Quality Control (QC) procedures for Australia’s National Reference Station’s Sensor Data – comparing semi-autonomous systems to an expert oceanographer

IMOS National Reference Stations: a continental scaled physical, chemical and biological coastal observing system
Tim P. Lynch, Elisabetta B. Morello, Karen Evans, Anthony J. Richardson, Craig R. Steinberg, Moninya Roughan, Peter Thompson, John F. Middleton, Ming Feng, Robert Sherrington, Vittorio Brando, Bronte Tilbrook, Ken Ridgway, Simon Allen, Peter Doherty, Katherine Hill, Tim C. Moltmann – submitted Plos One

Inter-fleet comparisons of FLNTU outputs to reduce uncertainties in bio-optical data streams in a sustained observing system
Tim P. Lynch, Lesley Clementson, Robert Kay, Moninya Roughan, Vittorio Brando, Martina Doblin, Karen Evans – in prep
Australian National Mooring Network
ANMN

Designed

77 moorings
15 Biogeochemical water sampling sites
9 National Reference Stations

Minimum scenario

35 moorings
12 Biogeochemical water sampling sites
8 (7) National Reference Stations
NRS system – 9 down to 7
NRS sensors

Institute

- SARDI
- SIMS
- AIMS
- AIMS
- AIMS
- AIMS
- CSIRO
- CSIRO
- CSIRO
- CSIRO

Nett Radiometer
CNR1
Vaisala
MRU

WETLabs ECO triplet B
Battelle pCO₂
SAMI2 pH
Aanderaa Oxygen Optode

Seabird 16+
WOM
NXIC-bio
AQUA logger 520T

AQUA logger 520PT
SBE 39
ADCP Workhorse 300kHz
ADCP Workhorse 600kHz

ADCP Nortek 600kHz

Dual acoustic releases
Cross validation – sensors/samples/satellite

![Graphs showing salinity and temperature variations over years.](image)
Cross validation of sensor to monthly BGC samples – Chlorophyll 20m Maria Island NRS

Comparisons sensor to bottle at 20 m

$y = 0.6506x + 0.3566$

$R^2 = 0.8285$
Chlorophyll 80m Maria Island NRS

Comparisons sensor to bottle @ 80 m

$y = 1.1654x + 0.0824$

$R^2 = 0.8994$
Chlorophyll 25m Port Hacking NRS – median of 3 bursts

Comparisons sensor to bottle at 25m

\[ y = 0.3618x + 0.0431 \]

\[ R^2 = 0.9782 \]
Temperature (°C) and salinity (psu) climatologies from the Maria Island NRS at each depth for each day of the year from the long–term monthly sampling (a and b, respectively) and from the regional climatology (c and d, respectively). The difference plots are shown in (e) and (f). The station climatologies are plotted against the regional climatologies in (g) temperature and (h) salinity, the lines are the linear best-fit to the data. The correlation and regression coefficients are also shown.
Climatology built with the 10 m temperature data (blue crosses) from Maria Island NRS. The number of observations contained in each fortnightly bin are in green, the mean and +/- 6 standard deviation are in red.
(a) WQM raw temperature data for the Maria Island site and (b) associated calculated climatologies at 20 m (red) and 90 m (blue)
Table 1. Data flag scheme used by the Intergovernmental Oceanographic Commission (IOC) of UNESCO

<table>
<thead>
<tr>
<th>Flag</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC performed</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Below detection limit</td>
</tr>
<tr>
<td>7</td>
<td>In excess of quoted value</td>
</tr>
<tr>
<td>8</td>
<td>Interpolated value</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
</tr>
<tr>
<td>A</td>
<td>Incomplete information</td>
</tr>
</tbody>
</table>
QC tests – qualitative flags

- 2.1.1 Impossible date
- 2.1.2 Impossible location test
- 2.1.3 In/out-water test
- 2.1.4 Global range test
- 2.1.5 Regional range
- 2.1.6 Impossible depth
- 2.1.7 Spike test
- 2.1.8 Rate of change test
- 2.1.9 Stationarity test
- 2.1.10 Climatology test
Schematic of the fuzzy quality assessment system for temperature observations at Maria Island NRS
Fuzzy logic membership and rules

Table 3. The fuzzy membership intervals for the quantitative tests for temperature and salinity observations at the Maria Island NRS.

<table>
<thead>
<tr>
<th>Variable</th>
<th>IMOS Test</th>
<th>Low</th>
<th>Low → Medium</th>
<th>Medium</th>
<th>Medium → High</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Vertical Spike (s)</td>
<td>$x_s &lt; 0.07$</td>
<td>$0.07 \leq x_s &lt; 0.2$</td>
<td>$0.2 \leq x_s \leq 2$</td>
<td>$2 &lt; x_s \leq 6$</td>
<td>$x_s &gt; 6$</td>
</tr>
<tr>
<td></td>
<td>Rate of Change (rc)</td>
<td>$x_{rc} &lt; 0.5 \sigma_{rc}$</td>
<td>$0.5 \sigma_{rc} \leq x_{rc} &lt; 1.5 \sigma_{rc}$</td>
<td>$1.5 \sigma_{rc} \leq x_{rc} &lt; 3 \sigma_{rc}$</td>
<td>$3 \sigma_{rc} \leq x_{rc} &lt; 4 \sigma_{rc}$</td>
<td>$x_{rc} &gt; 4 \sigma_{rc}$</td>
</tr>
<tr>
<td></td>
<td>Climatology (c)</td>
<td>$x_c &lt; 3 \sigma_c$</td>
<td>$3 \sigma_c \leq x_c &lt; 4 \sigma_c$</td>
<td>$4 \sigma_c \leq x_c &lt; 5 \sigma_c$</td>
<td>$5 \sigma_c \leq x_c &lt; 6 \sigma_c$</td>
<td>$x_c &gt; 6 \sigma_c$</td>
</tr>
<tr>
<td>Salinity</td>
<td>Vertical Spike (s)</td>
<td>$x_s &lt; 0.05$</td>
<td>$0.05 \leq x_s &lt; 0.15$</td>
<td>$0.15 \leq x_s \leq 0.5$</td>
<td>$0.5 &lt; x_s \leq 0.9$</td>
<td>$x_s &gt; 0.9$</td>
</tr>
<tr>
<td></td>
<td>Rate of Change (rc)</td>
<td>$x_{rc} &lt; \sigma_{rc}$</td>
<td>$\sigma_{rc} \leq x_{rc} &lt; 2 \sigma_{rc}$</td>
<td>$2 \sigma_{rc} \leq x_{rc} &lt; 3 \sigma_{rc}$</td>
<td>$3 \sigma_{rc} \leq x_{rc} &lt; 4 \sigma_{rc}$</td>
<td>$x_{rc} &gt; 4 \sigma_{rc}$</td>
</tr>
<tr>
<td></td>
<td>Climatology (c)</td>
<td>$x_c &lt; 3 \sigma_c$</td>
<td>$3 \sigma_c \leq x_c &lt; 4 \sigma_c$</td>
<td>$4 \sigma_c \leq x_c &lt; 5 \sigma_c$</td>
<td>$5 \sigma_c \leq x_c &lt; 6 \sigma_c$</td>
<td>$x_c &gt; 6 \sigma_c$</td>
</tr>
</tbody>
</table>

Table 4. Fuzzy rules used to map the quality test input functions to the output membership functions.

<table>
<thead>
<tr>
<th>Fuzzy Logic Rule</th>
<th>Mathematical operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $x_s$ is low and $x_{rc}$ is low and $x_c$ is low then $Q$ is low uncertainty</td>
<td>$Q_t = K_t(u_t)$</td>
</tr>
<tr>
<td></td>
<td>$u_t = \text{mean}(S_t(x_s), RC_t(x_{rc}), C_t(x_c))$</td>
</tr>
<tr>
<td>If $x_s$ is medium and $x_{rc}$ is medium and $x_c$ is medium then $Q$ is medium uncertainty</td>
<td>$Q_m = K_m(u_m)$</td>
</tr>
<tr>
<td></td>
<td>$u_m = \text{mean}(S_m(x_s), RC_m(x_{rc}), C_m(x_c))$</td>
</tr>
<tr>
<td>If $x_s$ is high or $x_{rc}$ is high or $x_c$ is high then $Q$ is high uncertainty</td>
<td>$Q_h = K_h(u_h)$</td>
</tr>
<tr>
<td></td>
<td>$u_h = \text{max}(S_h(x_s), RC_h(x_{rc}), C_h(x_c))$</td>
</tr>
</tbody>
</table>
Comparison of the quality assessment schemes against the expert assessment. (a) raw temperature observations acquired at Maria Island NRS at 20m from 24/11/2009 to 24/10/2011; (b) expert assessment; (c) qualitative assessment scheme (QC flag) (d) quantitative assessment scheme (Fuzzy logic CQ)
Percentage agreement between QC systems and manual expert QC

Table 5. A comparison of the quality assessments produced by two semi-automated QC schemes (qualitative and quantitative) with the benchmark quality assessments produced by an oceanographer expert for temperature and salinity observations acquired at Maria Island NRS (at 20m and 90m) between the 24/11/2009 and 24/10/2011. The number (%) of correctly assessed observations refers to the samples where the semi-automated quality assessments were in agreement with the corresponding expert based quality assessments of ‘good’ or ‘bad’.

<table>
<thead>
<tr>
<th></th>
<th>Qualitative tests (QC flags)</th>
<th>Quantitative tests (Fuzzy logic)</th>
<th>Expert Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number (%) of observations</td>
<td>Number (%) of observations</td>
<td>Number (%) of</td>
</tr>
<tr>
<td></td>
<td>correctly assessed as ‘bad’</td>
<td>correctly assessed as ‘good’</td>
<td>observations</td>
</tr>
<tr>
<td>Temperature (20m)</td>
<td>3962 (23.4%)</td>
<td>3939927 (99.9%)</td>
<td>4334 (25.6%)</td>
</tr>
<tr>
<td>Temperature (90m)</td>
<td>115810 (62.4%)</td>
<td>2495604 (66.3%)</td>
<td>103370 (55.7%)</td>
</tr>
<tr>
<td>Salinity (20m)</td>
<td>9227 (34.1%)</td>
<td>3930148 (100.0%)</td>
<td>9579 (35.4%)</td>
</tr>
<tr>
<td>Salinity(90m)</td>
<td>16631 (59.0%)</td>
<td>2667776 (68.0%)</td>
<td>15137 (53.7%)</td>
</tr>
</tbody>
</table>

Expert marks more data as ‘good’
- Regional ranges?
- Rate of change?
- Spikes?
- Climatologies?