Systematic quality control for long term ocean observations and applications

Daniel M. Toma[1], Albert Garcia-Benadí[2], Bernat-Joan Mànuel-González[1], Joaquín del-Río-Fernández[1]

[1] SARTI Research Group, Electronics Dept.,. Universitat Politècnica de Catalunya (UPC), Rambla Exposició 24, 08800, Vilanova i la Geltrú, Barcelona, Spain.
[2] Laboratori de Metrologia i Calibratge, Centre Tecnològic de Vilanova i la Geltrú, Universitat Politècnica de Catalunya (UPC), Rambla exposició, 24, 08800 Vilanova i la Geltrú, Barcelona, Spain

ABSTRACT
With the advances of last year's technologies many new observation platforms have been created and connected on networks for the diffusion of numerous and diverse observations, and also it provided a great possibility to connect all kind of people facilitating the creations of great scale and long-term studies. This paper is focused on the marine observations and platforms employed for this scope. Real time data and the big data have to accomplish some minimal quality of data requirements. Usually, the task to ensure these quality requirements is accomplished by the platforms responsible. The aim of this paper is to explain the design of these quality control systems and their implementation in an ocean observation platform.

Corresponding author: Daniel M. Toma / Albert Garcia-Benadí, e-mail: daniel.mihai.toma@upc.edu / albert.garcia-benadi@upc.edu

1. INTRODUCTION

Recently, marine observations have been rapidly growing. Some of the observed parameters that are currently monitored are the sea temperature, level of acidification and noise pollution. The value of each of these parameters gives very valuable information in different areas, for example, the temperature of seawater is currently used to perform better estimations of the climate change, the measure of the acidification of seawater is very important for fishing processes, etc. However, all these data are relevant if there is a continuous recording of them for long periods of time and the measurements have been performed under quality requirements. For this reason, in the last years many permanent seabed observatories as well as floating observatories such as moored or drifters buoys have been deployed to perform long term marine observations which are the basis for environmental modelling and assessment.

Close examination of these data often reveals a lack of quality that, frequently, happens for extended periods of time.

The growing needs for real-time processing of data and the sheer quantity of data produced by these observatories means that automated Quality Assurance/Quality Control (QA/QC) is necessary to ensure that the collected data has the quality required for its purpose [1]-[3].

This paper demonstrates the use of well-defined community adopted QA/QC tests and automated data quality assessments [4] and [5] to provide a continuous scale of data quality, the capture of information about the system provenance, sensor and data processing history, and the inclusion of the flag values in metadata stream.

An example of this system of implementation and testing the automated data quality assessments on a real time platform is the expandable seafloor observatory, OBSEA [6] and [7] that is deployed to monitor the Barcelona coast, Spain. Among the parameters observed at OBSEA, for which these automated data quality controls is applied are the seawater and air temperature, conductivity, underwater and air pressure.

2. DEVELOPMENT

Each data measured by the platforms has to pass different filters which are evaluated under some tests established for the customer. These tests have to be applied for every magnitude measured.

2.1. Qualification tests

The qualification tests are separated in format tests and behaviour tests. The automated application of these tests is rather straightforward using java programs. However, the estimation of the parameter "thresholds" of these tests poses the greatest challenge. The statistical assumptions dictate that these threshold parameters are ideally defined by having a distribution of values that are objectively considered "reasonable" for every sensor at every site [8]. In the OBSEA the following automatic quality control tests are implemented with the range, step, delta, sigma, null, and gap parameter thresholds determined by statistical distributions based on existing data over a period of three years:

- Platform identification. This criterion determinates the location of the equipment, for example: in the laboratory, connected to the OBSEA platform, or under test.

- Impossible Platform date/time. The date must be greater than the date of the OBSEA start, and the hour must be from 0 until 23, and the minutes must be from 0 until 59.

- Regional Impossible Parameter values. The thresholds of this test are defined for each magnitude. The thresholds establish the extreme values based on the minima and maxima observed for a given sample period (Table 1). For example, the seawater temperature of the Mediterranean Sea cannot exceed 28 degree Celsius and cannot be less than 10 degree Celsius.

- Spike test. The tendency between the previous and actual value has to be coherent. For example the step between the previous and actual value of the temperature of seawater cannot exceed 0.1 °C for a sampling interval of less than 1 minute. The threshold of this test must be in accordance with the parameter measured in the environment and the acquisition rate of the instrument.

- Gradient test. This test fails when the difference between vertically adjacent measurements is too steep. For example the gradient value of the temperature of seawater cannot exceed 0.2 °C for a sampling interval of less than 1 minute. Similar to the spike test the threshold of the gradient test must be in accordance with the parameter measured in the environment and the acquisition rate of the instrument.

The last test is the visual inspection of the data. This test is not automated, but it is a main test to ensure the quality of the adopted QA/QC. This test performs data verification by means of visual inspection; any flagged data from the previous tests is either verified as being an incorrect value or is accepted as correct data. This test minimizes the risk of inadvertently eliminating the observation of a rare and potentially interesting event for the sake of data quality [9].

2.2. Tools

The OBSEA automated Quality Assurance/Quality Control system was created using the java programming language. The

Table 1. The six tests employed in the OBSEA data quality control.

Problem to be identified Section	Test	Calculation
Data belongs to platform	Platform Identification Test	Defined in Metadata
Timestamp of Data in Observatory Range	Impossible Platform Date/Time	Year greater than 2008 Month in range 1 to 12 Day in range expected for month Hour in range 0 to 23 Minute in range 0 to 59
Data Outliers	Regional Impossible Parameter Values	Sea Water Temperature in range 10 °C to 28 °C Salinity in range 35 to 39 Sea Level Air Pressure in range 850 hPa to 1060 hPa (mbar) Air Temperature in range -10 °C + 40 °C Wind Speed in range 0 m/s to 60 m/s Wind Direction in range 0° to 360° Humidity in range 5 % to 95 % Current Speed in range 0 m/s to 3 m/s Current Direction in range 0° to 360° Wave Period in range 0 to 20 s Wave Height in range 0 m to 10 m Depth in range 18 m to 21 m Conductivity in range 3.5 S/m to 6.5 S/m Sound Velocity in range 1480 m/s to 1550 m/s
Jumps in Data Values	Spike test	$\| d_t - (d_{t+1} + d_{t-1})/2 \| - \| (d_{t+1} - d_{t-1}) / 2 \| > \sigma_s$ (where σ_s defined by sampling)
Change in Variance Structure	Gradient test	$\| d_t - (d_{t+1} + d_{t-1})/2 \| > \sigma_g$ (where σ_g defined by sampling)
A Dropped Data Point	Null Test	defined by sampling
Data belongs to platform	Platform Identification Test	Defined in Metadata

system block diagram in Figure 1 provides a general overview of the QA/QC system. The system is composed of three components and their interdependencies, as shown. From this figure, it is clear that dependencies between subsystems are simple and the communication between the subsystem and its dependencies are through their interfaces.

The system allows the building of application specific data structures assembled from standard building blocks allowing "on the fly" changes of parameters. Therefore, it will be easier to add new algorithms and new parameters to the application. This allows components to be generic and configurable and allows other modules to retrieve information required and query data structures.

2.3. Valuation

The result of these automatic quality control tests is a numeric value that is included inside the metadata in the acquired data. These qualifications are detailed in Table 2. When the data are collected and properly verified, usually the flag value is 1, but for long time acquisitions, in a hostile
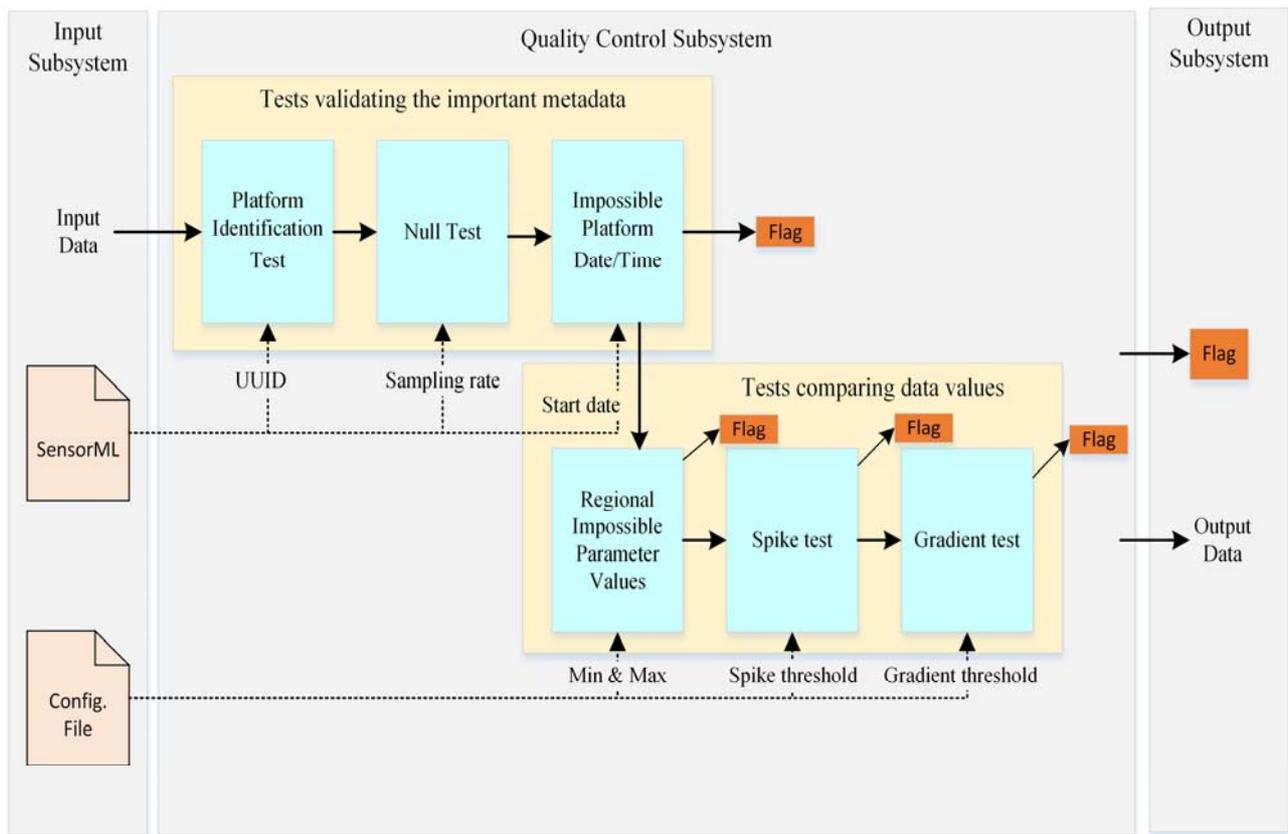
Figure 1. The Automated Data Quality Assurance system block diagram.

environment, in our case the sea, the equipment can suffer from undesirable variations.

3. RESULTS

The implementation of these automated quality control tests is illustrated using salinity and air temperature data from the OBSEA observatory. Figure 2 illustrates the time series of 2 days of salinity data sampled at 10 s intervals in March 2014 and Figure 3 illustrates the time series of 4 month of data sampled at 20 s intervals in February–June 2014. It should be noted that these data contain numerous known errors, which is useful for the purposes of this example. The salinity measurement is a function of conductivity, temperature and pressure measurements made with the SBE 37 SMP instrument installed in the OBSEA observatory at a constant depth of

approximately 20 m. As can be seen in Figure 2, the salinity measurements present numerous errors, which have been flagged as bad data (Flag=4) by the automated Quality Control system. The erroneous salinity measurements have been caused by the incorrect measurements of the conductivity cell or for an excess threshold in the criteria of acceptance. All these considerations are going to improve the automated Quality Control system.

The percentage of incorrect salinity and conductivity measurements of the SBE 37 SMP instrument for the 4 month of data sampled at 10 s intervals in February–June 2014 was 0.95 % as shown in Table 3. In the same period, there were

Table 3. Percentages of correct, incorrect and missing values for the OBSEA parameters between February 2014 and June 2014.

Parameter measured	Value seems correct	Value seems incorrect	Value missing
Sea Water Temperature	91.69 %	0.16 %	8.15 %
Salinity	90.90 %	0.95 %	8.15 %
Depth	91.68 %	0.17 %	8.15 %
Conductivity	91.74 %	0.11 %	8.15 %
Sound Velocity	91.78 %	0.07 %	8.15 %
Sea Level Air Pressure	93.59 %	0.00 %	6.41 %
Air Temperature	81,59 %	12.00 %	6.41 %
Wind Speed	93.59 %	0.00 %	6.41 %
Wind Direction	93.59 %	0.00 %	6.41 %

Table 2. Quality number included in metadata and its correspondence.

Flag	Meaning
0	No quality control
1	Value seems correct
2	Value appears inconsistent with other values
3	Value seems doubtful
4	Value seems erroneous
5	Value was modified
6	Flagged land test
7-8	Reserved for future use
9	Data is missing
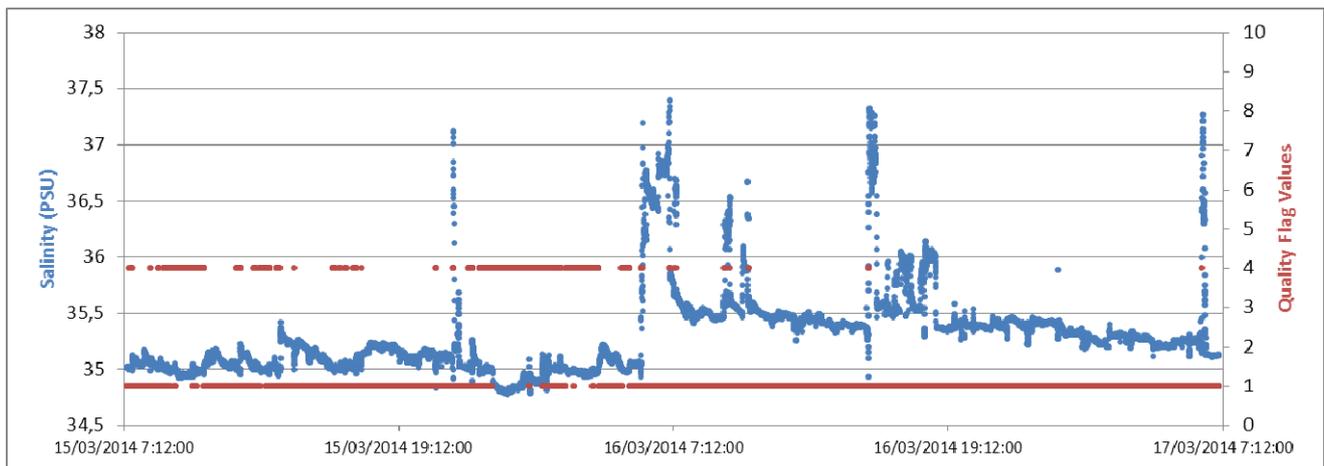
Figure 2. Time series of salinity observations between March 15 to March 17 2014 from OBSEA Sea-Bird CTD model (SBE 37 SMP) and the corresponding Quality Control Flags. The sampling rate of the instrument is 10 s. These data contain errors and missing value which are automatically flagged as incorrect data (4) and missing data (9), respectively.
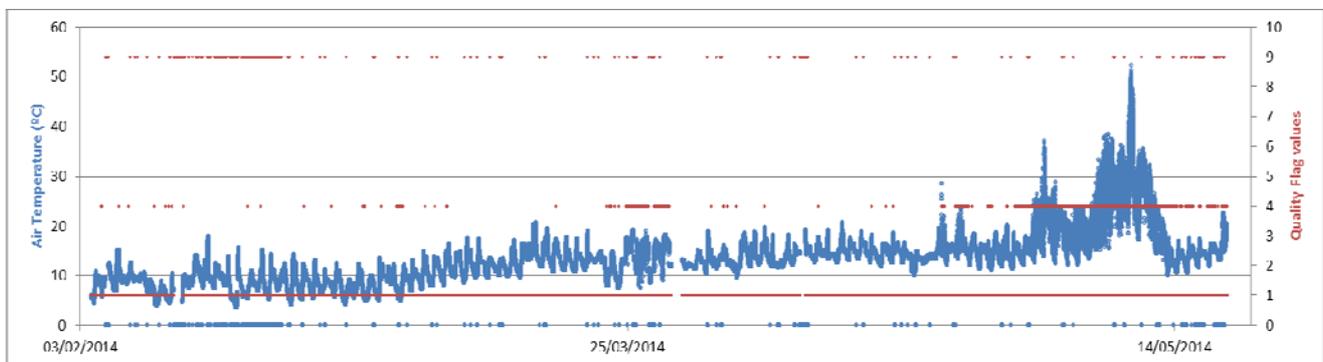
Figure 3. Time series of air temperature observations in February–June 2014 from OBSEA Airmar Weather Station model 150WX and the corresponding Quality Control Flags. The sampling rate of the instrument is 20 s. These data contain errors and missing value which are automatically flagged as incorrect data (4) and missing data (9), respectively.

8.15 % of missing data, caused by various factors such as communication problems between the land station and the OBSEA observatory and rarely caused by instrument outage.

4. CONCLUSION

The automated Quality Assurance/Quality Control provides truthfulness for long-term measures as well as the possibility of different states of the equipment, in calibration processes or out of services, among others, without cutting the link with the platform. Moreover, the metadata has to be standard for all instruments and sensors; the standardization improves the compatibility of the automatic quality control framework with different platform. It is only through the use of these standardized approach that global scale ecosystem questions can ever be addressed.

The future proposals include the uncertainty inside the metadata to know its uncertainty value, and other task will be to improve the top values of the different test to adapt at the place.

ACKNOWLEDGEMENT

REFERENCES

[1] Loescher, H. W., Ocheltree, T., Tanner, B., Swiatek, E., Dano, B., Wong, J., Zimmerman, G., Campbell, J. L., Stock, C., Jacobsen, L., Shiga, Y., Kollas, J., Liburdy, J., and Law, B. E.: Comparison of temperature and wind statistics in contrasting environments among different sonic anemometer-thermometers, Agr. Forest Meteorol., 133, pp.119–139.

[2] Ocheltree, T. O. and Loescher, H. W.: Design of the AmeriFlux portable eddy-covariance system and uncertainty analysis of carbon measurements, J. Atmos. Ocean. Tech., 24, pp.1389–1409, 2007.

[3] Taylor, J. R. and Loescher, H. W.: NEON's Fundamental Instrument Unit Dataflow and Quality Assurance Plan, NEON.011009, National Ecological Observatory Network, Boulder, Colorado, 2012.

[4] SeaDataNet, 2007: Data quality control procedures, Version 0.1, 6th Framework of EC DG Research.

[5] Sylvie Pouliquen and the DATA-MEQ working group, Recommendations for in-situ data Real Time Quality Control, EG10.19, December 2010.
URL: http://eurogoos.eu/download/publications/rtqc.pdf

[6] Daniel Toma, Ikram Bghiel, Joaquin del Rio, Alberto Hidalgo, Normandino Carreras, and Antoni Manuel 2014. Automated Data Quality Assurance using OGC Sensor Web Enablement

Frameworks for Marine Observatories, Geophysical Research Abstracts Vol. 16, EGU2014-11508, 2014 EGU General Assembly 2014.

[7] Jacopo Aguzzi, Antoni Mànuel, Fernando Condal, Jorge Guillén, Marc Nogueras, Joaquin Del Rio, Corrado Costa, Paolo Menesatti, Pere Puig, Francesc Sardà, Daniel Toma, Albert Palanques. The new Seafloor Observatory (OBSEA) for remote and long-term coastal ecosystem monitoring. In Sensors vol. 11, Issue 6, 2011.

[8] Taylor, J. R. and Loescher, H. L.: Automated quality control methods for sensor data: a novel observatory approach, Biogeosciences, 10, 4957-4971, doi:10.5194/bg-10-4957-2013, 2013.

[9] Essenwanger, O. M.: Analytical procedures for the quality control of meteorological data, Meteorological Observations and Instrumentation: Meteorological Monograph, Am. Meteorol. Soc., 33, 141–147, 1969.