Important Information for Users of NC DWR Ambient Water Quality Monitoring Data

August 2017, version 2.6

Introduction

The North Carolina Division of Water Resources (DWR) 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 329 locations across the state. The historical ambient water quality data set contains over six million individual results. Prior to 1998, data were stored in the U.S. Environmental Protection Agency's (USEPA) national STORET Legacy database. Since then, all historic and current data have been loaded into the modernized STORET database and are available for online querying and downloading at https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange. DWR maintains all data collected by the AMS surface water monitoring program using electronic databases and files, and is in the process of modernizing its data management system.

Data Format

Data are stored in a stacked (i.e. each record represents one result) format. The number of fields (columns) is minimized by having a single field (column) populated with all parameters for which AMS data are analyzed. An example of the file format is provided in Table 1, showing seven fields: LOCATION_CODE, COLLECTION_DATE, SAMPLE_DEPTH (in meters), ANALYTE_NAME, COMBINATION_RESULT (may be continuous, ordinal, or nominal, depending on the parameter), ANALYSIS_UNIT, and QUALIFIER. The complete database has additional fields for time of collection, collector name, comments, and other information. A group of records that have the same LOCATION_CODE and COLLECTION_DATE represent the results for one sampling event, which is a group of results (e.g., water temperature, dissolved oxygen, fecal coliform, and other parameters) measured during a single station visit. One sampling event can have multiple groups of samples and measurements differentiated by the depth of sample collection.

Interpreting the data format

The first line of data (not field names) in Table 1 can be interpreted as "The surface (depth = 0.1 meter) fecal coliform concentration measured at station B9050000 in the Cape Fear River Basin on March 20, 2014 was 37 CFU/100 mL. As indicated by the Q1 qualifier, the sample holding time was exceeded prior to receipt at the laboratory."

Additional details for selected fields are included below.

Table 1. General format of AMS databases.

LOCATION _CODE	COLLECTION _DATE	SAMPLE _DEPTH	ANALYTE_NAME	COMBINATION _RESULT	ANALYSIS _UNIT	QUALIFIER
B9050000	20-Mar-14	0.1	Coliform, MF Fecal in liquid	37	CFU/100ml	Q1
B9050000	20-Mar-14	5	Dissolved oxygen (DO)	9.7	mg/l	
B9050000	20-Mar-14	0.1	Dissolved oxygen (DO)	9.9	mg/l	
B9050000	20-Mar-14	10	Dissolved oxygen (DO)	9.7	mg/l	
B9050000	20-Mar-14	5	pH	6.8	None	
B9050000	20-Mar-14	10	pH	6.8	None	
B9050000	20-Mar-14	0.1	pH	6.7	None	
B9050000	20-Mar-14	10	Salinity	0.04	ppth	
B9050000	20-Mar-14	5	Salinity	0.04	ppth	
B9050000	20-Mar-14	0.1	Salinity	0.04	ppth	
B9050000	20-Mar-14	0.1	Turbidity	18	NTU	
B9050000	24-Apr-14	0.1	Coliform, MF Fecal in liquid	49	CFU/100ml	B7, Q1
B9050000	24-Apr-14	4.5	Dissolved oxygen (DO)	6.7	mg/l	
B9050000	24-Apr-14	0.1	Dissolved oxygen (DO)	6.7	mg/l	
B9050000	24-Apr-14	9	Dissolved oxygen (DO)	6.7	mg/l	
B9050000	24-Apr-14	9	pH	6.2	None	
B9050000	24-Apr-14	4.5	pH	6.3	None	
B9050000	24-Apr-14	0.1	pH	6.1	None	
B9050000	24-Apr-14	0.1	Salinity	0.03	ppth	
B9050000	24-Apr-14	4.5	Salinity	0.03	ppth	
B9050000	24-Apr-14	9	Salinity	0.03	ppth	
B9050000	24-Apr-14	0.1	Turbidity	14	NTU	

Location Codes

AMS 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 location codes are referred to as "STORET" codes since they are also used to identify AMS stations in the STORET data warehouse.

Table 2. North Carolina river basins and associated prefixes used for AMS station location codes.

NC River Basin	Station Code Prefix	NC 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		

Table 3. Links to additional information related to AMS data and stations.

INFORMATION	URL
River Basins; Water Resources Plans	http://deq.nc.gov/about/divisions/water-resources/planning/basin-planning
Current AMS Station Locations	http://deq.nc.gov/about/divisions/water-resources/water-resources-data/water-sciences-home-page/ecosystems-branch/ambient-monitoring-system
Ambient Monitoring Reports	http://deq.nc.gov/about/divisions/water-resources/water-resources-data/water-sciences-home-page/reports-publications-data

Remark Codes and Data Qualifiers

Often results have additional associated information that must be considered during any interpretation. The most common example is a non-detect: when the value is known to be less than the value reported (i.e., the value reported is the lower limit of the laboratory analytical reporting level). Historically, these Remark Codes were mandated and defined by the STORET data management system. In March 2001, the DWR Laboratory implemented a revised suite of data qualifiers. The following two tables provide a key to remark codes and data qualifiers.

Table 4. Legacy STORET remark codes (used until approximately 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.

Remark Code	Meaning
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")

Data validation is accomplished through a series of checks and reviews intended to assure that the reported results are of a verifiable and acceptable quality. The reported value always precedes the data qualifier code in DWR Laboratory results reporting. Table 5 lists the data qualification codes used for analytical results reported by the NC Department of Environmental Quality (DEQ) and DWR Chemistry Laboratories located in Raleigh and Asheville. The DWR Laboratory revises data qualifiers on an as-needed basis, so the web site should be consulted for the latest revision at https://deq.nc.gov/about/divisions/water-resources/water-resources-data/water-sciences-home-page/microbiology-inorganics-branch/methods-pqls-qa.

Table 5. NC Division of Water Resources Laboratory data qualifier codes (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). This code is not required for BOD, coliform, or acute/chronic metals reporting since averaging multiple results for these parameters is fundamental to those methods or manner of reporting. 1. The reported value is an average, where at least one result is qualified with a "U". The PQL is used for the qualified result(s) to calculate the average.
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 colonies exceeds the ideal ranges indicated by the method. These ideal ranges are defined in the method as: *Fecal coliform or Enterococcus 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 estimated or is a total of the counts on all filters reported per 100 ml. 5. Too many colonies were present; too numerous to count (TNTC). TNTC is generally defined as >150 colonies. The numeric value represents the maximum number of counts typically accepted on a filter membrane (60 for fecal or enterococcus 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 or non-enterococcus colonies or interfering non-coliform or non-enterococcus growth present. In this competitive situation, the reported value may under-represent actual density. Note: A "B" value shall be accompanied by justification for its use denoted by the numbers listed above (e.g., B1, B2, etc.). Note: A "J2" should be used for spiking failures.

Symbol	Definition
BB	This code applies to **most probable number (MPN)** microbiological tests. 1. No wells or tubes gave a positive reaction. Value based upon the appropriate MPN Index and reported as a less than "<" value. 2. All wells or tubes gave positive reactions. Value based upon the MPN Index and reported as a greater than ">" value. Note: A "BB" value shall be accompanied by justification for its use denoted by the numbers listed above (e.g., BB1, BB2, etc.).
C	Total residual chlorine was present in sample upon receipt in the laboratory; value is **estimated**. Generally applies to cyanide, phenol, NH3, TKN, coliform, and 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. The reported value is calculated from the highest dilution representing the maximum loading potential and should be considered an **estimated** value. 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 (e.g., 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. 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 or not verifiable (e.g., no temperature blank provided): 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 >6ºC) during storage, the data may not be accurate. 9. The reported value is determined by a **one-point estimation** rather than against a regression equation. The estimated concentration is less than the laboratory practical quantitation limit and greater than the laboratory method detection limit. 10. Unidentified peak; estimated value. 11. The reported value is determined by a **one-point estimation** rather than against a regression equation. The estimated concentration is less than the laboratory practical quantitation limit and greater than the instrument noise level. *This code is used when an MDL has not been established for the analyte in question.* 12. The calibration verification did not meet the calibration acceptance criterion for **field parameters**. Note: A "J" value shall be accompanied by justification for its use denoted by the numbers listed above (e.g., J1, J2, etc.). A "J" value shall not be used if another code applies (e.g., 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.
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.

Symbol	Definition
	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.* 4. 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 instrument noise level. *This code is used when an MDL has not been established for the analyte in question.* 5. The component has been tentatively identified based on a retention time standard.
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. The value does not meet NPDES requirements. 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 matrix spike 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*.
V	Indicates the analyte was detected in both the sample and the associated blank. Note: The value in the blank shall not be subtracted from the associated samples. 1. The analyte was detected in both the sample and the method blank. 2. The analyte was detected in both the sample and the field blank.
X	Sample not analyzed for this constituent. This code is to be used if: 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.
Y	Elevated PQL* due to insufficient sample size.
Z	The sample analysis/results are not reported due to: 1. Inability to analyze the sample. 2. Questions concerning data reliability. The presence or absence of the analyte cannot be verified.
Supporting definitions listed below	
MDL	A Method Detection Limit (MDL) is defined as the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the true value is greater than zero and is determined in accordance with 40 CFR Part 136, Appendix B.
ML	Minimum Levels are used in some EPA methods. A Minimum Level (ML) is the lowest level at which the entire analytical system must give a recognizable signal and acceptable calibration point for the analyte. It is equivalent to the concentration of the lowest calibration standard, assuming that all method-specific sample weights, volumes, and cleanup procedures have been employed. The ML is calculated by multiplying the MDL by 3.18 and rounding the result to the nearest factor of 10 multiple (i.e. 1, 2, or 5). For example, MDL = 1.4 mg/L x 3.18 = 4.45 rounded to the nearest factor of 10 multiple (i.e. 5) = 5.0 mg/L.
***PQL**	The Practical Quantitation Limit (PQL) is defined as the lowest concentration that can be reliably achieved within specified limits of precision and accuracy during routine laboratory operating conditions. PQLs are subjectively set at some multiple of typical MDLs for reagent water (generally 3 to 10 times the MDL depending upon the parameter or analyte and based on the analyst's best professional judgment, the quality and age of the instrument, and the nature of the samples) rather than explicitly determined. PQLs may be nominally chosen within these guidelines to simplify data reporting and, where applicable, are generally equal to the concentration of the lowest non-zero standard in the calibration curve. PQLs are adjusted for sample size, dilution, and % moisture. For parameters that are not amenable to MDL studies, the PQL may be defined by the sample volume and buret graduations for titrations or by the minimum measurement values set by the method for method-defined parameters (e.g. BOD requires a minimum DO depletion of 2.0 mg/L, fecal coliform requires a minimum plate count of 20 cfu, total suspended residue requires a minimum weight gain of 2.5 mg, etc.). Additionally, some EPA methods prescribe Minimum Levels (MLs) and the lab may set the PQL equal to this method-stated ML. Determination of PQL is fully described in the laboratory's analytical Standard Operating Procedure (SOP) document.

Potential for Errors

Given the enormous size of the databases and the long time period represented, as well as changes in technology, staff, and analytical methods, errors can be encountered in the AMS data set. Errors may include duplicate entries for which there is more than one result for a given station, date, depth, and parameter. These types of errors are known to occur in the data prior to 1998. After 1997, known data errors have been reduced by improved quality assurance practices. However, data screening and error checking should always be employed by the end user prior to any quantitative summaries or analyses.

Errors in interpretation can occur if qualifier codes and remarks are not considered. For example, results with K or U qualifiers are non-detects reported at the laboratory PQL, but can only be considered as less than the PQL. Changes in PQLs should not be interpreted as sudden upward or downward trends in parameter concentrations. The DWR Laboratory 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].

Comments on Specific Parameters

Nutrients

In early 2001 the DWR Laboratory Section reviewed their internal quality assurance/quality control (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 QA processes. As a result, the reporting levels quickly dropped back down to or near the previous reporting levels. Nutrients were especially effected by these changes, as shown in the following table.

	Reporting level by date (all values are mg/L)				
Parameter	Pre-2001	3/13/2001 to 3/29/2001	3/30/2001 to 7/24/2001	7/25/2001 to 7/2/2002	7/3/2002 to present
Ammonia	0.01	0.5	0.2	0.01	0.02
Total Kjeldahl Nitrogen	0.1	1.0	0.6	0.20	0.20
Nitrite + Nitrate	0.01	0.5	0.15	0.01	0.02
Total Phosphorus	0.01	0.5	0.1	0.02	0.02

Chlorophyll *a*

Historically, the DWR Laboratory reported 3 different chlorophyll values: uncorrected chlorophyll *a*, pheophytin, and corrected chlorophyll *a* (corrected for pheophytin). As a result of a number of QA/QC reviews, instrument evaluations, and contributions from the scientific community in the state, the DWR Laboratory 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

[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).

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*. The new analytical method compensates for pheophytin and chlorophyll *b* interferences so that only chlorophyll *a* is reported; correction is not necessary.

In November 2005, the NC DWR Laboratory identified a problem with chlorophyll *a* analyses performed 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.

Field Measurements

In November 2013, the NC DWR Environmental Sciences Section (now the Water Sciences Section) identified a problem with field meter calibrations performed in the Asheville and Fayetteville regional offices from April 26, 2010 to September 30, 2013 and April 26, 2010 to October 11, 2013, respectively. Standard operating procedures for the YSI Professional Plus field meter were not followed consistently, resulting in unknown data quality. Field measurements (i.e., dissolved oxygen, pH, water temperature, and specific conductance) for these respective time periods are not available for the two regions.