

US EPA ARCHIVE DOCUMENT

National Coastal Assessment-Northeast Year 2001 Summary Database

Revision date: 01/30/04

The Summary Database contains water quality, sediment, benthic community, and fish data collected by several partners in the Northeast Region estuaries in the summer of 2001. The database currently consists of fourteen summary data sets and three lookup tables.

This document is in three sections: (1) a database overview, (2) detailed dataset descriptions, and (3) a discussion of key fields and commonly used codes (QACODE, LABCODE)

1) Database Overview

The data sets that make up the database are:

| | | |
|-------------------------|---|--|
| Date/Location Data: | STATIONS EVENTS | Sampling Station Location Data Station Visit Data |
| Water Quality Data: | WATRPHYS ATTENCO NUTRNTS | Water Quality Physical Measurements Data Light Attenuation Data Water Quality Nutrients Data |
| Sediment Quality Data: | SEDGRAIN SEDTOX SEDCHEM | Sediment Grain Size Data Sediment Toxicity Test Data Sediment Chemistry Data |
| Benthic Community Data: | BEN_ABUN | Benthic Abundance Data |
| Fish Data: | FTRAWL FISH_CNT FISHLEN CRAB_LOB | Standard Trawl Data Fish Counts by Species per Trawl Fish Length and Pathology Data Crab and Lobster Measurements |
| Tissue Chemistry | TISSCHEM | Fish crab and lobster Tissue Chemistry Data |
| Lookup Tables | BEN_TAXA FISH_TAX ANALYTES | Benthic Taxonomy Information Fish Taxonomy Information Chemical Analyte Information |

2) Data Set Descriptions:

STATIONS - Sampling Station Location Data

This data set contains one record for each planned sampling station, with the planned latitude and longitude. The exact sampling location from each visit is recorded in the EVENTS data set, which may differ slightly from the planned locations. However, for analytical purposes, the station locations in the STATIONS data set should be used. The data set contains the following elements:

| | |
|----------|---|
| STATION | Coastal 2000 Station Name |
| STAT_ALT | Alternate Site Code (A,B,C) |
| STATE | State where Station is Located |
| ESTUARY | Estuary Name |
| STA_LAT | Latitude (decimal degrees, datum NAD83) |
| STA_LNG | Longitude (decimal degrees, datum NAD83) |
| ST_COOP | State Cooperative Agreement for Sampling |
| LOCAL_ID | Station Identifier Used by State |
| PROVINCE | EMAP Province Code |
| AREA | Statistical Area represented by Station (sq. m) |
| ST_AREA | Area Represented by Strata (sq. m) |
| STRATA | Strata Station is located in |
| SUBAREA | In-State Statistical Area (sq. km) |
| SUBSTRAT | Substrata Station is located in |
| SUBST_AR | Area Represented by Substrata (sq. km) |

The variable ST_COOP defines the Cooperative Agreement that the stations were sampled under and describes the purpose of the sampling. Note that some agreements involved sampling across state boundaries, so this code does not necessarily identify the state in which the station is located. The cods used are:

| ST_COOP | Description |
|---------|-------------------------|
| CT | Connecticut |
| CT-FSH | Connecticut Fish Survey |
| DE | Delaware Inland Bays |
| MA | Massachusetts |
| ME | Maine |
| NH | New Hampshire |
| NJ-C | New Jersey Coast |
| NJ-DB | New Jersey/Delaware Bay |
| NY | New York |
| RI | Rhode Island |

EVENTS - Station Visit Data

This data set contains one record for each sampling visit to a station. There may be multiple records per station. The actual latitude and longitude for a visit may differ slightly from the planned latitude and longitude included in the STATIONS data set. The two variables STATION and EVNTDATE make up the unique identifier for this data set.

Variables:

| | |
|----------|--|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |
| VISNUM | Number of Visit to this Station |
| STADEPTH | Depth of Water at Station (m) |
| EVNT_LAT | Event Latitude (decimal degrees, datum NAD83) |
| EVNT_LNG | Event Longitude (decimal degrees, datum NAD83) |
| SAV | Submerged Aquatic Vegetation visible |
| MACROALG | Macro-Algae present at Station |

WATRPHYS - Water Quality Data-Physical Measurements

This data set contains surface and bottom measurements of temperature, salinity, dissolved oxygen and pH collected in the field during sampling. At shallow stations, the surface and bottom values may be identical. The variables SL_DEPTH and BL_DEPTH provide the depth at where surface and bottom values were measured. Note: at shallow stations, the surface and bottom values may be from the same depth. Secchi Depth measurements are also present. This data set contains one record for each sampling event where CTD casts were performed.

Variables:

| | |
|----------|--|
| STATION | Station Name |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Date of Sampling Event |
| SECCHI_D | Secchi Depth (meters) |
| SECC_BOT | Secchi Disk on Bottom? |
| SL_TEMP | Surface Layer-Temperature from CTD (deg. C) |
| SL_SAL | Surface Layer-Salinity from CTD (ppt) |
| SL_OXY | Surface Layer-Dissolved Oxygen from CTD (mg/l) |
| SL_PH | Surface Layer-pH (pH units) |
| SL_DEPTH | Surface Reading Depth (m) |
| BL_TEMP | Bottom Layer-Temperature from CTD (deg. C) |
| BL_SAL | Bottom Layer-Salinity from CTD (ppt) |
| BL_OXY | Bottom Layer-Dissolved Oxygen from CTD (mg/l) |
| BL_PH | Bottom Layer-pH (pH units) |

| | |
|----------|--------------------------|
| BL_DEPTH | Bottom Reading Depth (m) |
| QACODE | QA Qualifier Code |

ATTENCO- Water Quality - Light Attenuation

This data set contains water a summary measure of PAR readings from the water column. The attenuation coefficient is computed from multiple Photo active Radiation (PAR) readings from different depths. PAR readings at depth should be normalized to surface PAR readings. In many cases this normalization was not possible because surface Par readings were not recorded. These cases are flagged with QACODE values of “PAR-A”.

The variable PAR_RECS indicates how many different measures were used to compute the coefficient. The coefficient can not computed for stations with less then two PAR readings. The coefficient is considered unreliable where there are fewer than four readings, so these cases are flagged with QACODE values of “PAR-B”. Attenuation coefficients less then zero are not valid. Cases where the regression equation resulted in a coeficient of less than zero are flagged with QACODE values of “PAR-C”. The coefficient was recoded to zero in these cases.

Variables:

| | |
|----------|-----------------------------|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |
| ATTENCO | PAR Attenuation Coefficient |
| PAR_RECS | Number of Par readings |
| QACODE | QA Qualifier Code |

NUTRNTS - Water Quality - Nutrients Data

This data set contains water chemistry measurements derived from laboratory analyses of water samples collected in the field. This dataset contains multiple records per station. Typical stations have both surface and bottom layer sample results and deep stations have additional mid-water sample results. The water layer is recorded in the variable LAYER Shallow stations have may have only one sample analyzed (LAYER=”Single”). Replicate samples collected for QA purposes have a REP_NUM value of “2”. There is one record for each analyte measured for each replicate at each water level. Most stations have at least 5 analytes measured.

Samples collected by different states were analyzed by different laboratories, for different analytes. The labs are identified by LABCODE.

Variables:

| | |
|---------|--------------------|
| STATION | Station Identifier |
|---------|--------------------|

| | |
|----------|---------------------------------|
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |
| LAYER | Water Layer of Nutrients Sample |
| REP_NUM | Replicate Sample Number |
| ANALYTE | Analyte Code |
| CONC | Concentration |
| UNITS | Unit of Measure |
| QACODE | QA Qualifier Code |
| MDL | Method Detection Limit |
| LABCODE | Lab identifier |

The following analyte codes are used in this dataset. Lab Code columns indicate which laboratories reported these analytes. Several of the analytes reported by the Connecticut lab are included even though they are not NCA core indicators.

| ANALYTE | Description | Core Indicator? | Labs reporting data | | | |
|---------|-------------------------------|-----------------|---------------------|----|----|-------|
| | | | NAT | DE | CT | NJ NY |
| SI | Dissolved Silica | Y | X | | | X |
| NH4 | Dissolved Ammonia | Y | X | X | X | X |
| NO23 | Dissolved Nitrite and Nitrate | Y | X | X | X | X |
| NO2 | Dissolved Nitrite | Y | X | X | | X |
| PO4F | Dissolved Orthophosphate | Y | X | X | X | X |
| TSS | Total Suspended Solids | Y | X | X | X | X |
| CHLA | Chlorophyll a | Y | X | X | X | X |
| TDN | Total Dissolved Nitrogen | | | | X | |
| TDP | Total Dissolved Phosphorous | | | | X | |
| PHOSP | Total Particulate Phosphorous | | | | X | |
| BIOSI | Biological Silica | | | | X | |
| DOC | Dissolved Organic Carbon | | | | X | |
| ORG-N | Organic Nitrogen | | | | X | |
| PC | Particulate Carbon | | | | X | |
| PN | Particulate Nitrogen | | | | X | |
| UREAP | Urea | | | | X | |

SEDGRAIN

This data set contains sediment grain size and moisture content measurements. The sediment analyzed for these measurements came from the same sediment homogenate that was analyzed for chemical contamination and toxicity.

| | |
|----------|-----------------------------|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |
| SILTCLAY | Silt/Clay Content (%) |

| | |
|----------|--------------------------|
| SAND | Sand Content (%) |
| MOISTURE | Moisture Content (%) |
| TOC | Total Organic Carbon (%) |
| LABCODE | Contract/Lab Identifier |

SEDTOX - Sediment Toxicity

This data set contains summary results from an ampelisca survival test. SRVPCCON represents the mean survival of animals in sample sediments as a percent of the survival of animals in clean control sediment. (Lower SRVPCCON values indicate higher toxicity). SRVPC_SG indicates whether the difference in survival between sample tests and controls is statistically significant. ATOX_SIG describes the biological significance of the results (“NT”= not toxic, “<80%”=toxic, “<60%”=very toxic).

| | |
|----------|--------------------------------------|
| STATION | Station Name |
| STAT_ALT | Alternate Site Code (A,B,C) |
| EVNTDATE | Date of Sampling Event |
| SRVPCCON | Ampelisca Survival as % of Control |
| SRVPC_SG | Statistical Significance (p<.05) |
| ATOX_SIG | Ampelisca Toxicity Test Significance |
| QACODE | QA Qualifier Code |
| LABCODE | Lab identifier |

SEDCHEM - Sediment Chemistry Data

This data set contains sediment chemistry measures. There are multiple records for each event - one record for each concentration measured per station visit. Three variables make up the unique identifier for records in this data set: STATION, STAT_ALT, and ANALYTE.

A concentration value is provided for every analyte unless the concentration could not be detected by the lab instruments. In these cases, the detection limit is present (MDL), and the QACODE is set to “CHM-A”. If the analyte was detected but at a level below the detection limit, the concentration is reported and the QACODE is set to “CHM-B”. The detection limit is provided in this case also.

| | |
|----------|-------------------------------|
| STATION | Station Name |
| STAT_ALT | Alternate Site Code (A,B,C) |
| EVNTDATE | Event Date |
| ANALYTE | Chemical Analyte Code |
| CONC | Concentration |
| QACODE | QA Qualifier Code |
| MDL | Detection Limit |
| CHMUNITS | Concentration Unit of Measure |

LABCODE

Contract/Lab Identifier

BEN_ABUN - Benthic Abundance Data

This data set contains benthic abundance measurements from a single sample collected at each station. It contains one record for each taxon found per grab. Three fields are needed to uniquely identify a record: STATION, STAT_ALT and LAT_NAME. The variable ID_LEVEL describes the level at which the organism was identified (SPECIES, GENUS, FAMILY, etc.) GRABSIZE defines the size of the grab sampler that collected the sample. Note that the results in this dataset have not been normalized to a uniform sample size. The TSN can be used to look up taxonomic information for the taxa in the table BEN_TAXA. This table includes the current scientific name for the taxa as used in the Integrated Taxonomic Information System (ITIS). The current name may differ from the Latin Name in this table if the organism has been renamed.

| | |
|----------|-----------------------------------|
| STATION | Station Name |
| STAT_ALT | Alternate Site Code (A,B,C) |
| EVNTDATE | Event Date |
| LAT_NAME | Taxa Latin Name |
| TSN | Taxonomic Serial Number |
| ABUNDANC | Taxa Abundance in sample |
| ID_LEVEL | Taxonomic Level of Identification |
| GRABSIZE | Size of Benthic Grab Sampler |
| LABCODE | Contract/Lab Identifier |

FTRAWL - Standard Trawl Data

This data set contains one record for each standard fish trawl. Only one standard trawl was performed per station visit. Additional, non-standard trawls were sometimes performed at stations to catch fish for chemistry samples but those trawls are not recorded. The different states performed different types of trawls. The trawl type is identified by the variable FTRLTYPE. If the trawl could not be completed successfully, its status is recorded in the variable FTRLFLAG

Two fields are needed to uniquely identify a record: STATION and EVNTDATE.

| | |
|----------|---------------------------------|
| STATION | Station Name |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |
| FTRLTYPE | Type of Fish Trawl |
| FTRLFLAG | Status of Completed Fish Trawl |
| FSPECCNT | Total Fish Species in Trawl (#) |
| FT_DUR | Duration of Fish Trawl (mmss) |

| | |
|----------|-------------------------------------|
| FWTR_SPD | Trawl Speed through Water (knots) |
| BEG_LAT | Fish Trawl Beginning Latitude (dd) |
| BEG_LNG | Fish Trawl Beginning Longitude (dd) |
| END_LAT | Fish Trawl End Latitude (dd) |
| END_LNG | Fish Trawl End Longitude (dd) |

The following FTRLTYPE codes are used:

| FTRLTYPE | Description | Planned Trawl Time |
|----------|---|--------------------|
| EPA | EMAP Standard trawl | 10 minutes |
| CT | Connecticut Fish Survey Trawl | 30 minutes |
| NH | New Hampshire (modified Standard Trawl) | 4 minutes |

FISH_CNT - Fish Counts by Species per Trawl

This data set contains one record for each species of fish caught in the trawl, and provides a count of individual fish, and their mean fork length. Four fields are needed to uniquely identify a record: STATION, EVNTDATE, FCOMNAME, and F_CLASS

| | |
|----------|---------------------------------|
| STATION | Station Name |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Date of Sampling Event |
| FCOMNAME | Fish Taxa Common Name |
| F_CLASS | Fish Size Classification |
| F_COUNT | Number of this Fish Taxa caught |

FISHLEN - Fish Length and Pathology Data

This dataset contains records on individual fish caught in standard trawls. Where large numbers of fish were caught, a random sub-sample were measured. This dataset does not contain a record for every fish caught. The record notes any pathologies noted by the field crew. Five fields are needed to uniquely identify a record: STATION, EVNTDATE, FCOMNAME, F_CLASS, AND FSEQNUM.

| | |
|----------|-------------------------------------|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Date of Sampling Event |
| FCOMNAME | Fish Taxa Common Name |
| F_CLASS | Size Classification |
| FSEQNUM | Fish Sequence Number |
| FLENGTH | Length (cm) |
| COMP_ID | Fish Chemistry Composite Identifier |
| LUMPS | Fish Pathology: Lumps |

| | |
|----------|------------------------------------|
| LUMP_LOC | Locations of Lumps |
| GROWTHS | Fish Pathology: Growths |
| GRWTHLOC | Locations of Growths |
| ULCERS | Fish Pathology: Ulcers |
| ULCERLOC | Locations of Ulcers |
| FINROT | Fish Pathology: Fin Erosion |
| FROT_LOC | Locations of Fin Erosion |
| GILL_ER | Fish Pathology: Gill Erosion |
| GE_LOC | Locations of Gill Erosion |
| GILL_DC | Fish Pathology: Gill Discoloration |
| GD_LOC | Locations of Gill Discoloration |

CRAB_LOB- Crab and Lobster Measurements

This dataset contains records on invertebrates caught at stations by trawling or in traps. This dataset may not contain a record for every invertebrate caught. . Five fields are needed to uniquely identify a record: STATION, EVNTDATE, FCOMNAME, SEX, and FSEQNUM

| | |
|----------|-----------------------------|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Date of Sampling Event |
| FCOMNAME | Taxa Common Name |
| FSEQNUM | Sequence Number |
| C_WIDTH | Carapace Width (mm) |
| SEX | Sex of Animal |

TISSCHEM - Tissue Chemistry Data

This data set contains chemistry measures from fish, crab, and lobster composite tissue samples. There are multiple records for each event - one record for each concentration measured per composite sample replicate. There were up to three composite samples (different species) collected at stations, and some samples were analyzed in replicate. Five variables make up the unique identifier for records in this data set: STATION, STAT_ALT, FCOMNAME, TISS_TYPE, REP_NUM and ANALYTE.

A concentration value is provided for every analyte unless the concentration could not be detected by the lab instruments. In these cases, the detection limit is present (MDL), and the QACODE is set to "CHM-A". If the analyte was detected but at a level below the detection limit, the concentration is reported and the QACODE is set to "CHM-B". The detection limit is provided in this case also.

| | |
|----------|-----------------------------|
| STATION | Station Identifier |
| STAT_ALT | Station Location (A,B or C) |
| EVNTDATE | Event Date |

| | |
|-----------|------------------------------------|
| FCOMNAME | Fish Taxa Common Name |
| TISS_TYPE | Type of Tissue analytzed |
| MN_SIZE | Mean Size of Animals in Homogenate |
| NUM_HOM | Number of Animals in Homogenate |
| PCTLIPID | Percent Lipid Content |
| WETWGHT | Sample Wet Weight |
| ANALYTE | Analyte Code |
| CONC | Concentration of Analyte in Sample |
| CHMUNITS | Unit of Measure |
| QACODE | QA Code |
| MDL | Detection Limit |
| ANALMETH | Analysis Method Code |
| LABID | Laboratory Identifier |

BEN_TAXA - Benthic Taxonomy Information

This dataset contains the most up-to-date taxonomic information available for benthic organisms identified by TSN in the BEN_ABUN dataset. The taxonomic name (TSN_NAME) in this dataset may diff from the latin name recorded in the BEN_ABUN dataset if a more current name is in use by ITIS.

| | |
|----------|-----------------------------------|
| TSN | ITIS Taxonomic Serial Number |
| TSN_NAME | Latin Name currently used by ITIS |
| PHYLUM | Taxonomic Phylum |
| CLASS | Taxonomic Class |
| ORDER | Taxonomic Order |
| FAMILY | Taxonomic Famiy |
| GENUS | Taxonomic Genus |

FISH_TAX - Fish Taxonomy Information

This dataset contains Latin names and Integrate Taxonomic Information System (ITIS) Taxonomic Serial Numbers for the common names of fish, crabs and lobsters found in the FISHLEN and CRAB_LOB datasets.

| | |
|----------|------------------------------|
| FCOMNAME | Common Name |
| FSCINAME | Scientific (Latin) Name |
| TSN | ITIS Taxonomic Serial Number |

ANALYTES Chemical Analyte Information

This dataset provides full chemical names an Chemical Abstract Society (CAS) Numbers for

the chemical analyte codes used in the SEDCHEM and TISSCHEM datasets.

| | |
|----------|-----------------------|
| ANALYTE | Chemical Analyte Code |
| CHEMNAME | Full Chemical Name |
| CASNUM | CAS Number |

3) Notes on Key fields and QACODE and LABCODE values

Key fields:

Planned station locations are identified by two fields: STATION, and STAT_ALT. For most stations only one of three possible alternates was sampled (A, B or C). However for certain stations, two different alternate sites were sampled at different times. Therefore it is necessary to use the alternate site code (STAT_ALT) to accurately identify the site.

Some stations were visited more than once. The key field EVNTDATE, in conjunction with STATION and STAT_ALT, uniquely identifies a visit to a site.

QACODE and LABCODE values

The variable QACODE is present in several datasets. Records may have more than one code separated by commas in the field. The codes are explained below:

SEDCHEM, TISSCHEM datasets:

| | |
|-------|--|
| CHM-A | Analyte not detected |
| CHM-B | Analyte detected but below limit |
| CHM-C | QA review indicates possible measurement problem |

NUTRNTS datasets:

| | |
|-------|----------------------|
| NUT-A | Analyte not detected |
|-------|----------------------|

WATRPHEYS dataset:

| | |
|-------|---|
| WTR-A | Shallow station: SL and BL measures are from same depth |
|-------|---|

The variable LABCODE is used to identify the source of different lab measurements. The code identifies the the contract or cooperative agreement the samples were processed under. The value "EPA" refers to national contracts administered by EPA and does not mean that samples were analyzed in EPA labs. This is also true of codes that refer to states- these states may have contracted the analyses out to private labs.

| | |
|-----|---|
| NAT | National Contract Lab Sediment and tissue chemistry: AD Little Sediment Toxicity: Trac Labs Nutrients: EPA GED Sediment Grain size: AD Little |
| CT | Connecticut Cooperative Agreement |
| NJ | New Jersey Cooperative Agreement |
| NY | New York Cooperative Agreement |
| MA | Massachusetts Cooperative Agreement |