The tool computes the maximum number of exceedances allowable for a given number of samples so that the user can have a user-defined level of confidence (presented as a percentage) that no more

than a specific percent of the samples in a given water body exceed the standard. For example, the user of the tool could define compliance as follows: a water body is compliant if there is 95% confidence that no more than 5% of all possible samples from the water body exceed the standard. The student t-test method, also presented in Section 6 of the main report, is another means by which compliance may be determined. The t-test determines of two values are statistically different, taking into account the variability of measurements. The t-test may be used to compare the sample mean with the water quality criteria defined by the standard. Compliance is met, for example, if there is 95% confidence that the sample mean is less than the standard.

To properly use the binomial test, samples should be spatially and temporally independent. The requirement of spatial independence will be met if all effluent outlets are sampled with equal frequency (or in proportion to discharge) and the samples are collected from well-mixed zones.

For temporal independence, there should be an effort to sample with even frequency through the year. Sample independence may not be satisfied if the samples are collected, for example, only during periods of high discharge, low discharge, or during particular seasons or weather conditions.

2.0 Permitting Guidelines for Nutrients

The EPA permitting guidance summarized above generally apply to all pollutants, including nonconventional pollutants such as nutrients (e.g., nitrogen and phosphorus). As with other pollutants, NPDES permits should specify effluent limits for nutrient that are protective of receiving waters, and are developed in consideration available technology-based limits. Technology-based effluent limits are not generally available as national ELGs for nutrients have not been established formally. Water-quality based limits are used more commonly in permits. Nutrient criteria may be utilized by permit writers in the same way as criteria for other pollutants. Like Montana, many states are in the process of establishing nutrient water quality criteria.[3] While some have numeric criteria, most rely on narrative criteria. In either case, the same method of back-calculating (see Section 1.2.4.) used to determine effluent limits for other pollutants is also used for nutrients.

Once a limit is established for a permit holder, the same federal compliance regulations apply. Nutrient limit exceedances must be reported to states on a quarterly basis just as are exceedances of limits for other pollutants, and states report the types of violations to EPA as described in Section 1.3. Assuming that a permit's nutrient limit is expressed in terms of a monthly average, a facility is in violation of a monthly average effluent limit if the monthly average effluent concentration exceeds or equals the limit times the TRC of 1.4 for two months during a six month period; or if the monthly average effluent concentration exceeds the limit (not multiplied by the technical review criteria) for four out of six months.

[3] The current status of numeric nutrient criteria development by states and territories may be found at the following EPA website: http://www.epa.gov/waterscience/criteria/nutrient/strategy/status.html.

One difference worthy of note is that annual or seasonal limits may be acceptable for nutrients as opposed to daily maximum or monthly averages (EPA 2008). EPA must approve such a decision on a case-by-case basis. For example, a memorandum from EPA's Director of the Office of Wastewater Management approved a proposal submitted by the Water Permits Division of EPA Region 3 and the Chesapeake Bay Program Office to implement annual nutrient limits as opposed to shorter-term limits (EPA 2004). The rationale behind this decision is discussed in greater detail below.

2.1 Chesapeake Bay Case Study

The Chesapeake Bay is a large watershed that is impacted by six different states. Since the 1970s, there have been various movements and programs established to protect the Bay. In 1983, the first Chesapeake Bay Agreement was signed that established the Chesapeake Bay Program (CBP) (CBA 1983). Three states including Maryland, Pennsylvania, and Virginia as well as the District of Columbia signed the agreement. Participants committed to protecting the Bay and laid the foundation for the Chesapeake Bay Water Quality Monitoring Program which began in 1984. A subsequent amendment to the Chesapeake Bay Agreement signed in 1992 called attention to pollution from nutrients and in 2000 (CBA 1992), Chesapeake 2000 committed Bay states to reducing nutrient and sediment pollution from the Bay enough so that the Bay and its tidal rivers would no longer be considered impaired water bodies by 2010 (Chesapeake 2000). Nitrogen and Phosphorus loadings to the Bay were to be reduced by 40%.

The Chesapeake Bay is an example of a watershed for which a general permit (i.e., covering multiple facilities) is in place. The six states that impact the Bay have identified 483 facilities that are considered "significant dischargers" of the Bay. A significant discharger is defined by the permit as either an existing facility that discharges 100,000 gallons or more per day directly into tidal waters or an existing facility that discharges 500,000 gallons or more per day directly into non-tidal waters (CBW 2007). All significant dischargers of the Bay are required to collectively meet nutrient load allocations that are established for the Chesapeake Bay subwatersheds by 2010 or 2011 (McFadden 2008). Based on the total annual nutrient load considered to be protective of the Bay, each significant discharger is allotted an annual effluent load. EPA agreed that annual limits for nitrogen and phosphorus for the Chesapeake Bay were more appropriate than daily maximum, weekly average, or monthly average effluent limits due to inherent differences between nutrients and conventional pollutants (EPA 2004). For example, the exposure period of concern for nutrients is longer than one month and may be up to several years. Thus, the average nutrient load over an extended time period is of greater concern than the maximum load. Nutrients do not have direct effects on the aquatic community but are processed incrementally by ecosystems. Nutrients are taken up by algae throughout the year and eventually settle to the bottom where they decay in warm summer waters. Here they contribute to anoxia, or lack of oxygen. This process introduces a delay between the time nutrient discharges occur and the emergence of effects on the water body. The Bay is able to integrate different monthly loads from dischargers, thus if the annual limit is met, any intra-annual nutrient variability should not affect the Bay adversely (EPA 2004).

Significant dischargers to the Bay are required to submit monthly reports that include the average monthly discharge as well as the total effluent load discharged since the beginning of the

calendar year (McFadden 2008). Some states such as West Virginia require that facilities consider the aggregate discharge and use that to project the total annual discharge for the calendar year. If the projected annual discharge exceeds the facility's allocated annual effluent limit, the facility must demonstrate that it is taking action to ensure that the annual limit will not be exceeded. If a facility is at risk of exceeding its annual limit, it may participate in a nutrient exchange program. The Chesapeake Bay permit establishes a credit trading system in which nutrient credits (i.e., the difference in pounds between the discharging facility's limit and the pounds actually discharged) may be exchanged or purchased by facilities that may exceed their wasteload allocations (EPA 2008). For example, facilities in Virginia may choose to participate in an exchange program implemented by the Virginia Nutrient Credit Association. The Exchange Association reviews member facilities' compliance plans, treatment capabilities, control costs, and potential expansions and ultimately develops compliance scenarios for each facility.

In order for a facility to remain in compliance with its annual effluent limit, it must demonstrate that it has not exceeded the annual limit in every monthly report (McFadden 2008). If a facility does exceed the limit, it is considered in violation of its limit for that month and every succeeding month that it is over the limit through the end of the calendar year. For example, if a facility exceeds its limit in October, it is considered out of compliance for October through December. A new compliance year begins on January 1st at which point the facility would be back in compliance. At this point, the states of the Chesapeake Bay watershed have not stipulated penalties for specific violations. About half of the 438 permits have been written and issued to significant dischargers, and these facilities have until 2010 or 2011 to come into compliance with nutrient limits. Specific enforcement strategies are currently under development by each state.

3.0 Conclusions and Recommendations

The process by which permit writers establish nutrient limits requires consideration of multiple factors. Permit writers must consider both technology-based effluent limits (where available) and water-quality based limits. The following conclusions may be drawn regarding the permit writing process for nutrients.

- Numeric effluent limits in NPDES permits must be based on the more stringent of two effluent limits (i.e., the technology-based effluent limit and the water-quality based limit).
- Although regulatory development for nutrients is behind other pollutants, the permitting, monitoring, and compliance framework developed for other pollutants may be applied to nutrients.
- States require permit holders to submit monitoring reports that include all exceedances of permit limits. Facilities violating their permit limits are subject to enforcement on state and federal levels. A facility is reported to EPA in a Quarterly Noncompliance Report if a pollutant's monthly average effluent concentration exceeds or equals the limit times the pollutant's TRC for two months during a six month period; or if the monthly average effluent concentration exceeds the limit (not multiplied by the technical review criteria) for four out of six months. For nutrients, the TRC is 1.4.

- Specific states such as those of the Chesapeake Bay have instituted a regulatory approach that is specific to nutrients. The Chesapeake Bay permit has unique elements such as an annual nutrient limit as opposed to a shorter-term limit due to the inherent differences between nutrients and other pollutants.
- Montana DEQ is in the process of developing compliance assessment tools for surface water quality criteria which may be used by MT point source permit writers to determine if nutrient limits are exceeded. EPA's TRC methodology may also be used to determine if nutrient limits are exceeded.

Specific steps can be taken to forward the process by which nutrient effluent limits are set and enforced. The following is a list of options that MT may consider in developing a protocol for setting nutrient limits and establishing a compliance framework. A recommended approach is described at the end.

3.1 Options for Establishing Point-source Effluent Limits for Nutrients

The options presented below potentially could be used to establish nutrient effluent limits for permitted point source dischargers in Montana. Each of these options would require further information gathering and development to implement.

Limits Option 1 – Technology-based Effluent Limits. Identify technology-based effluent limits achievable by available nutrient treatment technologies for wastewater discharges. This may be done using ELGs where available or best professional judgment. In addition, some information on achievable, technology-based concentrations was summarized in a previous report prepared for MT DEQ (ICF 2007). Next, use a probabilistic effluent discharge model to determine whether achievable effluent concentrations would meet instream water quality standards for the receiving water body (i.e., draft nutrient criteria). For example, EPA's Probabilistic Distribution Model (PDM)[4,5] estimates the number of days during a specified release period (i.e., annual, summer season) when instream conditions would be expected to exceed specified target concentrations. PDM could be used to determine whether the technology-based effluent concentrations would result in exceedances on no more than a specified percentage of days during the specified release period. Similarly, a model capable of providing a distribution of instream concentrations corresponding to technology-based effluent limits could be tested for compliance using the statistical compliance determination tools discussed in Section 6 of the main report.

Limits Option 2 – Back-calculated Water-Quality Based Effluent Limits. Identify the back-calculation models used by permit writers to determine effluent levels based on instream criteria. Determine limitations of using these models for nutrients and whether these limitations can be addressed.

[4] For a given number of days of discharge, PDM estimates the number of days during which the instream chemical concentration is expected to exceed a specified concentration of concern. PDM counts a day as having an exceedance of a concentration of concern if the concentration is exceeded for any part of a 24-hour day.
[5] PDM is a component within USEPA's Exposure and Fate Assessment Screening Tool Version 2.0 (E-FAST2), which is available from USEPA at http://www.epa.gov/opptintr/exposure/pubs/efastdl.htm and documentation is available at http://www.epa.gov/opptintr/exposure/pubs/efast2man.pdf.

Limits Option 3 – Forward-calculated Water-Quality Based Effluent Limits, Maximum Days with Exceedances. Use PDM to determine effluent limits corresponding to water quality standards. PDM could be run with a range of hypothetical effluent limits and a specified release period (e.g., the summer growing season) until a limit is identified that would be expected to result in the applicable surface water quality standard being exceeded on no more than a specified percentage of the release days.

Limits Option 4 – Forward-calculated Water-Quality Based Effluent Limits, Statistical Compliance Tools. Use a probabilistic effluent dilution model (to be identified) to estimate the downstream concentration associated with a range of potential end-of-the-pipe effluent limits. The distributions of estimated downstream concentrations predicted with each potential effluent limit would then be evaluated with the statistical compliance determination tools (i.e., the binomial test and the student t-test) discussed in Section 6 of the main report. The compliance tools would identify maximum potential effluent limit that resulted in complaint downstream concentrations

Limits Option 5 – Cumulative Nutrient Load Standards. Determine an annual or seasonal cumulative permissible nutrient load for specific water bodies. The methodology used by Chesapeake Bay permit writers to establish an annual limit for the Chesapeake Bay watershed could serve as a guideline for Montana. This option would be most applicable to point-source discharges going to lakes or to point-source discharges strongly influence a downstream standing waterbody like a lake.

3.2 Compliance Decision-making Options

Compliance Option 1 – Require daily effluent sampling and monthly averaging as required for other pollutants. Use TRC of 1.4 and consider a facility to be in violation if the limit is exceeded when the measurements multiplied by the TRC exceed the limit two out of six months or if the measurements alone exceed the limit four out of six months.

Compliance Option 2 – Require daily effluent sampling and monthly averaging as required for other pollutants. Define the required level of confidence required to establish compliance (e.g., 95%) as well as the percent of exceedances acceptable to maintain compliance (e.g., 10%). Use the binomial tool to determine if measurements meet the established compliance definition.

Compliance Option 3 – If a cumulative seasonal or annual nutrient load standard (Limits Option 5 above) is established, use a monitoring, reporting, and compliance system based on the one developed for the Chesapeake Bay.

3.3 Recommended Approach

With regard to setting point-source effluent limits for nutrients, the recommended option is one that is based on MT DEQ's draft water quality criteria for nutrients. Specifically, Limits Options 2, 3, and 4 are recommended for more detailed investigation. The draft water quality criteria would provide a technically-defensible and highly location-specific starting point for establishing

the effluent limits. In addition, these options would employ existing, EPA-accepted models, such as EPA's PDM tool. If it is determined that an averaging time greater than one month is warranted for a specific water body, Limits Option 6 should be considered.

With regard to compliance decision-making, the recommended approach is to compare results from the both the TRC methodology (i.e., Compliance Option 1) and the binomial tool (i.e., Compliance Option 2). If both methodologies suggest comparable results regarding which facilities remain in compliance, MT DEQ should choose to use the most cost-efficient tool. If the methodologies yield different results, ICF recommends that MT DEQ use the binomial tool as it allows the user greater flexibility in defining the terms of compliance.

4.0 References

40 CFR 123.45. Noncompliance and program reporting by the Director. 40 CFR Ch. 1 7-1-04 Edition. pp. 258-260.

CBA 1983. Chesapeake Bay Agreement. Available at: http://www.chesapeakebay.net/content/publications/cbp_12512.pdf.

CBA 1982. Chesapeake Bay Agreement: 1992 Amendments. Available at: http://www.chesapeakebay.net/content/publications/cbp_12507.pdf.

CBW 2007. Chesapeake Bay Watershed General Permit Case Study. Permit Effective Date: January 1, 2007. Available at: http://www.theexchangeassociation.org/Xchangewebdocs/News_Articles/5.%20EPA%20Case%20Study%20-%20Watershed%20Based%20Permitting%20in%20Chesapeake%20Bay.pdf.

Chesapeake 2000. Available at: http://www.chesapeakebay.net/content/publications/cbp_12081.pdf.

CWA 1972. Clean Water Act, a.k.a. Federal Water Pollution Control Amendments of 1972. 33 U.S.C. 1251 - 1376. Available at: http://epw.senate.gov/water.pdf.

Hair 2008. Personal Communication with David Hair, United States Environmental Protection Agency, Washington, D.C.

ICF 2007. Wastewater Treatment Performance and Cost Data to Support an Affordability Analysis for Water Quality Standard. Prepared for Montana Department of Environmental Quality, Helena, Montana. Prepared By ICF International, Lexington, Massachusetts, May 31, 2007.

McFadden 2008. Personal Communication with Angela McFadden, United States Environmental Protection Agency, Region 3.

Reid 2008. Personal Communication with Tom Reid, Montana Department of Environmental Quality. Water Permit Section.

U.S. EPA 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and their Uses. United States Environmental Protection Agency, Office of Research and Development.

U.S. EPA 1991. Technical Support Document for Water Quality-based Toxics Control. United States Environmental Protection Agency, Office of Water. EPA/505/2-90-001. Washington, D.C.

U.S. EPA, 1996. U.S. EPA NPDES Permit Writers' Manual. United States Environmental Protection Agency, Office of Water. EPA-833-B-96-003. Washington, D.C.

U.S. EPA 1997. Establishing Site Specific Aquatic Life Criteria Equal to Natural Background. United States Environmental Protection Agency, Office of Water.

U.S. EPA 1998. National Strategy for the Development of Regional Nutrient Criteria. United States Environmental Protection Agency, Office of Water. EPA 822-R-98-002. Washington, D.C.

U.S. EPA 2004. Annual Permit Limits for Nitrogen and Phosphorus for Permits Designed to Protect Chesapeake Bay and its Tidal Tributaries from Excess Nutrient Loading under the National Pollutant Discharge Elimination System. United States Environmental Protection Agency, Office of Water. Washington, DC. Available at: http://www.epa.gov/reg3wapd/npdes/pdf/ches_bay_nutrients_hanlon.pdf.

U.S. EPA 2006. Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act. United States Environmental Protection Agency, Office of Water, Watershed Branch.

U.S. EPA 2008. Nutrient Criteria Implementation. United States Environmental Protection Agency, Office of Water. Current as of 2008. Available at: http://cfpub.epa.gov/npdes/wqbasedpermitting/wspermitting.cfm#casestudy.