ID13- REAL-TIME DATA QUALITY CONTROL APPLIED AT OBSEA

ANDREU FORNÓS BAUTISTA[1], MARC MARTÍNEZ[1], DANIEL M. TOMA[1], ENOC MARTINEZ PADRO[1], JOAQUÍN DEL RÍO[1]

1 Universitat Politècnica de Catalunya, Electronics Department, SARTI Research Group, 08800 Vilanova i la Geltrú-Barcelona, Spain

ABSTRACT
The expandable seafloor observatory OBSEA (https://obsea.es) is located at 4 Km off Vilaonva i la Geltrú (Catalonia, Spain) at 20 m depth. This cabled observatory obtains different environmental data from a meteorological station, CTD sensor, hydrophone, ADCP, buoy and seismometer. Previously, the quality control of the acquired data was not in real-time and oversighted erroneous data values guiding to a misunderstanding of the environment.

Accounting that the data collected from the OBSEA observatory needs to comply with the Copernicus Marine Environment Monitoring Service (CMEMS) In Situ Thematic Assembly Center (INS TAC) procedures and the inefficiency of the previous data quality control processes, a new automatic real-time data quality control system has been developed. The quality control consists of a set of automatic tests according to the manual for real-time quality control of in-situ temperature and salinity data of the Quality Assurance/Quality Control of Real Time Oceanographic Data (QARTOD, https://ioos.noaa.gov/project/qartod) project of the Integrated Ocean Observing System (IOOS, https://ioos.noaa.gov). The follow-through of the QARTOD manual for the development of the tests was made considering the community acceptance, notability and previous work of the IOOS.

The developed quality control tests flag the data depending on its reliability to enrich its associated metadata (Table 1).

FLAG	0	1	3	4	9
SPECIFICS	No quality tests applied	Reliable	Suspicious	Bad	Missing

Table 1: Value of the flags attached at the data values by the tests and its interpretation.

The construction and implementation of the tests was orchestrated considering the characteristics of the data acquired by the sensors of the OBSEA, the scientific requirements and the QARTOD recommendations (Table 2).

TEST NAME	SPECIFICS	PROOF	RELEVANCE
No value	Proves if there is a value attached to the data set delivered by the sensor	Find sensor acquisition or delivering errors	Proposed by research group
Gross range	Checks if the value is within the sensor and reasonable range	Find outliers, sensor errors and suspicious data considering the region	Required
Climatology	Checks if the value is within a reasonable range considering the season its acquired	Find suspicious data considering the time of the year	Required
Rate of change	Correlate the value with the previous reliable value to spot if it is a spike	Find if the value is a spike considering the standard deviation of that variable for the season it is acquired	Strongly Recommended
Flat line	Match the value with previous values to detect if they change at all	Find if the sensor has some type of problem acquiring or sending the values	Strongly Recommended
Attenuated signal	Measure the variation of previous values to detect if their change is big enough	Find if the sensor has some type of problem acquiring the values	Suggested

Table 2: Test names with their specifics, importance and relevance.

Data generation algorithms were designated to study, evaluate and optimize the tests according to their particular needs and crafted to reveal their weak spots, lighten up their confidence level and, ultimately, find optimization solutions. Additionally, the tests were applied to all CTD data acquired by the OBSEA between 2016 and 2019 (Figure 1) and the results compared with the OBSEA data in the Pangaea databank (https://www.pangaea.de) to corroborate their trustworthiness and see the real importance of the quality control application.

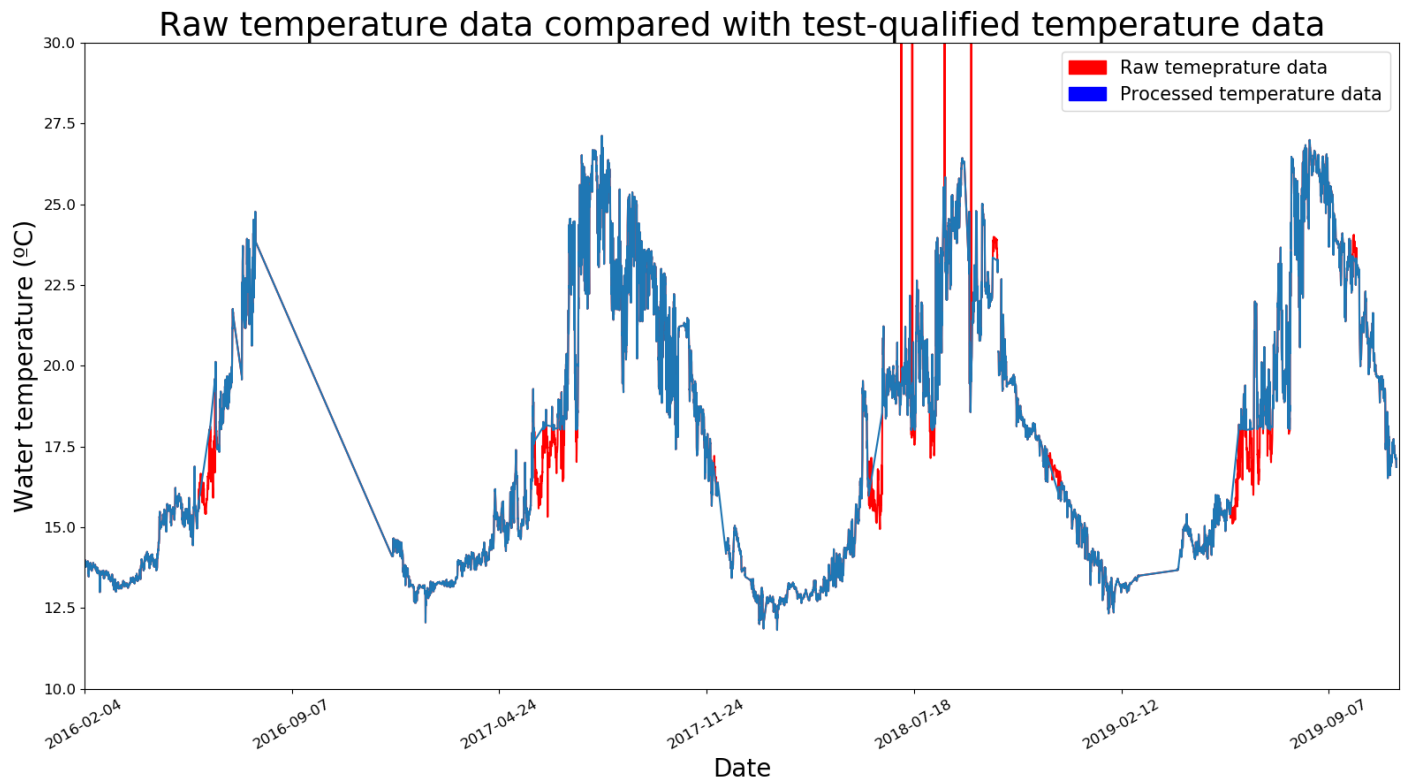Raw temperature data compared with test-qualified temperature data

Figure 1: Sea water temperature (ºC) between 2016 and 2019 acquired by the OBSEA: raw sea water temperature (red) and sea water temperature flagged with good quality by the implemented tests (blue) are shown in the graph.

After these procedures and an exhaustive look through the tests commenced their modification for their continuous application to the real time data acquisition.