Naming and content issues in BIO-sourced CIOOS files

Dan Kelley, Dalhousie University

2024-10-03

Abstract. I discovered files that (a) have duplicated values listed under similar names and (b) have Q (flag) values that are not named in the expected way, with respect to the associated measurements. Here, I illustrate this with one particular file, but I see the pattern across all the files that I've checked in this set.

Sample data

As a sample file, I downloaded a NetCDF file with the following unix command.

```
curl https://cioosatlantic.ca/erddap/files/\
bio_atlantic_zone_monitoring_program_ctd/\
Bedford%20Basin%20Monitoring%20Program/2023/\
CTD_BCD2023667_001_1_DN.ODF.nc > sample.nc
```

Problem 1: duplicated data

The variable names in this file are: `measurement_time`, `ScanNumber`, `QCNTR_01`, `PRESPR01`, `QPRES_01`, `TEMPS901`, `TEMPP901`, `TEMPPR01`, `QTEMP_01`, `CNDCST01`, `QCNDC_01`, `OXYOCPVL01`, `QOXYV_01`, `CPHLPR01`, `QCPHLPR01`, `PHXXZZ01`, `QPHPH_01`, `IRRDUV01`, `QPSAR_01`, `PSALST01`, `PSLTZZ01`, `QPSAL_01`, `POTMCV01`, `QPOTM_01`, `SIGTEQ01`, `QSIGP_01`, `DOXYZZ01`, `QDOXY_01`, `RecPerBin`, `QCNTR_02`, and `QCFF_01`.

Upon further investigation, I realized that there are 3 temperature variables, and 2 salinity variables. Using

```
ncdump -v TEMPS901 sample.nc
ncdump -v TEMPP901 sample.nc
ncdump -v TEMPPR01 sample.nc
```

I found all the temperatures to be identical, starting with 5.6177, 5.6178, 5.6185, 5.6188, and 5.6201. Similarly, the two salinity variables, `PSALST01` and `PSLTZZ01` hold the same values.

Comment. Having duplicate columns wastes space on the server, and is likely to confuse users.

Problem 2: QC flag names do not match measurement names

The normal oceanographic convention is that a measurement named, say `X` ought to have a flag named `QX`. This is not the case in this file. I find that e.g. `QPSAL_01` is the QC flag for the variable named `PSALT01` (and also for its identical twin, `PSLTZZ01`). Similarly, I infer that `QPSAR_01` refers to `IRRDUV01`. This later inference required me to examine the variable attributes named `ancillary_variables` and `legacy_gf3_code`.

Requiring users to examine files this deeply (by eye or by code) seems to be an unneccessary thing, particularly given the decades-long convention of naming QC flags by prepending the letter `Q`.

Comment. It would be very helpful to follow oceanographic convention for QC flags.