Emerging Methodologies Applied to Automated Data Quality Control

Mark Bushnell, NOAA/IOOS Consultant
Karen Grissom, NOAA/NDBC
Mayra Pazos, NOAA/AOML
October 23, 2018

This white paper is submitted to the JCOMM / DBCP Task Team for Data Management. We thank Shannon McArthur / NDBC for the suggestion.

Introduction

Oceanographers have faced three specific challenges for the past half-century: limited electrical power, challenging data telemetry, and biofouling. Amazing advances are being made in all three areas, and substantial benefits are already being realized for the more agile operators. These advances lead to more robust deployments with extended durations, which is a highly desired outcome since recovery and redeployment are usually expensive efforts that are generally associated with a data gap.

Still, the need for quality assurance (QA), quality control (QC), and best practices (BP) remains, and multiple entities within the oceanographic community are working to establish standards. In some cases, multiple standards already exist, and it can be challenging to choose just one. For example, over sixteen standards for data quality control flags exist (Schlitzer, 2013). Within the U.S., the Integrated Ocean Observing System (IOOS®), Quality Assurance / Quality Control of Real-Time Oceanographic Data (QARTOD) Project is working with communities that have variable-specific needs to develop and document standardized QC tests for real-time data. The process used by QARTOD to create these QC manuals engages manufacturers, data generators, integrators, and users, nationally and internationally.

The following paragraphs describe advances that have been made to address major challenges that oceanographers face, as well as how QA, QC, and BP can be used to ensure quality data.

Better power systems

The advantages of improved battery technology have become ubiquitous—supporting cell phones, cars, and storage of renewable energy. Larger capacities and better battery management are in place and continue to expand. At the same time, many sensors use available power more efficiently, leading to more capability, longer duration , or both. Some systems use supercapacitors (van Weeren, 2005), which can be recharged much more efficiently and many more times than batteries. Solar powered surface systems are common, and solar panel efficiencies continue to increase. Cabled systems provide an abundance of power, although they are quite expensive.

Better telemetry

Iridium® satellite telemetry has been a tremendous boon to oceanographers. Small antennas and low power enable global two-way communications, even though this system was originally designed for only voice communications. Iridium NEXT is now being deployed, bringing more bandwidth and greater telemetry speeds. Data communications are poised to improve as global satellite IP telemetry develops from several large constellations (Lawrence et al., 2017).

The International Telecommunications Union has received a half-dozen applications for massive constellations providing global broadband internet, as well as position, navigation, and timing (PNT). Some of the planned systems are outlined in red in Table 1. With several competing services, it seems likely that costs will be reduced.

Table 1. Existing and planned satellite systems providing PNT services. The planned low earth orbiting (LEO) constellations outlined in red are data telemetry systems that also will be capable of PNT. These massive satellite constellations will bring competitive broadband internet access globally.

System	Orbital regime	Number of satellites	Altitude [km]	Inclination [deg]	Year fully operational
Transit	LEO	5 - 10	1,100	90	1964
Parus / Tsikada	LEO	10	990	83	1976
GPS	MEO	31	20,200	55	1995
GLONASS	MEO	24	19,100	64	2011
Galileo	MEO	24	23,200	56	2020**
BeiDou	MEO IGSO	35	21,500 35,786	55	2020**
Globalstar	LEO	48	1400	52	2000
Iridium	LEO	66	780	87	1998
Iridium NEXT	LEO	66	780	87	2018
Teledesic	LEO	288	1400	98	Cancelled in 2002
OneWeb	LEO	648	1200	88	2019*
Boeing	LEO	2956	1200	45, 55, 88	?
SpaceX	LEO	4025	1100	?	2020*
Samsung	LEO	4600	<1500	?	?

Lawrence et al., GPS World, July 2017

Better biofouling protection
Enhancements to biofouling prevention include the use of UV irradiation, programmable mechanical wipers, advanced coatings, and simple awareness of the best use of materials such as copper and copper-bearing alloys (Bushnell, 2018). Decades of experience broadly shared among operators are resulting in best practices to improve component protection.

Other improvements
Continually improving computational processors are leading to more powerful capabilities in the field. Sensors and the supporting field components are getting smaller, leading to less expensive fabrication and shipping, as well as easier deployment from smaller ships. Autonomous airborne, surface, and subsurface platforms are increasingly common, further reducing deployment costs while expanding observations.

Future QC Impacts
Emergence of Niche Data Management Systems (DMS) - As these technology improvements are implemented, we expect to see higher quality data being generated, and subsequently fewer flagged data points during post-processing QC efforts. However, the examples cited above only scratch the surface of advances, which virtually ensure that much more data and more diverse observations will be collected by an increasingly broader community.

Sensor density will increase, and data volume should expand at the same rate or faster as more metadata are incorporated. Nevertheless, data transmitted from the field should be of higher quality, with some systems conducting more QC in the field (see below). Considering the economies of scaling, burdens on the conventional DMS (within the U.S., NCEI, NDBC, CO-OPS, CDIP, GDC, etc.) should not linearly increase with data volume. However, the burden will certainly increase faster for those embracing the expanding diversity of sensors, where new QC techniques must be developed. Conventional DMS have developed sophisticated, specialized QC processes that are unique to their mission, but they are unable to manage additional data from smaller operators who have systems that do not conform to their specific standards. The solidly established DMS are very good at what they now do, but with the limited resources they are receiving, their abilities to manage the increasing load imposed by the expanding observational community are bounded.

The logical evolution is the development of new niche DMS, such as eDNA and ocean acidification observations. Benefits of the emerging, focused DMS are expertise and enthusiasm that provide the most effective QC, by individuals close to the data source. What should be asked of the conventional DMS is that they share their expertise, providing guidance, suggesting practices and standards, and warning of potential difficulties.

Expanded use of multivariate QC tests - When manual evaluation of a suspect data point occurs, it is common to inspect other variables for correlated measurements. As these relationships are identified, they should be automated to improve timely QC and reduce dependency upon individual inspection. Figure 1 shows an example where a multivariate test could be employed. Questionable spikes in wind speed are seen to be uncorrelated with barometric pressure, but they do match air temperature spikes. These collaborating observations would result in both wind and air temperature observations being flagged as good data.

Field-capable real-time QC - We expect to see a continued and expanded international effort to develop standards of all types, including stable, broadly used QC standards. Manufacturers should begin implementing real-time QC processes and data flagging embedded within their sensors and field components as described in Bushnell (2017), so that data from new, small operators can obtain equal footing by data integrators. After all, our cars have been alerting operators of system component failures for decades; why can't our oceanographic systems do the same?

Indeed, many field systems already employ embedded QC processes. The Global Drifter Program buoys use onboard numerical filters for sea-surface temperature and barometric pressure measurements, and NDBC buoys also employ advanced filtering for met/ocean observations. It's not a large leap to establish filter thresholds and set real-time QC flags based upon the filtering results.

An informal survey of several drifter manufacturers provided interesting insights. Some saw the advantages of immediate QC for use by a diverse community, while others favored software/firmware management ashore; good arguments can be made for each viewpoint. All agreed there would be no progress towards onboard QC without market demand. Drifter manufacturers are just one of many hardware providers, and other technologies may have different perspectives.

As the Internet of Things emerges, we can anticipate fully networked systems that are remotely and autonomously configurable. We expect to see increases in sensor-to-sensor communications, making neighbor and multivariate QC analysis more viable.

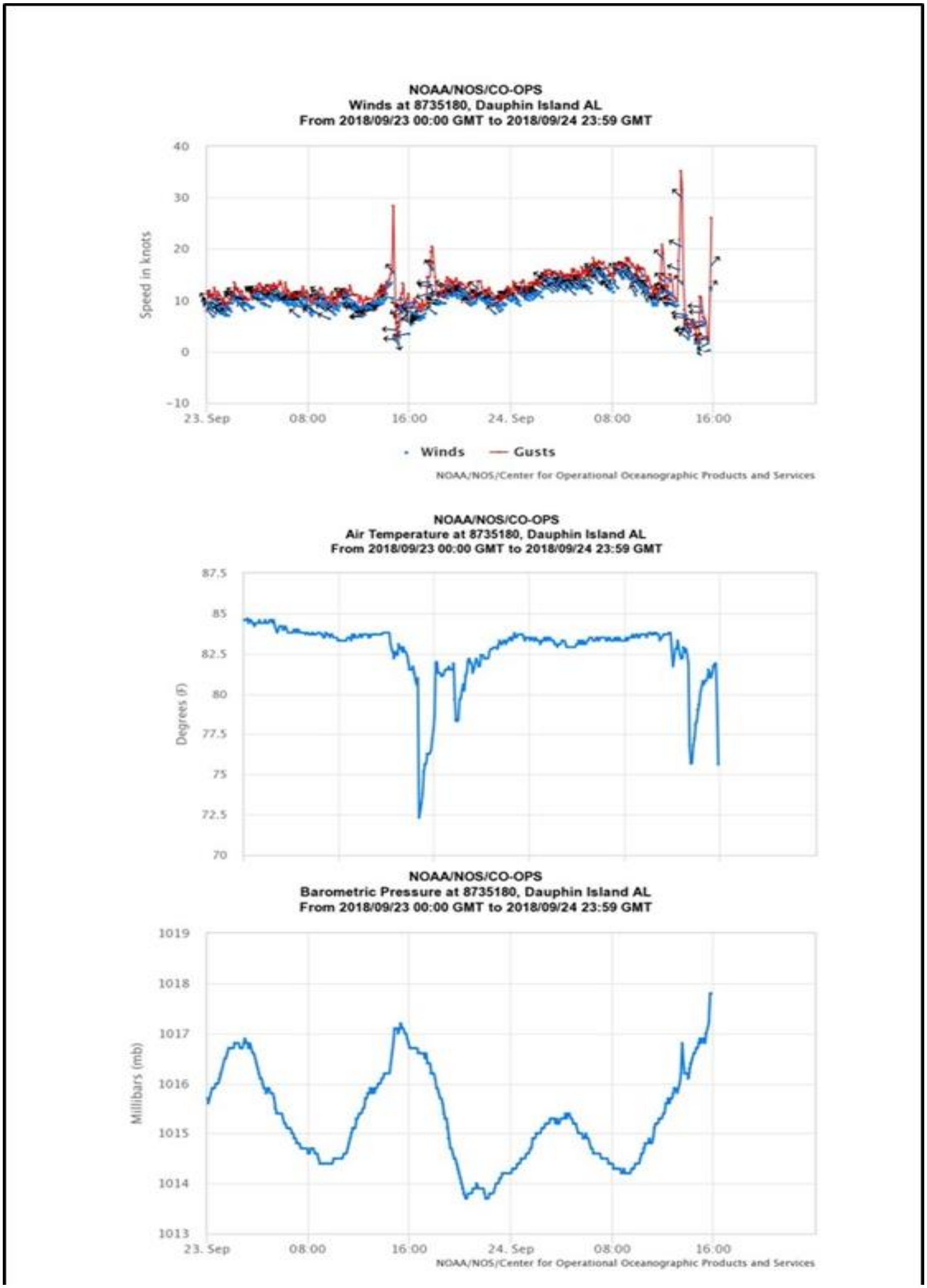

Figure 1 - Spikes in wind observations are seen to be highly related to air temperature spikes but poorly related to barometric pressure. The correlation between wind and air temperature spikes validates both observations.

QARTOD

The QARTOD Project stands ready to support and document emerging QC needs. The updated project plan (U.S. IOOS, 2017) describes the progress to date and the intentions for the next five years. This includes updates to the existing manuals, as well as the addition of new manuals as communities agree to the creation of standards for their specific variables.

QARTOD tests are grouped into required (most easily implemented), strongly recommended, and suggested categories. We expect manual updates to include documentation of the more challenging suggested tests, as the implementation of them expands. Examples of suggested tests include multivariate analysis, neighbor test, and attenuated signal test.

Summary

Ocean observations are becoming easier and less costly to obtain, and they are being gathered by an increasingly broader community. Established data management centers excel at specific functions but are hesitant to take on new ones, leading to emergent niche DMS. To address the growing QC challenge, we encourage: 1) increased cooperation to identify international standards, 2) formal guidance provided by established data management centers to emerging systems, and 3) manufacturer acceptance of established QC standards and incorporation of them within field systems.

References

Bushnell, Mark H., 2017. *Integration of QARTOD Tests Within a Sensor: Considerations for Sensor Manufacturers.* MTS/IEEE OCEANS'17 conference proceedings, 5 pp.

Bushnell, Mark H., 2018. *Evaluation of CuNiFer 10 Biofouling Resistance*. Virginia Beach, VA, CoastalObsTechServices LLC, 4pp. http://dx.doi.org/10.25607/OBP-79

Lawrence, David, H. Cobb, G. Gutt, M. O'Connor, T. Reid, T. Walter and D. Whelan, 2017. *Current Capability and Future Promise*, GPS World. http://gpsworld.com/innovation-navigation-from-leo

Schlitzer, R., 2013. Ocean Quality Flag Schemes and Mapping between Them, Version 1.4. Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany. 18 pp. http://odv.awi.de/fileadmin/user_upload/odv/misc/ODV4_QualityFlagSets.pdf

U.S. IOOS, 2017. *QARTOD Project Plan: Accomplishments for 2012-2016 and Update for 2017-2021*, 48 pp.

van Weeren, Dennis, H. Joosten, and R. Scrivens, 2005 *Hybrid power: the buoy has it*, International Ocean Systems, July/August 2005 http://datawell.nl/Portals/0/Documents/Publications/datawell_publication_ios_hybrid-power_2005-08.pdf