Methods in Oceanography 9 (2014) 17–33

Contents lists available at ScienceDirect

Methods in Oceanography

journal homepage: www.elsevier.com/locate/mio

Full length article

Quality Control (QC) procedures for Australia's National Reference Station's sensor data—Comparing semi-autonomous systems to an expert oceanographer

Elisabetta B. Morello[a,*], Guillaume Galibert[b], Daniel Smith[c], Ken R. Ridgway[d], Ben Howell[c], Dirk Slawinski[e], Greg P. Timms[c], Karen Evans[d], Timothy P. Lynch[d]

[a] *CSIRO Oceans & Atmosphere, 41 Boggo Road, Brisbane, 4001 QLD, Australia*
[b] *IMOS eMarine Information Infrastructure, University of Tasmania, private bag 21, Hobart TAS 7001 Australia*
[c] *CSIRO Digital Productivity & Services, Castray Esplanade, Hobart, 7001 TAS, Australia*
[d] *CSIRO Oceans & Atmosphere, Castray Esplanade, Hobart, 7001 TAS, Australia*
[e] *CSIRO Oceans & Atmosphere, Underwood avenue, Floreat, 6014 WA, Australia*

HIGHLIGHTS

- We describe a hybrid quality QC combining QC flags and a fuzzy logic approach.
- We apply the system to high frequency data from IMOS National Reference Stations.
- We compare the results to those produced by an independent manual QC (expert).
- The hybrid system flags 'bad' data well but did not accurately match expert QC.
- The system is a robust low-pass filter requiring further expert review of 'bad' data.

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form
22 August 2014
Accepted 8 September 2014

ABSTRACT

The National Reference Station (NRS) network, part of Australia's Integrated Marine Observing System (IMOS), is designed to provide the baseline multi-decadal time series required to understand how large-scale, long-term change and variability in the global ocean are affecting Australia's coastal ocean ecosystems. High temporal resolution observations of oceanographic variables are taken

* Corresponding author. Tel.: +61 7 3833 5947.
E-mail address: elisabetta.morello@csiro.au (E.B. Morello).

http://dx.doi.org/10.1016/j.mio.2014.09.001
2211-1220/

Keywords:
Sustained observing
Coastal oceanography
Climatology
Quality control
Fuzzy logic
IMOS

continuously across the network's nine moored stations using a Water Quality Monitor (WQM) multi-sensor. The data collected are made freely available and thus need to be assessed to ensure their consistency and fitness-for-use prior to release. Here, we describe a hybrid quality control system comprising a series of tests to provide QC flags for these data and an experimental 'fuzzy logic' approach to assessing data. This approach extends the qualitative pass/fail approach of the QC flags to a quantitative system that provides estimates of uncertainty around each data point. We compared the results obtained from running these two assessment schemes on a common dataset to those produced by an independent manual QC undertaken by an expert oceanographer. The qualitative flag and quantitative fuzzy logic QC assessments were shown to be highly correlated and capable of flagging samples that were clearly erroneous. In general, however, the quality assessments of the two QC schemes did not accurately match those of the oceanographer, with the semi-automated QC schemes being far more conservative in flagging samples as 'bad'. The conservative nature of the semi-automated systems does, however, provide a solution for QC with a known risk. Our software systems should thus be seen as robust low-pass filters of the data with subsequent expert review of data flagged as 'bad' to be recommended.

1. Introduction

Over the last decade, programs such as the Array for Real-time Geostrophic Oceanography (ARGO) (Busalacchi, 2010) and the Integrated Marine Observing System (IMOS) (Hill et al., 2010) have delivered substantially more data for oceanographic research than has ever been available before. The result is that the amount of data that needs to be quality controlled has increased substantially. For example, the National Reference Station (NRS) network, which is part of the Australian National Mooring Network (ANMN) facility of IMOS, grew between 2007 and 2010 from three monthly water quality monitoring sites, which were established in the 1940–1950s (Thompson et al., 2009), to nine highly instrumented sites (Fig. 1) (Lynch et al., 2001).

In addition to an expanded water sampling program, sensors at the NRS now collect a range of physical, chemical and biological data, including conductivity, temperature and pressure (CTD) and derived salinity, dissolved oxygen, fluorescence proxies for Chlorophyll *a*, turbidity and current velocities. Besides an increase in types of data collected, this instrumentation of sites also means that the temporal frequency of sampling has increased by up to five orders of magnitude for individual parameters. For example, temperature and salinity, which were historically measured a maximum of three times per season at each site, are now measured up to 480,000 times per season (Lynch et al., 2001). The establishment of multiple sites with sensors at multiple depths is progressively resulting in the generation of Big Data (i.e. datasets whose sizes make it impossible for commonly used software tools to manage, and process the data within a operative time frames) (Mayer-Schönberger and Cukier, 2013).

For the end-user to be able to assess the suitability of data collected by others, knowledge of any procedures performed on the data to assess quality is required. There are two ways in which the quality of data can be assessed:

- Applying a series of consecutive qualitative "gates" or flags for the data to pass through before classifying each data point; or
- Calculating quantifiable uncertainty estimates proving the goodness of the Quality Control (QC) carried out on the data in question.