|  |
| --- |
| US EPA Office of Compliance |
| Water Quality Indicators (WQI) Project Background and Technical Specifications |
|  |

|  |
| --- |
| OC Interagency WQI Workgroup13 September 2017  |

Previous Versions and Description of Changes

2 June 2017: first published version

13 Sept. 2017:

Add most-recent-48-month statistics

Extend search for speciation(units2) to the MethodSpecificationName column in the Portals "Biological” Result Retrieval

Water Quality Indicators (WQI) Project Background and Technical Specifications

**I Background:**

The Water Quality Indicators Project was originally developed as a component of EPA’s “Clean Water Action Plan.” The Action Plan sought to better align inspections with areas of poor water quality – particularly focusing on enforcement at facilities that have violations and contribute to water impairment. After reviewing information about watersheds that have been designated as having poor water quality, it became clear that a large number of watersheds have not been assessed. EPA was concerned that shifting resources into impaired watersheds might miss important areas that have not been designated as impaired. To address this data gap, EPA designed the WQI to identify potential hotspots using water station monitoring data, that when compared to water quality criteria, could yield a secondary data source showing possible problem areas. The WQI set out to use ambient water quality data in EPA database STORET (STOrage and RETrieval of Water Quality Data) and USGS database NWIS (National Water Information System) to flag “hotspots” which could then be evaluated by interested EPA, state and tribal staff. The development of the Water Quality Indicators (WQI) project coincided with an initiative by the Office of Enforcement and Compliance Assurance (OECA) to begin using data analytics to integrate and assess large data sets for problem identification. Under the WQI project, EPA used data analytic methods to bring together and compare these data sources. After compiling a data set and establishing a data refresh process, EPA developed mapping and data visualization tools that will help users display and assess the data. To develop the methods and techniques needed to produce the WQI, EPA convened a workgroup that included representatives from EPA Regions 2, 4 and 10, OAR, ORD, OEI, OW, OECA, OPP and the U.S. Geological Survey, all of whose contributions are gratefully acknowledged.

**II Scope:**

It was decided that hotspot identification would consist of comparison of water quality measurements with ambient water quality criteria based on national standards as indexed within defined ecoregions. EPA recognizes that states and tribes have developed water quality criteria that may differ from national standards; however, given resources, it was beyond the reach of the initial Indicators project to attempt to identify and mobilize all the appropriate state criteria. Instead, EPA designed the user interface so that users can input their own concern levels, while making default settings correspond to the CWA 304(a) national criteria. Data collected from before 2009 would not be used, and the initial pollutants included in the project were limited to the nutrients nitrogen and phosphorus – a choice which would support analysis of many sectors.

**III Approach and Data Use:**

The combined NWIS and STORET ambient water quality monitoring data were used. The period of interest was set to eight years (yearbegin=2009, yearend=2016) – with the intention of eventually increasing it to the most recent ten years - and relevant statistics (percentiles, counts) for each month and over all the months [yearend-yearbegin + 1) \*12 months] in the period -- would be determined for all monitoring locations nationally by pollutant for which there were data. These results were then associated with a graphical user interface which would visually identify nationally locations where the water quality measurements fell into one of several contiguous percentage ranges, extending from below the criteria to above, as desired by the user. This interface was made available to government users for exploration during a trial period held from April 2016 to December 2016 via ECHO Gov and the EPA GeoPlatform, and EPA’s Office of Enforcement and Compliance Assurance (OECA) collected comments from participating users. WQI Version 1.0 was released for government use in February 2017, and OECA is continuing to hold the comment period open and is [looking for comments](https://echo.epa.gov/maps/wqimap/docdata) on the technical methods used, the GIS interface, the overall usefulness of the data and tools, and the potential future release of this information to the public. In 2017, EPA is working to expand the project to encompass additional pollutants, and expects to engage with interested government stakeholders on this expansion. [The current WQI data file can be downloaded](https://echo.epa.gov/maps/wqimap/docdata).

Vetting the data was a major undertaking that is still considered a work in progress -- involving identifying protocols for

recognizing detection limits (for non-detects); for rejecting poor quality measurement data; for repairing poor quality or ambiguous reporting (e.g. multiple meanings of “total”); for recognizing duplicate samples; and for establishing a protocol for combining the particular population of measurements of those fractions of nitrogen and phosphorus found in the data into total nitrogen and total phosphorus, the measures used in the 304(a) criteria.

STORET and NWIS measurement submissions have available an extensive set of metadata questions, which, when answered properly, assist greatly in determining which measurements to use for a given analysis and which not to use. When crucial metadata are incorrect or incomplete, measurements which would otherwise be usable may be deemed unusable, as was the case with about half the nitrogen and phosphorus measurements in the Version 1.0 of the WQI released in February 2017 ([see Portals File Processing Summaries for particular refresh details](https://echo.epa.gov/maps/wqimap/docdata)).

Some of the more crucial fields ([Portals field names](https://echo.epa.gov/maps/wqimap/docdata)) are: OrganizationIdentifier, MonitoringLocationIdentifier, ActivityIdentifier, ActivityStartDate, CharacteristicName, ResultSampleFractionText, ResultDetectionConditionText, ResultMeasureValue, ResultMeasure/MeasureUnitCode(units1 and units2), DetectionQuantitationLimitTypeName, DetectionQuantitationLimitTypeMeasure/MeasureValue, DetectionQuantitationLimitTypeMeasure/MeasureUnitCode(units1 and units2). Units1 is the concentration, units2 is the reference species (e.g., in ”.2 mg/l as N”, units1=”mg/l” and units2=”N”). Agencies with incomplete or inaccurate crucial metadata can improve their agency representation rate in the WQI by adding and/or correcting the metadata. The STORET data in the Portals are updated every week. The refresh rate for the WQI is initially anticipated monthly.

[Find URLs for alpha or beta test sites, if any are currently active.](https://echo.epa.gov/maps/wqimap/docdata)

**IV Source Data and Water Quality Criteria**

The NWIS(USGS), STORET(EPA) and STEWARDS (USDA) ambient monitoring data are merged by USGS and are available using what is called the [“Portals” Web service](https://echo.epa.gov/maps/wqimap/docdata). The Portals data are served up in two files: [a Site file and a Result file](https://echo.epa.gov/maps/wqimap/docdata). The element MonitoringStationIdentifier is the key that links the two files. There are multiple Result records (measurements) for one Site record.

The present effort focuses on nitrogen and phosphorus. Ideally the state water quality criteria for N and P should be used, but [few states have total N and total P criteria](https://echo.epa.gov/maps/wqimap/docdata), or whatever criteria they do have are [described idiosyncratically by each state sometimes in unstructured text](https://echo.epa.gov/maps/wqimap/docdata). Consequently, any effort to extract and normalize state criteria was tabled, and instead, the [CWA 304(a) national criteria](https://echo.epa.gov/maps/wqimap/docdata) are used, which for nutrients including N and P defer to [Aggregate Nutrient Ecoregional (ANE) criteria](https://echo.epa.gov/maps/wqimap/docdata) The ANE criteria were published by EPA in the early 2000s and “represent conditions of surface waters that have minimal impacts caused by human activities, starting points to identify more precise numeric levels for nutrient parameters needed to protect aquatic life, recreational, or other uses on site-specific or subregion-specific [conditions](https://echo.epa.gov/maps/wqimap/docdata)” There are 14 ANEs, covering the conterminous 48 states only. The ANE consist of aggregates of “old” Level III ecoregions (not to be confused with later Level III ecoregions).

The N and P criteria applicable to waterbodies fall into one of two categories -- river/stream or lake/reservoir -- and are available only when the nutrient ecoregion is known

**V Output**

WQ Indicators CSV file format – March 2016 – revised Nov 2016

For initial implementation a CSV format record was decided upon, consisting of these elements taken directly from the [Portals Site file](https://echo.epa.gov/maps/wqimap/docdata):

MonitoringLocationIdentifier

OrganizationFormalName(OrganizationIdentifier)

MonitoringLocationName

LatitudeMeasure

LongitudeMeasure

CEightDigitCode

StateCode

CountyCode

MonitoringLocationTypeName,

These elements derived by WQI Site file protocols:

WaterBodyType ( “R”[iver/Stream] or “L”[ake/Reservoir])

Ecoregion Identifier ( I, II, III…XIV,noeco)

These elements derived by Result file processing:

Up to four groups of 12-values per group. If the first value in a group is:

N: The pollutant is nitrogen and the interval for the summary statistics is the WQI entire period of record, which has a length in months of (yearend-yearbegin+1)\*12 (from Jan yearbegin to Dec yearend).

N48: The pollutant is nitrogen and the interval for the summary statistics is the 48-month period ending on the date of the Portals retrieval. (The entire period of record is assumed to be greater than 48 months.)

P: The pollutant is phosphorus and the interval for the summary statistics is the WQI entire period of record, which has a length in months of (yearend-yearbegin+1)\*12 (from Jan yearbegin to Dec yearend).

P48: The pollutant is phosphorus and the interval for the summary statistics is the 48-month period ending on the date of the Portals retrieval. (The entire period of record is assumed to be greater than 48 months.)

Each time one of the above markers (N, N48, P, P48) occurs, it will be followed by the following eleven fields:

Number of samples

Number of samples rejected because of incomplete metadata

Samples maximum

Samples median

Samples time-weighted median

Samples 70th percentile

Samples time-weighted 70th percentile

Samples 90th percentile

Samples time-weighted 90th percentile

Number of sample values greater than the default 304(a) criteria

Maximum of monthly medians

When a month (yymm) appears instead of a marker as the first value in a group, the month will be followed by the first ten of these eleven fields and the period of record is the month. (“Maximum of monthly medians” does not apply to monthly data)

Only the months with data appear in the record. The sequence of fields in the record is N, N48, P, P48, months in ascending yymm order.

Here is an example record with 44 used and 0 rejected N samples, and 55 used and 0 rejected P samples (the white space added for clarity is not in the record):

21ARIZ\_WQX-SRTON053.87,ARIZONA DEPARTMENT OF ENVIRONMENTAL QUALITY(21ARIZ\_WQX),TONTO CREEK - BELOW BEAR FLATS,34.2833333,-111.0708333,15060105,04,007,River/Stream,R,II,

N,44,0,1.000,0.370,0.370,0.433,0.430,0.575,0.575,43,0.520,N48,22,0,0.910,0.360,0.350,0.410,0.410,0.460,0.460,21,0.412,1305,5,0,0.700,0.520,0.520,0.564,0.575,0.650,0.700,5,1306,6,0,1.000,0.375,0.430,0.502,0.575,0.788,1.000,6,1307,5,0,0.540,0.400,0.400,0.456,0.470,0.512,0.540,5,1308,6,0,0.550,0.317,0.325,0.347,0.370,0.460,0.550,6,1309,6,0,0.910,0.412,0.430,0.445,0.460,0.685,0.910,6,1407,5,0,0.785,0.410,0.410,0.450,0.460,0.655,0.785,5,1408,6,0,0.370,0.265,0.270,0.285,0.300,0.335,0.370,5,1409,5,0,0.420,0.290,0.290,0.314,0.320,0.380,0.420,5,

P,55,0,4100.000,50.000,50.000,85.800,92.000,153.000,155.000,55,2125.000,P48,22,0,230.000,91.000,90.000,130.000,130.000,154.500,155.000,22,127.250,0903,1,0,17.500,17.500,17.500,17.500,17.500,17.500,17.500,1,0904,2,0,25.000,25.000,25.000,25.000,25.000,25.000,25.000,2,0908,1,0,78.000,78.000,78.000,78.000,78.000,78.000,78.000,1,0909,1,0,77.000,77.000,77.000,77.000,77.000,77.000,77.000,1,1007,2,0,4100.000,2125.000,4100.000,2915.000,4100.000,3705.000,4100.000,2,1008,1,0,25.000,25.000,25.000,25.000,25.000,25.000,25.000,1,1009,1,0,25.000,25.000,25.000,25.000,25.000,25.000,25.000,1,1106,1,0,25.000,25.000,25.000,25.000,25.000,25.000,25.000,1,1107,1,0,15.000,15.000,15.000,15.000,15.000,15.000,15.000,1,1109,2,0,50.000,50.000,50.000,50.000,50.000,50.000,50.000,2,1305,4,0,36.000,32.500,34.000,34.200,34.000,35.400,36.000,4,1306,5,0,35.000,32.500,32.500,32.900,33.000,34.200,35.000,5,1307,5,0,500.000,49.000,49.000,361.800,440.000,476.000,500.000,5,1308,6,0,87.000,59.000,68.000,73.750,79.500,83.250,87.000,6,1309,6,0,230.000,127.250,150.000,152.500,155.000,192.500,230.000,6,1407,5,0,150.000,50.000,50.000,130.000,150.000,150.000,150.000,5,1408,6,0,190.000,88.000,120.000,125.000,130.000,160.000,190.000,6,1409,5,0,130.000,90.000,90.000,91.600,92.000,114.800,130.000,5

**VI Site File Processing**

As noted in **V**, eight original elements are needed from the Site file and two additional elements, waterbody type and ecoregion ID, must be derived.

**VIA Waterbody Type**

For waterbody type, the Site file is read and an inventory of MonitoringLocationTypeName (the Site file element name for waterbody type) was made. There were about 120 different types. The two columns of this inventory (count, waterbody type) were put into the first two columns of a spreadsheet, and two additional columns were added on the right, one for whether or not the named type could legitimately be compared to the criteria, and, if so, the second column a place to indicate whether the river(R) or lake(L) criterion should be used (see Figure T).

|  |  |  |  |
| --- | --- | --- | --- |
| Count | MonitoringLocationTypeName | Use? | Criterion type: R=rivers/streams, L=lakes/reservoirs |
| 120 | (no type present) | No |  |
| 1,473 | Aggregate groundwater use | No |  |
| 1,248 | Aggregate surface-water-use | No |  |
| 5,196 | Atmosphere | No |  |
| 2 | BEACH Program Site-Channelized stream | Yes | R |
| 1,551 | BEACH Program Site-Estuary | No |  |
| 821 | BEACH Program Site-Great Lake | Yes | L |
| 444 | BEACH Program Site-Lake | Yes | L |
| 2 | BEACH Program Site-Land | No |  |
| 6 | BEACH Program Site-Land runoff | No |  |
| 4,195 | BEACH Program Site-Ocean | No |  |
| 241 | BEACH Program Site-River/Stream | Yes | R |
| 1 | BEACH Program Site-Waste sewer | No |  |
| 797 | Borehole | No |  |
| 3,317 | Canal Drainage | Yes | R |
| 550 | Canal Irrigation | Yes | R |
| 4,624 | Canal Transport | Yes | R |
| 540 | Cave | No |  |
| 7,525 | CERCLA Superfund Site | No |  |
| 561 | Channelized Stream | Yes | R |
| 17 | Combined Sewer | No |  |
| 105 | Constructed Wetland | Yes | R |
| 68,660 | Estuary | No |  |
| 792 | Facility Industrial | No |  |
| 1,759 | Facility Municipal Sewage (POTW) | No |  |
| 1,867 | Facility Other | No |  |
| 1,082 | Facility Privately Owned Non-industrial | No |  |
| 266 | Facility Public Water Supply (PWS) | No |  |
| 9 | Facility: Cistern | No |  |
| 146 | Facility: Combined sewer | No |  |
| 4,485 | Facility: Diversion | No |  |
| 182 | Facility: Field, Pasture, Orchard, or Nursery | No |  |
| 114 | Facility: Golf course | No |  |
| 217 | Facility: Laboratory or sample-preparation area | No |  |
| 25 | Facility: Landfill | No |  |
| 1,377 | Facility: Outfall | No |  |
| 172 | Facility: Pavement | No |  |
| 49 | Facility: Septic system | No |  |
| 245 | Facility: Storm sewer | No |  |
| 93 | Facility: Waste injection well | No |  |
| 169 | Facility: Wastewater land application | No |  |
| 105 | Facility: Wastewater sewer | No |  |
| 254 | Facility: Water-distribution system | No |  |
| 2,719 | Facility: Water-use establishment | No |  |
| 3 | Floodwater | Yes | R |
| 1 | Floodwater non-Urban | Yes | R |
| 2,458 | Floodwater Urban | Yes | R |
| 11 | Gallery | No |  |
| 14 | Glacier | No |  |
| 844 | Great Lake | Yes | L |
| 78,406 | Lake | Yes | L |
| 24,752 | Lake, Reservoir, Impoundment | Yes | L |
| 5,681 | Land | No |  |
| 632 | Land Flood Plain | No |  |
| 800 | Land Runoff | No |  |
| 275 | Land: Excavation | No |  |
| 487 | Land: Outcrop | No |  |
| 1 | Land: Playa | No |  |
| 41 | Land: Shore | No |  |
| 456 | Land: Sinkhole | No |  |
| 985 | Land: Soil hole | No |  |
| 1 | Land: Volcanic vent | No |  |
| 347 | Landfill | No |  |
| 60 | Leachate-Lysimeter | No |  |
| 3 | Local Air Monitoring Station | No |  |
| 293 | Mine/Mine Discharge | No |  |
| 304 | Mine/Mine Discharge Adit (Mine Entrance) | No |  |
| 228 | Mine/Mine Discharge Tailings Pile | No |  |
| 148 | Mine/Mine Discharge Waste Rock Pile | No |  |
| 17,746 | Ocean | No |  |
| 466 | Ocean: Coastal | No |  |
| 5,065 | Other-Ground Water | No |  |
| 2,126 | Other-Surface Water | Yes | R |
| 591 | Pipe, Unspecified Source | No |  |
| 54 | Pond-Anchialine | Yes | L |
| 86 | Pond-Stormwater | Yes | L |
| 3 | Pond-Wastewater | Yes | L |
| 11,833 | Reservoir | Yes | L |
| 308,728 | River/Stream | Yes | R |
| 41 | River/Stream Ephemeral | Yes | R |
| 222 | River/Stream Intermittent | Yes | R |
| 1,728 | River/Stream Perennial | Yes | R |
| 1,098 | Riverine Impoundment | Yes | R |
| 210 | Seep | No |  |
| 5 | Spigot / Faucet | No |  |
| 36,684 | Spring | Yes | R |
| 5 | State/Local Air Monitoring Station | No |  |
| 1,032 | Storm Sewer | No |  |
| 134,125 | Stream | Yes | R |
| 2,634 | Stream: Canal | Yes | R |
| 2,055 | Stream: Ditch | Yes | R |
| 169 | Stream: Tidal stream | Yes | R |
| 100 | Subsurface | No |  |
| 20 | Subsurface: Cave | No |  |
| 794 | Subsurface: Groundwater drain | No |  |
| 1,586 | Subsurface: Tunnel, shaft, or mine | No |  |
| 816 | Subsurface: Unsaturated zone | No |  |
| 292 | Survey Monument | No |  |
| 268 | Test Pit | No |  |
| 61 | Waste Pit | No |  |
| 31 | Waste Sewer | No |  |
| 1,427,116 | Well | No |  |
| 392 | Well: Collector or Ranney type well | No |  |
| 45 | Well: Extensometer well | No |  |
| 911 | Well: Hyporheic-zone well | No |  |
| 643 | Well: Interconnected wells | No |  |
| 1,195 | Well: Multiple wells | No |  |
| 26,589 | Well: Test hole not completed as a well | No |  |
| 675 | Wetland | Yes | R |
| 683 | Wetland Estuarine-Emergent | No |  |
| 30 | Wetland Estuarine-Forested | No |  |
| 18 | Wetland Estuarine-Scrub-Shrub | No |  |
| 306 | Wetland Lacustrine-Emergent | Yes | R |
| 2,621 | Wetland Palustrine-Emergent | Yes | R |
| 57 | Wetland Palustrine-Forested | Yes | R |
| 7 | Wetland Palustrine-Moss-Lichen | Yes | R |
| 31 | Wetland Palustrine-Shrub-Scrub | Yes | R |
| 152 | Wetland Riverine-Emergent | Yes | R |
| 644 | Wetland Undifferentiated | Yes | R |

Figure T. MonitoringLocationType Names Found in Portals Site Data and Decision on Use

The WQI workgroup manually filled in the two rightmost columns (whether the waterbody type should be used and, if “yes”, whether the river/stream or lake/reservoir criteria should be applied). Some types were obviously inapplicable because they are not ambient surface water locations, but some – such as estuaries – were excluded over uncertainty of criteria applicability.

**VIB Obtaining ANE for Each Monitoring Location**

Once the Portals Site data have been retrieved, to obtain ecoregion a point-in-polygon search of the appropriate (ANE) [shapefiles](https://echo.epa.gov/maps/wqimap/docdata) is performed, and monitoring locations from the survivors of the waterbody-type screen that also had latitude and longitude placing them outside any nutrient ecoregion are assigned to “ecoregion noeco”.

The 14 nutrient ecoregions cover only the conterminous US, so locations in Alaska, Hawaii, Guam, etc. are “noeco”. Offshore monitoring locations with lat/long outside the ecoregions polygon boundaries are also included among the “noecos”. (There has been some interest in considering these locations as if they were inside the ecoregion they are geographically nearest to, but this was not done.)

A modified Site file was created containing:

From the original Site file:

MonitoringLocationIdentifier

MonitoringLocationName

LatitudeMeasure

LongitudeMeasure

CEightDigitCode

StateCode

CountyCode

MonitoringLocationTypeName,

And from the additions made above:

R or L criterion

ANE identifier or “noeco”

This file is called the “tagged sites” file.

[A processing summary for WQI refresh is available.](https://echo.epa.gov/maps/wqimap/docdata)

**VII Result File Processin g**

The Portals Result data are retrieved for media=water, starting date, characteristic type and characteristic names desired. This query overselects to prevent inadvertently missing characteristic names wanted but not in a characteristic group.

The Portals Result file has one record for each measurement.

Locations where there are no nutrient measurements after the start date of the period involved are dropped.

A sample is defined as all records with the same value for MonitoringLocationIdentifier, ActivityIdentifier and ActivityStartDate. OrganizationIdentifier was not used as all the MonitoringLocationIdentifiers were unique.

Multiple determinations of the same pollutant on the same day are regarded as QC exercises, and are averaged to produce only one measurement. Multiple determinations are defined as records with the same values for CharacteristicName and ResultSampleFractionText, in addition to being in the same sample.

Depth was not included in the test for duplicate determinations, with the result that determinations differing only by depth in the same sample will be averaged.

For each record

Screen

ActivityMediaName: “Water”

ActivityStartDate: yearbegin-01-01 to yearend-12-31

CharacteristicName After removing “ as xxx” or “xxx”, if present as rightmost word(s), one of these values not

preceded by the // symbol:

 "Ammonia",

 "Ammonia and ammonium",

 "Ammonia-nitrogen",

 "Ammonium",

 // "Ammonium-Nitrogen",

 "Inorganic nitrogen (ammonia, nitrate and nitrite)",

 "Inorganic nitrogen (nitrate and nitrite)",

 "Kjeldahl nitrogen",

 "Nitrate",

 // "Nitrate-Nitrite",

 // "Nitrate-Nitrogen",

 "Nitrite",

 "Nitrogen",

 "Nitrogen Kjeldahl",

 // "Nitrogen, ammonia (NH3)",

 "Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3)",

 // "Nutrient-nitrogen",

 "Organic Nitrogen",

 "Organic nitrogen"

For nitrogen, the // names do not appear in the SRS substance search and were not used, although they show up in the data (about 65,000 samples). It was not clear how these measurements would figure in the total N derivation.

 // "Inorganic phosphorus",

 // "Organic phosphorus",

 // "Ortho-Phosphate-Phosphorus",

 // "Orthophosphate",

 // "Phosphate",

 "Phosphate-phosphorus",

 "Phosphorus",

 // "Polyphosphate",

 // "Soluble Reactive Phosphorus (SRP)"

 For phosphorus, the // names cannot be used in combination to derive total P.

[CharacteristicName specificies the pollutant](https://echo.epa.gov/maps/wqimap/docdata). The Characteristic names shown above were determined to be relevant for identifying all the N and P data needed.

**Vetting the measurement**

Abbreviations

 AE ResultDetectionConditionText

 AH ResultMeasureValue

 AI ResultMeasure/MeasureUnitCode (units1 and units2)

BG DetectionQuantitationLimitTypeName

BH DetectionQuantitationLimitTypeMeasure/MeasureValue

BI DetectionQuantitationLimitTypeMeasure/MeasureUnitCode(units1 and units2)

BR MethodSpecificationName (units2)

Units1 is the concentration, units2 is the reference species (e.g., in ”.2 mg/l as N”, units1=”mg/l” and units2=”N”).

Check ResultDetectionConditionText(AE)

ResultDetectionConditionText(AE) Action

\*Present >QL Automatic Hotspot flag

Present Above Quantification Limit Automatic Hotspot flag

\*Present If AH is null then don’t use measurement

Detected Not Quantified If AH is null then don’t use measurement

\*Not Reported Don’t use measurement

Not Reported Don’t use measurement

\*< Use detection limit protocols

\*Non-detect Use detection limit protocols

\*Not Detected Use detection limit protocols

\*Present <QL Use detection limit protocols

< Use detection limit protocols

ND Use detection limit protocols

Not Detected Use detection limit protocols

Present Below Quantification Limit Use detection limit protocols

Any other text or no text Not a detection limit

There are two, possibly different, detection limit protocols, one for combining subspecies (“subspecies detection limit protocol”) and the other for determining summary statistics (“statistics detection limits protocol”).

**Definition of a valid detection limit measurement:**

If AE indicates measurement is a detection limit, AH is null and BH and BI (units1 and units2) are present, use BH/BI as the detection limit. If BI units2 is null, seek units2 at the end of Characteristic Name. If units2 is still not determined, look for a units2 in the table in S**ubspecies Combination Protocol – Nitrogen** (infra). If units2 remains undetermined, look for units2 in BR. If either units1 or units2 remain(s) undetermined, or if BG is ”Drinking Water Maximum” or “Water Quality Standard or Criteria”, drop the measurement.

BG DetectionQuantitationLimitTypeName

Drinking Water Maximum

Estimated Detection Level

Historical Lower Reporting Limit

Instrument Detection Level

Laboratory Reporting Level

Long Term Method Detection Level

Lower Quantitation Limit

Lower Reporting Limit

Method Detection Level

Method Detection Limit (MDL)

Practical Quantitation Limit

Sample-specific min detect conc

Upper Quantitation Limit

Upper Reporting Limit

Water Quality Standard or Criteria

**Definition of a valid quantification measurement:**

If AE is null, AH is present, and AI (units1) [preferred] or BI (units1) are present, and units2 is present in (preferred order) AI, Characteristic Name or BR , the record is used (and is not a detection limit). Otherwise, drop the measurement.

Convert nitrogen concentration to mg/l as N and phosphorus concentration to ug/l as P, if not already in these units.

Continue until all N and P species in the sample have been processed.

 **Determination of Total N and/or Total P**

 **If Total N as N is present**, it is used.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Characteristic** | **Fraction** | **Speciation** | **Note** | **Comment** |
| Total N | Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3) | Total | ? | Total N, No Combo Needed | Assumed as N |
| Nitrogen | Total | NS |   |
| Nitrogen | Total | N |   |
| Nitrogen | Total | ? | Assumed as N |
|   |   |   |   |   |   |

 **Otherwise**

Derive values for total N from N subspecies, if possible, as follows:

Subspecies Combination Protocol – Nitrogen

The subspecies in the following table found in the Result file have been identified as combinable:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Row** | **Characteristic Name** | **Fraction** | **Species** | **Remarks** |  |
| 1 | Kjeldahl Nitrogen | Total | N |  |  |
| 1 | Nitrogen Kjeldahl  | Total | N |  |  |
| 2 | Ammonia-nitrogen  | Total | N |  |  |
| 3 | Ammonia-nitrogen  | Total | ? | Assume as N |
| 4 | Ammonia and ammonium | Total | N |  |  |
| 5 | Ammonia | Total | N |  |  |
| 6 | Nitrate | Total | N |  |  |
| 7 | Nitrogen | Dissolved | ? | Assume as N |
| 8 | Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3) | Dissolved | ? | Assume as N |
| 9 | Kjeldahl Nitrogen | Dissolved | N |  |  |
| 9 | Nitrogen Kjeldahl | Dissolved | N |  |  |
| 10, 11 | Ammonia | Dissolved | N |  |  |
| 12 | Ammonium  | Filterable | NH4 | Convert to as N |
| 13 | Ammonium  | Filterable | N |  |  |
| 14 | Inorganic Nitrogen (Nitrate & nitrite) | Dissolved | N |  |  |
| 15 | Ammonia and ammonium | Dissolved | N |  |  |
| 16 | Nitrate  | Filterable | NO3 | Convert to as N |
| 17 | Nitrate | Dissolved | N |  |  |
| 18 | Nitrite | Dissolved | N |  |  |
| 19 | Organic Nitrogen | Dissolved | ? | Assume as N |
| 20 | Nitrogen | Suspended | ? | Assume as N |
|  |  |  |  |  |  |
|  |  |  Filterable=Dissolved?=Not Specified |  |  |

The combinations yielding total N are these:

7 + 20

8 + 20

9 + 14 + 20

4 + 6 + 18 + 19 + 20

1 + 6 + 18

2/3/5 + 6 + 18 + 19 + 20

4 + 6 + 18 + 19 + 20

10/11 + 14 + 12/13 + 19 + 20

15 + 16/17 + 18 + 19 + 20

1 + 14

1 + 6/16/17 + 18

If Phosphorus or Phosphate-phosphorus is present, use this value as total P.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Characteristic** | **Fraction** | **Speciation** | **Note** | **Comment** |
| Complete | Phosphorus | Total | as P | Total P. No conversion needed. |   |
| Phosphorus | Total | NS |
| Phosphorus | Total ? | NS |
| Phosphate-phosphorus | Total | as P |
| Phosphate-phosphorus | Total | NS | As P assumed  |
|  |  |  |  |  |  |
| A review of the phosphorus subspecies available in the Portals data did not identify any combination of them which would produce total P. |  |  |  |  |
|   | Phosphorus | Dissolved | as P | Dissolved/Soluble P fraction only. Needs Suspended/Particulate P |   |
| NS | NS |   |   |
| NS | as P |   |   |
|   | Phosphate-phosphorus | Dissolved | NS | Dissolved/Soluble P fraction only. Needs Suspended/Particulate P |   |
| Dissolved | as P |   |
| NS | NS |   |   |
| NS | as P |   |   |
|   | Phosphate | Dissolved | as P | Dissolved/Soluble P fraction only. Needs Suspended/Particulate P |   |
| NS | Assumed reported as PO4, may require conversion factor of 0.33 to P |
|   | Orthophosphate | Total | as P | Total Inorganic P. Needs Organic Fraction |   |
| Dissolved | NS | Dissolved/ Soluble Inorganic P only. Less than dissolved P fraction | May need conversion if reported as PO4 |
| Dissolved | as P |   |
| NS | as P | Not Specified. Inorganic P. Could be Total or Dissolved orthophosphate but not Total P. |   |
| NS | as PO4 | Need conversion of value with 0.33 to P |
| NS | NS |   |

Figure P. Combination of available P subspecies are insufficient to derive Total P.

There may be additional, less obvious, combinations of fraction, subspecies and units in the Portals data which were overlooked in the initial effort to identify them.

**Non-Detect Handling**

**When Combining Species to Arrive at Total N**

Deriving total N from N subspecies requires adding together a number of subspecies values, some of which may be detection limits.

**When Deriving Summary Statistics**

A summary statistic (e.g. maximum, percentile) requires a number of sample values be reduced to a single number. Some of these values may be detection limits.

Measurements below the detection limit(DL) of the measurement method, often referred to as “non-detects”, are typically reported as “<DL”. But how to combine or summarize non-detects and detections is a choice made anew by every investigator. It seems to be widely recognized that substitution methods (removing the “<” and using the DL, zero, ½ DL or a random value between zero and the DL) are inadequate, but are still in use.

Studies cited above determined that simple substitution methods performed poorly in comparison to other procedures. Substitution of zero produced estimates of mean and median which were biased low, while substituting the reporting limit resulted in estimates above the true value. Results for the standard deviation and IQR, and for substituting one-half the reporting limit, were also far less desirable than alternative methods. With the advent of powerful desktop computers to perform more complex calculations there appears to be no reason to use simple substitution methods. As the choice of value to be substituted is essentially arbitrary without some knowledge of instrument readings below the reporting limit, and as large differences may occur in the resulting estimates, simple substitution methods are not defensible ([Helsel and Hirst, USGS, 2002](https://echo.epa.gov/maps/wqimap/docdata)).

[Antweiler and Taylor](https://echo.epa.gov/maps/wqimap/docdata) have this to say:

The main classes of statistical treatment of below-detection limit (left-censored) environmental data for the determination of basic statistics that have been used in the literature are substitution methods, maximum likelihood, regression on order statistics (ROS), and nonparametric techniques. These treatments, along with using all instrument-generated data (even those below detection), were evaluated by examining data sets in which the true values of the censored data were known. It was found that for data sets with less than 70%censored data, the best technique overall for determination of summary statistics was the nonparametric Kaplan–Meier technique. ROS and the two substitution methods of assigning one-half the detection limit value to censored data or assigning a random number between zero and the detection limit to censored data were adequate alternatives. The use of these two substitution methods, however, requires a thorough understanding of how the laboratory censored the data. The technique of employing all instrument-generated data--including numbers below the detection limit--was found to be less adequate than the above techniques. At high degrees of censoring (greater than 70% censored data), no technique provided good estimates of summary statistics. Maximum likelihood techniques were found to be far inferior to all other treatments except substituting zero or the detection limit value to censored data.

These and other studies argue persuasively that Kaplan-Meir or ROS should be used for non-detects; however, some time would be required to study the implementation and implications of these methods, and, in the interest of arriving at an initial version of the WQI product ASAP, the time-honored but non-defensible substitution of ½ the censored value(BH) as the actual value when computing the statistics was used. Perhaps the next version of the WQI file will do better. A consequence of this choice is that the DL is removed as soon as the measurement is read and so there is no need for methods for combining non-detects in this scenario because there are no non-detects to combine.

**Statistics, Addition of Site file data and Output**

At this point all the (usable) values for N or P for date range yearbegin to yearend for a monitoring location are available, statistics for the location (median, max, etc.) are computed, the “tagged sites” file data for the location are prepended to form the entire output record (see **V** supra), and the record is written.

The statistics chosen are conventional (maximum, various percentiles including median, but not interquartile range), and exploratory (the time-weighted percentiles). Time weighting applied a weight of w=(yearend-yearbegin+1)\*12 to Dec yearend, w-1 to Nov yearend,…,1 to Jan yearbegin. The problem was how to avoid giving as much visibility to older data as to newer, unless, of course, equal visibility was desired. The weighting method is still open for discussion.

The presence of only a small number of samples can produce artificially large differences in percentiles.

S[ee Portals File Processing Summaries for particular refresh details](https://echo.epa.gov/maps/wqimap/docdata)