

IMPLEMENTATION OF REAL-TIME QAQC ON A CABLED OBSERVATORY

Application of QAQC Methods to Real-Time Scalar Data

M.A. Jeffries

With large quantities of data made available on the internet within minutes of collection, cabled observatories face the problem of how to determine the initial quality of the data in near real-time. At Ocean Networks Canada, we have designed and implemented a database model allowing for quality control of scalar data in real-time that, combined with delayed assessment tests and manual quality control, provide users with accurately assessed data.

QAQC of ONC Scalar Data

- * Automated QAQC Tests on scalar data:
 - Real-Time (before archival)
 - Instrument, Regional, Site level range tests
 - Dual sensor tests: for example, conductivity-temperature curves tests
 - Delayed Assessment (dependent upon other values from same sensor)
 - Spike, gradient, stuck values tests

Manual QAQC of scalar data

- All data are reviewed by a Data Specialist on a regular basis
- Correction of QAQC flags from automated system as needed
- Addition of annotation as needed

QAQC Test and Flag Details

- * Each test assigned a 'priority' level
- * Multiple tests applied to one sensor, varying priority levels
- * For every QAQC test, a 1 or 4 is assigned to the data point (pass or fail respectively)
- * Output flags in data products
 - Determined by combining all the applicable QAQC test values and their priorities
 - Follows ARGO QC Flag convention (Argo quality control manual V 2.8, Jan 3, 2013):

QC Flag Value = 0: no quality control applied to data
 QC Flag Value = 1: data quality controlled and deemed good
 QC Flag Value = 2: data quality controlled and probably good
 QC Flag Value = 3: data quality controlled and probably bad
 QC Flag Value = 4: data quality controlled and deemed bad
 QC Flag Value = 7: averaged value (ONC defined)
 QC Flag Value = 8: interpolated value
 QC Flag Value = 9: missing value: NaN

- * Manual QAQC values override all Automated QAQC values regardless of priority



Data QC in the ONC-Digital Infrastructure Data Acquisition Model

- * Digital Infrastructure QC Model supports all ONC Observatories
- * Real-time Automated data QC analysis is performed on a shorestation node, prior to the archiving of the data
- * Delayed QC analysis is batch processed on a task node at set intervals
- * Real-time and Delayed QC test results stored in the database by sensor
- * Model supports cascading flags for derived and/or dependent data
- * Model supports the application of multiple automated QC tests applied to one data point

QAQC in ONC Data Products

- * Clean and Raw options for all scalar data products (CSV, mat, PNG/PDF, ODV)
 - Clean: All 'bad' data removed in plots or set to NaN in other data products
 - Raw: All data represented in data products
- * Averaged data are 'cleaned' before averaging algorithm is applied
- * Visualization of 'flagged' data in products using flags
- * Automatic and Manual QAQC test details returned in metadata file

QAQC Future Considerations

- * Application of QC Model to Complex data products
- * Assessment of raw file quality

References

* Argo quality control manual V 2.8, Jan 3, 2013, <http://www.argodatamgt.org/content/download/15699/102401/file/argo-quality-control-manual-version2.8.pdf>

Acknowledgements: Ocean Networks Canada is supported by the Canada Foundation for Innovation, the Province of British Columbia, the Networks of Centres of Excellence of Canada, the Natural Sciences and Engineering Research Council of Canada and the University of Victoria.

The figures to the right show examples of clean and raw PNG data products from Ocean Network Canada's NEPTUNE (top) and VENUS (bottom) Observatories. 'Bad' data are represented with a flag in the 'raw' data plots (left plots) and are removed in the 'clean' data plots (right plots).

