Appendix 5: Data assessment document

Important Information for Users of North Carolina Ambient Water Quality Monitoring Data, version 2.4

Important Information for Users of North Carolina Ambient Water Quality Monitoring Data

North Carolina operates one of the most extensive surface water quality monitoring networks in the nation. Surface water quality data have been collected for over 40 years. The current Ambient Monitoring System (AMS) network collects monthly samples at 323 locations across the state. The historical water quality database currently contains 5.3 million individual results. Prior to 1998, data were stored in the U.S. Environmental Protection Agency's national STORET (Legacy) database. The North Carolina Division of Water Quality currently maintains all data including those previously stored in EPA's Legacy STORET database using micro-computer databases and files. These databases only include information collected by the NC Division of Water Quality ambient surface water monitoring program. Since 2003, all historic and current data have also been loaded into the modern STORET database and are available for online quering and downloading at http://www.epa.gov/storet/dw_home.html.

Given the enormous size of the database and the long period of time which is presented, changes in technology, staff changes, changes in analytical methods, errors can be encountered. Errors may include instances where there is more than one result for a Station, Date, Depth, and Parameter (i.e. "duplicate" entries). These types of errors are known to occur in the data prior to 1998. After 1997, known data errors have been reduced because of improved error checking. However, data screening and error checking should always be employed by the end user prior to any quantitative summaries or analyses.

Data Format

Data are stored in a format that is succinct. The number of fields (columns) is minimized by having a single field (column) represent all parameters for which we have results or remarks. A small example of the file format is provided in the following table (the first row represents the name of the fields). The example file shows six fields – Station_Code, Date, Depth, Method_Code, Result, and Remark. The complete database has additional fields for collector, comments, and other information. Note the "stacked" arrangement of the data. That is, each record represents one result. A group of lines that have the same Station_Code and Date represent the results for one sampling event. A sampling event is a group of results – e.g. temperature, dissolved oxygen, ammonia-nitrogen, etc. measured during a single station visit. One sampling event can have multiple groups of measurements and/or samples, each differentiated by the depth.

Table 1. Format of data (Result) file.

STATION_CODE	DATE	DEPTH	METHOD_CODE	RESULT	RMK
A1510000	05/09/2000	0.1	10	21	
A1510000	05/09/2000	0.1	94	38	
A1510000	05/09/2000	0.1	610	0.01	K
A1510000	05/09/2000	0.1	625	0.3	
A1510000	05/09/2000	0.1	630	0.03	
A1510000	05/09/2000	0.1	665	0.01	
A1510000	05/09/2000	0.1	900	8	
A1510000	05/09/2000	0.1	82079	4.2	
A1520000	01/13/1997	0.1	10	5	
A1520000	01/13/1997	0.1	300	12.6	
A1520000	01/13/1997	0.1	400	7.4	
A1520000	01/13/1997	0.1	1002	10	K
A1520000	01/13/1997	0.1	1027	2	K
A1520000	01/13/1997	0.1	1034	25	K
A1520000	01/13/1997	0.1	1067	10	K
A1520000	01/13/1997	0.1	71900	0.2	K
A1520000	01/13/1997	0.1	82079	3.6	

Field names are described on the following pages.

Interpreting the data format

The first line of data (not field names) in the table above is:

STATION_CODE	DATE	DEPTH	METHOD_CODE	RESULT	RMK
A1510000	05/09/2000	0.1	10	21	

This can be interpreted as "the surface (DEPTH =0.1 meters) water temperature (METHOD_CODE = 10) measured at station A1510000 in the Broad River Basin, on May 9, 2000 was 21°C.

A second example is the concentration of arsenic (METHOD_CODE = 1002) measured at station A1520000 on January 12, 1997 was below (RMK = K) 10 µg/L.

STATION_CODE	DATE	DEPTH	METHOD_CODE	RESULT	RMK
A1520000	01/13/1997	0.1	1002	10	K

Overview of the fields (Detailed information for some fields is found after this overview)

1. **STATION_CODE** Represents the sample location. Additional details are found below.

2. **DATE** Represents when the sample or measurement was taken.

3. **DEPTH** The Depth field indicates the water depth (meters) from which a sample was taken. Coastal and lake stations may have measurements from many depths including the surface. Surface measurements generally have a depth of "0" or "0.1 meters."

4. **METHOD_CODE** is used to distinguish the parameters in the database instead of names. A separate table is used to link Method Codes with the appropriate parameter name. This table includes the measurement units for the result. Additional details are found below.

5. **RESULT** The numerical value of the parameter that has been measured or observed (Results can be continuous or nominal depending on the parameter).

6. **RMK** (Remark Code) Results may have additional associated information that must be taken into account during any interpretation. The most common example is when the value is know to be less than the value reported--the value reported being the lower limit of the laboratory analytical reporting level. Additional details are found below.

Station Codes

All stations are usually defined by an eight character code (e.g. B4000000, Q4120000). The first character is a letter that indicates the major river basin in North Carolina. The remaining 7 characters are numbers; generally numbered from upstream (smaller numbers) to downstream (larger numbers). Sometimes these station codes are referred to as "STORET" codes.

Table 2. Links to additional information.

Information Type	URL
Watersheds (basins) in N.C.	http://www.eenorthcarolina.org/public/ecoaddress/riverbasins/riverbasinmapinteractive.htm
Basin maps	http://portal.ncdenr.org/web/wq/ps/bpu
Current Active station locations:	http://portal.ncdenr.org/web/wq/ess/eco/ams
Basin assessment reports	http://portal.ncdenr.org/web/wq/ess/reports
Basin management plans	http://portal.ncdenr.org/web/wq/ps/bpu/basin

Table 3. North Carolina river basins and the prefix code used for station numbers.

North Carolina River Basin	Station Code Prefix	North Carolina River Basin	Station Code Prefix
Broad	A	New	K
Cape Fear	B	Pasquotank	M
Catawba	C	Roanoke	N
Chowan	D	Savannah	H
French Broad	E	Tar-Pamlico	O
Hiwassee	F	Watauga	L
Little Tennessee	G	White Oak	P
Lumber	I	Yadkin	Q
Neuse	J		

Method Codes

Method codes are a numerical code used to define the parameters measured. They have been inherited from the Legacy STORET system. For example, a method code of 300 is translated as Dissolved Oxygen measured in mg/L. A text file containing a key to all the method codes is available, however the most common are found in Table 4.

Table 4. Method codes for commonly sampled parameters.

Parameter	Method Code
Fecal Coliform (# colonies/100 ml)	31616
Turbidity (NTU)	82079
Solids: Total Suspended (mg/L)	530
Total Dissolved (mg/L)	70300
Chlorophyll *a* (µg/L)	70953
NH3 as N (mg/L)	610
TKN as N (mg/L)	625
NO2 plus NO3 as N (mg/L)	630
P: Total as P (mg/L)	665
Color: True (c.u.)	80
(pH) (c.u.)	83
pH 7.6 (c.u.)	82
Chloride (mg/L)	940
Fluoride (mg/L)	951
Sulfate (mg/L)	945
Grease and Oil (mg/L)	556
Al- Aluminum, total (µg/L)	1105
As- Arsenic, total (µg/L)	1002
Cd- Cadmium, total (µg/L)	1027
Cr- Chromium, total (µg/L)	1034
Cu- Copper, total (µg/L)	1042
Fe- Iron, total (µg/L)	1045
Hg- Mercury, total (µg/L)	71900
Mn- Manganese, total (µg/L)	1055
Ni- Nickel ,total (µg/L)	1067
Pb- Lead, total (µg/L)	1051
Zn- Zinc, total (µg/L)	1092
Water Temperature (^{o}C)	10
Field- Specific Conductance (µS/cm)	94
Dissolved Oxygen- DO (mg/L)	300
Field- pH (s.u.)	400
Salinity (ppt)	480
Secchi Depth (m)	78
Air Temperature (^{o}C)	20
Cloud Cover %	32
Wind Velocity (mph)	35
Wind Direction-Degrees from North	36
Precipitation- (inches/day)	45

The Laboratory Section's submission form (DM-1) has additional (but not all) method codes

Remark Codes and Data Qualifiers

Often results have additional associated information that must be taken into account during any interpretation. The most common example is when the value is know to be less than the value reported--the value reported being the lower limit of the laboratory analytical reporting level. Historically, these Remark Codes were mandated and defined by the Environmental Protection Agency's Legacy STORET data management system. In March of 2001 the NC Division of Water Quality Laboratory Section implemented a revised suite of data qualifiers. The following two tables provide a key to the remark codes and data qualifiers. The lab revises data qualifiers on an as-needed basis, so their web site (http://portal.ncdenr.org/web/wq/lab/qualityassurance) should be consulted for the latest revision.

Table 5. Legacy STORET remark codes (used up to about March 2001).

Remark Code	Meaning
A	Value reported is the mean of two or more determinations.
B	Results based upon colony counts outside the acceptable range.
C	Value calculated. (Also see "$")
D	Indicates field measurement.
E	Indicates extra samples taken at composite stations.
F	In the case of species, F indicates female sex.
G	Value reported is the maximum of two or more determinations.
H	Value based on field kit determination; results may or may not be accurate.
J	Estimated value; value not accurate.
K	Actual value is known to be less than value given.
L	Actual value is known to be greater than value given.
M	Presence of material verified but not quantified. In the case of temperature or oxygen reduction potential, M indicates a negative value. In the case of species, M indicates male sex.
N	Presumptive evidence of presence of material.
O	Sampled, but analysis lost or not performed.
P	Too numerous to count.
Q	Sample held beyond normal holding time.
R	Significant rain in the past 48 hours.
S	Laboratory test.
T	Value reported is less than criteria of detection.
U	Indicates material was analyzed for, but not detected. In case of species, U indicates undetermined sex.
V	Indicates the analyte was detected in the sample and associated blank method.
X	Value is quasi vertically integrated sample.
Y	Laboratory analysis from unpreserved data may not be accurate.
Z	Too many colonies were present to count (TMTC), the numeric value represents the filtration volume.
NULL	No remark.
$	Calculated value. (also see "C")

NC DENR/DWQ CHEMISTRY LABORATORY DATA QUALIFIER CODES (11/01/2001)

http://portal.ncdenr.org/web/wq/lab/qualityassurance
Check this site for revisions and/or previous versions

The table below lists the data qualification codes used for analytical results reported by the NC DENR/DWQ Chemistry Laboratories located in Raleigh and Asheville. Data validation is accomplished through a series of checks and reviews that are intended to assure that the reported results are of a verifiable and acceptable quality. The reported value always precedes the data qualifier code. Combination Letter and number….

Table 6. Data qualifiers (initiated in March 2001).

SYMBOL	DEFINITION
A	Value reported is the mean (average) of two or more determinations. This code is to be used if the results of two or more discrete and separate samples are averaged. These samples shall have been processed and analyzed independently (e.g. field duplicates, different dilutions of the same sample).
B	Results based upon colony counts outside the acceptable range and should be used with caution. This code applies to microbiological tests and specifically to membrane filter (MF) colony counts. It is to be used if less than 100% sample was analyzed and the colony count is generated from a plate in which the number of coliform colonies exceeds the ideal ranges indicated by the method. These ideal ranges are defined in the method as: *Fecal coliform bacteria: 20-60 colonies Total coliform bacteria: 20-80 colonies* 1. Countable membranes with less than 20 colonies. Reported value is estimated or is a total of the counts on all filters reported per 100 ml. 2. Counts from all filters were zero. The value reported is based on the number of colonies per 100 ml that would have been reported if there had been one colony on the filter representing the largest filtration volume (reported as a less than "<" value). 3. Countable membranes with more than 60 or 80 colonies. The value reported is calculated using the count from the smallest volume filtered and reported as a greater than ">" value. 4. Filters have counts of both >60 or 80 and <20. Reported value is a total of the counts from all countable filters reported per 100 ml. 5. Too many colonies were present; too numerous to count (TNTC), the numeric value represents the maximum number of counts typically accepted on a filter membrane (60 for fecal and 80 for total), multiplied by 100 and then divided by the smallest filtration volume analyzed. This number is reported as a greater than value. 6. Estimated Value. Blank contamination evident. 7. Many non-coliform colonies or interfering non-coliform growth present. In this competitive situation, the reported coliform value may under-represent actual coliform density. <u>Note</u>: A "B" value shall be accompanied by justification for its use denoted by the numbers listed above (ex. B1, B2, etc.)

Table 6 continued. Data qualifiers

SYMBOL	DEFINITION
C	Total residual chlorine was present in sample upon receipt in the laboratory; value not accurate (cyanide, phenol, NH3, TKN, coliform, organics)
G	A single quality control failure occurred during biochemical oxygen demand (BOD) analysis. The sample results should be used with caution. 1. The dissolved oxygen (DO) depletion of the dilution water blank exceeded 0.2 mg/L. 2. The bacterial seed controls did not meet the requirement of a DO depletion of at least 2.0 mg/L and/or a DO residual of at least 1.0 mg/L. 3. No sample dilution met the requirement of a DO depletion of at least 2.0 mg/L and/or a DO residual of at least 1.0 mg/L. 4. Evidence of toxicity was present. This is generally characterized by a significant increase in the BOD value as the sample concentration decreases. 5. The glucose/glutamic acid standard exceeded the range of 198 ± 30.5 mg/L. 6. The calculated seed correction exceeded the range of 0.6 to 1.0 mg/L. 7. Less than 1 mg/L DO remained for all dilutions set. The reported value is an **estimated** greater than value and is calculated for the dilution using the least amount of sample. 8. Oxygen usage is less than 2 mg/L for all dilutions set. The reported value is an **estimated** less than value and is calculated for the dilution using the most amount of sample. 9. The DO depletion of the dilution water blank produced a negative value. Note: A "G" value shall be accompanied by justification for its use denoted by the numbers listed above (ex. G1, G2, etc.)
J	**Estimated** value; value may not be accurate. This code is to be used in the following instances: 1. surrogate recovery limits have been exceeded; 2. the reported value failed to meet the established quality control criteria for either precision or accuracy; 3. the sample matrix interfered with the ability to make any accurate determination; or 4. the data is questionable because of improper laboratory or field protocols (e.g. composite sample was collected instead of grab, plastic instead of glass container, etc.). 5. temperature limits exceeded (samples frozen or >6° C) during transport, non-reportable for NPDES compliance monitoring. 6. the laboratory analysis was from an unpreserved or improperly chemically preserved sample. The data may not be accurate. 7. This qualifier is used to identify analyte concentration exceeding the upper calibration range of the analytical instrument/method. The reported value should be considered estimated. 8. Temperature limits exceeded (samples frozen or >6º C) during storage, the data may not be accurate. Note: A "J" value shall be accompanied by justification for its use denoted by the numbers listed above (ex. J1, J2, etc.). A "J" value shall not be used if another code applies (ex. N, V, M).
M	Sample and duplicate results are "out of control". The sample is non-homogenous (e.g. VOA soil). The reported value is the lower value of duplicate analyses of a sample.

Table 6 continued. Data qualifiers

SYMBOL	DEFINITION
N	Presumptive evidence of presence of material; **estimated** value. This code is to be used if: 1. The component has been tentatively identified based on mass spectral library search; 2. There is an indication that the analyte is present, but quality control requirements for confirmation were not met (i.e., presence of analyte was not confirmed by alternate procedures). 3. This code shall be used if the level is too low to permit accurate quantification, but the **estimated** concentration is less than the laboratory practical quantitation limit and greater than the laboratory method detection limit. *This code is not routinely used for most analyses.*
P	Elevated PQL* due to matrix interference and/or sample dilution.
Q	Holding time exceeded. These codes shall be used if the value is derived from a sample that was received, prepared and/or analyzed after the approved holding time restrictions for sample preparation and analysis. 1. Holding time exceeded prior to receipt by lab 2. Holding time exceeded following receipt by lab
S	Not enough sample provided to prepare and/or analyze a method-required matrix spike (MS) and/or duplicate (MSD).
U	Indicates that the analyte was analyzed for but not detected above the reported practical quantitation limit*. The number value reported with the "U" qualifier is equal to the laboratory's practical quantitation limit*.
X	Sample not analyzed for this constituent 1. Sample not screened for this compound. 2. Sampled, but analysis lost or not performed-field error 3. Sampled, but analysis lost or not performed-lab error Note: an "X" value shall be accompanied by justification for its use by the numbers listed.
V	Indicates the analyte was detected in both the sample and the associated method blank. Note: The value in the blank shall not be subtracted from the associated samples.
Z	The sample analysis/results are not reported due to : 1. Inablility to analyze the sample. 2. Questions concerning data reliability. The presence or absence of the analyte cannot be verified.
Y	Elevated PQL* due to insufficient sample size

***PQL** The Practical Quantitation Limit (PQL) is defined and proposed as "the lowest level achievable among laboratories within specified limits during routine laboratory operation". The PQL is about three to five times the calculated Method Detection Limit (MDL) and represents a practical and routinely achievable detection limit with a relatively good certainty that any reported value is reliable".

March 10, 2011/dbs

Comments on Specific Parameters

Nutrients

In early 2001 the NC Division of Water Quality Laboratory Section reviewed their internal QA/QC programs and some of their analytical methods. This effort resulted in a marked increase in reporting levels for certain parameters. New analytical equipment and methods were subsequently acquired to establish new lower reporting levels and more scientifically supportable quality assurance. As a result, the reporting levels quickly dropped back down to or near the previous reporting levels. Nutrients were especially affected by these changes, as shown in the following table:

Table 7. Changes in reporting levels for nutrients during 2001.

		Reporting level by date (all values are mg/L)			
Parameter	Method Code	Pre-2001	3/13/2001 to 3/29/2001	3/30/2001 to 7/24/2001	7/25/2001 to present
NH_3	610	0.01	0.5	0.2	0.01
TKN	625	0.1	1.0	0.6	0.20
NO_2+NO_3	630	0.01	0.5	0.15	0.01
TP	665	0.01	0.5	0.1	0.02

Note:
Pay special attention to remark codes ("K" and "U" can be interpreted as "<" /non-detect) and don't let increased reporting levels be interpreted as a sudden upward trend. The Laboratory Section cautions that the establishment of minimum reporting levels may have been inconsistent and undocumented prior to those established in July 2001. An excellent discussion on method detection levels and considerations for interpretations of water quality data is provided by the US Geological Survey[1].

Chlorophyll *a*

As a result of a number of quality assurance reviews, instrument evaluations and contributions from the scientific community in the state, the NC Division of Water Quality Laboratory Section revised chlorophyll data for samples analyzed from 1996 forward. Values reported prior to 1996 using spectrophotometric methods are believed to be accurate. The revised values from 1996 to 2001 were reported as uncorrected chlorophyll (i.e. uncorrected for pheophytin). Corrected chlorophyll values are unavailable for all analyses performed from 1996 through early 2001. Since early 2001 the Laboratory has used new equipment and methods to determine chlorophyll *a* and have been reporting this change under a new method code (70953) to differentiate it from the previous analytical method (32209). Historically, the laboratory reported 3 different values: uncorrected chlorophyll *a* (32217), pheophytin (32213), and corrected chlorophyll a (corrected for pheophytin; 32209). The new analytical method compensates for pheophytin and chlorophyll *b* interferences so now only chlorophyll *a* is reported; correction is not necessary.

In November 2005, the NC DWQ Laboratory Section identified a problem with chlorophyll *a* analyses preformed from April 11, 2005 through August 23, 2005. Standard operating procedures per the method were not followed during analysis of the samples. Therefore, chlorophyll *a* data for this time period are not available.

Zinc

Metal samples collected between April 1995 and April 1999 may have been contaminated with zinc. Results for zinc during this period are high and caution is warranted if these results are used.

[1] Oblinger Childress, Carolyn J., William T. Foreman, Brooke F. Conner, and Thomas J. Mahoney. 1999. New Reporting Procedures Based on Long-Term Method Detection Levels and Some Considerations for Interpretations of Water Quality Data Provided by the U.S. Geological Survey National Water Quality Laboratory. U.S. Geological Survey, Open-File Report 99-123; (http://water.usgs.gov/owq/OFR_99-193/ofr99_193.pdf)