Three-Level QA Review of Coastal 2000 Northeast Database
Updated 20 July 2001

Atlantic Ecology Division
U.S. Environmental Protection Agency
Narragansett, Rhode Island 02882

This document describes the QA review process performed on Coastal 2000 data in the Northeast Region, coordinated by the Atlantic Ecology Division (AED). Each state or Cooperative Agreement recipient is measuring a suite of field data and collecting samples of water, sediments, and fish for laboratory analysis. The states may elect to forward the samples to a national contract laboratory or conduct the analytical analyses themselves. The results of the field and laboratory analyses are sent to AED for incorporation into a regional database. These data are subjected by AED to the three levels of QA review described below.

The states or contract laboratories provide the data in electronic form to the project officer at EPA-AED. A regional database manager at the AED combines all of the states' data into a "d1-database", organized into separate data files by similarity and by states. For example, all nutrient-related data are entered into the NUTRNTS file. In turn, each data file contains several parameters; for example, the NUTRNTS file includes the nutrient parameters: nitrate, ammonium, phosphate, *etc.*

The d1-database contains many parameters that are administrative in nature or descriptive of the sampling event, for example, the identity of the sampling vessel and crew, the weather conditions at the time of sampling, *etc.* The AED database manager constructs a "summary database", or d2-database, consisting of parameters that have been identified to be the most useful to data users. Table 1 lists all the data files and associated parameters in the d2-database.

Level 1 QA Review

A Level 1 review examines the d1-database for completeness, format compatibility, and internal consistency. The checks listed below are simple and can be performed without detailed knowledge of the nature of the parameters. A Level 1 review is complete when all data gaps are filled or explained and obvious errors have been corrected. Records are kept of any changes made to the database.

1.1 A completeness check is performed on all data submitted by states and laboratories. This check simply involves comparing the number of data entries in each file to the number of stations sampled. The database manager notes and investigates any missing data.

1.2 A range check of each parameter is performed to highlight records falling outside an expected range. The database manager simply notes outliers and corrects any obvious errors, such as data submitted with incorrect units. Persistent outliers are highlighted for a Level 2 review.

1.3 Simple consistency checks are performed by comparing independent records of closely related parameters. For instance, records of latitudes and longitudes are compared with planned locations, water depths measured by independent methods are compared, *etc.* Table 2 contains a list of the checks performed.

The AED database manager submits to the Project Officer any questions/corrections that have been identified and suggested database changes. The Project Officer transmits these questions/corrections to the Cooperative Agreement Program Manager, who resolves the concerns, concurs/non-concurs with suggested changes, and submits revised data file(s) if necessary. Once the Cooperative Agreement recipient concurs with the changes to the database, Level 1 review is complete. The data files passing Level 1 QA Review are made available on the password-protected Coastal 2000 Northeast web site.

Level 2 QA Review

A Level 2 review is performed on the summary database, or d2-database, parameters only. The review highlights values that are unusual enough to raise the suspicions of a data user. Anomalous data include values that are especially large or small, or are noteworthy in other ways. Focus is on rare extreme values since outliers usually merit most attention by users and may affect statistical quantities such as averages and standard deviations.

2.1 Extreme values are flagged by highlighting any record deviating from the average by more than 3 standard deviations.

2.2 Extreme values are also highlighted visually by plotting parameter values vs station ID. The benefit of such a plot is that the outliers can be compared with nearby stations or with associated parameters. For example, if several stations in an estuary are exceptionally high or low, we would suspect that the data may be reliable. Similarly, if several closely associated parameters are extreme at a station (*e.g.*, consistently high nutrients, or consistently high organic compounds, *etc.*), we would suspect that the records may be valid.

2.3 Correlations among the parameters are examined. An array of miniature x-y plots is generated, one plot for each combination of associated parameters (for example, a standard application of SAS Insight). For instance, a matrix of 5 water quality parameters would generate a 5X5 array of plots systematically varying in variables for the x- and y-axes. Typical plots show a regular relationship between the plotted parameters. Anomalous data are readily evident on these plots. Examination of closely related parameters may resolve questions regarding the accuracy of anomalous data.

Documentation of suspicious data identified is prepared, with invalid data flagged. This documentation becomes part of the metadata. Level 2 data are made available on the same web site as the Level 1 data.

Level 3 QA Review

A Level 3 review is conducted to evaluate whether data submitted by the states or laboratories are comparable across areas, recognizing that the magnitudes of the values may indeed be different in the various geographic areas.

3.1 A regional map is prepared for each measured parameter. Discrete map symbols denote station location and the magnitude of the parameter (*e.g.,* low, moderate, or high). The maps are examined for noteworthy patterns that may be attributed to database errors.

3.2 A bar chart is prepared for each measured parameter. The chart shows the percent area of each state's water designated by a condition category (*e.g.,* low, moderate, or high). The charts are also examined for anomalous patterns that may indicate database irregularities.

3.3 A distribution graph is prepared for each parameter, grouping data by estuarine system to compare the range and distribution of measured values cross the states.

3.4 A table is prepared for each parameter summarizing the descriptive statistics of parameters by state. While the magnitude of a parameter may vary by state, it is expected that the coefficient of variation should be roughly equivalent across the states.

 A summary report is prepared, utilizing the maps, charts, and tables developed in the Level 3 review. This report is made available on the same web site that the data are available on.

Records are maintained of all data files examined and entries considered anomalous. The Project Officer reports the anomalies to Cooperative Agreement recipient or contract laboratory data managers, who correct and resubmit the data. All changes to the original database are documented.

References

USEPA. 2000. Coastal 2000 Northeast Component Information Management Plan. Draft, 5/24/2000.
Atlantic Ecology Division, National Health and Environmental Effects Research Laboratory
Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, Rhode Island.

Table 1. Data files and parameters included in the Summary of d2 Database. (Refer to USEPA (2000) for details).

STATIONS - Sampling Station Location Data

STATION	Coastal 2000 Station Name
STAT_ALT	Station Alternate Site Code (A, B, or C)
STATE	State station is located in
STA_TYPE	Station Type
ESTUARY	Estuary Name
STA_LAT	Latitude (decimal degrees)
STA_LNG	Longitude (decimal degrees)
ST_COOP	State Cooperative Agreement Station sampled by
LOCAL_ID	Alternate Station Identifier used by State

EVENTS - Station Visit Data

STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
STAT_ALT	Station Alternate Site Code (A,B or C)
EVENT_ID	Partner's Event Identifier
CREW_ID	Sampling Crew Identifier
EVNT_LAT	Latitude (decimal degrees)
EVNT_LNG	Longitude (decimal degrees)
DEPTH	Water Depth (meters)
SECCHI_D	Secchi Depth (meters)
SAV	Submerged Aquatic Vegetation Present at site (Y/N)
TRASH	Trash Present at site (Y/N)
MACROALG	Macroalgae present at site (Y/N)

WATRPROF - Water Profile Data-Physical Measurements

STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
W_DEPTH	Water Depth of measurements
TEMP	Temperature (Deg C)
SAL	Salinity (ppt)
OXY	Dissolved Oxygen (mg/l)
PH	pH (pH units)
PAR_SRF	PAR at Surface (mE/m2/s)
PAR_DPTH	PAR at Depth (mE/m2/s)

FTRAWL - Standard Trawl Data

STATION	Station Identifier
EVNTDATE	Date of Sampling Event
FTRAWLID	Trawl Identifier
TRLTYPE	Standard or Non Standard Fish Trawl? (S/N)
BEG_LAT	Trawl Beginning Latitude (decimal degrees)
BEG_LNG	Trawl Beginning Longitude (decimal degrees)
END_LAT	Trawl End Latitude (decimal degrees)
END_LNG	Trawl End Longitude (decimal degrees
FTRL_DUR	Duration of Fish Trawl (minutes:seconds)
FTRL_SPD	Average Speed over Bottom (knots)
FSPECCNT	Number of Unique Species in Trawl
FISHCNT	Number of Individual Fish in Trawl
GEARCODE	Gear Code

FISHSPEC - Fish Counts by Species per Trawl
    STATION   Coastal 2000 Station Name
    EVNTDATE   Date of Sampling Event
    FTRAWLID   Trawl Identifier
     FSCINAME   Fish Species-Scientific Name
    TAX_CNT   Number of Fish of this species

FISHPATH - Fish Length and Pathology Data
    STATION   Station Identifier
    EVNTDATE   Date of Sampling Event
    FTRAWLID   Trawl Identifier
     FSCINAME   Fish Species-Scientific Name
    FSEQNUM   Fish Sequence Number
    F_LENGTH   Fish Length (mm)
    LUMPS   Fish Pathology: Lumps (Y/N)
    LUMPLOC   Locations of Lumps
    GROWTHS   Fish Pathology: Growths (Y/N)
    GRTHLOC   Locations of Growths
    ULCERS   Fish Pathology: Ulcers (Y/N)
    ULC_LOC   Locations of Ulcers
    FINROT   Fish Pathology: Fin Erosion (Y/N)
    FROTLOC   Locations of Finrot
    GILL_ERO   Fish Pathology: Gill Erosion (Y/N)
    GERO_LOC   Gill Erosion Location
    GILL_DC   Fish Pathology: Gill Discoloration (Y/N)
    GDC_LOC   Gill Discoloration Location

NUTRNTS - Water Quality-Nutrients Data
    STATION   Station Identifier
    EVNTDATE   Date of sampling event
    LAYER   Water Layer (SRF, BOT, MID)
    REP_NUM   Replicate Number (>1 indicates replicate sample)
    ANALTYE   Code to identify analyte measured
    CONC   Concentration of analyte
    UNIT   Unit of Measure
    MDL   Method Detection Limit
    QACODE   QA Qualifier
    LABCODE   Laboratory Identification Code

The following analyte codes are used to define the measurements in the NUTRNTS data file.

ANALYTE	Description	Core Indicator
SI	Dissolved Silica	Y
NH4	Dissolved Ammonia	Y
NO23	Dissolved Nitrite and Nitrate	Y
NO2	Dissolved Nitrite	Y
PO4F	Dissolved Orthophosphate	Y
TSS	Total Suspended Solids	Y
CHLA	Chlorophyll a	Y
PON	Particulate Organic Nitrogen	N
TDN	Total Dissolved Nitrogen	N
TDP	Total Dissolved Phosphorous	N
PHAE	Phaeophytin	N
PHOSP	Total Particulate Phosphorous	N
POC	Particulate Organic Carbon	N

SEDGRAIN - Sediment Grain Size Data
STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
REP_NUM	Replicate Number (>1 indicates replicate sample)
SAND	Sand Content (%)
SILTCLAY	Silt/Clay Content (%)
MOISTURE	Moisture Content (%)
TOC	Total Organic Carbon (%)
QACODE	QA Qualifier
LABCODE	Laboratory Identification Code

TOXICITY - Sediment Toxicity and Microtox Test Data
STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
REP_NUM	Replicate Number (>1 indicates replicate sample)
SRVPCCON	Ampelisca Survival as % of Control
SRVPC_SG	Ampelisca Survival- Stat. Significance
ATOX_SIG	Ampelisca Survival- Significance
EC50_MC	Microtox Moisture Corrected Mean EC50 (%)
MTOX_SIG	Microtox Test Significance
QACODE	QA Qualifier
LABCODE	Laboratory Identification Code

SEDCHEM - Sediment Chemistry Data
STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
BATCH_ID	Laboratory Batch Name
ANALYTE	Code for Analyte Measured
CONC	Concentration of Analyte in Sample
CHMUNITS	Concentration Units of Measure
MDL	Method Detection Limit
QACODE	QA Qualifier
LABCODE	Laboratory Identification Code

BEN_ABUN - Benthic Abundance Data
STATION	Coastal 2000 Station Name
EVNTDATE	Date of sampling event
BENGRAB	Grab Associate with Infauna Sample (# per sq. meter)
TAXNAME	Taxa Name
ABUNDANC	Species Abundance in Sample (#)
ID_LEVEL	Level of Taxonomic ID
QACODE	QA Qualifier
LABCODE	Laboratory Identification Code

CHEM_QA - Chemistry QA Data
BATCH_ID	Laboratory Batch Name
SAMPTYPE	QA Sample Type
ANALYTE	Code for Analyte Measured
CONC	Concentration of Analyte in Sample
CHMUNITS	Concentration Units of Measure
MDL	Method Detection Limit
QACODE	QA Qualifier
LCMNAME	Lab Control Material Name
LABCODE	Laboratory Identification Code

The following SAMPTYPE codes will be used to define QA samples in CHEM_QA data file.

SAMPTYPE	Description	Unit of Measure
LRB	Lab Reagent Blank	varies
LCM	Lab Control Material	ug/g or ng/g Dry W't
LCMPR	Lab Control Material % Rec.	Percent Recovery
LF1	Lab Spiked Sample- 1st Member	ug/g or ng/g Dry W't
LF1PR	Lab Spiked Sample- 1st Mem. % Rec.	Percent Recovery
LF2	Lab Spiked Sample- 2nd Member	ug/g or ng/g Dry W't
LF2PR	Lab Spiked Sample- 2nd Mem. % Rec.	Percent Recovery
MSDRPD	Rel % Difference: LF1 to LF2	Percent
LFB	Lab Fortified Blank	Percent Recovery
LSFPR	Lab Spiked Sample % Rec.	Percent Recovery
LDRPD	Lab Duplicate Relative % Diff.	Percent

TISSCHEM - Fish and crustacean tissue chemistry data.

STATION	Coastal 2000 Station Name
EVNTDATE	Date of Sampling Event
SAMPTYPE	Sample Type (Identifies QA Samples)
BATCH_ID	Laboratory Batch Name
TAXNAME	Scientific name of organisms analyzed
TISUTYPE	Tissue Type Code
NUM_HOM	Number of Individuals in Homogenate
MN_WGHT	Mean Weight of Individuals in Homogenate
MN_SIZE	Mean length (fish) or width (Crabs) in Homogenate (mm)
WETWGHT	Wet Weight of Sample Analyzed (g)
PCTMOIST	Percent Moisture of sample Analyzed (%)
PCTLIPID	Percent Lipid Content of Tissue Examined
ANALYTE	Code for Analyte Measured
CONC	Concentration of Analyte in Sample
CHMUNITS	Concentration Units of Measure
MDL	Method Detection Limit
QACODE	QA Code
LABCODE	Laboratory Identification Code

Table 2. Comparison Checks for Level 1 QA review of Coastal 2000 Northeast data.

Compare event latitude	to planned station latitude
Compare event longitude	to planned station longitude
Compare depth from fathometer	to CTD cast depth
Compare depth from fathometer	to bottom layer depth
Compare oxygen from CTD (surface)	to ambient dissolved oxygen (surface)
Compare salinity from CTD (surface)	to ambient salinity (surface)
Compare temperature from CTD (surface)	to ambient temperature (surface)
Compare number of fish species	to actual count of fish species data sheets
Compare number of individual fish	to actual count of recorded fish lengths